PSYCHOMETRIKA—VOL. 58, NO. 1, 53–69
MARCH 1993

THE ROWWISE CORRELATION BETWEEN TWO PROXIMITY MATRICES AND THE PARTIAL ROWWISE CORRELATION

HAN DE VRIES

PROJECTGROUP OF ETHOLOGY & SOCIO-ECOLOGY
STATE UNIVERSITY OF UTRECHT

This paper discusses rowwise matrix correlation, based on the weighted sum of correlations between all pairs of corresponding rows of two proximity matrices, which may both be square (symmetric or asymmetric) or rectangular. Using the correlation coefficients usually associated with Pearson, Spearman, and Kendall, three different rowwise test statistics and their normalized coefficients are discussed, and subsequently compared with their nonrowwise alternatives like Mantel's Z. It is shown that the rowwise matrix correlation coefficient between two matrices **X** and **Y** is the partial correlation between the entries of **X** and **Y** controlled for the nominal variable that has the row objects as categories. Given this fact, partial rowwise correlations (as well as multiple regression extensions in the case of Pearson's approach) can be easily developed.

Key words: matrix permutation tests, rowwise matrix correlation, partial matrix correlation, Mantel's Z statistic, nonparametric statistics.

Introduction

Matrix association/correlation methods are used increasingly in widely different research disciplines, for example in geography, psychometrics (Hubert, 1987), population biology (Smouse, Long, & Sokal, 1986), systematic zoology (Cheverud, Wagner, & Dow, 1989), animal behavior (de Waal & Luttrell, 1988; Hemelrijk, 1990a, 1990b; Schnell, Watt, & Douglas, 1985). Most applications involve two square matrices containing for each pair of distinct objects the value of some measure of relationship, such as correlation, distance, interaction frequency, flow, and so on. Such matrices will be called proximity matrices, following Hubert (p. 121) in his use of the term "proximity" as referring to any measure of relationship, symmetric or asymmetric, that is specified for object pairs. The correlation between two proximity matrices is assessed using Mantel's statistic, defined simply as the sum of all crossproducts between the two matrices (Mantel, 1967). To evaluate the significance of the statistic, a permutation test is employed, originally described by Mantel. This inferential procedure respects the interdependencies of observations within rows and columns of the matrices (see, e.g., Krackhardt, 1988).

Extensions of Mantel's approach have been developed in different directions: (a) besides Mantel's statistic, which is really an unnormalized Pearson product-moment correlation coefficient, statistics have been introduced based on Spearman and Kendall rank-order correlation coefficients (Dietz, 1983); (b) Mantel's bivariate matrix correlation test has been extended with multiple regression and partial correlation methods (Dow, Cheverud, & Friedlaender, 1987; Smouse et al., 1986); (c) matrix correlation methods have been developed for rectangular proximity matrices for which the set of

The author wishes to thank the Editor, two referees, Jan van Hooff, and Ruud Derix for their useful comments, and E. J. Dietz for a copy of the algorithm of the Mantel permutation test.

Requests for reprints should be sent to Han de Vries, Projectgroup of Ethology & Socio-ecology, The University of Utrecht, Postbox 80.086, 3508 TB Utrecht, The Netherlands.

0033-3123/93/0300-91365$00.75/0

row objects is different from the set of column objects (Hemelrijk, 1990a; Hubert, 1987; Klauber, 1971). In this paper an alternative matrix correlation coefficient will be discussed, based upon a weighted sum of the correlations between all pairs of corresponding rows of the two matrices. This *rowwise* matrix correlation coefficient is similarly defined using the indices usually associated with Pearson, Spearman, and Kendall; it can be extended with multiple regression and partial correlation methods, and used with square (symmetric or asymmetric) or rectangular proximity matrices.

The basic feature of rowwise correlation measures is that they only involve comparisons of pairs of cells within the same row. They can therefore be applied in cases where the data has been obtained in a way that only allows for within-row comparisons, such as conditional proximity matrices in the sense of Shepard (1972). Thus, Hubert (1987, p. 274) applied a rowwise index based on Kendall's scoring function to the Miller-Nicely confusion matrices containing conditional probabilities. As has been shown by Hemelrijk (1990a), the application of rowwise statistics is, however, not limited to these cases only, but is also useful with matrices containing entries that may all be compared to each other, such as those containing frequencies of interactions flowing from the row objects (individuals initiating interactions) to the column objects (individuals receiving interactions); see the section on Examples. It will be shown that the rowwise correlation coefficient between two matrices **X** and **Y** is equal to the partial correlation between the entries of **X** and **Y** controlled for the nominal variable R that has the row objects as categories. This means that whenever one wishes to control for the differences among the row objects, a rowwise measure can be profitably used in our matrix comparison task. For instance, when there are differences among the individuals initiating interactions in their tendencies to interact (showing in different row means), these differences can be controlled for by using a rowwise measure (Hemelrijk, 1990a), thus partialling out from the total correlation between the two matrices that which is due to across-row internal comparisons. A rowwise index based on Kendall's approach has been introduced by Hubert (1987). Rowwise indexes based on Kendall's and Spearman's approaches have been introduced by Hemelrijk (1990a, 1990b) in the context of models of reciprocity/interchange of behavior.

To provide some background terminology that will prove helpful in the way of analogy to the matrix comparison context, note that the correlation coefficients of Pearson, Spearman, and Kendall are all specific forms of the generalized correlation coefficient (see Daniels, 1944; Kendall, 1962):

$$\Gamma = \frac{\sum \gamma(x_i, x_j)\gamma(y_i, y_j)}{(\sum \gamma(x_i, x_j)^2 \sum \gamma(y_i, y_j)^2)^{1/2}},$$

where x and y are two numerical measures and γ is a function that assigns to each pair of x-scores, (x_i, x_j), a new value, $\gamma(x_i, x_j)$, subject only to the condition that $\gamma(x_i, x_j) = -\gamma(x_j, x_i)$ (and analogously for the y-scores). This pairwise comparison function γ can have different forms. If $\gamma(x_i, x_j) = x_i - x_j$, Pearson's product-moment correlation coefficient results; when $\gamma(x_i, x_j) = \text{rank}\,(x_i) - \text{rank}\,(x_j)$, Γ is Spearman's correlation coefficient, and with $\gamma(x_i, x_j) = \text{sign}\,(x_i - x_j)$, one has Kendall's tau$_b$ coefficient. In the case of two matrices **X** and **Y** (with elements X_{ij} and Y_{ij} respectively),

$$\Gamma = \frac{\sum \gamma(X_{ij}, X_{kl})\gamma(Y_{ij}, Y_{kl})}{(\sum \gamma(X_{ij}, X_{kl})^2 \sum \gamma(Y_{ij}, Y_{kl})^2)^{1/2}},$$

can be considered a generalized matrix correlation coefficient for rows i, k and columns j, l. Note that in this formula the summation is over all pairs of cells within the two matrices (excluding the diagonal cells in the case of square matrices **X** and **Y** with undefined diagonal). When the summation is restricted to all those pairs of cells that belong to the same row, one arrives at the generalized *rowwise* correlation coefficient between two matrices:

