RR 95-04

TRIADIC DISTANCE MODELS FOR TRIADIC DISSIMILARITY DATA

Willem J. Heiser[1]
Mohammed Bennani[2]

[1]Department of Data Theory
Faculty of Social and Behavioural Sciences
Leiden University
Leiden, The Netherlands

[2]Département d'Analyse des Données
Universitee de Haut Bretagne Rennes II,
Rennes, France

Abstract

Triadic proximity data are empirical relations defined on triples of objects, brought together in a three-way, one mode table. A theoretical framework is given for characterizing triadic dissimilarity, triadic similarity, and triadic distance. Three specific distance models, the perimeter model, the Minkowski-*p* model, and Daws' model, are shown to satisfy the tetrahedral inequality, a condition that is characteristic for the present approach. Convergent algorithms are described for finding least squares representations in Euclidean space, and an analysis of dispersion is derived, which enables a scale-free evaluation of the quality of the fit. Several indices of (dis)similarity defined on presence-absence variables are shown to satisfy the given definition of triadic distance. Distance analysis of three-way, three-mode tables is treated as an important special case.

Keywords: three-way one-mode data, three-way symmetry, tetrahedral inequality, analysis of dispersion, multidimensional scaling, three-way unfolding, iterative majorization, Minkowski distance, Pijk model, perimeter model, surface area model, presence-absence variables, free sorting, cross-modality matching

Triadic Distance Models for Triadic Dissimilarity Data

Introduction

A data element in a three-way table refers to a relation among a triad of objects, each object being selected from one of three sets of objects. The growing interest in the analysis of such tables, and in the mathematical structure of the models used (Law, Snyder, Hattie, and McDonald, 1984; Coppi and Bolasco, 1989), can largely be explained by the richness of the information that they contain. A ternary relation is much richer than a binary relation. When we go from "two-way" to "three-way", the level of complexity grows considerably, and in order to get a good understanding of the situation, we are forced to bring in more efforts and more skills to maximally exploit all characteristics of the table. It is often felt that analysis techniques that do not act upon all three ways simultaneously run serious risks of being insufficient by masking a part of the three-way information.

A large number of different three-way tables exists, and a multitude of analysis techniques. Yet we can group them into three classes, characterized by an increasing amount of symmetry in the entries of the table. In doing so, we employ Tucker's (1964) terminology, which distinguishes between "ways" and "modes": while the term *way* simply refers to the dimensions of a table, the term *mode* refers to the identity of each way; the number of modes indicates whether one set of objects is identical to anyone of the others, or distinct.

The first class consists of *three-way three-mode* tables, where the modes typically are a set of individuals O_1, a set of variables O_2, and a set of conditions O_3 (or stimuli, or occasions). Most methods to analyse this type of table are special cases of Tucker's (1963, 1966) three-mode factor analysis model, also called the Tucker-3 model (Kroonenberg, 1983). This approach tries to reduce the information in the three-mode table as a weighted trilinear function of a small number of parameters for typical individuals, latent variables, and prototypical conditions. The elements of the representation are mutually related in a smaller three-way three-mode table, called the *core*, which contains the weights of the trilinear function. If the core is superdiagonal, we obtain the Parafac/Candecomp model, due independently to Harshman (1970) and Carroll and Chang (1970), which is the most prominent and best studied case (Kruskal, 1977, 1989). When columns or pairs of columns of the parameter matrices are restricted to unity, we obtain models with additive components (Gower, 1977).

The second class consists of *three-way two-mode* tables, defined on $O_1 \times O_1 \times O_3$, $O_3 \times O_2 \times O_3$, or any other combination obtained by setting two of the three sets of objects equal to each other. For fixed values of the mode that is associated to a single way (in the case $O_1 \times O_1 \times O_3$, this is O_3), the two-way table defines a symmetric relation on the Cartesian product of the other mode (in the example, O_1). We assume symmetry, because the possibility of asymmetry is better subsumed under the previous class of tables. The symmetric relation conditional on O_3 is expressed in the data as a measure of dissimilarity or similarity among pairs of elements from O_1, collected from different data sources indexed by O_3. Such data are also known in the literature as three-way proximity data, and several generalizations of the Euclidean distance model have been considered, best known of which is the INDSCAL model (Carroll and Chang, 1970).

This paper will be concerned with the third class of tables, obtained by identifying each way with the same mode, that is, with *three-way one-mode* tables. Generally, we will again assume some similarity or dissimilarity relation, this time defined on triads instead of dyads. To clearly distinguish this case from the previous one, we call it *triadic proximity data*, and the three-way two-mode case (replicated, or conditioned) *dyadic proximity* data. Triadic proximities are not often directly collected. Daws (1993) used the free sorting method, in which subjects are instructed to produce a partition of the set of stimuli into any number of classes, according to any self-selected criterion. One of the usual procedures in this method is to construct dyadic proximity data from the partitions by counting the number of times that each pair of stimuli has been placed in the same class. However, Daws showed that reduction to dyadic proximities implied a serious loss of information with respect to the original classifications, and therefore he proposed to study the number of times that each triple of stimuli has been placed in the same class as a measure of triadic similarity.

Hayashi (1972) has been the first to consider a triadic model for the dissimilarity among three elements from the same set; he proposed the *surface area model*, using a special quantification method in two dimensions for modelling the data by the squared area of the triangle formed by three points (also see Hayashi, 1989). A similar method has been developed by Pan and Harris (1991), who used the area model to define triadic similarities in high-dimensional space, and then modelled these in low-dimensional space with the square of what we will call the *generalized Euclidean model*, or M_2 model. Independently, Cox, Cox and Branco (1991) developed a least squares method for the M_2 model, and extended it to the non-metric case and to K-adic relations. In an important unpublished paper, Joly and Le Calvé (1989) initiated the axiomatic study of distance models for triads, introduced the *three-way ultrametric* and the *perimeter model*, and stated the necessary and sufficient conditions for Euclidean embeddability of the M_2 model.

Our discussion starts with a similar axiomatic framework, in which triadic similarity and dissimilarity are defined as mappings from the three-way Cartesian product of the set of objects into the non-negative reals that satisfy very general conditions, the most important of which is *three-way symmetry*. Then the ordinary concept of distance between pairs is extended to the case of triads by posing an inequality to be called the *tetrahedral inequality*, which bounds the triadic distance associated with one face of a tetrahedron in terms of the sum of the triadic distances associated with the other three faces, and we establish some properties of the resultant class of functions. Three specific models are discussed in detail: the perimeter model, the *Minkowski-p* model, of which M_2 model is a special case, and Daws' model. For all three models it is shown that they satisfy the stated conditions for being a triadic distance. The next section describes two convergent algorithms – one for the Euclidean perimeter model, the other for the M_2 model – to find least squares representations of triadic dissimilarities in any prescribed dimensionality, and demonstrates how the total dispersion can be broken down into a residual component and a percentage dispersion-accounted-for. Since they incorporate weights and are based on a general algorithmic strategy, the proposed procedures are very versatile; as shown in one of the examples, they can be used for three-way three-mode data as well. A separate section is devoted to dissimilarities defined on presence-absence variables, in which we show that a number of generalized indices, representative for the most common indices in the dyadic case, satisfy our definition of triadic distance.

Triadic Dissimilarity, Triadic Similarity, and Triadic Distance

Let O be a finite set of n elements, denoted by $O = \{1, 2, \dots, i, j, k, \dots, n\}$, which are the labels of the modelling units, or *objects*. We start by defining the concept of triadic dissimilarity, provide a dual treatment of similarity, and then move to triadic distance by adding two more properties to the dissimilarity definition, one shared by all semi-metrics, and another comparable to the triangle inequality for ordinary metrics.

A triadic dissimilarity on O measures the lack of resemblance between objects in O taken three at a time, which is zero if there is no lack of resemblance, and positive otherwise. Mathematically, it is represented as a *bounded function*: a mapping T of $O \times O \times O$ into $\mathbb{R}^+$, the non-negative reals. Thus the first property that a triadic dissimilarity $\tau_{ijk} = T(i, j, k)$ has to satisfy is *non-negativity*:

$$\tau_{ijk} \geq 0 , \tag{1a}$$

for all i, j, and k. It is conceptually undesirable if the resemblance or lack of resemblance between three objects would depend on the order in which they are listed. Therefore, the second natural requirement is *three-way symmetry*, that is, for all permutations π of $\{i, j, k\}$ we must have

$$\tau_{ijk} = \tau_{\pi(i)\pi(j)\pi(k)} , \tag{1b}$$

for all i, j, and k. Lack of resemblance of an object with itself should not be different from zero, so we require *minimality*:

$$\tau_{ijk} = 0 \quad \text{if } i = j = k \,. \tag{1c}$$

When one of the objects is identical to one of the others, the lack of resemblance between the two non-identical objects should remain invariant regardless of which two are the same, so that a last natural requirement is

$$\tau_{iij} = \tau_{iji} = \tau_{ijj} \,, \tag{1d}$$

for all i and j. Requirement (1d) is called *diagonal-plane equality*, because it requires equality of the three matrices $\{\tau_{iij}\}$, $\{\tau_{iji}\}$, and $\{\tau_{ijj}\}$, which are formed by cutting the three-way block diagonally, starting at one of the three edges joining at the corner τ_{111}. By symmetry, we must also have $\tau_{iij} = \tau_{jji}$, $\tau_{iji} = \tau_{jij}$, and $\tau_{ijj} = \tau_{jii}$. The quantities in the diagonal planes of the three-way table simply measure the *dyadic dissimilarity*, defined by $\delta_{ij} = \tau_{iij}$. Note that where the three planes intersect we have the elements of the (three-way) diagonal $\delta_{ii} = \tau_{iii} = 0$. Summarizing, a triadic dissimilarity on O is a mapping of the three-fold Cartesian product of O into the real numbers that satisfies (1a-1d).

Triadic dissimilarity is defined here on triples of elements selected from a single set O. The case where the three elements are selected from three distinct sets (mentioned in the introduction) is a special case, obtained by partitioning O into $\{O_1, O_2, O_3\}$ and considering triples from the subset $O_1 \times O_2 \times O_3$ only. Thus three-way three-mode data are regarded as a special case of three-way one-mode data, just as two-way two-mode data (e.g., individual preferences) can be regarded as a special case of two-way one-mode data (in the example: similarities between two distinct groups of objects). It is also important to note that the notion of triadic dissimilarity is entirely different from the notion of three-way two-mode dissimilarity (used in the INDSCAL model), because the latter generally lacks three-way symmetry.

It is possible to define in a dual way the notion of *triadic similarity* as a measure of resemblance on triples of objects (Bennani, 1993). Formally, a triadic similarity function is a mapping R of $O \times O \times O$ into $\mathbb{R}^+$ such that, $\forall\ i, j, k \in O$, and $\forall\ \pi$ on $\{i, j, k\}$, $\rho_{ijk} = R(i, j, k)$ satisfies

$$\rho_{ijk} \geq 0\ , \tag{2a}$$

$$\rho_{ijk} = \rho_{\pi(i)\pi(j)\pi(k)}\ , \tag{2b}$$

$$\rho_{iii} = \rho_{jjj} = \rho_{kkk} \geq \rho_{ijk}\ , \tag{2c}$$

$$\rho_{iji} = \rho_{ijj}\ . \tag{2d}$$

The notions of dissimilarity and similarity play opposite roles. Like in the two-way case, we may associate each triadic similarity R with one triadic dissimilarity T by specifying a decreasing function. As an extension of classic transformations (Gower, 1986) we could specify, for example, $\tau_{ijk} = \rho_{iii} - \rho_{ijk}$, $\tau_{ijk} = \rho_{iii} / (1 + \rho_{ijk})$, or $\tau_{ijk} = (\rho^2_{iii} - \rho^2_{ijk})^{1/2}$.

Triadic dissimilarity functions can be embedded in a vector space of dimension $n(n^2 - 1)/6$, and the set of all such functions is a convex cone, in fact the positive orthant of this vector space (Bennani, 1993). Not all these functions allow a simple representation, and therefore it is often desirable to require a stronger, metric structure for modelling purposes. In the notation we will distinguish triadic distances from triadic dissimilarities by writing t_{ijk} for the former and τ_{ijk} for the latter. In analogy to the dyadic case, where a basic requirement already for semi-metric structure is *definiteness*, a first addition is

$$t_{ijk} = 0 \quad \textit{only} \text{ if } i = j = k\ . \tag{3a}$$

Thus, while zero triadic dissimilarity is not excluded in triads of different objects, the notion of a semi-metric space implies that coinciding points are regarded as identical. In the role of the *triangle inequality*, we propose the following *tetrahedral inequality* as a second additional constraint:

$$2t_{ijk} \leq t_{ikl} + t_{jkl} + t_{ijl}\ . \tag{3b}$$

It compares, in the tetrahedron formed by four corners i, j, k and l, the size of the face made up by i, j, and k with the sum of the sizes of the faces that join at corner l. The corresponding empirical principle is: three objects that all resemble a fourth object cannot be very different, while three objects being very different implies that at least two of them are very different from the fourth. Of course, if neither of them resembles the fourth object, they still could be arbitrarily similar among themselves. As we shall see shortly, if l coincides with i, j, or k, the tetrahedral inequality reduces to the triangular inequality.

Thus a *triadic distance function* is defined as a mapping T of O^3 into $\mathbb{R}^+$ satisfying (1*a*-*d*) and (3*a*-*b*). Certain axioms in the definition are redundant; for example, three-way symmetry (1*b*) and the tetrahedral inequality (3*b*) imply positivity, as can been seen by adding the three inequalitities

$$2t_{ikl} \le t_{ijk} + t_{ijl} + t_{jkl} ,$$

$$2t_{jkl} \le t_{ijk} + t_{ikl} + t_{ijl} ,$$

$$2t_{ijl} \le t_{ijk} + t_{ikl} + t_{jkl} .$$

member by member, which yields $t_{ijk} \ge 0$. We also have the following lemma.

