A COMPARISON OF TWO PROXIMITY MATRICES*

Min Ning Yu

余　民　寧**

摘　要

本文的目的旨在提出一種資料分析策略的一般性理論架構稱作二次式分配準則（quadratic assignment paradigm, 簡寫爲QAP），以考驗兩個方形近似矩陣（proximity matrices）之間的相似性。本文同時提出如何運用QAP技術的詳細計算過程與兩個應用實例，對於QAP技術的應用範圍亦加以扼要地評論。

Abstract

The purpose of this paper is to present a general framework of data-analytic strategies, referred to as the quadratic assignment paradigm (QAP), that can be used to test a hypothesis that two or more proximity matrices on the same set of objects reflecting a similar patterning of high and low entries. The detailed computation procedure for executing the QAP technique is provided. Two empirical examples showing how the QAP technique is applied are given. Comments on the applications of the QAP technique are also summarized.

I. Introduction

One difficulty faced by all behavioral and social scientists in analysing data concerns the choice of formal techniques that are intended to be of aid in developing reasonable substantive interpretations. Most of the case the final selection of a statistical tool is guided either by tradition in the researcher's field or by one particular procedure happens to be in vague. In either case, the chosen methodology may not be the most appropriate way to answer the specific questions proposed by the behavioral and social scientists. The most difficult in chosing a statistical tool becomes even more acute when a research problem cannot be easily delineated

---

* Based on a paper presented at the Seminar of Graduate School of Education, National Chengchi University, Taipei, Taiwan, R.O.C., held on December 11, 1991.

**作者現為教育系副教授及兼代實驗小學校長

within an omnipotent general linear model (cf. Cohen, 1968a, 1982), since there are very few alternative paradigms that are broad enough to formulate sufficiently powerful analyses. Consequently, because of the general inflexibility of statistical schemes that do not rely on rather strong parametric assumptions, novice researchers tend to limit the questions they ask to those that fit neatly within the analysis of variance context and its derivations, or, alternatively, embrace some other familiar strategy that may not be suitable for the particular application at hand.

Readers may be familiar with many, if not most, data-analysis strategies which are based on sums of squared differences. As is well-known, sums of this type appear in the definition of a variance, in least-squares regression, in chi-square goodness-of-fit test, and so on. In fact, concepts derived from the sums of squared differences are the basic forms of cross-products which used in comparing among two to k numerical sequences (Cohen, 1968b; Fleiss, Cohen, & Everitt, 1969; Friedman, 1937; Hubert, 1977, 1978, 1979a, 1979b; Kendall, 1970; Page, 1963). For a general framework of the sequence-comparison, this data-analytic strategy can be expanded to comparisons among two to k matrices. Particularly, such matrices containing correlational measures on the same set of variables are encountered frequently in education, biometrics, geography, social and behavioral sciences. Those matrices are usually referred to proximity or preference matrices (Golledge & Rayner, 1982; Schultz & Hubert, 1976), similarity or distance matrices (Borg & Lingoes, 1987), and sociometric matrices or sociograms (Hubert & Baker, 1978). Depending upon the researcher's statistical preferences, these comparisons may take the form of searching for comparable common factors or dimensions (Young, 1987), a related analysis (Kendall, 1970), a multivariate parametric test (Morrison, 1990), and so on. Usually, the entries are compared only informally with some heuristic evaluation made as to whether the same general pattern of high and low values appears in both or many datasets.

The purpose of this paper is to present a detailed strategy, referred to as the quadratic assignment paradigm (QAP), that can be used to test a hypothesis that two or more proximity matrices on the same set of variables or objects reflecting a similar pattern of high and low entries (Hubert, 1987; Hubert & Arabie, 1989; Hubert & Schultz, 1976; Mantel, 1967). Although the actual strategy has been available for some time in the biometrics and geography literature, this technique described below is more generally applicable than suggested by its original biometric and geographic motivations and deserves a wider recognition in the behavioral and social sciences or in the educational researches. The approach taken below appears flexible enough to give a overview of many problems a researcher faces in chosing

an appropriate methodology, and, more importantly, is broad enough to provide a general framework of organizing principles for an extensive theoretical analysis of structure within a proximity matrix. For the parsimonious reason, the scope of this paper is limited only in the comparison between two n x n symmetric or asymmetric proximity matrices, where n is the number of objects in a given set. Those n x r rectangular matrices are not our concerns.

II. Formulation of the Quadratic Assignment Paradigm

Background of theory

It is assumed that the analytic data is collected on n objects that are denoted by $o_1, o_2, \ldots, o_n$. The term "object" is meant to be extremely general and could refer to individuals, stimuli, societies, tests, and so on. Furthermore, it is assumed that data collected from each ordered object pair of these objects can be reduced to a single numerical value, a statistic, or an index. Our task is to look for an appropriate procedure to maximize such an index which contains numerical measures of relationship or proximity between each pair of objects from the set of S:$\{o_1, o_2, \ldots, o_n\}$. The concept of "proximity" is intended to be very broad and will be used as a generic term for a host of possible measurements of relationship that could be considered. For technical and operational convenience, proximity is assumed to be positive symmetric measure for which the larger numerical values are assigned to the more dissimilar (or distant) object pairs, or an non-negative real-valued function for which elements are defined by the researcher; e.g., in the usual correlational context, proximity could be defined by one minus the standard Pearson product-moment correlation coefficient.