$$\Gamma_{rw} = \frac{\sum \gamma(X_{ij}, X_{ik})\gamma(Y_{ij}, Y_{ik})}{(\sum \gamma(X_{ij}, X_{ik})^2 \sum \gamma(Y_{ij}, Y_{ik})^2)^{1/2}},$$

for rows i and columns j, k. For each of the three different γ's, a different rowwise correlation coefficient results. (When using Spearman's comparison function, the transformation to ranks is understood to take place within each row; otherwise, the pairwise comparisons in Γ_{rw} do not depend exclusively on the values of the cells within the same row.) In this same type of generalized matrix correlation context, one might also consider alternative normalizations such as

$$\Gamma_{rw,av} = \frac{\sum \gamma(X_{ij}, X_{ik})\gamma(Y_{ij}, Y_{ik})}{\sum_i (\sum_{j,k} \gamma(X_{ij}, X_{ik})^2 \sum_{j,k} \gamma(Y_{ij}, Y_{ik})^2)^{1/2}},$$

which turns out to be a simple weighted average of correlations between pairs of rows. Both of these measures will be discussed in their various specific forms in the sequel.

In this paper rowwise correlation between two proximity matrices is discussed. Of course, by restricting the summation in Γ to all pairs of cells that belong to the same column, one obtains a columnwise correlation. The same can be achieved by transposing both matrices and calculating the rowwise correlation between these transposed matrices. When comparing two rectangular or asymmetric proximity matrices, one can often fruitfully distinguish between two different conjectures of a similar patterning of entries across the two matrices: rowwise and columnwise (see Examples). Rowwise (and columnwise) correlation measures are pre-eminently the statistics to be used for evaluating these two different conjectures. Obviously, when the entries in the matrices may only be compared within the same row, there is necessarily only one conjecture of similar patterning to be evaluated.

The Rowwise Correlation Between Two Matrices

Pearson's Product-Moment Approach

Several authors (e.g., Dietz, 1983, Hubert, 1987, plus references therein) have discussed the raw crossproduct statistic, originally introduced by Mantel (1967), for use in assessing the correlation (or association) between two proximity matrices, say **X** and **Y**. Mantel's statistic, denoted by Z, can be defined as

$$Z = \sum_{i,j} X_{ij} Y_{ij},$$

for rows $i = 1, \ldots, n$ and columns $j = 1, \ldots, n$; $i \neq j$ (if **X** and **Y** are square); for rows $i = 1, \ldots, p$ and columns $j = 1, \ldots, q$ (if **X** and **Y** are rectangular). The statistic Z is an unnormalized Pearson's product-moment correlation coefficient. For two rectangular matrices each with p rows and q columns, it can be shown (see Kendall, 1962, p. 21) that

$$\sum_{i,j,k,l} (X_{ij} - X_{kl})(Y_{ij} - Y_{kl}) = 2\{pq\sum_{i,j} X_{ij} Y_{ij} - (\sum_{i,j} X_{ij})(\sum_{i,j} Y_{ij})\},$$

$i, k = 1, \ldots, p; j, l = 1, \ldots, q$. Similarly, for two square matrices each with n rows and n columns,

$$\sum_{i,j,k,l} (X_{ij} - X_{kl})(Y_{ij} - Y_{kl}) = 2\{n(n-1) \sum_{i,j} X_{ij} Y_{ij} - (\sum_{i,j} X_{ij})(\sum_{i,j} Y_{ij})\},$$

$i, k = 1, \ldots, n; j, l = 1, \ldots, n; i \neq j$ and $k \neq l$. Thus, the left-hand statistic is equivalent to the Z statistic, each being a constant linear transformation of the other under any permutation of the rows and the columns of one of the two matrices. From this formula it can be deduced that the Z statistic is based upon all comparisons of pairs of cells across the two matrices.

The rowwise statistic Z_r is defined by restricting the summation to those pairs of cells that belong to the same row. Thus,

$$Z_r = \sum_{i,j,k} (X_{ij} - X_{ik})(Y_{ij} - Y_{ik}),$$

$i = 1, \ldots, n; j, k = 1, \ldots, n; j < k; i \neq j$ and $i \neq k$ (if **X** and **Y** are square); or $i = 1, \ldots, p; j, k = 1, \ldots, q; j < k$ (if **X** and **Y** are rectangular). The measure Z_r can be seen to be a weighted sum of the correlations between all pairs of corresponding rows, with each weight being proportional to the standard deviations of the two rows. In fact, the weight of the i-th pair of rows (w_i) is equal to the product of the standard deviation of row i in the **X**-matrix and the standard deviation of row i in the **Y**-matrix multiplied by the number of columns (in case **X** and **Y** are rectangular), or by the number of columns minus one (in case **X** and **Y** are square). Let $w_i = (\sum_{j,k}(X_{ij} - X_{ik})^2 \sum_{j,k}(Y_{ij} - Y_{ik})^2)^{1/2}$. Then,

$$Z_r = \sum_i w_i [\sum_{j,k} (X_{ij} - X_{ik})(Y_{ij} - Y_{ik})/w_i] = \sum_i w_i r_i,$$

where r_i is the product-moment correlation of the i-th pair of rows (see Kendall, 1962, sec. 2.5). The value of $w_i r_i$ is the contribution of the i-th pair of rows to the total value of Z_r.

The Z_r statistic can be normalized to lie between -1 and 1 in at least two possible ways, giving rise to two different coefficients that only differ in the denominator:

$$r_{rw,av} = \frac{\sum_{i,j,k} (X_{ij} - X_{ik})(Y_{ij} - Y_{ik})}{\sum_i (\sum_{j,k} (X_{ij} - X_{ik})^2 \sum_{j,k} (Y_{ij} - Y_{ik})^2)^{1/2}},$$

$$r_{rw} = \frac{\sum_{i,j,k} (X_{ij} - X_{ik})(Y_{ij} - Y_{ik})}{(\sum_{i,j,k} (X_{ij} - X_{ik})^2 \sum_{i,j,k} (Y_{ij} - Y_{ik})^2)^{1/2}}.$$

The first is a *weighted average of the correlations* between all pairs of corresponding rows. With w_i defined as above,

$$r_{rw,av} = \frac{Z_r}{\sum_i w_i} = \frac{\sum_i w_i r_i}{\sum_i w_i},$$

which is the Pearson form of the generalized coefficient $\Gamma_{rw,av}$ of the introduction. The second, called the *rowwise matrix correlation coefficient*, is the Pearson form of the generalized rowwise correlation coefficient Γ_{rw} of the introduction.