Lemma 1. *The tetrahedral inequality implies* $t_{ijk} \le t_{ikl} + t_{jkl}$ *for all* $i, j, k, l \in O$.

<u>*Proof*</u>. Suppose that the assertion were not true; then there would exist a tetrad $i, j, k, l \in O$ such that $t_{ijk} > t_{ikl} + t_{jkl}$, and since $t_{ikl} + t_{jkl} + t_{ijl} \ge 2t_{ijk}$, we would have $t_{ijl} > t_{ijk}$. However, by three-way symmetry it is also true that $t_{ijk} + t_{ikl} + t_{jkl} \ge 2\, t_{ijl}$, from which it would follow that $t_{ijl} < t_{ijk}$, and so our supposition leads to a contradiction. ❐

What is called triadic distance here was called *strong three-way distance* in Bennani (1993), distinguishing it from the concept of *three-way distance* used in Joly and Le Calvé (1989). The Joly-Le Calvé three-way distance is based on the inequalities $t_{iik} \le t_{ijk}$ and $\max(t_{ijk}, t_{ijl}) \le t_{ikl} + t_{jkl}$. It is not hard to prove that a triadic distance is a Joly-Le Calvé three-way distance, but not conversely. Bennani (1993) showed that, for four points, each triadic distance can be written as

$$t_{ijk} = d_{ij} + d_{ik} + d_{jk} , \qquad (4)$$

where the non-negative, symmetric quantities d_{ij} satisfy the triangle inequality $d_{ij} \leq d_{ik} + d_{jk}$, i.e., they are distances. This result does not hold for more then four points, but representation (4), called the *perimeter model*, more of which will be said in the section on special cases, is a rather fundamental one. Inserting (4) into (3*b*) and simplifying we obtain

$$t_{ijk} \leq 2d_{il} + 2d_{jl} + 2d_{kl} \,, \tag{5}$$

where the triadic distance is bounded in terms of the dyadic distances with respect to a fourth object. The combination of (4) and (5) proves the next lemma.

Lemma 2. *The tetrahedral inequality reduces to the triangular inequality if l coincides with i, j, or k. It becomes an equality only if, in addition, the coinciding point is on the line segment connecting the others.*

The following proposition gives a way to construct a new triadic distance from a given one.

Proposition 1. *Let c be a strictly positive real number and T a triadic distance function. Then T* defined as T*(i, j, k) = T(i, j, k) / (c + T(i, j, k)) is also a triadic distance function.*

Proof. Obviously, T^* is a triadic dissimilarity and is definite when T is a triadic distance. Let the quantity a be defined as

$$a = (c + t_{ijk})\,(c + t_{ikl})\,(c + t_{jkl})\,(c + t_{ijl})\,(t^*_{ikl} + t^*_{jkl} + t^*_{ijl} - 2t^*_{ijk}) \,, \tag{6}$$

then T^* satisfies the tetrahedral inequality provided that a is non-negative. Expanding (6) into polynomial form, we obtain

$$\begin{aligned} a \; &= c^3\,(t_{ikl} + t_{jkl} + t_{ijl} - 2\,t_{ijk}) \\ &\quad + c^2\,(2\,t_{ikl}t_{ijl} + 2\,t_{ikl}t_{jkl} + 2\,t_{ijl}t_{jkl} - t_{ijk}t_{ijl} - t_{ijk}t_{ikl} - t_{ijk}t_{jkl}) \\ &\quad + 3c\,t_{ikl}t_{jkl}t_{ijl} + t_{ikl}t_{jkl}t_{ijl}t_{ijk} \,. \end{aligned}$$

Because the cubic coefficient and the linear terms are non-negative by the definition of T, we only have to verify that the coefficient of c^2, denoted as b, is non-negative. We have

$$\begin{aligned}2b &= t_{ijl}\,(t_{ikl} + t_{jkl} + t_{ijl} - 2\,t_{ijk}) + t_{ikl}\,(t_{ikl} + t_{jkl} + t_{ijl} - 2\,t_{ijk}) + t_{jkl}\,(t_{ikl} + t_{jkl} + t_{ijl} - 2\,t_{ijk}) \\ &\quad + 2\,t_{ikl}t_{ijl} + 2\,t_{ikl}t_{jkl} + 2\,t_{ijl}t_{jkl} - t_{ijl}^2 - t_{ikl}^2 - t_{jkl}^2\,.\end{aligned}$$

Since the first three terms of this expression are non-negative because T is a triadic distance, it is sufficient to check the sign of the remaining terms. By repeated use of Lemma 1, with the role of t_{ijk} taken by t_{ijl} and t_{ikl}, respectively, and rearranging the inequalities, we may write

$$\begin{aligned}t_{ikl}^2 &\geq (t_{ijl} - t_{jkl})^2\,, \\ t_{jkl}^2 &\geq (t_{ijl} - t_{ikl})^2\,, \\ t_{ijl}^2 &\geq (t_{ikl} - t_{jkl})^2\,.\end{aligned}$$

Adding these inequalities member by member, and simplifying, we obtain

$$t_{ijl}^2 + t_{jkl}^2 + t_{ikl}^2 \leq 2\,t_{ijl}t_{jkl} + 2t_{ijl}t_{ikl} + 2\,t_{ikl}t_{jkl}\,,$$

which proves that $b \geq 0$, and hence $a \geq 0$. ❐

This result closes our discussion of the general concept of triadic distance function. Let us now turn to a number of special cases.

Triadic Distance Models

In this section we will give three specific examples of classes of functions with the property that all their members satisfy the tetrahedral inequality. Since each class is generated by varying some set of parameters, they can be called models. The first two models are constructed on the basis of dyadic distances, while the last one is built in a different manner.

Perimeter model

This model was already introduced in the previous section. Given dyadic distances $\{d_{ij}\}$ among all pairs of points in some metric space, (4) measures the lack of resemblance in each triangle by its perimeter. More generally, given dyadic dissimilarities $\delta_{ij} = \Delta(i, j) = \Delta(j, i)$, the perimeter dissimilarity is defined by $\tau_{ijk} = \delta_{ij} + \delta_{ik} + \delta_{jk}$. The following result establishes when the perimeter model is a triadic distance model.

Proposition 2. *Let D be a mapping of $O \times O$ into $\mathbb{R}^+$, with $d_{ij} = D(i, j) = D(j, i)$ and $d_{ij} = 0$ if and only if $i = j$. If D satisfies the triangle inequality, then the mapping T defined by $t_{ijk} = d_{ij} + d_{ik} + d_{jk}$ is a triadic distance function.*

Proof. It is easily checked that T satisfies (1*a-d*) and (3*a*). Inserting the definition of T into the tetrahedral inequality (3*b*), we must have

$$2\, d_{il} + 2\, d_{jl} + 2\, d_{kl} \geq d_{ij} + d_{ik} + d_{jk}\, .$$

This inequality follows immediately if we add the three triangle inequalities that are related to each side of the triangle i, j, k

$$d_{il} + d_{jl} \geq d_{ij}\, ,$$
$$d_{il} + d_{kl} \geq d_{ik}\, ,$$
$$d_{jl} + d_{kl} \geq d_{jk}\, ,$$

member by member. ❐

Although in general a triadic distance need not have perimeter distance form, we have already pointed to the result that, for four points, it must have that property. This fact can be demonstrated by showing that the system of 12 triangle inequalities in the 6 unknown dyadic distances always has at least one solution. With more than four points, the result does not hold. In that case, we might be interested in the following optimization problem: given a set of triadic dissimilarities $\{\tau_{ijk}\}$, find a set of dyadic distances $D = \{d_{ij}\}$ that minimizes the least squares criterion

$$\sigma_A^2(D) = \sum_{(i,j,k)\in L} (\tau_{ijk} - d_{ij} - d_{ik} - d_{jk})^2. \tag{7}$$

The summation in (7) is over the index list L, containing the $\binom{n}{3}$ off-diagonal triplets (i, j, k) with

$i < j < k$ and the $\binom{n}{2}$ diagonal-plane triplets (i, i, j) with $i < j$. This list is sufficient if $\{\tau_{ijk}\}$ is a proper set of triadic dissimilarities. For suppose the latter is not the case, and let $\{\delta_{ijk}\}$ denote a general three-way array of dissimilarities that only satisfy non-negativity. Then a simplifcation is possible, as shown by Proposition 3, which is stated without proof.

Proposition 3. *When the loss function in (7) is adjusted to include general three-way arrays δ_{ijk}, with summation over all i, j, and k, it can be decomposed into six components:*

$$[\,\Sigma_i \Sigma_j \Sigma_k\, (\delta_{ijk} - d_{ij} - d_{ik} - d_{jk})^2\,] / 6 =$$

$$[\,\Sigma_{i=j=k}\, \delta^2_{ijk}] / 6 \qquad \textit{(lack of minimality)}$$

$$+ \;\Sigma_{i<j<k}\, [\,\Sigma_\pi (\delta_{\pi(i)\pi(j)\pi(k)} - \hat{\tau}_{ijk})^2\,] / 6 \qquad \textit{(lack of three-way symmetry)}$$

$$+ \;\Sigma_{i \neq j}\, [\,\Sigma_\kappa (\delta_{i\kappa(i)\kappa(j)} - \hat{\tau}_{iij})^2\,] / 6 \qquad \textit{(lack of diagonal-plane equality)}$$

$$+ \;\Sigma_{i<j}\, (\hat{\tau}_{iij} - \hat{\tau}_{jji})^2 / 4 \qquad \textit{(lack of symmetry within diagonal planes)}$$

$$+ \;\Sigma_{i<j<k}\, (\hat{\tau}_{ijk} - d_{ij} - d_{ik} - d_{jk})^2 \qquad \textit{(off-diagonal loss)}$$

$$+ \;\Sigma_{i<j}\, (\,\hat{\delta}_{ij} - 2d_{ij})^2, \qquad \textit{(diagonal-plane loss)}$$

where π indexes the six permutations of (i, j, k), κ indexes the three combinations (i, j), (j, i) and (j, j), and where the quantities with a hat are defined as:

$$\hat{\tau}_{ijk} = [\,\Sigma_\pi\, \delta_{\pi(i)\pi(j)\pi(k)}\,] / 6\,,$$

$$\hat{\tau}_{iij} = [\,\Sigma_\kappa\, \delta_{i\kappa(i)\kappa(j)}\,] / 3\,,$$

$$\hat{\delta}_{ij} = (\hat{\tau}_{iij} + \hat{\tau}_{jji}) / 2\,.$$

Only the off-diagonal loss and the diagonal-plane loss are dependent upon the dyadic distances, and therefore the summation in (7) only needs to involve triplets in the list L, as indicated. The quantities $\{\hat{\tau}_{ijk}\}$, $\{\hat{\tau}_{iij}\}$, and $\{\hat{\delta}_{ij}\}$ are the closest approximations of the relevant parts of $\{\delta_{ijk}\}$ that satify three-way symmetry, diagonal-plane equality and two-way symmetry, respectively, in the least squares sense. Due to least squares, corresponding pairs of quantities in the decomposition are orthogonal: the off-diagonal residuals $(\delta_{\pi(i)\pi(j)\pi(k)} - \hat{\tau}_{ijk})$ with $(\hat{\tau}_{ijk} - d_{ij} - d_{ik} - d_{jk})$, the diagonal-plane residuals $(\delta_{i\kappa(i)\kappa(j)} - \hat{\tau}_{iij})$ with the asymmetric averages $\hat{\tau}_{iij}$, and the skewsymmetric residuals ($\hat{\tau}_{iij} - \hat{\tau}_{jji}$) with the symmetric residuals ($\hat{\delta}_{ij} - 2d_{ij}$).

The minimization of (7) is a quadratic optimization problem in $\{d_{ij}\}$ with inequality constraints, which can be shown to have a unique solution, and which can be solved with Uzawa's method (see Ciarlet, 1989). The constraint set is an intersection of convex cones, and therefore Dykstra's (1983) model algorithm applies, too.

Minkowski-p model

Let Δ be a dyadic dissimilarity on O and let T be a mapping defined by, for all $p \geq 1$,

$$\tau^{p}_{ijk} = \delta^{p}_{ij} + \delta^{p}_{ik} + \delta^{p}_{jk} \ . \tag{8}$$

This class of triadic dissimilarity functions generalizes the perimeter model, which has $p = 1$, and is called the *Minkowski-p dissimilarity*, or *M_p dissimilarity* for short.