To formulate an index of numerical measures of relationship between two matrices, two n x n square matrices are defined on the Cartesian product S x S, where S is the set $\{o_1, o_2, \ldots, o_n\}$. One is the *data matrix* **X** whose i^{th} row and i^{th} column refer to object o_i, and the entry in row i and column j is defined by the researcher as X_{ij}. For convenience, it is assumed that $X_{ii} \equiv 0$ for $1 \leq i \leq n$, and usually, $X_{ij} \geq 0$ for $1 \leq i, j \leq n$. In general, X_{ij} is defined to be a non-negative real-valued function on the Cartesian product S x S, where S is the set $\{o_1, o_2, \ldots, o_n\}$; furthermore, if the two arguments in X_{ij} are the same, the function value is defined to be zero. The other is the *structure matrix* **Y** whose rows and columns are labelled by the integers 1, 2,..., n, and if N: {1, 2,..., n}, the

entries in **Y** are the values defined by a non-negative real-valued function on N x N. In particular, Y_{ij} is the entry in row i and column j of **Y** and $Y_{ii} \equiv 0$ for $1 \leq i \leq n$.

These two basic matrices may be two symmetric (Hubert & Levin, 1976a) or asymmetric (Hubert & Schultz, 1976) n x n matrices which provide numerical measures between each pair of objects. The data matrix **X** is the user-collected matrix. The structure matrix **Y** is the matrix that represents the type of hypothesis the researcher wishes to evaluate *against* his/her data, or, alternatively, the type of structure he/she may wish to identify *in* his/her data. In other words, the structure matrix **Y** is defined in some *a priori* manner by the researcher from theorectical considerations regarding the structure of the elements in S. As technical assumptions, the diagonal elements in **Y** and **X** (i.e., Y_{ii} and X_{ii} for all o_i) are assumed to be zero. Typically, all off-diagonal entries in **Y** and **X** will be non-negative. Our first task is to index the correspondence between Y and X; the second is to develop a statistical technique for evaluating whether the two basic matrices have a similar patterning of high and low entries.

As an introductory example of how the data and structure matrices can be interpreted, assume that a researcher has a set of n statements reflecting various psychological symptoms, and he/she believes that these statements can be ordered along a hypothetical continuum from neurotic to psychotic. A subject is given each pair of statements once and is asked to rate their similarity from, say, 1 to 5, with 5 being the least similar. Consequently, the data from this single subject can be put into a data matrix **X**, where $X_{ij}=X_{ji}$=rating given to the pair (o_i, o_j). The structure matrix **Y** is defined to reflect the contention that the n items or objects can be ordered along a continuum, and where distances between the items are reflected perfectly in the ratings. To be more explicit, assume that the hypothesized ordering is now represented by the sequence $\{o_i,\ldots,o_n\}$. Then, if the proposed seriation is reasonable, the entries in **X** should satisfy the relation

$$X_{ij} \leq X_{ij'} \text{ and } X_{ij} \leq X_{i'j},$$

whenever $i \leq j$; $j \leq j'$; $i' \leq i$. Similarly, a structure matrix **Y** is defined by a particular idealization of this necessary property of the measure X_{ij}; for instance, in the seriation case,it would be possible to define $Y_{ij}=|i-j|$, since $Y_{ij} \leq Y_{ij'}$ and $Y_{ij} \leq Y_{i'j}$, whenever $i \leq j$; $j \leq j'$; $i' \leq i$. The same approach taken for seriation is easily generalized to other situations by hypothesizing a specific underlying organization of the objects $o_1,\ldots,o_n$ that in turn implies some patterning of the entries in the data matrix **X**. The same structure is then related to the matrix **Y** by

identifying an idealized pattern of entries that satisfies perfectly the necessary implication on the **X** matrix.

The most obvious measure of relationship between **Y** and **X** would be a simple Pearson product-moment correlation between the corresponding n(n-1)/2 (if symmetric) or n(n-1)(if asymmetric) off-diagonal entries in **Y** and **X**. This expression, however, is rather cumbersome to deal with entirely and for our purpose it is sufficient (and equivalent statistically) to consider only the cross-product term from the numerator. Particularly, if we define the index by Γ (see also Hubert & Levin, 1976a, 1976b; Hubert & Schultz, 1976; Schultz & Hubert, 1976):

$$\Gamma = \Sigma\Sigma X_{ij} Y_{ij} \tag{2.1}$$

then Γ means the sum of the cross-products of the corresponding elements between **X** and **Y**. Thereafter, Γ can be used as an index of correspondence between **X** and **Y**. Γ may be interpreted as an "unnormalized correlation coefficient".