Each of these two normalizations is a particular type of partial correlation coefficient between the entries of **X** and **Y** controlled for the nominal variable R that has values as the row categories (i.e., the row objects). As noted by Quade (1974), the most basic concept of control is to actually hold the control variable constant and define partial correlation as the (weighted) average of conditional correlations, where a conditional correlation is the correlation between values of **X** and **Y** for which the control variable is constant. The first normalization $r_{rw,av}$ fits this definition of partial correlation perfectly, with the nominal variable R as the control variable. A second concept of control, also discussed by Quade, involves adjusting for the control variable, and as shown below, r_{rw} is the partial correlation between the entries of **X** and **Y** controlled for the nominal variable R by adjusting for it.

r_{rw} as a Partial Correlation

By way of background to the demonstration that r_{rw} is a partial correlation, suppose there are three interval scaled variables X, Y, and Z; the partial correlation between X and Y controlled for Z can be defined as the correlation between X' and Y', where X' is the residual $X - f(Z)$ and Y' is the residual $Y - g(Z)$ with f and g prediction functions. If f and g are the usual linear regression functions, the partial correlation based on this concept of control can also be obtained from the well-known formula:

$$r_{XY.Z} = \frac{r_{XY} - r_{XZ}r_{YZ}}{((1 - r_{XZ}^2)(1 - r_{YZ}^2))^{1/2}}.$$

In the case the control variable Z is nominal having p categories, one can introduce $p - 1$ dummy variables D_i, each of which is a binary variable taking on only the values 0 and 1: $D_i = 1$ if Z = category i, otherwise $D_i = 0$ (see Dunn & Clark, 1987, pp. 344 ff). The partial correlation between X and Y controlled for Z can then be written as $r_{XY.D_1D_2 \cdots D_{p-1}}$ and calculated in the following way: determine the residual $X' = X - f(D_1, D_2, \ldots, D_{p-1})$, where f is the multiple linear regression function obtained from regressing X on the variables $D_1, \ldots, D_{p-1}$; similarly, determine the residual $Y' = Y - g(D_1, D_2, \ldots, D_{p-1})$, where g is the multiple linear regression function obtained from regressing Y on the variables $D_1, \ldots, D_{p-1}$. The partial correlation $r_{XY.D_1D_2 \cdots D_{p-1}}$ is then the correlation between X' and Y', denoted by $r_{X'Y'}$ or $r(X', Y')$.

An alternative, but equivalent method to partial out the nominal Z using the second concept of control, is by directly adjusting the values of the observations on X and Y for the nominal variable Z (without introducing a set of dummy variables) by subtracting the respective category means ($\bar{X}_i$ or $\bar{Y}_i$, being the mean X-value (Y-value) of all observations with Z = category i) from each X-value or each Y-value. Explicitly, since $f(D_1, D_2, \ldots, D_{p-1}) = a + b_1D_1 + \cdots + b_{p-1}D_{p-1}$, where $a = \bar{X}_p$, $b_1 = \bar{X}_1 - \bar{X}_p, \ldots, b_{p-1} = \bar{X}_{p-1} - \bar{X}_p$, $f(Z) = \bar{X}_i$ if Z = category i. A similar equivalence holds for $g(D_1, D_2, \ldots, D_{p-1})$.

To show that the second normalization, r_{rw}, is the partial correlation between the entries of the two rectangular $p \times q$ matrices **X** and **Y** controlled for the nominal variable R, it is proven that $r_{rw} = r(\mathbf{X}', \mathbf{Y}')$, where $\mathbf{X}'$ is the residual $\mathbf{X} - f(R)$ and $\mathbf{Y}'$ is the residual $\mathbf{Y} - g(R)$; $f(R)$ and $g(R)$ are defined for the nominal variable R similarly as done above for the nominal Z, where R is now the nominal variable with the p row objects as categories. For each category i (i.e., for each row object i) there are q observations on X and on Y, the total number of observations $N = pq$, and $\bar{X}_i(\bar{Y}_i)$ denotes the mean X-value (Y-value) of all the q observations with R = category i. Note, that for each category i, the mean of the residuals is equal to zero, $\bar{X}'_i = 0$ and $\bar{Y}'_i = 0$; therefore, the mean of all the residuals is also zero, $\bar{X}' = 0$ and $\bar{Y}' = 0$.

For each category i the following identities hold (see Kendall, 1962, p. 21):

$$\frac{1}{2}\sum_{j,k=1}^{q}(X'_{ij}-X'_{ik})(Y'_{ij}-Y'_{ik}) = q\sum_{j=1}^{q}X'_{ij}Y'_{ij} - \sum_{j,k=1}^{q}X'_{ij}Y'_{ik}$$

and

$$\frac{1}{2}\sum_{j,k=1}^{q}(X'_{ij}-X'_{ik})^2 = q\sum_{j=1}^{q}(X'_{ij})^2 - \left(\sum_{j=1}^{q}X'_{ij}\right)^2.$$

For both equations the last right-hand term is equal to zero, because X'_{ij} and Y'_{ik} are residuals from their respective means in category i. Now, using these identities and the fact that $\bar{X}' = \bar{Y}' = 0$.

$$r(\mathbf{X}', \mathbf{Y}') = \frac{\sum_{i=1}^{p}\sum_{j=1}^{q}(X'_{ij}-\bar{X}')(Y'_{ij}-\bar{Y}')}{\left(\sum_{i=1}^{p}\sum_{j=1}^{q}(X'_{ij}-\bar{X}')^2\sum_{i=1}^{p}\sum_{j=1}^{q}(Y'_{ij}-\bar{Y}')^2\right)^{1/2}}$$

$$= \frac{\frac{1}{2q}\sum_{i=1}^{p}\sum_{j,k=1}^{q}(X'_{ij}-X'_{ik})(Y'_{ij}-Y'_{ik})}{\left(\left[\frac{1}{2q}\sum_{i=1}^{p}\sum_{j,k=1}^{q}(X'_{ij}-X'_{ik})^2\right]\left[\frac{1}{2q}\sum_{i=1}^{p}\sum_{j,k=1}^{q}(Y'_{ij}-Y'_{ik})^2\right]\right)^{1/2}}.$$

Because $X'_{ij} - X'_{ik} = X_{ij} - \bar{X}_i - (X_{ik} - \bar{X}_i) = X_{ij} - X_{ik}$, and similarly, $Y'_{ij} - Y'_{ik} = Y_{ij} - Y_{ik}$, one can drop the quotes and rewrite as:

$$\frac{\sum_{i=1}^{p}\sum_{j<k}^{q}(X_{ij}-X_{ik})(Y_{ij}-Y_{ik})}{\left(\sum_{i=1}^{p}\sum_{j<k}^{q}(X_{ij}-X_{ik})^2\sum_{i=1}^{p}\sum_{j<k}^{q}(Y_{ij}-Y_{ik})^2\right)^{1/2}} = r_{rw}.$$

The derivation is exactly the same for the case that $\mathbf{X}$ and $\mathbf{Y}$ are square $n \times n$ matrices with undefined diagonals. The number of observations for row category i excluding the diagonal, equals $n - 1$ for every i, and the total number of observations N equals $n(n - 1)$.