Theorem 1. *If Δ is a dyadic distance D, then the M_p dissimilarity is a triadic distance function, called the M_p distance.*

<u>*Proof*</u>. From the definition of the M_p distance it is evident that it satisfies three-way symmetry (1*b*), diagonal-plane equality (1*d*), and is positive-definite (1*a*, 1*c*, and 3*a*). To show that the tetrahedral inequality (3*b*) holds, we first note that, adding the base distance to both sides of the triangle equality, we have

$$2\,d_{ij} \leq d_{ij} + d_{il} + d_{jl}\ ,$$
$$2\,d_{ik} \leq d_{ik} + d_{il} + d_{kl}\ ,$$
$$2\,d_{jk} \leq d_{jk} + d_{jl} + d_{kl}\ .$$

Raising both sides of each of these inequalities to the pth power and adding, we obtain

$$2^{p}\,(d^{p}_{ij} + d^{p}_{ik} + d^{p}_{jk}) \leq (d_{ij} + d_{il} + d_{jl})^{p} + (d_{ik} + d_{il} + d_{kl})^{p} + (d_{jk} + d_{jl} + d_{kl})^{p}\ ,$$

which implies

$$2\,(d_{ij}^{\,p} + d_{ik}^{\,p} + d_{jk}^{\,p})^{1/p} \le \left[(d_{ij} + d_{il} + d_{jl})^p + (d_{ik} + d_{il} + d_{kl})^p + (d_{jk} + d_{jl} + d_{kl})^p\right]^{1/p} . \qquad (9)$$

The left-hand side of inequality (9) equals two times the M_p distance between *i, j*, and *k*, and the right-hand side is always smaller than the sum of the other three triadic distances that form the right-hand side of the tetrahedral inequality. That the latter statement is correct follows from Minkowski's inequality, here extended to three terms, which asserts that

$$[\,\Sigma_k\,(a_k + b_k + c_k)^p]^{1/p} \le [\,\Sigma_k\, a_k^p]^{1/p} + [\,\Sigma_k\, b_k^p]^{1/p} + [\,\Sigma_k\, c_k^p]^{1/p} .$$

With the substitution

$a_1 = d_{ij}$,	$a_2 = d_{il}$,	$a_3 = d_{jl}$
$b_1 = d_{il}$	$b_2 = d_{ik}$	$b_3 = d_{kl}$
$c_1 = d_{jl}$	$c_2 = d_{kl}$	$c_3 = d_{jk}$

inequality (9) and Minkowski's inequality can be combined, showing that the M_p distance satisfies the tetrahedral inequality, which completes the proof. ❐

Apart from the perimeter model for $p = 1$, other special cases of the M_p distance of particular interest are the *generalized Euclidean model* for $p = 2$ and with D Euclidean, and the *maximum model* for $p = \infty$, where the latter becomes equivalent to $\tau_{ijk} = \max\,(d_{ij}, d_{ik}, d_{jk})$. Note that the triadic dissimilarity $\tau_{ijk} = \min\,(d_{ij}, d_{ik}, d_{jk})$ is *not* an M_p distance, nor even a triadic distance. If D is Euclidean, all M_p distance models are invariant under rotation, translation and reflection, because in that case they are all built up from invariant dyadic parts; so this invariance property is not restricted to the generalized Euclidean model, as suggested by Cox *et al.* (1991). Hayashi (1989) considered what he calls a distance measure H defined by

$$H(i, j, k) = d_{ij}^2 + d_{ik}^2 + d_{jk}^2 ,$$

with D Euclidean. Thus H is the square of a generalized Euclidean triadic distance. Hayashi dismissed H with the argument that it does not model the three-way information in the data beyond the two-way marginals. It is not hard to show by counter-example that H also does not satisfy the present definition of triadic distance. In fact, it is proportional to the *inertia* of triads of points, defined as the sum of squared distances towards their center of gravity.

Daws' model: degree of triadic distinguishability

As mentioned in the Introduction, Daws (1993) studied dyadic and triadic similarities in the context of the free sorting method of data collection, where m individuals classify n objects into any number of mutually exclusive classes; i.e., the data consists of m partitions. We can say that two objects in different classes are distinguishable, while two objects in the same class are indistinguishable. Let ρ_{ij} be the number of individuals that have classified object i and object j into the same subset of their partition. Corresponding to ρ_{ij}, the degree of *dyadic distinguishability* Δ is defined as $\delta_{ij} = m - \rho_{ij}$.

The following example, due to Daws, shows how the reduction from a distribution over all subset patterns to a dyadic (dis)similarity implies loss of information about the way in which the individuals have classified the objects. Suppose that free sorting by two groups of subjects yields the two distributions given in Table 1. Here (123) – (4) indicates that objects 1, 2 and 3 have been

--- Insert Table 1 about here ---

put together, but 4 has been left apart, and so on. It is easily verified that for both groups we obtain the dyadic similarities $\rho_{12} = 6$, $\rho_{13} = 6$, $\rho_{23} = 7$, $\rho_{14} = 0$, $\rho_{24} = 4$, $\rho_{34} = 3$. So, if one uses only dyadic information, the two groups cannot be recognized, because they do not distinguish pairs differently. Now, let ρ_{ijk} denote the number of individuals that have classified objects i , j, and k into the same subset of their partition, and define the triadic dissimilarity $\tau_{ijk} = m - \rho_{ijk}$. In the example, we obtain

for *Group* 1: $\rho_{123} = 5$, $\rho_{124} = 0$, $\rho_{134} = 0$, $\rho_{234} = 1$,

for *Group* 2: $\rho_{123} = 1$, $\rho_{124} = 0$, $\rho_{134} = 0$, $\rho_{234} = 2$.

Use of the triadic similarities clearly leads to the conclusion that two groups have classified the objects in a different fashion.

The *degree of triadic distinguishability* τ_{ijk} is an empirical measure of the difference between the three objects i, j, and k. Remarkably enough, it *also* satisfies the model properties of a triadic distance model.

Theorem 2. *Let ρ_{ijk} be the number of times that the triplet (i, j, k) is contained in one of the subsets generated by m partitions, where a triplet of the form (i, i, j) is identified with the pair (i, j), and a triplet of the form (i, i, i) is identified with the singleton (i). Then the mapping T defined as $t_{ijk} = m - \rho_{ijk}$ is a triadic distance function.*

Proof. Since the elements of the subsets generated by m partitions are unordered, it is evident that ρ_{ijk}, and hence t_{ijk}, satisfies three-way symmetry (1*b*). The diagonal plane equality (1*d*) holds, because $(i, i, j) = (i, j, i) = (i, j) = (i, j, j)$. The mapping T is positive-definite (1*a*, 1*c*, and 3*a*), because $m \geq \rho_{ijk}$ and $\rho_{iii} = \rho_i = m$. The case that $\rho_{ijk} = m$ with $i \neq j \neq k$ can be excluded by relabelling and reducing the problem from n to $n - 2$ objects. Showing that the tetrahedral inequality (3*b*) holds is equivalent to showing that

$$m + 2\,\rho_{ijk} \geq \rho_{ikl} + \rho_{jkl} + \rho_{ijl}\,. \tag{10}$$

Subsets containing the triplet (i, j, k) must consist of subsets including l, and subsets excluding l, so that, if ρ_{ijkl} denotes the count of quadruples and $\rho_{ijk/l}$ the count of triples excluding l,

$$\rho_{ijk} = \rho_{ijkl} + \rho_{ijk/l}\,,$$

and, in a similar fashion,

$$\rho_{ikl} = \rho_{ijkl} + \rho_{ikl/j}\,,$$

$$\rho_{jkl} = \rho_{ijkl} + \rho_{jkl/i}\,,$$

$$\rho_{ijl} = \rho_{ijkl} + \rho_{ijl/k}\,.$$

Inequality (10) thus becomes

$$m + 2\,\rho_{ijk/l} \geq \rho_{ijkl} + \rho_{ikl/j} + \rho_{jkl/i} + \rho_{ijl/k}\,.$$

Now, the actual number of partitions can be counted by enumerating the complete set of partitions (or subset patterns), as follows

$$m = \rho_{ijkl} + \rho_{ijk/l} + \rho_{ikl/j} + \rho_{jkl/i} + \rho_{ijl/k} + \rho_{ij/kl} + \dots\,,$$

where the summation continues over all other subset patterns listed in Table 1. Adding 2 $\rho_{ijk/l}$ to both sides of the last equation, and dropping the relevant non-negative terms at the right-hand side yields the result. ❐

Euclidean Representation of Triadic Dissimilarities

In a broad sense of the term, multidimensional scaling is concerned with the problem of finding an embedding or representation of the set O as n points in some prechosen metric space, in such a way that the associated distances D approximate the given dissimilarities Δ, where the quality of the approximation is measured by a badness-of-fit function (De Leeuw and Heiser, 1982). The representation is a Euclidean space if the chosen distance is Euclidean, a hierarchical tree if it is ultrametric, or an additive tree if it is quadrangular (an additive tree distance). Since in general the dissimilarities cannot be expected to be exactly representable by any of these type of distances, approximate representations are obtained by minimizing the badness-of-fit function.

In this section we will propose multidimensional scaling methods for triadic dissimilarity data. We will restrict ourselves to Euclidean representations, but ultrametric or additive trees could be fitted as well. Given a triadic dissimilarity T defined on O, we seek to represent the elements of O by n points in a Euclidean space of given dimensionality r, with coordinates collected in the $n \times r$ matrix $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_n)'$. The ordinary (dyadic) Euclidean distance is defined as

$$d^2_{ij}(\mathbf{X}) = \mathrm{tr}\,\mathbf{X}'\mathbf{E}_{ij}\mathbf{X}\,, \qquad (11)$$

with the matrix $\mathbf{E}_{ij} = (\mathbf{e}_i - \mathbf{e}_j)(\mathbf{e}_i - \mathbf{e}_j)'$, in which $\mathbf{e}_i$ is the ith column of the identity matrix. Since $\mathbf{e}'\mathbf{E}_{ij} = \mathbf{0}$, with $\mathbf{e}$ the unit element vector and $\mathbf{0}$ the vector of zeroes, we may assume without loss of generality that the matrix $\mathbf{X}$ is centered. A triadic distance is called Euclidean if it can be written as a function of $d_{ij}(\mathbf{X})$, $d_{ik}(\mathbf{X})$, and $d_{jk}(\mathbf{X})$; this is expressed in the notation by writing it as $t_{ijk}(\mathbf{X})$. Least squares triadic multidimensional scaling finds a representation $\mathbf{X}$ by minimizing the (weighted) sum of squared deviations of $t_{ijk}(\mathbf{X})$ with respect to τ_{ijk}.

Weights w_{ijk} will be introduced in the badness-of-fit function for greater generality. They are assumed to be given in advance, and can be used to accomodate missing data (by setting $w_{ijk} = 0$ if τ_{ijk} is missing), to control the influence of the residuals as a function of the estimated standard error of each dissimilarity, or to mimic the behavior of other badness-of-fit functions (Heiser, 1988). It is also assumed that the weights have all the properties of triadic dissimilarities, that is, non-negativity (1a), three-way symmetry (1b), minimality (1c), and diagonal plane equality (1d).

Approximation with the Euclidean perimeter model

In case of the perimeter model, we already studied the optimization of $\sigma_A^2(D)$ defined in (7), i.e., finding the least squares set of perimeter distances with respect to given τ_{ijk} over the whole set of dyadic distances. The Euclidean perimeter model poses the restriction $t_{ijk}(\mathbf{X}) = d_{ij}(\mathbf{X}) + d_{ik}(\mathbf{X}) + d_{jk}(\mathbf{X})$, so – including the weights – the problem becomes one of minimizing

$$\sigma_B^2(\mathbf{X}) = \sum_{(i,j,k)\in L} w_{ijk}\,(\tau_{ijk} - d_{ij}(\mathbf{X}) - d_{ik}(\mathbf{X}) - d_{jk}(\mathbf{X}))^2. \tag{12}$$

As before, the minimization only needs to regard the off-diagonal loss and the diagonal-plane loss, involving the residuals listed in the list L, containing off-diagonal triplets (i, j, k) with $i < j < k$ and diagonal-plane triplets (i, i, j) with $i < j$, since Proposition 3 applies regardless of the Euclidean restriction.

Developing expression (12), we obtain

$$\sigma_B^2(\mathbf{X}) = SSQ_\tau + \alpha(\mathbf{X}) + \gamma(\mathbf{X}) - 2\,\beta(\mathbf{X}) \,,$$

where

$$SSQ_\tau = \textstyle\sum_{i<j<k} w_{ijk}\, \tau^2_{ijk} + \sum_{i<j} w_{iij}\, \tau^2_{iij}\ ,$$
$$\alpha(\mathbf{X}) = \textstyle\sum_{i<j} w_{ij.} d^2_{ij}(\mathbf{X})\ ,$$
$$\beta(\mathbf{X}) = \textstyle\sum_{i<j} \delta_{ij.} d_{ij}(\mathbf{X})\ ,$$
$$\gamma(\mathbf{X}) = \textstyle\sum_{i<j} \sum_k w_{ijk}\, d_{ij}(\mathbf{X})\, [\, d_{ik}(\mathbf{X}) + d_{jk}(\mathbf{X})\,]\ ,$$

with the dyadic quantities $w_{ij.}$ and $\delta_{ij.}$ defined as

$$w_{ij.} = \textstyle\sum_k w_{ijk}\ ,$$
$$\delta_{ij.} = \textstyle\sum_k w_{ijk}\, \tau_{ijk}\ .$$

The components $\alpha(\mathbf{X})$, $\beta(\mathbf{X})$, and $\gamma(\mathbf{X})$ are all convex functions of **X**, since $d_{ij}(\mathbf{X})$ is convex in **X**, and non-negative linear combinations of convex functions again satisfy convexity. Thus $\sigma^2_B(\cdot)$ is the difference between two convex functions, and can be minimized by the iterative majorization (IM) approach. This computational strategy was developed in the ordinary multidimensional scaling context by De Leeuw and Heiser (1980) under the name SMACOF, and since then it has been successfully applied in several other approximation and equilibrium problems (Heiser, 1995). To characterize the stationary points of $\sigma^2_B(\cdot)$, and to show how IM can be applied to find them, it is convenient to express $\sigma^2_B(\cdot)$ in matrix form. The following lemma's are stated without proof.

Lemma 3. *Let* **V** *be defined as the order-n symmetric matrix with elements*

$$v_{ij} = -\, w_{ij.} \qquad \text{if} \quad i \neq j,$$
$$v_{ii} = \textstyle\sum_{l\neq i} w_{il.}\ .$$

Then $\alpha(\mathbf{X}) = \mathrm{tr}\, \mathbf{X'VX}$ *and the matrix of partial derivatives is* $\nabla\alpha(\mathbf{X}) = 2\, \mathbf{VX}$.