A rather simple approach can be developed through a randomization model common in nonparametric statistics that is analogous to the justification for Fisher's exact test (Fisher, 1934) in contingency table analysis. Specifically, the matrix **Y** is assumed to be fixed as it is; the rows and simultaneously the columns of **X**, however, are reordered in all possible ways, generating n! possible matrices that contain the same entries as **X** but possibly in different positions. Furthermore, the Γ statistic is computed for each of the n! reordered matrices developed from **X** by using the original fixed **Y** matrix as a target. The index Γ measures the degree to which the patterning of the entries in **Y** corresponds to the patterning of the entries in **X**. Hence, Γ attains a maximum value when the elements in **X** and the corresponding elements in **Y** are in a perfect monotone relationship. Specifically, if it is true that $X_{ij} \geq X_{i'j'}$ if and only if $Y_{ij} \geq Y_{i'j'}$, then the index Γ is as large as possible.

As an illustration of these ideas using the seriation example, if the particular ordering of the objects $\{o_1, \ldots, o_n\}$ is hypothesized *a priori*, then the index Γ measures the degree to which this same ordering is reflected in the data; obviously, the hypothesis is perfectly validated through the **Y** matrix if the corresponding elements in **X** and **Y** are monotonically related, but in any event, the statistic Γ indexes the closeness of the relationship in the units of measurement implicitly defined by the functions X_{ij} and Y_{ij}.

From now on, it is obvious that our task is to find a permutation such that Γ is as large as possible. With this idea in mind, we are looking for a reordering

of the **X** matrix that fits closely to the fixed Y matrix. In any event, the **Y** matrix numerically characterizes a necessary property of the **X** matrix and, therefore, searching for a permutation that maximizes Γ is an attempt to force a necessary condition to hold as close as possible, where the index of "closeness" is given by the formula Γ. The search for these optimal permutations is called the quadratic assignment (QA) problem, and is an extremely active area of operations research introduced by Koopmans & Beckmann (1957).

Since Γ is interpreted as an unnormalized correlation coefficient, it may be useful to have a normalized statistic Z of the form

$$Z = \frac{\Gamma - E(\Gamma)}{Var(\Gamma)^{1/2}} \tag{2.2}$$

under the standard nonparametric hypothesis of randomness (i.e., the rows and the columns of the **X** matrix have the same chance of been permuting), where $E(\Gamma)$ is the expected value of Γ and $Var(\Gamma)$ is the expected variance of Γ. Because the type of relationship between **X** and **Y** indexed by Γ, Γ can be calculated and tabled to form an actual reference distribution from each of the n! permutations of the rows (and simultaneously, the columns) of the **X** matrix with the fixed **Y** matrix. Thus, if the original value of Γ is sufficiently extreme with respect to the probability distribution for Γ, the hypothesis of a random permutation is rejected and the value of the Γ statistic is assumed to denote a nontrivial degree of common patterning of high and low entries between the two original matrices **X** and **Y**. Furthermore, if a large-sample normal approximation were assumed sufficiently accurate, Z could be compared in the usual manner to a normal distribution with mean 0 and variance 1 to find the appropriate *p*-value for the observed index Γ. Therefore, Z has some of the same characteristics of an ordinary measure of association. But a great caution should be taken by any researcher who wishes to rely on the conjectured adequacy of a normal approximation when n is small.

Derivation of the permutational mean and variance of Γ

For whatever closeness (or similarity) or distance (or proximity) scales we may employ, let X_{ij} be some special measure between objects i and j (i.e., o_i and o_j), and Y_{ij} another measure. Our test statistic is then

$$\Gamma = \Sigma \ \Sigma \ X_{ij} \ Y_{ij}$$

which can be compared with its null expectation. With one restriction $i \neq j$ on the index Γ, we can replace X_{ii} and Y_{ii} by an arbitrary constants. Zeros are the most conveniently specified values for these constants. These zero constants (i.e., the diagonal elements in matrices **X** and **Y**) mean that the closeness or distance between object i and itself (i.e., o_i and o_i) is none and zeros are the appropriate indicators.

The null distribution of Γ can be obtained by a finite population approach. We have n objects in matrix **X** and n objects in matrix **Y**. The hypothesis of no closeness or correspondence between **X** and **Y** is equivalent to one that the objects in matrix **X** are matched at random with the objects in matrix **Y**, there being a total of n! equiprobable sets of matchings. In prin ciple we can list the n! possible permutations of our data, compute the Z statistic for each permutation, and obtain the null distribution of Γ against which the observed value of Γ can be judged.

When n is too large for the null permutational approach to be practicable, a Monte Carlo approach may be more feasible. In this approach one simulates the randomization scheme enough times to get an empirical distribution of Γ adequate for significance testing purposes. For a larger n an approach can be taken which is predicted on the assumption that Γ is approximately normally distributed, so that its deviation from its null expectation can be tested relative to its null standard deviation. Therefore, the formulas for the permutational expectation and variance of Γ are developed and a computational procedure for the routine evaluation of the departure of Γ from its null expectation is presented as follows (Mantel, 1967).