Spearman and Kendall Rank-Correlation Approaches

Within Spearman's approach the following rowwise correlation statistic can be introduced:

$$R_r = \sum_{i,j,k} (S_{ij} - S_{ik})(T_{ij} - T_{ik}),$$

$i = 1, \ldots, n; j, k = 1, \ldots, n; j < k; i \neq j$ and $i \neq k$ (if **X** and **Y** are square), or $i = 1, \ldots, p; j, k = 1, \ldots, q; j < k$ (if **X** and **Y** are rectangular). $S_{ij}(T_{ij})$ is the rank number of $X_{ij}(Y_{ij})$ ranked within row i, and thus, R_r is identical to Z_r calculated on the matrices **S** and **T**.

If R_r is divided by the normalization factor, $\Sigma_i(\Sigma_{j,k}(S_{ij} - S_{ik})^2 \Sigma_{j,k}(T_{ij} - T_{ik})^2)^{1/2}$, one arrives at the coefficient $\rho_{rw,av}$, which is a *weighted average of the rank correlations* between all pairs of matching rows, and equal to the partial rank-correlation coefficient between the entries of **X** and **Y** controlling for the nominal variable R by holding it constant (a special case of $\Gamma_{rw,av}$ in the introduction). If R_r is divided by the normalization factor $(\Sigma_{i,j,k}(S_{ij} - S_{ik})^2 \Sigma_{j,k}(T_{ij} - T_{ik})^2)^{1/2}$, one obtains the *rowwise rank-correlation coefficient* ρ_{rw}, which can be shown as in the last section to be the partial correlation between the rank-transformed entries of **S** and **T** controlled for R by adjusting for it (and a special case of Γ_{rw} in the introduction).

In an analogous way using the Kendall scoring function, one obtains

$$K_r = \sum_{i,j,k} \text{sign}\,(X_{ij} - X_{ik})\ \text{sign}\,(Y_{ij} - Y_{ik}),$$

$i = 1, \ldots, n; j, k = 1, \ldots, n; j < k; i \neq j$ and $i \neq k$ (if X and Y are square) or $i = 1, \ldots, p; j, k = 1, \ldots, q; j < k$ (if **X** and **Y** are rectangular) (see Hemelrijk, 1990a, and Hubert, 1987, pp. 267 ff.). K_r can be seen to be a weighted sum of the correlations between all pairs of corresponding rows. Let $w_i = (\Sigma_{j,k}\ \text{sign}\,(X_{ij} - X_{ik})^2\ \Sigma_{j,k}\ \text{sign}\,(Y_{ij} - Y_{ik})^2)^{1/2}$. Then

$$K_r = \sum_i w_i \left[\sum_{j,k} \text{sign}\,(X_{ij} - X_{ik})\ \text{sign}\,(Y_{ij} - Y_{ik})/w_i \right] = \sum_i w_i \tau_i,$$

where τ_i is Kendall's tau$_b$ for the i-th pair of rows (see Kendall, 1962, sec. 2.3). The value of $w_i\tau_i$ is the contribution of the i-th pair of rows to the total value of K_r.

The first type of normalization ($\tau_{rw,av}$) of K_r is equal to

$$\frac{\sum_{i,j,k} \text{sign}\,(X_{ij} - X_{ik})\ \text{sign}\,(Y_{ij} - Y_{ik})}{\sum_i \left(\sum_{j,k} \text{sign}\,(X_{ij} - X_{ik})^2 \sum_{j,k} \text{sign}\,(Y_{ij} - Y_{ik})^2 \right)^{1/2}},$$

and is, just as in the Z_r and R_r case, a *weighted average of conditional correlations*, and therefore a partial correlation coefficient based on the concept of control of holding the control variable constant. As for the second normalization of K_r, τ_{rw}, which is equal to

$$\frac{\sum_{i,j,k} \text{sign}\,(X_{ij} - X_{ik})\ \text{sign}\,(Y_{ij} - Y_{ik})}{\left(\sum_{i,j,k} \text{sign}\,(X_{ij} - X_{ik})^2 \sum_{i,j,k} \text{sign}\,(Y_{ij} - Y_{ik})^2 \right)^{1/2}},$$

it appears impossible to give a natural adjustment interpretation similar to the one demonstrated above for the second normalization of Z_r, given the absence of a way of defining a Kendall partial correlation between two ordinal variables X and Y controlling for a *nominal* variable Z by adjusting for it.

Significance Testing

For the case of square matrices, Mantel (1967) has described a method to assess the significance of a matrix correlation statistic by means of a permutational method (also, see Dietz, 1983; and Hubert, 1987). The reference distribution of the statistic under the conjecture of an absence of a similar patterning between the entries of **X** and **Y** is found by generating a set of permutations of the rows (and simultaneously of the columns) for one of the two matrices and calculating for each permutation the value of the statistic. The significance of the observed value of the statistic is then assessed by calculating the proportion of values as large as or larger than the observed value (i.e., the right-tail probability P_r) and the proportion of values as small as or smaller than the observed value (i.e., the left-tail probability P_l). For square matrices this same permutational method can be used to assess the significance of the rowwise statistics Z_r (or equivalently r_{rw}), R_r (or equivalently ρ_{rw}) and K_r (or equivalently τ_{rw}). With this permutation method, however, Z_r is not (necessarily) equivalent to $r_{rw,av}$, because the denominator of $r_{rw,av}$ can vary across the permutations. Therefore, if one wishes to assess the significance of this weighted average of conditional correlations, $r_{rw,av}$, one has to use this coefficient itself instead of the raw statistic Z_r. The same is true for the normalized coefficient $\rho_{rw,av}$ and the raw statistic R_r, and also for $\tau_{rw,av}$ and K_r.

The permutational method to be used for rectangular matrices depends on the type of data these matrices contain and, in the case of proximity data, on whether the row and column categories are considered fixed or random. First, if the rows correspond to p objects and the columns correspond to q variables (or vice versa), the rectangular matrices contain profile data (Shepard, 1972). In this case either only the rows or only the columns, whichever contain the profiles, should be permuted (see also Hubert, 1987, sec. 2.3.2).

If the rectangular matrices are proximity matrices, then the pairs of objects (i.e., the dyads) are the observational units. In any randomization test the measurements made on the observational units have to be randomized. In case of proximity matrices therefore, the observations on the dyads must be randomized, but under the restriction that the interdependence existing among these observations is respected. That is to say, only those random assignments of the observations to the dyads are allowed that keep the rows and columns intact. So, for rectangular proximity matrices each randomization can be obtained by randomly permuting the rows and independently the columns. In this way, a random subset (possibly containing replications) of the total $p!q!$ allowed randomizations is constructed. By giving consideration to a specific aspect of the row and column objects, the set of allowed randomizations may have to be restricted even more. Thus, several authors (Hemelrijk, 1990a; Hubert, 1987; Klauber, 1971) have stated that the permutational method depends on whether the row and column categories are considered fixed or random. If the row objects as well as the column objects are considered random, both rows and columns should be permuted independently. If, on the other hand, the row objects are considered fixed and the column objects random (or vice versa), only the columns (or only the rows) should be permuted. Finally, according to Shepard (1972), a proximity matrix can also be treated as a matrix of profile data. So, if the columns are regarded as containing the profiles, these should be permuted and not the rows.