Lemma 4. *Let* $\mathbf{B} = B(\mathbf{X})$ *be the order-n symmetric matrix-valued function defined as*

$$b_{ij}(\mathbf{X}) = 0 \qquad \text{if} \quad i \neq j \text{ and } d_{ij}(\mathbf{X}) = 0$$
$$b_{ij}(\mathbf{X}) = -\, \delta_{ij.} / d_{ij}(\mathbf{X}) \qquad \text{if} \quad i \neq j \text{ and } d_{ij}(\mathbf{X}) \neq 0$$
$$b_{ii}(\mathbf{X}) = -\textstyle\sum_{l\neq i} b_{il}(\mathbf{X})\ .$$

Then $\beta(\mathbf{X}) = \text{tr}\ \mathbf{X}'B(\mathbf{X})\mathbf{X}$ *and the matrix of partial derivatives is* $\nabla\beta(\mathbf{X}) = B(\mathbf{X})\mathbf{X}$.

Lemma 5. *Let* $\mathbf{C} = C(\mathbf{X})$ *be the order-n symmetric matrix-valued function defined as*

$$c_{ij}(\mathbf{X}) = 0 \qquad \text{if} \quad i \neq j \text{ and } d_{ij}(\mathbf{X}) = 0$$

$$c_{ij}(\mathbf{X}) = -\textstyle\sum_k w_{ijk}\,[\,d_{ik}(\mathbf{X}) + d_{jk}(\mathbf{X})\,]\,/\,d_{ij}(\mathbf{X}) \qquad \text{if} \quad i \neq j \text{ and } d_{ij}(\mathbf{X}) \neq 0$$

$$c_{ii}(\mathbf{X}) = -\textstyle\sum_{l \neq i} c_{il}(\mathbf{X})\,.$$

Then $\gamma(\mathbf{X}) = \text{tr}\ \mathbf{X}'C(\mathbf{X})\mathbf{X}$ *and the matrix of partial derivatives is* $\nabla\gamma(\mathbf{X}) = 2\ C(\mathbf{X})\mathbf{X}$.

Using Lemma's 3-5, and defining the matrix $A(\mathbf{X}) = \mathbf{V} + C(\mathbf{X})$, we can re-express $\sigma_B^2(\cdot)$ as

$$\sigma_B^2(\mathbf{X}) = SSQ_\tau + \text{tr}\ \mathbf{X}'A(\mathbf{X})\mathbf{X} - 2\,\text{tr}\ \mathbf{X}'B(\mathbf{X})\mathbf{X}\ . \tag{13}$$

From setting the partial derivatives equal to zero it follows that a stationary point $\hat{\mathbf{X}}$ must satisfy

$$A(\hat{\mathbf{X}})\hat{\mathbf{X}} = B(\hat{\mathbf{X}})\hat{\mathbf{X}}\ . \tag{14}$$

The stationary equation in (14) implies that $\text{tr}\ \hat{\mathbf{X}}'A(\hat{\mathbf{X}})\hat{\mathbf{X}} = \text{tr}\ \hat{\mathbf{X}}'B(\hat{\mathbf{X}})\hat{\mathbf{X}}$, and hence that $\sigma_B^2(\hat{\mathbf{X}}) = SSQ_\tau - \text{tr}\ \hat{\mathbf{X}}'A(\hat{\mathbf{X}})\hat{\mathbf{X}}$. Recalling that $\text{tr}\ \mathbf{X}'A(\mathbf{X})\mathbf{X} = \alpha(\mathbf{X}) + \gamma(\mathbf{X})$ contains the two parts of the total sum of squares of the Euclidean triadic distances $t_{ijk}(\mathbf{X})$, we arrive at the result stated in the next proposition.

Proposition 4. *Given a set of triadic dissimilarities, suppose that the configuration* $\hat{\mathbf{X}}$ *satisfies the nonlinear equation* $\mathbf{V}\hat{\mathbf{X}} + C(\hat{\mathbf{X}})\hat{\mathbf{X}} = B(\hat{\mathbf{X}})\hat{\mathbf{X}}$ *for the Euclidean perimeter model, where the matrices* $\mathbf{V}$, $C(\cdot)$, *and* $B(\cdot)$ *are as defined in Lemma's 3, 5, and 4, respectively. Then the total weighted sum of squares of the triadic dissimilarities* SSQ_τ *can be decomposed as*

$$SSQ_\tau = \textstyle\sum_{(i,j,k)\in L} w_{ijk}\,\tau_{ijk}^2 = \sigma_B^2(\hat{\mathbf{X}}) + \sum_{(i,j,k)\in L}\ w_{ijk}\,t_{ijk}^2(\hat{\mathbf{X}})\ .$$

The decomposition in Proposition 4 shows a basic optimality property of weighted least squares approximation with the Euclidean perimeter model, and allows an evaluation of the stationary point, by division of both terms with SSQ_τ, in terms of a *relative loss* and a percentage of *Dispersion-Accounted-For* (%DAF). The latter quantity is analogous to the diagnostic *Variance-*

Accounted-For in regression analysis. Thus, the relative loss is $100 \times \sigma_B^2(\hat{\mathbf{X}}) \,/\, SSQ_\tau$, and the %DAF is $100 \times \mathrm{tr}\, \hat{\mathbf{X}}'A(\hat{\mathbf{X}})\hat{\mathbf{X}} \,/\, SSQ_\tau$.

To arrive at an iterative algorithm for solving the nonlinear equation (14), it is useful to know some properties of the matrix $A(\mathbf{X})$. It turns out that, under mild conditions, this matrix is positive semi-definite, has rank $n - 1$, and that the eigenvector corresponding to the vanishing eigenvalue is the unit vector $\mathbf{e}$ (see Bennani, 1993). From these properties it follows that $A(\mathbf{X})$ has generalized inverse $A^+(\mathbf{X}) = [A(\mathbf{X}) + \mathbf{ee}'/\mathbf{e}'\mathbf{e}]^{-1} - \mathbf{ee}'/\mathbf{e}'\mathbf{e}$, which can be easily shown to satisfy the Moore-Penrose conditions. The basic IM algorithm for this case is to build up a sequence of configurations $\mathbf{X}_0, \mathbf{X}_1, \dots, \mathbf{X}_s$ by applying the mapping

$$\mathbf{X}_{s+1} = A^+(\mathbf{X}_s)B(\mathbf{X}_s)\mathbf{X}_s = G_B(\mathbf{X}_s)\ . \tag{15}$$

A fixed point $\mathbf{X}_* = G_B(\mathbf{X}_*)$ of the mapping $G_B(\cdot)$ defined in (15) will solve (14), and hence it will be a stationary point of $\sigma_B^2(\cdot)$. As in all fixed-point algorithms, the initial configuration $\mathbf{X}_0$ could be selected arbitrarily, but it is preferable to start the process close to a local minimum, e.g., by ordinary MDS of the diagonal plane dissimilarities τ_{iij}. Not only does a fixed point exist, but IM guarantees that the sequence defined by $G_B(\cdot)$ is monotonically convergent.

Theorem 3. *The sequence* $\{\sigma_B^2(\mathbf{X}_s) \mid s = 0, 1, 2, \dots\}$ *is monotonically decreasing and converges.*

Proof. The proof of this theorem is based upon two inequalities:

$$\beta(\mathbf{X}) \geq \mathrm{tr}\, \mathbf{X}'B(\mathbf{Y})\mathbf{Y} \qquad \forall\ \mathbf{X}, \mathbf{Y}; \tag{16}$$

$$\gamma(\mathbf{X}) \leq \mathrm{tr}\, \mathbf{X}'C(\mathbf{Y})\mathbf{X} \qquad \forall\ \mathbf{X}, \mathbf{Y}. \tag{17}$$

Suppose that these inequalities do hold (this will be verified in a short while), and consider the family of functions $\mu_B(\mathbf{X} \mid \mathbf{Y})$, indexed by $\mathbf{Y}$ and defined as

$$\mu_B(\mathbf{X} \mid \mathbf{Y}) = SSQ_\tau + \mathrm{tr}\, \mathbf{X}'\mathbf{V}\mathbf{X} + \mathrm{tr}\, \mathbf{X}'C(\mathbf{Y})\mathbf{X} - 2\, \mathrm{tr}\, \mathbf{X}'B(\mathbf{Y})\mathbf{Y}\ , \tag{18}$$

which is quadratic in $\mathbf{X}$. It follows from (16) and (17) that each member of the family (18) satisfies, $\forall\ \mathbf{X}$,

$$\sigma_B^2(\mathbf{X}) \leq \mu_B(\mathbf{X} \mid \mathbf{Y}),$$
$$\sigma_B^2(\mathbf{Y}) = \mu_B(\mathbf{Y} \mid \mathbf{Y}),$$
$$\nabla\sigma_B^2(\mathbf{Y}) = \nabla\mu_B(\mathbf{Y} \mid \mathbf{Y}),$$

that is, $\mu_B(\cdot \mid \mathbf{Y})$ *majorizes* the loss function, coincides with it at $\mathbf{Y}$, and has the same partial derivatives at the point of coincidence. From setting $\mathbf{Y} = \mathbf{X}_s$ it appears that the mapping $G_B(\cdot)$ in (15) defines $\mathbf{X}_{s+1}$ as the argument that minimizes the majorizing function $\mu_B(\cdot \mid \mathbf{X}_s)$, and therefore

$$\sigma_B^2(\mathbf{X}_{s+1}) \leq \mu_B(\mathbf{X}_{s+1} \mid \mathbf{X}_s) = \min\ \mu_B(\mathbf{X} \mid \mathbf{X}_s). \tag{19}$$

But since $\mathbf{X}_{s+1}$ is a minimizer, we also have

$$\mu_B(\mathbf{X}_{s+1} \mid \mathbf{X}_s) \leq \mu_B(\mathbf{X}_s \mid \mathbf{X}_s) = \sigma_B^2(\mathbf{X}_s)\ , \tag{20}$$

with equality if and only if $\mathbf{X}_{s+1} = \mathbf{X}_s$, in which case we may conclude that we have found a configuration satisfying $\nabla\sigma_B^2(\mathbf{X}_s) = \nabla\mu_B(\mathbf{X}_s \mid \mathbf{X}_s) = 0$, that is, a stationary point, and we stop. It is clear from combining (19) and (20) that the sequence is monotonically decreasing; because it is also bounded below by zero, it must converge. To finish the proof, we have to verify (16) and (17); as shown in De Leeuw and Heiser (1977), the first of these follows from the Cauchy-Schwarz inequality written in the form

$$d_{ij}(\mathbf{X})d_{ij}(\mathbf{Y}) \geq \mathrm{tr}\ \mathbf{X}'\mathbf{E}_{ij}\mathbf{Y}\ .$$

Both sides of this inequality can be divided by $d_{ij}(\mathbf{Y})$ if it is non-zero, while if $d_{ij}(\mathbf{Y}) = 0$, the inequality is replaced by $d_{ij}(\mathbf{X}) \geq 0$; then the result is obtained by multiplying both sides with $\delta_{ij.}$, and using the definition of $\beta(\mathbf{X})$ in Lemma 4. To derive expression (17) we expand, following an idea first used in Heiser (1987, 1991) and extended in Groenen and Heiser (1991), the inequality

$$[d_{ij}(\mathbf{X})d_{ik}(\mathbf{Y}) - d_{ij}(\mathbf{Y})d_{ik}(\mathbf{X})]^2 \geq 0\ ,$$

which is an equality if $\mathbf{X} = \mathbf{Y}$. Using (11) and assuming that $d_{ij}(\mathbf{Y}) \neq 0$ and $d_{ik}(\mathbf{Y}) \neq 0$, we have

$$d_{ij}(\mathbf{X})d_{ik}(\mathbf{X}) \leq (1/2) \text{ tr } \mathbf{X}'[\{d_{ik}(\mathbf{Y})/d_{ij}(\mathbf{Y})\}\mathbf{E}_{ij} + \{d_{ij}(\mathbf{Y})/d_{ik}(\mathbf{Y})\}\mathbf{E}_{ik}]\mathbf{X} .$$

(whenever $d_{ij}(\mathbf{Y}) < \varepsilon$, it is replaced by ε). Addition of a similar expression for $d_{ij}(\mathbf{X})d_{jk}(\mathbf{X})$ yields

$$\begin{aligned} d_{ij}(\mathbf{X})[d_{ik}(\mathbf{X}) + d_{jk}(\mathbf{X})] \leq\ & (1/2) \text{ tr } \mathbf{X}'[\{[d_{ik}(\mathbf{Y}) + d_{jk}(\mathbf{Y})]/d_{ij}(\mathbf{Y})\}\mathbf{E}_{ij}]\mathbf{X} \\ +\ & (1/2) \text{ tr } \mathbf{X}'[\{d_{ij}(\mathbf{Y})/d_{ik}(\mathbf{Y})\}\mathbf{E}_{ik} + \{d_{ij}(\mathbf{Y})/d_{jk}(\mathbf{Y})\}\mathbf{E}_{jk}]\mathbf{X} . \end{aligned} \tag{21}$$

After multiplying (21) with w_{ijk}, and summing over $i < j$ and all k, we obtain two equal terms on the right-hand side of the summation, because, due to the symmetry of all quantities involved,

$$\begin{aligned} &\Sigma_{i<j} \Sigma_k w_{ijk} \{[d_{ik}(\mathbf{Y}) + d_{jk}(\mathbf{Y})]/d_{ij}(\mathbf{Y})\}\mathbf{E}_{ij} \\ &= \Sigma_{i<j} \Sigma_k w_{ijk} \{d_{ij}(\mathbf{Y})/d_{ik}(\mathbf{Y})\}\mathbf{E}_{ik} + \Sigma_{i<j} \Sigma_k w_{ijk} \{d_{ij}(\mathbf{Y})/d_{jk}(\mathbf{Y})\}\mathbf{E}_{jk} . \end{aligned}$$

Thus both terms are equal to (1/2) tr $\mathbf{X}'C(\mathbf{Y})\mathbf{X}$, which establishes the correctness of (17). ❐

Theorem 3 shows an important characteristic of the IM algorithm based on (15), but it should be noted that it does not guarantee convergence of the sequence of configurations itself (we would need to establish properties of $\|\mathbf{X}_{s+1} - \mathbf{X}_s\|$ for this purpose, which would lead us outside the scope of this paper). The major difference with the standard SMACOF algorithm is the fact that the matrix $A^+(\mathbf{X}_s)$ is variable, rather than fixed, during the iterations, as it depends on the previous update $\mathbf{X}_s$. Thus $A(\mathbf{X}_s)$ must be inverted each iteration, which burdens the computations; note, however, that we have a three-way problem that can be solved by a two-way process.