So now, we consider two n x n matrices, one of X_{ij}'s, the other of Y_{ij}'s, and both with zero diagonals. The quantities X_{ij} and Y_{ij} are two different measures relating the ith element of a sample (i.e., o_i in the set S) to the jth element of the same sample (i.e., o_j in the same set S). For purposes of obtaining the permutational distribution of $\Gamma = \Sigma\Sigma X_{ij}Y_{ij}$, we may arbitrarily keep the **X** matrix fixed while permuting only the rows and the columns of the **Y** matrix, or vice versa. But any permutation must follow the requirement that if any two rows are permuted, the corresponding two columns are also permuted so that, for each i, the ith row and the jth column will correspond to the same case. Under such permutations diagonal elements will remain diagonal elements, but off-diagonal elements will appear with equal probability in each of the n(n-1) off-diagonal positions. Regarding the **X** matrix as fixed so that only the Y_{ij}'s change from permutation to permutation, we can write

$$\begin{aligned} E(\Gamma) &= E(\Sigma\Sigma X_{ij}Y_{ij}) \\ &= \Sigma\Sigma X_{ij}E(Y_{ij}) \\ &= \Sigma\Sigma X_{ij}\Sigma\Sigma Y_{ij}/n(n-1) \end{aligned} \tag{2.3}$$

(Had the diagonal elements been constants other than zero, both Γ and $E(\Gamma)$ would have increased by identical amounts, leaving the deviation of Γ from its null expectation unchanged.)

We can write for the variance of Γ, the shorthand

$$Var(\Gamma) = \Sigma\Sigma X_{ij}X_{kl}Cov(Y_{ij}, Y_{kl}) \tag{2.4}$$

where summation is over-all permutations of the subscripts, although the cases $i=j$ or $k=l$ can be ignored.

We can write that, in general,

$$Cov\ (Y_{ij}, Y_{kl}) = E(Y_{ij}Y_{kl}) - E(Y_{ij})E(Y_{kl}) \tag{2.5}$$

where

$$E(Y_{ij}) = E(Y_{kl}) = \Sigma\Sigma Y_{ij}/n(n-1) \tag{2.6}$$

Using Equations (2.5) and (2.6) we can rewrite Equation (2.4) as

$$\begin{aligned} Var(\Gamma) &= \Sigma\Sigma X_{ij}X_{kl}Cov(Y_{ij}, Y_{kl}) \\ &= \Sigma\Sigma X_{ij}X_{kl}E(Y_{ij}Y_{kl}) - [\Sigma\Sigma X_{ij}X_{kl}(\Sigma\Sigma Y_{ij})^2/n^2(n-1)^2] \\ &= \Sigma\Sigma X_{ij}X_{kl}E(Y_{ij}Y_{kl}) - [(\Sigma\Sigma X_{ij})^2(\Sigma\Sigma Y_{ij})^2/n^2(n-1)^2] \end{aligned}$$

leaving yet to be resolved the value of

$$E(Y_{ij}Y_{kl}).$$

So our task becomes simple. We just need to consider seven possible cases and their corresponding computations as shown in Table 1 which may arise.

Case 1. Suppose $i=k$, $j=l$ so that $Y_{ij}Y_{kl}=Y_{ij}^2$. Since, on permutation, any of the n(n-1)(see Column 1 of Table1) off-diagnoal elements can appear with equal probability in a off-diagonal position, the average value of Y_{ij}^2 equals the total, $\Sigma\Sigma Y_{ij}^2$ (see Column 2 of Table 1) divided by n(n-1). The contribution of this case

to the first term of Var (Γ) as shown by Equation (2.7) is then $\Sigma\Sigma X_{ij}^2$ (see Column 3 of Table 1) times $\Sigma\Sigma Y_{ij}^2$ (Column 2) divided by n(n-1)(Column 1).

Case 3. Suppose $i=k$ and $j \neq 1$. For any of the n(n-1) off-diagonal Y_{ij} values, there are (n-2) off-diagonal, Y_{il}, elements in the same row, for a total of n(n-1)(n-2) (see Column 1) possible products of the form $Y_{ij}Y_{il}$. The total of the products for a given Y_{ij} is given by multiplying Y_{ij} by its own row total, the ith, and subtracting Y_{ij}^2. Overall i and j this sums to the total of the squares of the row sums less the sum of the squares of the individual Y_{ij}'s, that is, $\Sigma_i(\Sigma_j Y_{ij})^2 - \Sigma\Sigma Y_{ij}^2$ (Column 2). The corresponding **X** total is shown in Column 3 and again the contribution to Equation (2.7) is Column 2 x Column 3/Column 1.

Similarly, for each of the other cases (Case 2, $i=1$, $j=k$; Case 4, $i \neq k$, $j=1$; Case 5, $i=1$, $j \neq k$; Case 6, $i \neq 1$, $j=k$; Case 7, Y_{ij} and Y_{kl} have no subscripts in common), Table 1 shows the possible number of product pairs (Column 1), the sum of all possible product pairs (Column 2 for **Y**, Column 3 for **X**), these permitting determining the contribution of each case to the final variance.