If the permutational method is used in which only the columns are permuted, then each of the raw statistics Z_r, R_r and K_r is equivalent to its corresponding normalizations under this permutation scheme: Z_r is equivalent to both $r_{rw,av}$ and r_{rw}; R_r is equivalent to both $\rho_{rw,av}$ and ρ_{rw}; and K_r is equivalent to both $\tau_{rw,av}$ and τ_{rw}, because all of the respective denominators in the normalized coefficients are constant under any permutation of the columns. If, on the other hand, rows as well as columns are permuted, then the same restrictions as discussed above for square matrices are imposed on the use of the raw and normalized statistics.

Examples

To show the usefulness of rowwise indexes in matrix comparison tasks, some examples will be presented and the rowwise indexes will be compared with the non-rowwise alternatives Z, R, and K. R is Mantel's statistic calculated on the within-matrix ranks of the two matrices, and $K = \sum \text{sign}(X_{ij} - X_{kl}) \text{ sign}(Y_{ij} - Y_{kl})$, where summation is over all distinct pairs of cells ($i < k$ or ($i = k$ and $j < l$)), excluding the diagonal cells when the matrices are square with undefined diagonal (see Dietz, 1983). It is helpful to distinguish between the following types of proximity data (slightly modified from Shepard, 1972):

- I. Proximity data, based on a symmetric measure.
  - A. Square matrix; rows and columns correspond to the same n objects. The matrix is symmetric.
  - B. Rectangular matrix; rows and columns correspond to different objects.
- II. Proximity data, based on an asymmetric measure. (Shepard used the term "dominance data" for this category; Hubert, 1987, p. 121 is followed in the use of the term "proximity" as referring to any measure of relationship for object pairs.)
  - A. Square matrix; rows and columns correspond to the same n objects. The matrix is in general asymmetric.
  - B. Rectangular matrix; rows and columns correspond to different objects. (This category is not distinguished by Shepard.)

The rowwise approach is especially useful if both matrices belong to one of the categories IB, IIA, or IIB, or if one of the two matrices belongs to category IIA and the other to category IA, because in all these cases we can fruitfully distinguish between a rowwise conjecture of similar patterning and one specified columnwise.

Rectangular matrices. It is instructive to start with an example in which both matrices belong to category IIB. Suppose we are studying the social behavior of nine monkeys living in a social group comprising five males a, . . . , e, and four females f, . . . , i. Let **X** denote the artificial matrix containing frequencies of aggressive acts directed from the males a, . . . , e towards the females f, . . . , i. Similarly, matrix **Y** contains the frequencies of grooming acts. Consider two possible conjectures of a similar patterning of the matrix entries: (a) whether aggressive behavior shown by each of the males towards the females correlates with the grooming behavior shown by these males towards the females (i.e., whether the rows of **X** are correlated with the corresponding rows of **Y**). This is the rowwise conjecture. (b) Whether aggressive behavior received by each of the females from the males correlates with the grooming behavior received by these females from the males (i.e., whether the columns of **X** are correlated with the corresponding columns of **Y**). This is the columnwise conjecture. To test the rowwise conjecture, the three different test statistics (Z_r, R_r, and K_r) have been applied to the matrices **X** and **Y**:

$$\mathbf{X} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array}\begin{array}{cccc} f & g & h & i \\ \hline 11 & 3 & 6 & 7 \\ 20 & 1 & 10 & 15 \\ 22 & 2 & 1 & 3 \\ 6 & 1 & 5 & 2 \\ 12 & 0 & 11 & 10 \end{array}, \quad \mathbf{Y} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array}\begin{array}{cccc} f & g & h & i \\ \hline 13 & 1 & 2 & 3 \\ 10 & 1 & 1 & 5 \\ 8 & 2 & 1 & 6 \\ 11 & 3 & 4 & 7 \\ 12 & 4 & 0 & 5 \end{array}.$$

To apply the R_r test, the matrices **X** and **Y** are first transformed to rank numbers, ranked within the rows, giving the matrices **S** and **T**. Subsequently a Z_r test is applied to this pair of matrices.

$$\mathbf{S} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array}\begin{array}{cccc} f & g & h & i \\ \hline 4 & 1 & 2 & 3 \\ 4 & 1 & 2 & 3 \\ 4 & 2 & 1 & 3 \\ 4 & 1 & 3 & 2 \\ 4 & 1 & 3 & 2 \end{array}, \quad \mathbf{T} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array}\begin{array}{cccc} f & g & h & i \\ \hline 4 & 1 & 2 & 3 \\ 4 & 1.5 & 1.5 & 3 \\ 4 & 2 & 1 & 3 \\ 4 & 1 & 2 & 3 \\ 4 & 2 & 1 & 3 \end{array}.$$

The results of the Z_r, R_r, and K_r test are presented in Table 1. Calculated for each pair of corresponding rows in the two matrices is the weight, the correlation, and the product of weight and correlation, which is the contribution of this pair of rows to the statistic. Thus, it is directly visible how much each row subject contributes to the total value of the statistic. ($\tau_{rw,av}$ and τ_{rw} coincidentally have the same value; this is not true in general).

Comparing the different tests, the following can be noted. The weights w_i in the K_r test are more or less proportional to the weights in the R_r test, the last being about 3.3 times as large as the former. Oppositely, the Z_r weights are clearly disproportional to those in the R_r and K_r test. These differences reflect the ways in which the different statistics incorporate the variabilities in the pairs of rows. It is up to the investigator to judge the plausibility of the different weightings employed implicitly by the statistics. Note, for instance, that with the Z_r test, and in spite of the correlation between the **e**-rows being much smaller than the one between the **d**-rows, female **e** contributes more to the value of Z_r than female **d**, since the weight (the variance) of the **e**-rows is much higher than for the **d**-rows. Based on the outcomes of the R_r and K_r test, a positive rowwise correlation is established at a one-tailed significance level of .04 (two-tailed: .08), and one may thus conclude (if the test was one-tailed): males show more frequently aggression towards those females whom they groom more often.

The columnwise conjecture was also evaluated: the values of the three columnwise statistics and their respective right-tail and left-tail probabilities (based on 2000 permutations of the rows and columns) are: $Z_c = -332$ ($P_r = 0.99$, $P_l = .010$); $R_c = -128.75$ ($P_r = 1.00$, $P_l = .006$); $K_c = -22$ ($P_r = 1.00$, $P_l = .005$). All three statistics indicate a negative columnwise correlation at a one-tailed significance level smaller than .01 (two-tailed: .02), thus allowing us to conclude: females receive more aggression from those males by whom they are less frequently groomed.