Approximation with the M_2 model

The generalized Euclidean model, obtained by setting $p = 2$ in the Minkowski-p model (8), poses the restriction $t_{ijk}(\mathbf{X}) = [d^2_{ij}(\mathbf{X}) + d^2_{ik}(\mathbf{X}) + d^2_{jk}(\mathbf{X})]^{1/2}$. So, due to the square root, and in contrast to the perimeter model, the M_2 model is not additive in the dyadic distances (or in their squares). Defining $\mathbf{E}_{ijk} = \mathbf{E}_{ij} + \mathbf{E}_{ik} + \mathbf{E}_{jk}$, we can write the square of the M_2 distance as

$$t^2_{ijk}(\mathbf{X}) = \text{tr } \mathbf{X}'\mathbf{E}_{ij}\mathbf{X} + \text{tr } \mathbf{X}'\mathbf{E}_{ik}\mathbf{X} + \text{tr } \mathbf{X}'\mathbf{E}_{jk}\mathbf{X} = \text{tr } \mathbf{X}'\mathbf{E}_{ijk}\mathbf{X} . \tag{22}$$

The approximation problem becomes one of minimizing

$$\sigma_C^2(\mathbf{X}) = \Sigma_{(i,j,k)\in L}\ w_{ijk}\,(\tau_{ijk} - [\mathrm{tr}\ \mathbf{X}'\mathbf{E}_{ijk}\mathbf{X}]^{1/2})^2 \tag{23}$$

over $\mathbf{X}$, with $\mathbf{E}_{ijk}$ having the required additive form, and under the same symmetry assumptions as before (although Proposition 3 was given for the additive case only, a similar decomposition holds for non-additive triadic distances). Developing expression (23), we obtain

$$\sigma_C^2(\mathbf{X}) = SSQ_\tau + \alpha(\mathbf{X}) - 2\,\phi(\mathbf{X})\ ,$$

where SSQ_τ and $\alpha(\mathbf{X})$ are as defined earlier, and the cross-product term $\phi(\mathbf{X})$ becomes

$$\phi(\mathbf{X}) \ = \{(1/3)\ \Sigma_{i<j}\ \Sigma_{k\neq i,j}\ w_{ijk}\tau_{ijk}t_{ijk}(\mathbf{X})\} + \Sigma_{i<j}\ w_{iij}\tau_{iij}t_{iij}(\mathbf{X}).$$

Since $t_{ijk}(\mathbf{X})$, and hence $\phi(\mathbf{X})$, is convex in $\mathbf{X}$, we still have a difference between two convex functions. The matrix form of $\phi(\cdot)$ is again stated in a lemma.

Lemma 6. *Let the quantities* $h_{ijk}(\mathbf{X})$ *be defined as*

$$h_{ijk}(\mathbf{X}) = 0 \qquad \textit{if}\ \ t_{ijk}(\mathbf{X}) = 0,$$
$$h_{ijk}(\mathbf{X}) = w_{ijk}\tau_{ijk}\,/\,t_{ijk}(\mathbf{X}) \qquad \textit{if}\ \ t_{ijk}(\mathbf{X}) \neq 0,$$

and let $\mathbf{F} = F(\mathbf{X})$ *be the order-n symmetric matrix-valued function defined as*

$$f_{ij}(\mathbf{X}) = -\,\Sigma_k\, h_{ijk}(\mathbf{X}) \qquad \textit{if}\ \ i \neq j,$$
$$f_{ii}(\mathbf{X}) = -\,\Sigma_{l\neq i}\, f_{il}(\mathbf{X})\ .$$

Then $\phi(\mathbf{X}) = \mathrm{tr}\ \mathbf{X}'F(\mathbf{X})\mathbf{X}$ *and the matrix of partial derivatives is* $\nabla F(\mathbf{X}) = F(\mathbf{X})\mathbf{X}$.

Proof. By the definition of $t_{ijk}(\mathbf{X})$ and $h_{ijk}(\mathbf{X})$, and using symmetry, we may rewrite $\phi(\mathbf{X})$ as

$$\phi(\mathbf{X}) = \Sigma_{i<j}\ \{(1/3)\ \Sigma_{k\neq i,j}\, h_{ijk}(\mathbf{X})\}\, t_{ijk}^2(\mathbf{X}) + \Sigma_{i<j}\ \{\Sigma_{k=i,j}\, h_{ijk}(\mathbf{X})\}\, d_{ij}^2(\mathbf{X}).$$

The matrix expression follows from inserting the additive decomposition (22), substitution of (11), and simplifying. At configurations $\mathbf{X}$ where $t_{ijk}(\mathbf{X}) = 0$ for some i,j and k, the partial derivatives are not defined, but the definition of $F(\mathbf{X})$ is such that $\nabla F(\mathbf{X})$ satisfies the definition of a subdifferential (Rockafellar, 1970). ❐

Thus the matrix expression for the loss function of the M_2 model becomes

$$\sigma_C^2(\mathbf{X}) = SSQ_\tau + \text{tr } \mathbf{X}'\mathbf{V}\mathbf{X} - 2 \text{ tr } \mathbf{X}'F(\mathbf{X})\mathbf{X} \;, \tag{24}$$

and the stationary equation for an M_2 representation is

$$\mathbf{V}\hat{\mathbf{X}} = F(\hat{\mathbf{X}})\hat{\mathbf{X}} \;. \tag{25}$$

Analogous to a similar implication for the perimeter model, the stationary configuration $\hat{\mathbf{X}}$ will satisfy tr $\hat{\mathbf{X}}'\mathbf{V}\hat{\mathbf{X}}$ = tr $\hat{\mathbf{X}}'F(\hat{\mathbf{X}})\hat{\mathbf{X}}$, and therefore a result similar to Proposition 4 can be obtained by merely adjusting its condition.

Proposition 5. *Given a set of triadic dissimilarities, suppose that the configuration* $\hat{\mathbf{X}}$ *satisfies the nonlinear equation* $\mathbf{V}\hat{\mathbf{X}} = F(\hat{\mathbf{X}})\hat{\mathbf{X}}$ *for the* M_2 *model, where the matrices* $\mathbf{V}$ *and* $F(\cdot)$ *are as defined in Lemma's 3 and 6, respectively. Then the total weighted sum of squares of the triadic dissimilarities* SSQ_τ *can be decomposed as*

$$SSQ_\tau = \textstyle\sum_{(i,j,k)\in L} w_{ijk}\, \tau_{ijk}^2 = \sigma_C^2(\hat{\mathbf{X}}) + \textstyle\sum_{(i,j,k)\in L} w_{ijk}\, t_{ijk}^2(\hat{\mathbf{X}}) \;.$$

The M_2 model may be fitted to a given set of triadic dissimilarities by an IM algorithm consisting of repeatedly applying the mapping

$$\mathbf{X}_{s+1} = \mathbf{V}^+F(\mathbf{X}_s)\mathbf{X}_s = G_C(\mathbf{X}_s) \;. \tag{26}$$

A fixed point $\mathbf{X}_* = G_C(\mathbf{X}_*)$ of the mapping $G_C(\cdot)$ defined in (26) will solve (25), and hence it will be a stationary point of $\sigma_C^2(\cdot)$. Comparing the mappings $G_B(\cdot)$ and $G_C(\cdot)$ for the Euclidean perimeter model and the M_2 model, respectively, we see two differences. First, instead of recalculating the generalized inverse $A^+(\mathbf{X}_s)$ in each iteration, we need to calculate $\mathbf{V}^+$ once (where $\mathbf{V}$ has the same formal matrix properties as $A(\mathbf{X})$, and hence the calculation of $\mathbf{V}^+$ is easy). Second, instead of premultiplying $\mathbf{X}_s$ in $G_B(\cdot)$ with a correction matrix $B(\mathbf{X}_s)$ having generic element $b_{ij}(\mathbf{X}_s) = -\{\sum_k w_{ijk}\tau_{ijk}\} / d_{ij}(\mathbf{X}_s)$, in $G_C(\cdot)$ we premultiply the previous configuration with a matrix $F(\mathbf{X}_s)$ having generic element $f_{ij}(\mathbf{X}_s) = -\sum_k \{w_{ijk}\tau_{ijk} / t_{ijk}(\mathbf{X}_s)\}$. Thus in $G_B(\cdot)$ the positions of points i and j are corrected on the basis of the disparity of the current dyadic distance $d_{ij}(\mathbf{X}_s)$ with the (weighted) total dissimilarity involving i and j, while in $G_C(\cdot)$ it is the (weighted) total disparity of the current triadic distance $t_{ijk}(\mathbf{X}_s)$ with each separate dissimilarity that determines the correction.

A convergence theorem for the $G_C(\cdot)$ mapping can be proven along the same lines as the proof of Theorem 3; the only adjustment needed would be the replacement of inequality (16) by

$$\phi(\mathbf{X}) \geq \text{tr } \mathbf{X}'F(\mathbf{Y})\mathbf{Y} \qquad \forall\ \mathbf{X}, \mathbf{Y}. \tag{27}$$

That this inequality holds can be derived from the Cauchy-Schwarz inequality written in the form

$$[\text{tr } \mathbf{X}'\mathbf{E}_{ijk}\mathbf{X}]^{1/2}\ [\text{tr } \mathbf{Y}'\mathbf{E}_{ijk}\mathbf{Y}]^{1/2} \geq \text{tr } \mathbf{X}'\mathbf{E}_{ijk}\mathbf{Y}\ , \tag{28}$$

multiplying both sides of (28) with $h_{ijk}(\mathbf{Y})$ defined in Lemma 6, summing over $i < j$ and k, and simplifying.

Triadic Distances Defined on Presence-Absence Variables

Before turning to some examples of fitting triadic distance models, we theoretically discuss a class of applications in which a three-way one-mode table is constructed from a two-mode table with presence-absence data (usually rectangular, of the form $O_1 \times O_2$, with O_1 a set of individuals and O_2 a set of variables). Our primary interest will be to characterize the entries of the newly constructed table as genuine triadic distances. In this situation, the raw observations on each individual i from $O_1 = \{a_1, a_2, \ldots, a_n\}$ consist of a list of m binary signals, indicating the presence or absence of each of m attributes from some collection $O_2 = \{b_1, b_2, \ldots, b_m\}$. The whole set of observations is collected in an $n \times m$ matrix $\mathbf{U}$, with elements u_{it} $(t = 1, \ldots, m)$, defined as $u_{it} = 1$ if attribute b_t is present in individual a_i, and $u_{it} = 0$ if it is absent.

A large number of (dis)similarity indices exists that aim to measure the resemblance (or the lack of it) between individuals (or between attributes). Detailed studies of the most important of such indices and their properties may be found in Fichet and Le Calvé (1984) or Gower and Legendre (1986). As convincingly argued by Cox et al. (1991), it may very well be that objects are equi-(dis)similar when compared two at a time, while they are clearly different when compared three at a time. The parallel phenomenon, possible irreducibility of multivariate association between categorical variables, i.e., the presence of interaction between attributes taken three at a time, is well-known in statistics ever since Bartlett (1935), but triadic resemblance lags behind. Our treatment will be such that all of these indices – classically defined on pairs of objects – can be considered as two-way restricted versions of more general indices defined on triples of objects.

Since, typically, the "present" category of an attribute is given a different role compared to the "absent" category, we need a somewhat more elaborate notation compared to what we used when discussing Daws' model, which also defined triadic similarity and distance on the basis of two-mode categorical observations, but which was fully symmetric in its treatment of the categories. Identifying O_1 with the set of analysis objects O, we will use, for three objects i, j, and k from O, in that order, the following count measures across all attributes from O_2:

n_{111}^{ijk} = number of positive matches between i, j, and k (count of attributes present in all three),
n_{000}^{ijk} = number of negative matches between i, j, and k (count of attributes absent in all three),
q_{ijk} = number of mismatches between i, j, and k (count of attributes different in one of three).

These counts can be expressed directly in terms of the observations **U** as follows:

$$n_{111}^{ijk} = \sum_t u_{it}u_{jt}u_{kt} ,$$
$$n_{000}^{ijk} = \sum_t (1 - u_{it})(1 - u_{jt})(1 - u_{kt}) ,$$
$$q_{ijk} = m - n_{111}^{ijk} - n_{000}^{ijk} .$$

By writing q_{ijk} this way, we avoid to express it as the total of six sums, which – in the current notation – would be denoted as n_{011}^{ijk}, n_{101}^{ijk}, n_{110}^{ijk}, n_{001}^{ijk}, n_{010}^{ijk}, and n_{100}^{ijk}. In a similar way we can express n^{ijkl}_{1110}, the number of attributes shared by *i, j*, and *k*, but not by *l*, and n^{ijkl}_{0001}, the number of attributes absent in both *i, j*, and *k*, but present in *l* as

$$n^{ijkl}_{1110} = \sum_t u_{it}u_{jt}u_{kt}(1 - u_{lt}) \,,$$
$$n^{ijkl}_{0001} = \sum_t (1 - u_{it})(1 - u_{jt})(1 - u_{kt})u_{lt} \,.$$

In this notation, we have the useful relationships

$$n^{ijk}_{111} = n^{ijkl}_{1111} + n^{ijkl}_{1110} \,, \tag{29}$$
$$n^{ijk}_{000} = n^{ijkl}_{0000} + n^{ijkl}_{0001} \,, \tag{30}$$
$$\rho_{ijk/l} = n^{ijkl}_{1110} + n^{ijkl}_{0001} \,, \tag{31}$$
$$n^{ijk}_{111} + q_{ijk} = m - n^{ijk}_{000} = m - n^{ijkl}_{0000} - n^{ijkl}_{0001} \,, \tag{32}$$

where $\rho_{ijk/l}$ is the subset pattern similarity measure used in the discussion of Daws' model, that is, the number of variables with *i, j* and *k* in one subset and *l* in another, and where (32) follows from the definition of q_{ijk} and (30). It is also important to observe that counts on the diagonal planes satisfy $\sum_t u_{it}u_{it}u_{jt} = \sum_t u_{it}u_{jt}$, that is, the marginals of a given type of count are contained in the higher-order tables. This relationship is shared by all other quantities defined above. With these notational preliminaries, we now turn to a number of specific (dis)similarity measures.