In the bottom section of Table 1 the variance of Γ is sketched out mathematically and in terms of the totals necessary to be determined for each of the seven cases. By footnote, Table 1 also defines symbolically by various letter totals needed in the variance calculation. For a single matrix, most of these letters are the same as would be necessary for the analysis of variance of a two-way classification of data. Thus, A is the grand total of the matrix elements; B is the total of the squares of the elements; D is the total of the squares of the row sums; E is the total of the squares of the column sums; and G is the square of the grand total. Only two new totals are required in excess of these ordinarily arising in the analysis of variance. These are: C, the sum of the products of each element, Y_{ij}, with its transpose element Y_{ji}; F, the sum of the products of each row total with its transpose column total. These quantities, A to G, must be obtained separately for the X and Y matrices and from them, in turn, four additional quantities, H to K, are simply calculated as shown in the symbols column of Table 1.

Special Computations for the symmetric and asymmetric matrices

Table 1 also shows the simplications which occur for specific kinds of matrices. For the symmetric matrix, $Y_{ij}=Y_{ji}$, it is true that $B=C$, and that $D=E=F$. For the seven cases shown in Table 1, it is then true that 1 and 2 show identical totals ($B=C$), and that Cases 3 to 6 show identical totals ($H=I=J$). For the skew symmetric matrix, $Y_{ij}=-Y_{ji}$, we have that $A=G=0$, $B=-C$, $D=E=-F$. For this situation,

in Table 1, Cases 1 and 2 are inverse to each other ($B=-C$), Cases 3 and 4 show identical totals but are inverse to Case 5 and 6 ($H=I=-J$), and Case 7 disappears ($K=0$).

Table 2 shows the simpler computations that would obtain if the X_{ij}, and Y_{ij} matrices were both symmetric or both skew symmetric. Where only X_{ij} or only Y_{ij} shows symmetry, the general procedure may be desirable. The case where one matrix is symmetric, the other skew symmetric, needs not be considered since then Γ is identical zero. Such cases may not be too obvious where skew symmetric is concealed by an additive constant.

III. Two Illustrative Examples

In this section, two empirical examples are introduced to show how the usages of the quadratic assignment paradigm (QAP) are applied. The first example is to compare and to confirm similarities between two symmetric matrices with one similarity rating task and the other sorting task. The second is to compare and to identify differences between two asymmetric proximity matrices with one expert's knowledge structure and the other novice's knowledge structure.

Symmetric case

This example is taken from Gliner, Goldman,& Hubert (1983). Its purpose is to confirm the similarity from a multidimensional scaling of two proximity matrices with one similarity rating task and the other sorting task. In the similarity rating task, forty subjects were asked to rate each animal pair (of nine animal names) a number between 1 and 7, where 1 meant the animal pairs were very similar and 7 meant they were very different. Based on the matrix of similarity ratings of each subject, the total group similarity rating matrix was formed by adding the ratings for each subject. Therefore, a proximity matrix was formed with each entry ranged between 40 and 280, where numbers close to 40 meant the group saw the animal pair as very similar and larger numbers close to 280 meant the group viewed the animals in the pair as very different. In the sorting task, these forty subjects put animals into as many groups as they liked, based on similarity. Similarly, a proximity matrix for the sorting task was formed with ones or zeros on each entry, where one meant the animal pair was grouped together and zero otherwise. Hence the entries for the group matrix in the sorting task represented how many subjects sorted the animal pairs into the same group. The group matrices for the similarity rating task and sorting task are given in Figure 1.

Table 1. Symbolic Development of the the Variable of the Sum of Cross Products Based on All Possible Pairs

Case	Particular instances of $Y_{ij}Y_{kl}$ All indicated subscripts different	No. of such instances (1)	Symbolic representation of total for such instances*	Total of $Y_{ij}Y_{kl}$ for such instances (2)
1	$Y_{ij}Y_{ij} = Y_{ij}^2$	$n(n-1)$	B	$\Sigma\Sigma Y_{ij}^2$
2	$Y_{ij}Y_{ji}$	$n(n-1)$	C	$\Sigma\Sigma Y_{ij}Y_{ji}$
3	$Y_{ij}Y_{il}$	$n(n-1)(n-2)$	H = D - B	$\Sigma_i(\Sigma_j Y_{ij})^2 - \Sigma\Sigma Y_{ij}^2$
4	$Y_{ij}Y_{kj}$	$n(n-1)(n-2)$	I = E - B	$\Sigma_i(\Sigma_j Y_{ji})^2 - \Sigma\Sigma Y_{ij}^2$
5	$Y_{ij}Y_{ki}$	$n(n-1)(n-2)$	J = F - C	$\Sigma_i(\Sigma_j Y_{ij})(\Sigma_j Y_{ji}) - \Sigma\Sigma Y_{ij}Y_{ji}$
6	$Y_{ij}Y_{jl}$	$n(n-1)(n-2)$	J = F - C	$\Sigma_i(\Sigma_j Y_{ij})(\Sigma_j Y_{ji}) - \Sigma\Sigma Y_{ij}Y_{ji}$
7	$Y_{ij}Y_{kl}$	$n(n-1)(n-2) \cdot (n-3)$	K = G + B + C - D - E - 2F = G - B - C - H - I - 2J	$(\Sigma\Sigma Y_{ij})^2 + \Sigma\Sigma Y_{ij}^2 + \Sigma\Sigma Y_{ij}Y_{ji} - \Sigma_i(\Sigma_j Y_{ij})^2 - \Sigma_i(\Sigma_j Y_{ji})^2 - 2\Sigma_i(\Sigma_j Y_{ij})(\Sigma_j Y_{ji})$
Totals		$n^2(n-1)^2$	$G = A^2$	$(\Sigma\Sigma Y_{ij})^2$