How does this application of rowwise (and columnwise) indices compare to the use of the nonrowwise indices Z, R, and K? If the permutational method is used in which only the columns are permuted, the Z index is equivalent to the Z_r index, because Z is a linear transformation of Z_r under any permutation of the columns, which can be shown as follows. For each row i, the following identity holds (see Kendall, 1962, p. 21):

TABLE 1

Results of Three Different Rowwise Matrix Correlation Tests Applied to the Matrices **X** and **Y** in the Text. All Significance Levels are Based on 2000 Permutations of the Rows and Columns Independently.

	Z_r test			R_r test			K_r test		
rows	w_i	r_i	$w_i r_i$	w_i	ρ_i	$w_i \rho_i$	w_i	τ_i	$w_i \tau_i$
a	220.5	.92	203	20	1.00	20	6	1.00	6
b	415.4	.87	362	19	.95	18	5.5	.91	5
c	397.8	.80	320	20	1.00	20	6	1.00	6
d	102.7	.60	62	20	.80	16	6	.67	4
e	333.1	.25	83	20	.40	8	6	.33	2

	Z_r test	R_r test	K_r test
Sum of the contributions	Z_r =1030	R_r =82	K_r = 23
Weighted average of the correlations	$r_{rw,av}$ =0.701	$\rho_{rw,av}$ =0.829	$\tau_{rw,av}$ = 0.780
Rowwise matrix correlation	r_{rw} =0.593	ρ_{rw} =0.828	τ_{rw} = 0.780
Right-tail probability	P_r =0.153	P_r =0.035	P_r = 0.037
Left-tail probability	P_l =0.848	P_l =0.979	P_l = 1.00

$$\sum_{j<k} (X_{ij} - X_{ik})(Y_{ij} - Y_{ik}) = q\sum_{j} X_{ij} Y_{ij} - X_{i.} Y_{i.},$$

where $X_{i.}$ is the total of row i. By summing across all rows,

$$\sum_{i}\sum_{j<k} (X_{ij} - X_{ik})(Y_{ij} - Y_{ik}) = q\sum_{i}\sum_{j} X_{ij} Y_{ij} - \sum_{i} X_{i.} Y_{i.},$$

which can be rewritten as

$$Z_r = qZ - Z_{rtot},$$

where Z_{rtot} denotes the sum of crossproducts of the row totals (i.e., the unnormalized Pearson correlation between the row totals). The equivalence between Z and Z_r follows from the observation that the last right-hand term is constant under any permutation of the columns. It follows directly that Z is not equivalent with Z_r when both the rows and columns are permuted. When this permutational procedure is used in assessing the significance of Z, two different types of correlational information that can be extracted from the two matrices are confounded, because $Z = 1/q\,(Z_r + Z_{rtot})$, namely the rowwise correlation, Z_r, and the correlation between the row totals, Z_{rtot}. In fact, at the same time, the columnwise correlation, Z_c, is confounded with the correlation between the column totals, Z_{ctot}, because $Z = 1/p\,(Z_c + Z_{ctot})$. Thus, Z

TABLE 2

Results of Three Different Rowwise and Columnwise Matrix Correlation Tests Applied to Two Asymmetric Matrices.

Rowwise conjecture	$Z_r = 1159$		$R_r = 86$		$K_r = 25$	
	.164	.840	.007	1.00	.007	1.00
Columnwise conjecture	$Z_c = 365$		$R_c = -36$		$K_c = -11$	
	.233	.771	.933	.082	.989	.042

Note: the right-tail and left-tail probabilities are given below the value of each statistic.

confounds four different types of correlational information when the permutational method of permuting both rows and columns is used.

For R and K, similar identities as shown for Z above do not hold, and therefore, these indexes are not equivalent with their rowwise counterparts even when only the columns are permuted. For instance, when the significance of R is assessed (which is equal to 2492.75) by generating 2000 permutations of the columns, a p_r-value of .143 results, which clearly differs from the p_r-value of .039 resulting from the R_r test. Other examples of applying the K_r index to rectangular matrices can be found in Hemelrijk (1990a).

Square, asymmetric matrices. A second example involves two proximity matrices that both belong to category IIA. For the same group of monkeys as above, compare aggressive with grooming behavior among all the males. Matrices **X** and **Y** are now two square, asymmetric proximity matrices containing interaction frequencies of the two types of behavior performed by each male (the row individuals) towards all other males (the column individuals).

$$\mathbf{X} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \left[\begin{array}{ccccc} * & 1 & 2 & 3 & 4 \\ 40 & * & 4 & 20 & 12 \\ 8 & 2 & * & 6 & 7 \\ 5 & 3 & 4 & * & 5 \\ 12 & 4 & 0 & 5 & * \end{array}\right] \end{array}, \quad \mathbf{Y} = \begin{array}{c} \\ a \\ b \\ c \\ d \\ e \end{array} \begin{array}{c} \begin{array}{ccccc} a & b & c & d & e \end{array} \\ \left[\begin{array}{ccccc} * & 3 & 6 & 7 & 8 \\ 20 & * & 10 & 15 & 13 \\ 22 & 2 & * & 3 & 4 \\ 23 & 1 & 5 & * & 6 \\ 12 & 0 & 11 & 10 & * \end{array}\right] \end{array}.$$

Again, two different conjectures can be distinguished: (a) rowwise: is aggressive behavior shown by each of the males towards the other males correlated with the grooming behavior shown by these males towards the other males; and (b) columnwise: is aggressive behavior received by each of the males from the other males correlated with the grooming behavior received by these males from the other males. Table 2 presents the p-values for the different test statistics, based on 2000 permutations of the rows and columns simultaneously.

As noted above in the former example, when the permutational method is used in which both the rows and the columns are permuted, Z confounds four different types of correlational information: the rowwise correlation Z_r, the correlation between the row totals Z_{rtot}, the columnwise correlation Z_c, and the correlation between the column totals Z_{ctot}. Indeed, for square $n \times n$ matrices, a multiple of Z is equal to the sum

TABLE 3

Values and Probabilities of Z and Four Different Types of Correlational Information Extracted From Two Asymmetric Matrices.

	Z	Z_r	Z_{rtot}	Z_c	Z_{ctot}
Value	1952	1159	6649	365	7443
P_r	.041	.164	.051	.233	.049

Note: right-tail probabilities are based on 2000 permutations of rows and columns simultaneously.

of these four different indices: $2(n - 1)Z = Z_r + Z_{rtot} + Z_c + Z_{ctot}$. Table 3 presents the values and right-tail probabilities for all these statistics. The relatively extreme value of Z is seen to be due to the strong correlation between the row totals as well as between the column totals.