Triadic Hamming dissimilarity as a special case of Daws' model

The dyadic Hamming dissimilarity is defined as the total number of mismatches across all attributes. Generalized to the triadic case, we have

$$T_{Ham}(i, j, k) = q_{ijk} \,. \tag{33}$$

Normed by the number of variables, the corresponding similarity index $(m - q_{ijk}) / m$ would be the triadic Sokal-Michener similarity. It may not be obvious how to express q_{ijk} as an additive distance defined on the rows of **U**, as in the dyadic case. However, since the *m* binary attributes considered here are a special case of the *m* partitions considered in Daws' model, in which each partition is a division into two groups, the following proposition is an immediate corollary of Theorem 2.

Proposition 6. *The triadic Hamming dissimilarity is a triadic distance function.*

Another way of stating this result is that for two categories we have the relationship $T_{Daws}(i, j, k) = m - \rho_{ijk} = m - (n_{111}^{ijk} + n_{000}^{ijk}) = q_{ijk}$. For a proof of the stronger result that $T_{Ham}(\cdot)$ is in fact a demi-perimeter distance in terms of the dyadic Hamming distance, we refer to Bennani (1993).

Triadic Rogers-Tanimoto dissimilarity

The Rogers-Tanimoto similarity index is $m - q_{ij}$ relative to $m + q_{ij}$. Thus if there are no mismatches, it is equal to one, and it becomes equal to zero only if there are neither positive matches nor negative matches. One minus the Rogers-Tanimoto similarity equals

$$1 - (m - q_{ij}) \,/\, (m + q_{ij}) = 2q_{ij} \,/\, (m + q_{ij}) = q_{ij} \,/\, \{[(m - n_{11}^{ij}) + (m - n_{00}^{ij})]/2\} \;,$$

that is, it can be regarded as a normalized version of the Hamming dissimilarity, where the normalization is the average of $m - n_{11}^{ij}$, the number of attributes that do not match positively, and $m - n_{00}^{ij}$, the number of attributes that do not match negatively. For the triadic case we define

$$T_{R\text{-}T}(i, j, k) = 2q_{ijk} \,/\, (m + q_{ijk}) \;. \tag{34}$$

Since the Hamming dissimilarity is a triadic distance, application of Proposition 1 directly leads to the next result.

Theorem 4. *The triadic Rogers-Tanimoto dissimilarity is a triadic distance function.*

Triadic Jaccard dissimilarity

The Jaccard index normalizes the Hamming similarity or distance by the sum of the number of positive matches and the number of mismatches. Thus the triadic Jaccard dissimilarity is defined as

$$T_{Jac}(i, j, k) = q_{ijk} \,/\, (n_{111}^{ijk} + q_{ijk}) \;. \tag{35}$$

Proving that $T_{Jac}(\cdot)$ possesses all the properties of the triadic distance function is not as direct as in the previous cases.

Theorem 5. *The triadic Jaccard dissimilarity is a triadic distance function.*

Proof. It is evident that $T_{Jac}(\cdot)$ is non-negative and satisfies three-way symmetry. Since q_{ijk} is a triadic distance, we have $q_{ijk} = 0 \Leftrightarrow i = j = k$; hence $T_{Jac}(i, j, k) = 0 \Leftrightarrow i = j = k$. The diagonal plane equality (1*d*) holds, because both q_{ijk} and n_{111}^{ijk} have this property. To show that the tetrahedral inequality (3*b*) holds, we first note that it follows from the last part of the proof of Theorem 2 that we must have

$$m + 2\rho_{ijk/l} - \rho_{ijkl} - \rho_{ikl/j} - \rho_{jkl/i} - \rho_{ijl/k} \geq 3\,\rho_{ijk/l}\,.$$

Using the relationships $\rho_{ijk/l} = \rho_{ijk} - \rho_{ijkl}$ and $\rho_{ijk} = m - q_{ijk}$ on the left-hand side, and substituting (31) on the right-hand side, we obtain

$$q_{ikl} + q_{jkl} + q_{ijl} - 2q_{ijk} \geq 3n_{1110}^{ijkl} + 3n_{0001}^{ijkl}\,. \tag{36}$$

From the definition of $T_{Jac}(\cdot)$ we deduce the relationship $q_{ijk} = [n_{111}^{ijk} + q_{ijk}]\,\tau_{ijk}$, where for convenience $T_{Jac}(i, j, k)$ is denoted by τ_{ijk}. Substituting this equation into inequality (36), using (32), and simplifying, we find

$$(m - n_{0000}^{ijkl})\{\tau_{ikl} + \tau_{jkl} + \tau_{ijl} - 2\tau_{ijk}\} \geq \{3n_{1110}^{ijkl} + n_{0100}^{ijkl}\tau_{ikl} + n_{1000}^{ijkl}\tau_{jkl} + n_{0010}^{ijkl}\tau_{ijl}\} + n_{0001}^{ijkl}(3 - 2\tau_{ijk}).$$

Since $(m - n_{0000}^{ijkl}) \geq 0$, and $(3 - 2\tau_{ijk}) > 0$ because τ_{ijk} is bounded above by 1, we conclude that τ_{ijk} satisfies the tetrahedral inequality. ❐

Triadic Fichet-Gower dissimilarity

Let us now consider a family of indices that includes a (positive) parameter ϑ modifying the importance of the number of positive matches in the triadic Jaccard index:

$$T_{F\text{-}G}(i, j, k \mid \vartheta) = q_{ijk} \,/\, (\vartheta n_{111}^{ijk} + q_{ijk})\,. \tag{37}$$

Metric properties of the dyadic version of the family defined in (37) were discussed by Fichet

(1986) and Gower (1986); hence the name coined here. The Fichet-Gower dissimilarity generalizes the indices of Sokal & Sneath and Anderberg ($\vartheta = 1/2$), Jaccard ($\vartheta = 1$) and Czekanowski & Dice ($\vartheta = 2$). To prove under what condition $T_{F\text{-}G}(i, j, k \mid \vartheta)$ is a triadic distance it turns out to be useful to work with a reparametrization.

Theorem 6. *The triadic Fichet-Gower dissimilarity is a triadic distance function if and only if $\vartheta \leq 1$.*

Proof. For $\vartheta = 1$, the triadic Jaccard dissimilarity $T_{F\text{-}G}(i, j, k \mid \vartheta = 1)$ is obtained, which is a triadic distance by Theorem 5. For $0 < \vartheta < 1$, we reparametrize by choosing a strictly positive real number α so that $\vartheta = \alpha / (\alpha + 1)$. Upon substitution in (37),

$$\begin{aligned} T_{F\text{-}G}(i, j, k \mid \vartheta) &= (\alpha + 1)q_{ijk} / [\alpha n_{111}^{ijk} + (\alpha + 1)q_{ijk}] \\ &= (\alpha + 1)T_{F\text{-}G}(i, j, k \mid \vartheta = 1) / [\alpha + T_{F\text{-}G}(i, j, k \mid \vartheta = 1)] \,. \end{aligned}$$

Since $T_{F\text{-}G}(i, j, k \mid \vartheta = 1)$ is a triadic distance, the result follows from Proposition 1. ❐

It should be noted that, for $\vartheta > 1$, $T_{F\text{-}G}(i, j, k \mid \vartheta)$ cannot be a triadic distance, because under this condition the dyadic dissimilarities on the diagonal plane are not metric (Fichet, 1986).

Triadic Russel -Rao dissimilarity

The Russel-Rao index is based on the positive matches only, normed by the number of attributes; thus for $i \neq j \neq k$ the triadic version of this dissimilarity is

$$T_{R\text{-}R}(i, j, k) = (m - n_{111}^{ijk}) / m \;, \tag{38}$$

where it should be noted that the definition must include $T_{R\text{-}R}(i, j, k) = 0$ if $i = j = k$ (this proviso was not necessary in the previous cases, as these all have only q_{ijk} in the nominator).

Theorem 7. *The triadic Russel-Rao dissimilarity is a triadic distance function.*

Proof. It will only be proven that the tetrahedral inequality holds. Writing τ_{ijk} for $T_{R\text{-}R}(i, j, k)$, and using equation (29), we see that

$$m\ \{\tau_{ikl} + \tau_{jkl} + \tau_{ijl} - 2\tau_{ijk}\} = (m - n_{111}^{ikl}) + (m - n_{111}^{jkl}) + (m - n_{111}^{ijl}) - 2(m - n_{111}^{ijk})$$

$$= m - (n_{1111}^{ijkl} + n_{1011}^{ijkl} + n_{0111}^{ijkl} + n_{1101}^{ijkl}) + 2n_{1110}^{ijkl}\ .$$

The right-hand side is non-negative, because the term in brackets cannot be larger than m. ❐

This result concludes our discussion of the metric properties of triadic dissimilarities defined on presence-absence variables. Proposition 6 and Theorems 4-7 indicate, which indices are preferably used to model the triadic relationships between the row elements of a two-mode table.

Examples

We now turn to some examples of triadic model fitting. The first two examples concern the approximation of a one-mode triadic dissimilarity table with an M_2 distance model. In the third example, a triadic unfolding method is proposed; i.e., a method analyzing a three-mode table, in which the (dis)similarities are defined on $O_1 \times O_2 \times O_3$, the Cartesian product of three different sets of objects. It is shown how this analysis can be embedded within the algorithmic framework presented, and the approach is illustrated for the Euclidean perimeter model.

Comparison of M_2 model with Hayashi's surface area model

Hayashi (1972) studied an example of three-mode data indicating the unproductivity of teams of three individuals as a function of their composition (Table 2). Note that the values on the diagonal

--- Insert Table 2 about here ---

plane are not given. In view of the small number of objects (n = 6), this example serves only to illustrate a key difference between the M_2 model and Hayashi's model, in which each dissimilarity τ_{ijk} is represented as the *surface area* of the triangle formed by the points *i, j*, and *k*. It should be noted that this quantity does not satisfy the requirements of a triadic distance; a major drawback of the area model would seem to be that three collinear points get surface area zero even when they are arbitrarily far apart. Figure 1 gives the configuration as obtained by Hayashi, and Figure 2

--- Insert Figure 1 and Figure 2 about here ---

the configuration obtained from fitting the M_2 model by minimizing $\sigma_C^2(\mathbf{X})$ in (23) with **X** two-dimensional, which gives an optimal D.A.F. of 96.6%. The M_2 configuration in Figure 2 clearly falls apart into two groups, {1, 2, 3} and {4, 5, 6}, accounting quite well for the data in Table 2, which has $\tau_{123} = \tau_{456} = 1$ and the highest values for triads combining one object from the first group with object 6. Once we realize that the area model predicts small dissimilarity for points on a line, the two groups are apparent in Figure 1 as well, but in a less natural way.

Free sorting of kinship terms

Our next example concerns free sorting data collected by Rosenberg (1982)(Note). The stimuli in this experiment were 15 kinship terms (given in Table 3), representing the most common genetic

--- Insert Table 3 about here ---

relationships; 85 undergraduates classified these terms by similarity. Neither the number of classes nor the number of elements within classes were restricted beforehand, so the data consist of 85 partitions of 15 objects.

A classical two-dimensional multidimensional scaling analysis of the degree of distinguishability m - ρ_{ij} between pairs of the kinship terms, calculated across subjects, accounts for 67% of the dispersion and tightly groups them into three clusters:

{1, 2, 3, 4} – kins two generations removed from the self,

{5, 6, 7 , 8, 9, 10} – the nuclear family, and

{11, 12, 13, 14, 15} – collaterals,

at about equal distance. Using Daw's triadic index of distinguishability m - ρ_{ijk}, we fitted a M_2 distance model in three dimensions (see Figure 3), which accounts for 99.14% of the dispersion.

--- Insert Figure 3a,b about here ---

From Figure 3b it is clear that the third dimension distinguishes the terms by sex, combined with collaterality, while dimension one and two (Figure 3a) both capture the remoteness of the genetic relationship in terms of the same three clusters as described above. Since the first two axes are globally redundant, detailed relations between the kinship terms can best be seen in the plot of dimension one and three (Figure 3b). A subdivision of the three main clusters is formed by the same-sex pairs {1, 3}, {2, 4}, {7, 9}, {8, 10}, {11, 13}, and {12, 14}, which are close together in the whole space.

Cross-modality similarities

Joly and Le Calvé (1989) reported a sensory experiment, in which 60 undergraduates were asked to associate a color with a taste and a sound; that is, subjects had to choose triadic combinations from three different stimulus sets O_1, O_2, and O_3, each one associated with one sense modality. Every modality was represented by four terms (see Table 4), and the subjects were instructed to select the 5 combinations that to them seemed to be most similar. If the raw

--- Insert Table 4 about here ---

observations are collected in a binary 12×300 table (3×4 rows for the stimuli and 60×5 columns for the subjects, where each entry in a column indicates membership of a triad), we may calculate any of the triadic indices discussed in the previous section, but note that the design of the experiment is such that only a $4 \times 4 \times 4$ subtable of the entire $12 \times 12 \times 12$ table contains empirical information.