* Definitions of symbols (shown for Y's only; apply correspondingly for X's).

$A_y = \Sigma\Sigma Y_{ij}$ = Grand total of the Y_{ij}'s.

$B_y = \Sigma\Sigma Y_{ij}^2$ = Sum of squares of the Y_{ij}'s.

$C_y = \Sigma\Sigma Y_{ij}Y_{ji}$ = Sum of products of each Y_{ij} with its transpose element.

$D_y = \Sigma_i(\Sigma_j Y_{ij})^2$ = Sum of squares of the row totals.

$E_y = \Sigma_i(\Sigma_j Y_{ji})^2$ = Sum of squares of the column totals.

$F_y = \Sigma_i(\Sigma_j Y_{ij})(\Sigma_j Y_{ji})$ = Sum of products of each row total with its transpose column total.

$G_y = A_y^2 = (\Sigma\Sigma Y_{ij})^2$ = Square of the grand total.

$E(\Sigma\Sigma X_{ij}Y_{ij}) = \Sigma\Sigma X_{ij}\Sigma\Sigma Y_{ij} / n(n-1) = A_x A_y / n(n-1)$.

$Var(\Sigma\Sigma X_{ij}Y_{ij}) = \Sigma\Sigma\Sigma\Sigma X_{ij}X_{kl} \cdot Cov(Y_{ij}, Y_{kl})$.

(Table 1 continued).

Case	Corresponding total of $X_{ij}X_{kl}$. (3)	Special cases of such totals (Shown for Y variable only)		
		Symmetric case $Y_{ij} = Y_{ji}$	Skew symmetric case $Y_{ij} = - Y_{ji}$	Symmetric 1's and 0's only $r_i = \Sigma_j Y_{ij}$; $r = \Sigma r_i$
1	$\Sigma\Sigma X_{ij}^2$	$B_y = \Sigma\Sigma Y_{ij}^2$	$B_y = \Sigma\Sigma Y_{ij}^2$	r
2	$\Sigma\Sigma X_{ij}X_{ji}$	B_y	$- B_y$	r
3	$\Sigma_i(\Sigma_j X_{ij})^2 - \Sigma\Sigma X_{ij}^2$	$H_y = D_y - B_y$ $= \Sigma_i(\Sigma_j Y_{ij})^2 - \Sigma\Sigma Y_{ij}^2$	H_y	$\Sigma r_i^2 - r$
4	$\Sigma_i(\Sigma_j X_{ji})^2 - \Sigma\Sigma X_{ij}^2$	H_y	H_y	$\Sigma r_i^2 - r$
5	$\Sigma_i(\Sigma_j X_{ij})(\Sigma_j X_{ji}) - \Sigma\Sigma X_{ij}X_{ji}$	H_y	$- H_y$	$\Sigma r_i^2 - r$
6	$\Sigma_i(\Sigma_j X_{ij})(\Sigma_j X_{ji}) - \Sigma\Sigma X_{ij}X_{ji}$	H_y	$- H_y$	$\Sigma r_i^2 - r$
7	$(\Sigma\Sigma X_{ij})^2 + \Sigma\Sigma X_{ij}^2 + \Sigma\Sigma X_{ij}X_{ji} - \Sigma_i(\Sigma_j X_{ij})^2 \Sigma_i(\Sigma_j X_{ji})^2 - 2\Sigma_i(\Sigma_j X_{ij}) \cdot (\Sigma_j X_{ji})$	$K_y = G_y + 2B_y - 4D_y$ $= (\Sigma\Sigma Y_{ij})^2 + 2(\Sigma\Sigma Y_{ij}^2) - 4[\Sigma_i(\Sigma_j Y_{ij})^2]$	$K_y = 0$	$r^2 + 2r - 4\Sigma r_i^2$
	$(\Sigma\Sigma X_{ij})^2$	$G_y = (\Sigma\Sigma Y_{ij})^2$	$G_y = 0$	$G_y = r^2$

$Cov(Y_{ij}, Y_{kl}) = E(Y_{ij}Y_{kl}) - E(Y_{ij})E(Y_{kl})$.