It is worth noting that there are yet other types of information that can be extracted from two square proximity matrices. Each square matrix, say **X**, can be decomposed into the sum of a symmetric and skew-symmetric matrix:

$$\mathbf{X} = \mathbf{X}^+ + \mathbf{X}^-,$$

where $X_{ij}^+ = (X_{ij} + X_{ji})/2$ and $X_{ij}^- = (X_{ij} - X_{ji})/2$ for all i and j (see Hubert, 1987). As shown by Hubert (p. 229),

$$Z_{\mathbf{XY}} = Z_{\mathbf{X}^+\mathbf{Y}^+} + Z_{\mathbf{X}^-\mathbf{Y}^-},$$

from which it follows that $Z_{\mathbf{XY}}$ confounds the symmetric and skew-symmetric information when **X** and **Y** are both asymmetric. In that case, the symmetric and skew-symmetric components should be dealt with separately, just as for the rowwise and columnwise components.

Square matrices one of which is asymmetric. Our next example is quite similar to the one presented by Hubert (1987, pp. 130–131) and involves two proximity matrices, one of which belongs to category IIA and the other to category IA. Matrix **X** contains the frequencies of initiating an interaction among five persons seated around a table. This matrix is to be compared with the hypothesized structure matrix **Y**, in which 0's identify adjacencies and 1's identify nonadjacencies. The (artificial) matrix **X** shows fairly large differences in tendencies among the five persons to initiate an interaction; these differences can be controlled for by using a rowwise statistic.

Person receiving interaction

Person initiating interaction	1	2	3	4	5
1	*	21	22	24	23
2	22	*	23	24	25
3	45	34	*	30	46
4	13	14	15	*	14
5	22	23	24	19	*

,

	1	2	3	4	5
1	*	0	1	1	1
2	0	*	0	1	1
3	1	0	*	0	1
4	1	1	0	*	0
5	1	1	1	0	*

.

TABLE 4

Results of Three Different Tests to Evaluate Three Types of Conjectures of Similar Patterning of Entries Across Two Square Matrices One of Which is Asymmetric.

Rowwise conjecture	$Z_r = 76$		$R_r = 44$		$K_r = 11$	
	.011	1.00	.043	1.00	.043	1.00
Columnwise conjecture	$Z_c = 44$		$R_c = 12$		$K_c = 3$	
	.288	.738	.308	.804	.308	.804
Symmetric conjecture	$Z_{X+Y+} = Z_{XY} = 305$		$R_{X+Y+} = 2345$		$K_{X+Y+} = 28$	
	.215	.820	.213	.860	.222	.837

Note: the right-tail and left-tail probabilities are given below the value of each statistic.

Here also, one can distinguish between a rowwise and a columnwise conjecture: (a) persons initiating interactions will do so more to persons who sit opposite than to persons next to them; and (b) persons receiving interactions will receive more interactions from persons who sit opposite than from persons next to them. A third conjecture, which will be called *symmetric* to distinguish it from the rowwise and columnwise conjectures, states that persons interacting with each other will more often sit opposite than next to each other, and can be tested with the Z statistic if the matrix **X** contains the frequencies of interactions without taking the direction of the interaction into account. Of course, this matrix can easily be computed by summing all pairs of corresponding entries across the main diagonal and thus, equals twice the $\mathbf{X}^+$ matrix. To evaluate this symmetric conjecture one may as well calculate the Z index on the original matrices (as did Hubert, 1987), because Z_{XY} is equal to $Z_{X^+Y^+}$ for pairs of matrices, one of which is symmetric (Hubert, 1987, p. 229). Analogous identities for R and K do not hold: $R_{XY} \neq R_{X^+Y^+}$ and $K_{XY} \neq K_{X^+Y^+}$ (for the two matrices above: $R_{XY} = 2330$, $R_{X^+Y^+} = 2345$; $K_{XY} = 25$, $K_{X^+Y^+} = 28$). To test the symmetric conjecture using one of these indexes calculate R and K on the matrices $\mathbf{X}^+$ and $\mathbf{Y}^+$, but not on the original matrices X and Y. Table 4 presents the outcomes of the different statistics used to test each of these three different conjectures.

Square, symmetric matrices. In the final example we apply rowwise statistics to a pair of square, symmetric matrices. Dietz (1983) applied several matrix association tests to a pair of genetic and anthropometric distance matrices presented by Spielman (1973). Besides the Z, R, and K tests, she also applied the K_r test (in fact, she used a statistic called K_c, that is equivalent to K_r because the matrices are symmetric), and established a large difference between the p-value of the K_r test (0.006) and the p-value of the K test (0.077). This difference is apparently due to the fact that the K_r test deals with the large differences that exist among the row means, which the K test does not. We performed tests employing the two other rowwise statistics, yielding also significant outcomes: the estimated p-values for Z_r and R_r were 0.041 and 0.004, respectively. Here also, we note large differences with the outcomes of the nonrowwise Z and R tests (p-values of 0.201 and 0.065, respectively). An improved matrix correlation method that controls for the differences among the object totals, while explicitly reckoning with the symmetry of the matrices, is proposed in de Vries (1992).

The Partial Rowwise Correlation Between Two Matrices Controlled for a Third

Smouse et al. (1986) have recently developed multiple regression and correlation extensions of Mantel's statistic Z, utilizing the concept of control of "adjusting for the control variable". All these extensions apply to the rowwise Z_r as well, because the rowwise matrix correlation r_{rw} is equal to $r(\mathbf{X}', \mathbf{Y}')$, which is in fact the normalized Mantel's Z calculated on the matrices $\mathbf{X}'$ and $\mathbf{Y}'$ containing the deviates from the row-means ($X'_{ij} = X_{ij} - \bar{X}_i$; $Y'_{ij} = Y_{ij} - \bar{Y}_i$).

With Mantel's Z approach, the partial correlation between the matrices $\mathbf{X}$ and $\mathbf{Y}$ controlling for the matrix $\mathbf{Q}$ can be calculated according to the usual partial correlation formula, and the same is true for the partial rowwise correlation $r_{rw;\mathbf{XY.Q}}$. Moreover, this partial rowwise correlation can easily be seen to be the partial correlation between the entries of $\mathbf{X}$ and $\mathbf{Y}$ controlling for both R and $\mathbf{Q}$ (R being the nominal variable with the row objects as categories):

$$r_{rw;\mathbf{XY.Q}} = \frac{r_{rw;\mathbf{XY}} - r_{rw;\mathbf{XQ}} r_{rw;\mathbf{YQ}}}{((1 - r^2_{rw;\mathbf{XQ}})(1 - r^2_{rw;\mathbf{YQ}}))^{1/2}}$$

$$= \frac{r_{\mathbf{XY}.R} - r_{\mathbf{XQ}.R} r_{\mathbf{YQ}.R}}{((1 - r^2_{\mathbf{XQ}.R})(1 - r^2_{\mathbf{YQ}.R}))^{1/2}}$$

$$= r_{\mathbf{XY}.R\mathbf{Q}}.$$

From this observation we can conclude that the *partial rowwise* correlation coefficient is indeed appropriately defined, and can be generalized when additional matrices have to be partialled out.