This observation can be generalized: each three-way three-mode table can be regarded as a subtable of a larger three-way one-mode table, in which the elements of the single mode are the union of the three original modes. Then any triadic one-mode method may be used to find a Euclidean (or other) representation of the three-mode table, provided that a weighted algorithm is available in which we set $w_{ijk} = 1$ if $(i, j, k) \in O_1 \times O_2 \times O_3$ and $w_{ijk} = 0$ otherwise.

For the Joly-Le Calvé data we used the Russel-Rao dissimilarity, that is, a constant minus the number of positive matches, and fitted a Euclidean perimeter representation in three dimensions.

--- Insert Figure 4a,b about here ---

The resulting configuration accounted for 96.3 % of the dispersion, and is shown in Figure 4. The points in this figure are labelled with the $\mathcal{T}_1$ ... $\mathcal{T}_4$, $\mathcal{C}_1$ ... $\mathcal{C}_4$, and $\mathcal{S}_1$ ... $\mathcal{S}_4$ labels given in Table 4 to simplify the comparison of terms from different sets. For instance, the triad $\{\mathcal{T}_2, \mathcal{C}_1, \mathcal{S}_1\}$, which is {sweet, white, silent} has the smallest perimeter in the three-dimensional configuration as a whole, followed by $\{\mathcal{T}_3, \mathcal{C}_4, \mathcal{S}_4\}$ = {bitter, black, deep}. There are other triads, such as $\{\mathcal{T}_4, \mathcal{C}_4, \mathcal{S}_3\}$ = {sour, black, strident}, that are closer together in Figure 4b, but further apart in Figure 4a, indicating less strong association. Generally, the color and sound elements are more spread out in space than the taste elements. The first dimension contrasts {red, green, shrill, strident} with {white, silent}, that is, activity versus rest. In the second dimension, the most characteristic difference appears to be between {black, deep} and {sour, strident}, and could be interpreted as heaviness-lightness. The third dimension is a contrast between {white, red, deep, sweet} and {salt, green, bitter}, which seems to indicate attraction/satisfaction versus aversion/danger.

Discussion

Several triadic dissimilarity and distance models have been brought together by this paper in a common theoretical framework. Triadic dissimilarity was characterized by three-way symmetry and diagonal plane equality, two properties which clearly distinguish this concept from the more usual three-way two-mode dyadic proximities. To define triadic distance, we proposed two additional requirements: definiteness and the tetrahedral inequality. The latter is stronger than the "generalized triangle inequality" proposed by Joly and Le Calvé (1989); it is our feeling, however, that these more general models are less fruitful for theoretical purposes, and less useful for understanding psychological phenomena.

The models and algorithms presented in this paper can easily be generalized to the *K*-adic case, but exactly how the tetrahedral inequality generalizes is less straightforward. Compared to the Cox et al. (1991) algorithm, which is based on the gradient method, the present algorithm for the

M_2 model has two advantages: it includes weights, and it is guaranteed to converge. The weights can be used in iteratively reweighted least squares procedures to adapt to different assumptions about the residuals. Also, other structural representations, such as hierarchical and additive trees, easily fit into our framework. To our knowledge, there was no least squares procedure available for the perimeter model.

Whether or not fitting triadic models really adds a lot to what dyadic models have to tell us is a difficult question. From simulation work of Cox et al. (1991) with the Jaccard index it appears that dyadic MDS is best at picking out pairs of individuals, triadic MDS is best at picking out triples, and so on for higher-way analyses. Pan and Harris (1991) concluded in their example that the results of a dyadic analysis and a triadic analysis were globally similar, although the latter high-lighted some interesting triple associations. However, we note that, if the two analyses would essentially give identical information, this circumstance should perhaps not be regarded as a negative result, since it implies that a single spatial configuration accounts for a lot more data than is usual. In a statistical analogy, it implies that the three-way interaction is a simple function of the two-way interactions. The three-mode analysis for triadic proximity data, as illustated in the third example, certainly adds something valuable, since it extends our experimental repertoire with the possibility to study relationships with a whole extra set of points compared to a standard unfolding paradigm.

We have been successful in establishing the distance properties of several of the most common dissimilarity coefficients. The fact that these satisfy not only the generalized triangle inequality, but also the tetrahedral inequality provides additional evidence for the importance of the stronger condition put forward here.

REFERENCES

Bartlett, M.S. (1935). Contingency table interactions. *Journal of the Royal Statistical Society, Supplement 2*, 248-252.

Bennani, M. (1993). *Analyses Métriques à Trois Voies*. Unpublished doctoral thesis, University of Rennes 2.

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.

Ciarlet, P.J. (1989). *Introduction to numerical linear algebra and optimisation.* Cambridge: Cambridge University Press.

Coppi, R. & Bolasco, S. (Eds.).(1989). *Multiway Data Analysis.* Amsterdam: North-Holland.

Cox, T.F., Cox, M.A.A., & Branco, J.A. (1991). Multidimensional scaling for *n*-tuples. *British Journal of Mathematical and Statistical Psychology, 44*, 195-206.

Daws, J.T. (1993). *The analysis of free-sorting data: Beyond pairwise co-ocurrences.* Unpublished doctoral dissertation, University of Illinois at Urbana-Champaign, Illinois.

De Leeuw, J., & Heiser, W.J. (1977). Convergence of correction matrix algorithms for multidimensional scaling. In J.C. Lingoes, E.E. Roskam, & I. Borg (Eds.), *Geometric Representations of Relational Data* (pp. 735-752). Ann Arbor, MI: Mathesis Press.

De Leeuw, J., & Heiser, W.J. (1980). Multidimensional scaling with restrictions on the configuration. In P.R. Krishnaiah (Ed.), *Multivariate Analysis, Vol. V.* (pp. 501-522). Amsterdam: North-Holland.

De Leeuw, J., & Heiser, W.J. (1982). Theory of multidimensional scaling. In L. Kanal, & P.R. Krishnaiah (Eds.), *Handbook of Statistics, Volume II* (pp. 285-317). Amsterdam: North-Holland.

Dykstra, R.L. (1983). An algorithm for restricted least squares regression. *Journal of the American Statistical Association, 78*, 839-842.

Fichet, B. (1986). Distances and Euclidean distances for presence-absence characters and their application to factor analysis. In J. de Leeuw, W.J. Heiser, J.J. Meulman, & F. Critchley (Eds.), *Multidimensional Data Analysis* (pp. 23-46). Leiden: DWSO Press.

Fichet, B., & Le Calvé, G. (1984). Structure géométrique des principaux indices de dissimilarité sur signe de présence-absence. *Statistique et Analyse des Données, 9*, 11-64.

Gower, J.C. (1977). The analysis of three-way grids. In P. Slater (Ed.), *Dimensions of Intrapersonal Space, Vol.2* (pp. 163-173). New York: Wiley.

Gower, J.C. (1986). Euclidean distance matrices. In J. de Leeuw, W.J. Heiser, J.J. Meulman, & F. Critchley (Eds.), *Multidimensional Data Analysis* (pp. 11-22). Leiden: DWSO Press.

Gower, J.C., & Legendre, P. (1986). Metric and Euclidean properties of dissimilarity coefficients. *Journal of Classification, 3*, 5-48.

Groenen, P., & Heiser, W.J. (1991). *An improved tunneling function for finding a decreasing series of local minima in MDS* (Research Report RR-91-06). Leiden, The Netherlands: University of Leiden, Department of Data Theory.

Harshman, R.A. (1970). *Foundations of the PARAFAC procedure: Models and conditions for an "explanatory" multimodal factor analysis.* (Working Papers in Phonetics No. 16). Los Angeles, CA: University of California at Los Angeles, Phonetics Laboratory.

Hayashi, C. (1972). Two dimensional quantification based on the measure of dissimilarity among three elements. *Annals of the Institute of Statistical Mathematics, 24*, 251-257.

Hayashi, C. (1989). Multiway data matrices and method of quantification of qualitative data as a strategy of data analysis. In R. Coppi, & S. Bolasco (Eds.), *Multiway Data Analysis* (pp. 131-142). Amsterdam: North-Holland.

Heiser, W.J. (1987). Correspondence analysis with least absolute residuals. *Computational Statistics and Data Analysis, 5*, 337-356.

Heiser, W.J. (1988). Multidimensional scaling with least absolute residuals. In H.- H. Bock (Ed.), *Classification and Related Methods of Data Analysis* (pp.455-462). Amsterdam: North-Holland.

Heiser, W.J. (1991). A generalized majorization method for least squares multidimensional scaling of pseudodistances that may be negative. *Psychometrika, 55*, 7-27.

Heiser, W.J. (1995). Convergent computation by iterative majorization: Theory and applications in multidimensional data analysis. In W.J. Krzanowski (Ed.), *Recent Advances in Descriptive Multivariate Analysis* (pp. 157-189). Oxford: Oxford University Press.

Joly, S., & Le Calvé, G. (1989). *Three-way distances*. Manuscript in preparation.

Kroonenberg, P.M. (1983). *Three-mode principal components analysis: Theory and applications*. Leiden: DSWO Press.

Kruskal, J.B. (1977). Three-way arrays: Rank and uniqueness of trilinear decompositions, with applications to arithmetic complexity and statistics. *Linear Algebra and Its Applications, 18*, 95-138.

Kruskal, J.B. (1989). Rank, decomposition, and uniqueness for 3-way and *N*-way arrays. In R. Coppi, & S. Bolasco (Eds.), *Multiway Data Analysis* (pp. 7-18). Amsterdam: North-Holland.

Law, H.G., Snyder, C.W., Hattie, J.A., & McDonald, R.P. (Eds.).(1984). *Research methods for multimode data analysis*. New York: Praeger.

Pan, G., & Harris, D.P. (1991). A new multidimensional scaling technique based upon associations of triple objects Pijk and its application to the analysis of geochemical data. *Mathematical Geology, 23*, 861-886.

Rockafellar, R.T. (1970). *Convex Analysis*. Princeton, NJ: Princeton University Press.

Rosenberg, S. (1982). The method of sorting in multivariate research with applications selected from cognitive psychology and person perception. In N. Hirschberg & L.G. Humphreys (Eds.), *Multivariate Applications in the Social Sciences* (pp. 117-142). Hillsdale, NJ: Erlbaum.

Tucker, L.R. (1963). Implications of factor analysis of three-way matrices for measurement of change. In C.W. Harris (Ed.), *Problems in Measuring Change* (pp.122-137). Madison, Wisconsin: University of Wisconsin Press.

Tucker, L.R. (1964). The extension of factor analysis to three-dimensional matrices. In N. Fredericksen & H. Gulliksen (Eds.), *Contributions to Mathematical Psychology* (pp.109-127).

New York: Holt, Rinehart, and Winston.

Tucker, L.R. (1966). Some mathematical notes on three-mode factor analysis. *Psychometrika, 31*, 279-311.

Footnote

The free-sorting data were kindly made available to us by dr. John Daws of the Department of Psychology, Columbia University, New York.

FIGURE CAPTIONS

Figure 1. Surface model solution for the Hayashi (1972) data.

Figure 2. Generalized Euclidean solution for the Hayashi (1972) data.

Figure 3. Three-dimensional generalized Euclidean solution for the Rosenberg (1982) data; at the left, axis 2 against axis 1 (a), at the right, axis 3 against axis 1 (b).

Figure 4. Three-dimensional solution under the Euclidean perimeter model for the Joly & Le Calvé (1989) data; at the left, axis 2 against axis 1 (a), at the right, axis 3 against axis 1 (b).

TABLE 1

Frequency of subset choice in two groups of subjects (Daws' example)

Subset pattern	*Group 1*	*Group 2*
(1234)	0	0
(123) – (4)	5	1
(124) – (3)	0	0
(134) – (2)	0	0
(1) – (234)	1	2
(12) – (34)	0	1
(13) – (24)	1	2
(14) – (23)	0	0
(12) – (3) – (4)	1	4
(13) – (2) – (4)	0	3
(1) – (23) – (4)	1	4
(14) – (2) – (3)	0	0
(1) – (24) – (3)	2	0
(1) – (2) – (34)	2	0
(1) – (2) – (3) – (4)	5	1
Total	18	18

TABLE 2

Unproductivity of triadic teams (Hayashi, 1972)

$\tau_{123} = 1$			
$\tau_{124} = 7$			
$\tau_{125} = 6$			
$\tau_{126} = 9$			
$\tau_{134} = 7$	$\tau_{234} = 8$		
$\tau_{135} = 6$	$\tau_{235} = 7$		
$\tau_{136} = 9$	$\tau_{236} = 9$		
$\tau_{145} = 4$	$\tau_{245} = 6$	$\tau_{345} = 3$	
$\tau_{146} = 9$	$\tau_{246} = 8$	$\tau_{346} = 5$	
$\tau_{156} = 6$	$\tau_{256} = 7$	$\tau_{356} = 3$	$\tau_{456} = 1$

TABLE 3

Kinship terms used in Rosenberg's experiment

1. grandfather	5. brother	9. son	13. uncle
2. grandmother	6. sister	10. daughter	14. aunt
3. grandson	7. father	11. nephew	15. cousin
4. granddaughter	8. mother	12. niece	

TABLE 4

Sensory terms used in Joly & Le Calvé's experiment

Taste	*Color*	*Sound*
$\mathcal{T}_1$. salt	$\mathcal{C}_1$. white	$\mathcal{S}_1$. silent
$\mathcal{T}_2$. sweet	$\mathcal{C}_2$. red	$\mathcal{S}_2$. shrill
$\mathcal{T}_3$. bitter	$\mathcal{C}_3$. green	$\mathcal{S}_3$. strident
$\mathcal{T}_4$. sour	$\mathcal{C}_4$. black	$\mathcal{S}_4$. deep