For all cases $E(Y_{ij})E(Y_{kl}) = [\Sigma\Sigma Y_{ij} / n(n - 1)]^2 = G_y / n^2(n - 1)^2 = \Sigma$ Column 2 / Σ Column 1.

For specific cases :

$E(Y_{ij}Y_{kl})$ = Total of such cases, Column 2 / No. of such cases, Column 1.

$Cov(Y_{ij}, Y_{kl})$ = Column 2 / Column 1 - Σ Column 2 / Σ Column 1.

$\Sigma X_{ij}X_{kl} \cdot Cov(Y_{ij}, Y_{kl})$ = Column 3 · [Column 2 / Column 1 - Σ Column 2 / Σ Column 1].

Totaling over all cases :

$$Var(\Sigma\Sigma X_{ij}Y_{ij}) = \Sigma\Sigma\Sigma\Sigma X_{ij}X_{kl} \cdot Cov(Y_{ij}, Y_{kl})$$
$$= \Sigma \text{ Column 2 Column 3 / Column 1} - \Sigma \text{ Column 2 } \Sigma \text{ Column 3 / } \Sigma \text{ Column 1.}$$
$$= [1/n(n - 1)] \{B_xB_y + C_xC_y + [H_xH_y + I_xI_y + 2J_xJ_y] / (n - 2) + [K_xK_y / (n - 2)(n - 3)] - [G_xG_y / n(n - 1)]\}.$$

Table 2. Calculating Procedures for the Symmetric and Skew Symmetric Cases

For both the symmetric and skew symmetric case, obtain the following 5 columns :

1. Row sum of X_{ij}
2. Row sum of X_{ij}^2
3. Row sum of Y_{ij}
4. Row sum of Y_{ij}^2
5. Row sum of $X_{ij}Y_{ji}$

Obtain the following derived statistics

Symmetric case	Skew symmetric case
$A_x = \Sigma$ Column (1), $A_y = \Sigma$ Column (3)	
$B_x = \Sigma$ Column (2), $B_y = \Sigma$ Column (4)	$B_x = \Sigma$ Column (3), $B_y = \Sigma$ Column (4)
$\Gamma = \Sigma$ Column (5)	$\Gamma = \Sigma$ Column (5)
$D_x = \Sigma$ Column $(1)^2$, $D_y = \Sigma$ Column $(3)^2$	$D_x = \Sigma$ Column $(1)^2$, $D_y = \Sigma$ Column $(3)^2$
$G_x = [\Sigma \text{ Column } (1)]^2 = A_x^2$, $G_y = [\Sigma \text{ Column } (3)]^2 = A_y^2$	
$H_x = D_x - B_x$, $H_y = D_y - B_y$	$H_x = D_x - B_x$, $H_y = D_y - B_y$
$K_x = G_x - 2B_x - 4H_x$, $K_y = G_y - 2B_y - 4H_y$	
$L = 2B_xB_y$	$L = 2B_xB_y$
$O = 4H_xH_y / (n - 2)$	$O = 4H_xH_y / (n - 2)$
$P = K_xK_y / (n - 2)(n - 3)$	
$Q = G_xG_y / n(n - 1)$	
$R = L + O + P - Q$	$R = L + O$
$S = R / n(n - 1) = Var(\Gamma)$	$S = R / n(n - 1) = Var(\Gamma)$
$T = \sqrt{S}$	$T = \sqrt{S}$
$U = A_xA_y / n(n - 1) = E(\Gamma)$	
$V = \Gamma - U = \Gamma - E(\Gamma)$	$V = \Gamma - U = \Gamma - E(\Gamma)$
$W = V / T = Z$	$W = V / T = Z$

(a) Similarity rating tasks

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lion	(1)		258	49	240	198	261	198	168	209
Mouse	(2)			261	135	257	53	262	218	224
Tiger	(3)				239	201	258	203	168	215
Rabbit	(4)					241	139	250	189	187
Horse	(5)						259	163	211	157
Rat	(6)							266	207	233
Elephant	(7)		(Symmetric)						246	228
Fox	(8)									203
Sheep	(9)									

(b) Sorting tasks

		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Lion	(1)		0	38	1	1	3	3	9	5
Mouse	(2)			0	20	4	34	3	2	2
Tiger	(3)				0	0	4	3	8	4
Rabbit	(4)					5	16	4	7	11
Horse	(5)						0	16	3	12
Rat	(6)							0	2	4
Elephant	(7)		(Symmetric)						3	2
Fox	(8)									6
Sheep	(9)									

Figure 1. Group (n = 40) matrices for similarity rating tasks (a) and for sorting tasks (b).

The comparison of these two matrices using the QAP esentially involves testing for the significance of the correlation between the two matrices, in which the null hypothesis is that there is no similarity between the patterns of entries in both matrices. In the QAP scheme, the groups proximity matrix from the sorting task is compared to the group proximity matrix from the similarity rating task to assess wether the pattern represented in one matrix is also present in the other matrix. The index, Γ, defined on Equation (2.1), representing the sum of products of the corresponding elements between the two matrices is computed. If the probability of the particular value occuring given the distribution of Γ based on all equally likely permutations of rows and corresponding columns of one of the matrices is sufficiently small, then the structure in the second matrix is considered mirrored in the first.