Because the rowwise rank-correlation ρ_{rw} is identical to r_{rw} calculated on the matrices $\mathbf{S}$ and $\mathbf{T}$ (which contain the rank-transformed values of $\mathbf{X}$ and $\mathbf{Y}$, ranked within the rows), the partial rowwise correlation coefficient $\rho_{rw;\mathbf{XY.Q}}$ is identical to $r_{rw;\mathbf{ST.U}}$ (with $\mathbf{U}$ being the rowwise rank-transformed matrix of $\mathbf{Q}$), and is thus appropriately defined. Continuing, the formula for Kendall's rowwise correlation τ_{rw} is similar in form to the one for Pearson's rowwise correlation r_{rw} (in fact, both are special forms of the generalized correlation coefficient Γ_{rw} of the Introduction), and because Kendall's partial correlation formula is identical in form to Pearson's partial correlation formula, Kendall's partial rowwise correlation coefficient, $\tau_{rw;\mathbf{XY.Q}}$, is also identical in form to Pearson's; thus:

$$\tau_{rw;\mathbf{XY.Q}} = \frac{\tau_{rw;\mathbf{XY}} - \tau_{rw;\mathbf{XQ}} \tau_{rw;\mathbf{YQ}}}{((1 - \tau^2_{rw;\mathbf{XQ}})(1 - \tau^2_{rw;\mathbf{YQ}}))^{1/2}}.$$

Significance Testing

To evaluate the significance of the different partial rowwise correlations, the procedure described by Hemelrijk (1990b, appendix; see also Maghsoodloo, 1975) can be used. Thus, for each random permutation of the columns and, depending on the type of the matrices (see above), possibly also the rows of $\mathbf{X}$ and independently also of $\mathbf{Y}$, while keeping the control matrix $\mathbf{Q}$ fixed, the values of the correlation coefficients $r_{rw;\mathbf{XY}}$, $r_{rw;\mathbf{XQ}}$ and $r_{rw;\mathbf{YQ}}$ are computed and entered into the partial correlation formula. Whenever $r^2_{rw;\mathbf{XQ}} = 1$ or $r^2_{rw;\mathbf{YQ}} = 1$ the partial correlation is undefined and therefore cannot be included into the reference set of partial correlation values. Finally, the significance of the observed value of the partial correlation is evaluated in the usual way by determining how extreme this value is relative to the values in the reference set. Exactly the same procedure can be used for the other two partial rowwise correlation coefficients $\rho_{rw;\mathbf{XY.Q}}$ and $\tau_{rw;\mathbf{XY.Q}}$.

Conclusion

In this paper three rowwise matrix correlation coefficients (r_{rw}, ρ_{rw} and τ_{rw}), each a specific form of the generalized rowwise correlation coefficient Γ_{rw}, are proposed as alternatives to the nonrowwise indexes Z, R, and K in matrix comparison tasks involving two proximity matrices (square or rectangular). For testing purposes, the raw indexes Z_r, R_r, and K_r, being the numerators of the respective rowwise correlation coefficients, may be used as well, because the denominators of the coefficients are constant under the permutation procedures applied (the simultaneous permutation of rows and columns for square matrices and the permutation of columns or the independent permutation of rows and columns for rectangular matrices).

In contrast with the nonrowwise indices Z, R, and K, which are based on all comparisons of pairs of cells across the two matrices, their rowwise alternatives are based on comparisons of pairs of cells within the rows only, and as such provide a means of partialling out from the total correlation between the entries of the two matrices that which is due to across-row comparisons of pairs of cells. Thus, rowwise correlation measures provide a means to control for the differences among the row objects reflected in their row totals. When applied to asymmetric or rectangular proximity matrices, rowwise and columnwise (i.e., rowwise in the transposed matrices) indexes extract two different types of information from the pair of matrices, each referring to a clearly distinct conjecture of a similar patterning of entries across the two matrices. Rowwise and columnwise indices are pre-eminently suited for evaluating these two different types of conjectures.

References

Cheverud, J. M., Wagner, G. P., & Dow, M. M. (1989). Methods for the comparative analysis of variation patterns. *Systematic Zoology*, *38*, 201–213.

Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika*, *33*, 129–135.

de Vries (1992). *The object-controlled correlation between two symmetric proximity matrices*. Manuscript submitted for publication.

de Waal, F. B. M., & Luttrell, L. M. (1988). Mechanisms of social reciprocity in three primate species: Symmetrical relationship characteristics or cognition? *Ethology and Sociobiology*, *9*, 101–118.

Dietz, E. J. (1983). Permutation tests for association between two distance matrices. *Systematic Zoology*, *32*, 21–26.

Dow, M. M., Cheverud, J. M., & Friedlander, J. S. (1987). Partial correlation of distance matrices in studies of population structure. *American Journal of Physiological Anthropology*, *72*, 343–352.

Dunn, O. J., & Clark, V. A. (1987). *Applied statistics: Analysis of variance and regression* (2nd ed.). New York: Wiley.

Hemelrijk, C. K. (1990a). Models of, and tests for, reciprocity, unidirectionality and other social interaction patterns at a group level. *Animal Behaviour*, *39*, 1013–1029.

Hemelrijk, C. K. (1990b). A matrix partial correlation test used in investigations of reciprocity and other social interaction patterns at group level. *Journal of Theoretical Biology*, *143*, 405–420.

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

Kendall, M. G. (1962). *Rank correlation methods* (3rd ed.). London: Griffin.

Klauber, M. R. (1971). Two-sample randomization tests for space-time clustering. *Biometrics*, *27*, 129–142.

Krackhardt, D. (1988). Predicting with networks: Nonparametric multiple regression analysis of dyadic data. *Social Networks*, *10*, 359–381.

Maghsoodloo, S. (1975). Estimates of the quantiles of Kendall's partial rank correlation coefficient. *Journal of Statistical Computation and Simulation*, *4*, 155–164.

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*, 209–220.

Quade, D. (1974). Nonparametric partial correlation. In H. M. Blalock (Ed.), *Measurement in the Social Sciences: Theories and strategies* (pp. 369–398). London: Macmillan Press.

Schnell, G. D., Watt, D. J., & Douglas, M. (1985). Statistical comparison of proximity matrices: Applications in animal behaviour. *Animal Behaviour*, *33*, 239–253.

Shepard, R. N. (1972). A taxonomy of some principal types of data and of multidimensional methods for their analysis. In R. N. Shepard, A. K. Romney, & S. B. Nerlove (Eds.), *Multidimensional scaling: Theory and applications in the behavioral sciences Volume I* (pp. 21–47). New York: Seminar Press.

Smouse, P. E., Long, J. C., & Sokal, R. R. (1986). Multiple regression and correlation extensions of the Mantel test of matrix correspondence. *Systematic Zoology*, *35*, 627–632.

Spielman, R. S. (1973). Differences among Yanomama Indian villages: Do the patterns of allele frequencies, anthropometrics and map locations correspond? *American Journal of Physiological Anthropology*, *39*, 461–480.

Manuscript received 7/23/91
Final version received 3/20/92