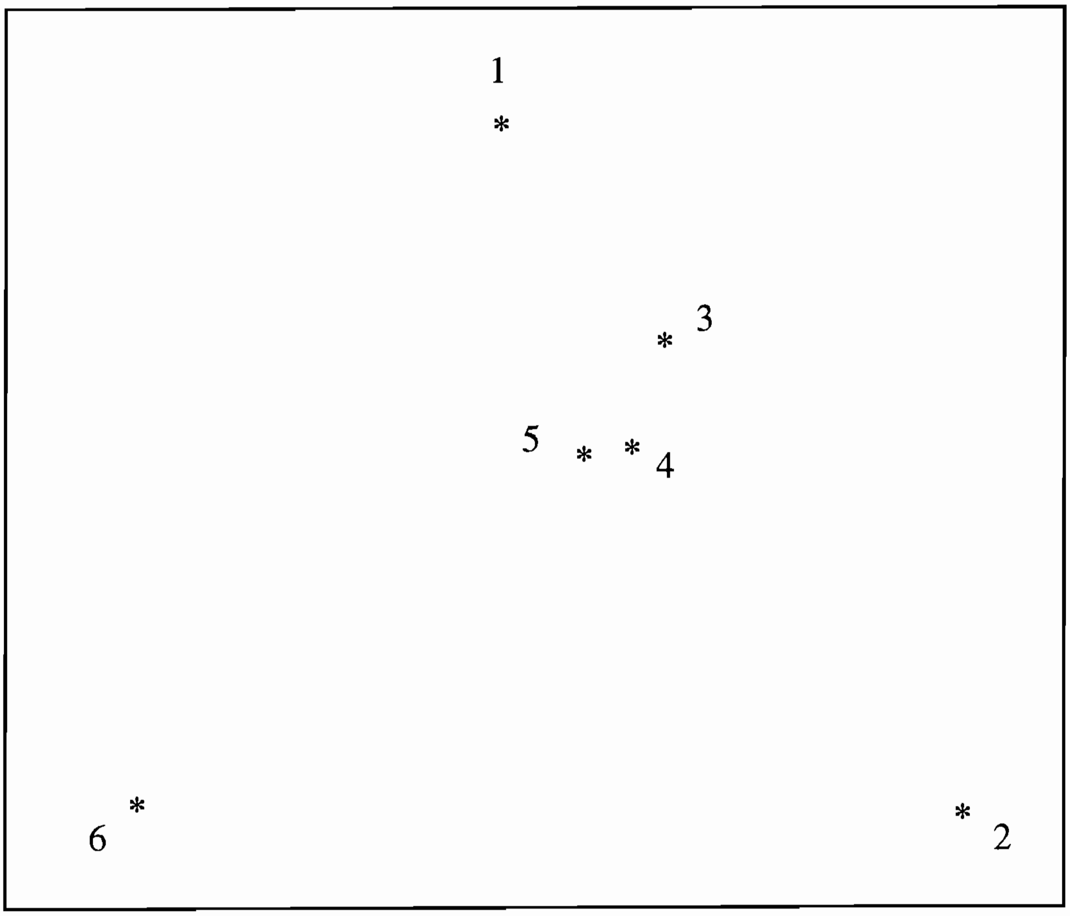

Figure 1

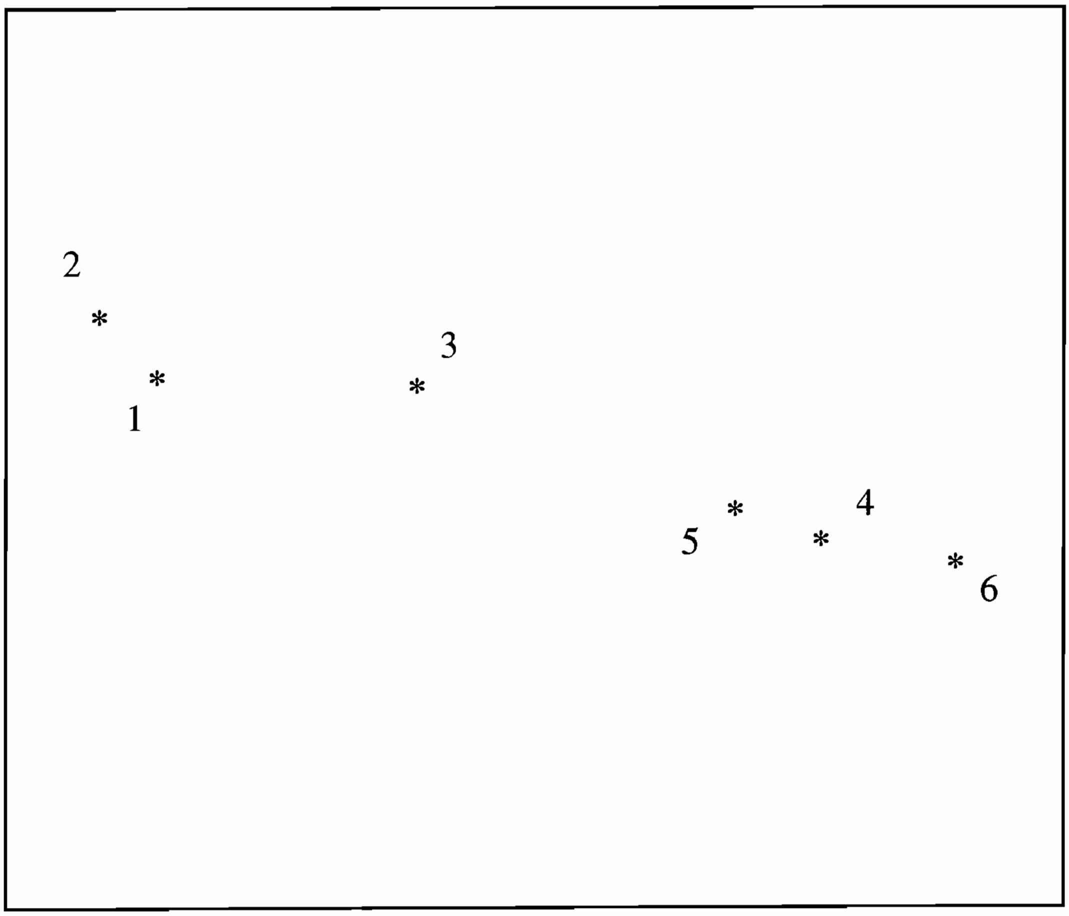

Figure 2

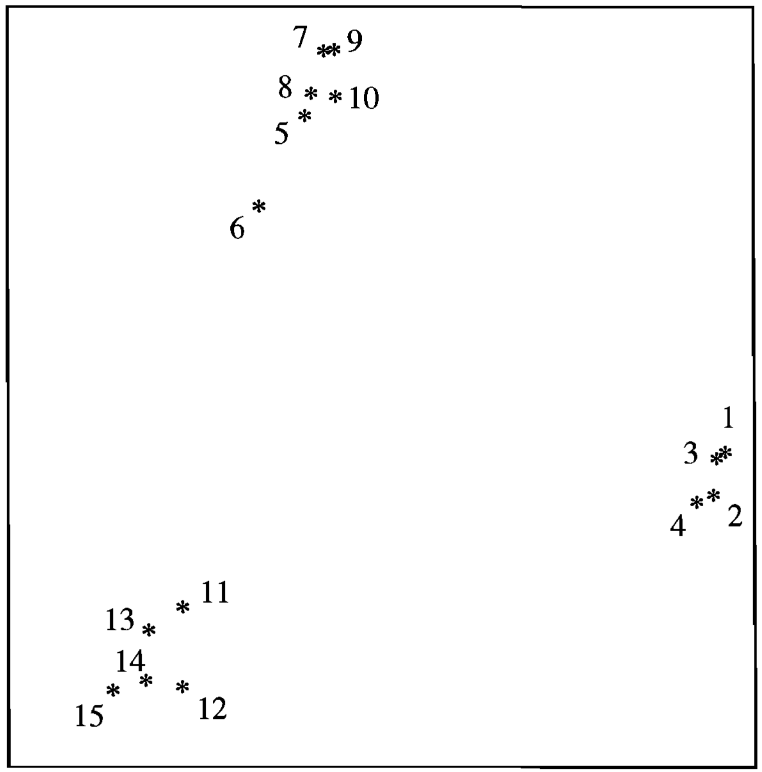

Figure 3a

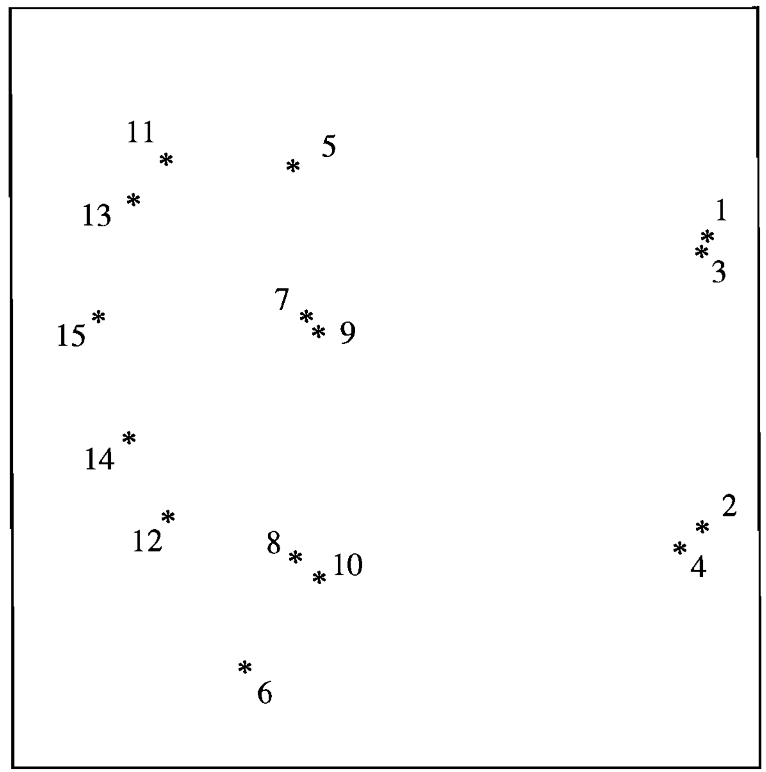

Figure 3b

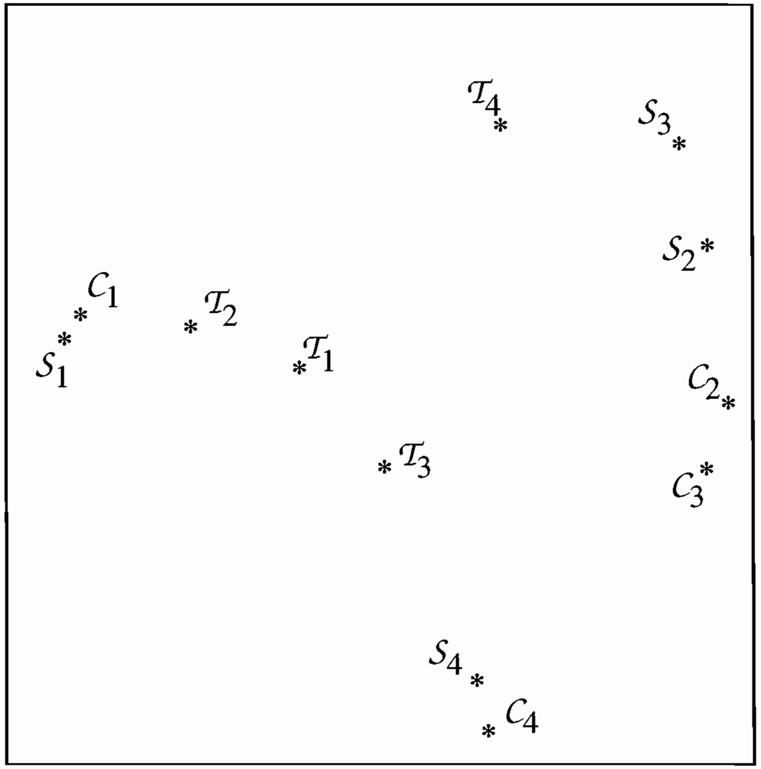

Figure 4a

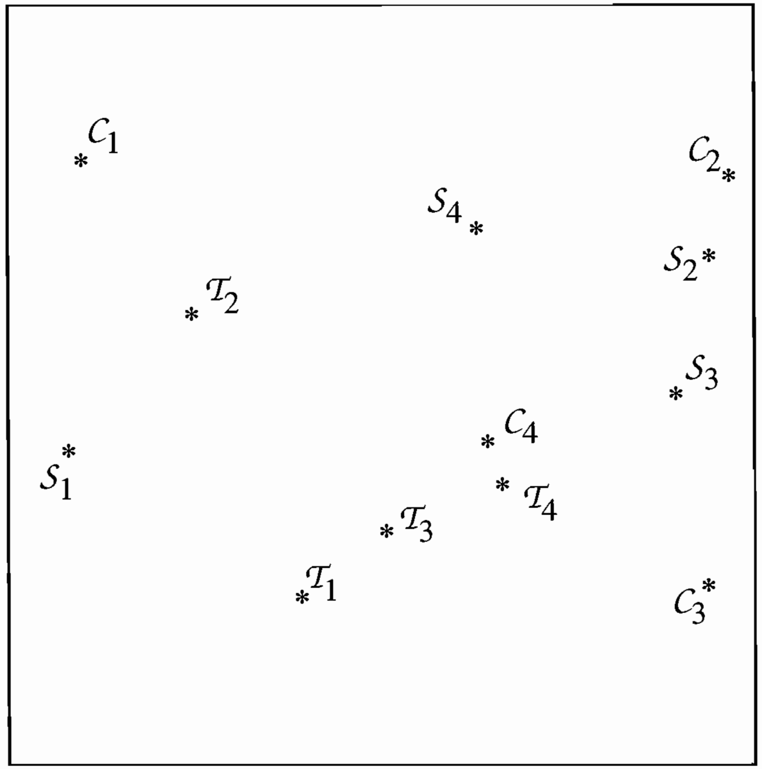

Figure 4b