For the matrices used in this illustrative example, the permutation distribution are large (i.e., 9! for the 9 x 9 matrices of animal comparisons). The expectation and variance of Γ using Equations (2.3) and (2.4) and simpler formulas in Table 2 can be computed and a Z statistic (i.e., Equation (2.2)) can be obtained to test for the significance of the correspondence between the two matrices. It should be noted that Γ is merely an "unnormalized" Pearson product-moment correlation coefficient, and therefore, comparing the index value for Γ to a sampling distribution is equivalent to performing the same procedure on the correlation coefficient for the entries in the two matrices. The only quantity in the Pearson correlation coefficient that varies, depending upon the order of the rows and columns, is the index Γ.

Using the QAP technique, the following indexes for this example are provided:

$$r_{xy} = -.91$$

$$\Gamma = 33776$$

$$E(\Gamma) = 97316.11$$

$$Var(\Gamma) = 162844812.20$$

$$Z = -4.98 \ (p \leq 0.1)$$

The Z statistic indicates that the test for the significance of the correspondence between the two matrices achieves the significant level .01, which means that the large values in the similarity rating task meant the animal pair was perceived as dissimilar, and large values for the sorting task meant the pair was similar. Therefore, the use of the QAP has provided a statistical test of the significance of this apparent similarity in organization, which confirms the similarity between the two matrices from an exploratory multidimensional scaling analysis.

Asymmetric case

This example is taken from Sato, Kurata, Shimada, & Harnisch (1990). Its purpose is to identify students' knowledge structures and to compare with a teacher's knowledge structure to debug students' ill-structurally derived misconceptions. Twenty six college students were asked to identify the network structure among forth

chemical concepts after a briefly introductory instruction of College Chemistry. Their structures were used to compare with an instructor's structure about the same subject of College Chemistry. The students' responses represent thinkings of the novices and the teacher's responses represent an expert's ideas. Our jobs are to compare students' knowledge structures with the teacher's knowledge structures and to locate the misconceptions where occurred in students' ill-structural maps of concepts. For illustrative purposes, two students' and the teacher's knowledge structures of forty chemical concepts are selected for demonstration.

The correlation coefficients of these twenty six comparisons are ranged between .237 and .930. The students whose correlation coefficients are lowest and highest are chosen to be examples. The data structures of the lowest ability student (abbreviated for L student), the highest ability student (abbreviated for H student) and the teacher are shown in Figure 2. The knowledge structures of the L student, the H student and the teacher are shown in Figure 3. The QAP teachnique is applied to this dataset.

From the analysis of QAP, the following indexes for this example are given:

L student vs. Teacher	H student vs. Teacher
$r_{LT} = .237$	$r_{HT} = .930$
$\Gamma = 12$	$\Gamma = 48$
$E(\Gamma) = 1.37$	$E(\Gamma) = 1.70$
$Var(\Gamma) = 1.31$	$Var(\Gamma) = 1.62$
$Z = 9.30$ $(p<.001)$	$Z = 36.37$ $(p<.001)$

The Γ index for the comparison between the L student and the teacher is 12, whose Z statistic shows that the test for the significance of the correspondence between these two matrices is significant at .001 level. The Γ index for the comparison between the H student and the teacher is 48, whose Z statistic indicates that the p value is significant at .001 level. The correlation between the H student and the teacher is larger than that of the L student and the teacher. This means that the knowledge structure of forty chemical concepts of the H student is more close to that of the teacher than the L student does. The H student has more similar entries

(a) L student's structures

```
0000010000000001000000000000000010000000
0000000100000000000000000000000000000000
0000000001000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000001100000000000
0000001000000000001000001000000000000000
0000000000000000000000000000000000000010
0000000000000000000000001000000000000000
0000000000000000000000001000000000000000
0000000000000010000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000001000000
0000000000000000000000000000000001000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000010100000000000000000000
0000000000000000000000000000000000000001
0000000000000000000001100000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000001101110000
0000100000000000000000000000000000000000
0000000000000000000000000000100000000000
0000000000000000000000000100000000000000
0000000000000000000000000000000000000000
0010000000111100000000000000010000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000010000000
0000000000000000000000000000000010000000
1100000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000010000000
0000000000000000000000000000000010000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0001000000000000000000000000000000000000
```

(b) H student's structures

```
0000110000000001000000000000000000001010
0001100100000000100000000000000000000101
0000000000000000000100000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000101000000000
0000001000000000001000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000111110000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000001000000000000000000000000000000
0000010000000000000000000000000000000000
0000010000000000000000000000000000000000
0000000000000010001000000000000000000000
0000000000000010000000000000000000000000
0000000000000000011000000000000000000000
0000000000000000000001100000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000110110000
0000000010000000000000000000000000000000
0000000000000000000000001010000000000000
0000000000000000000000000000000000000000
0000000000000000000100000100000000000000
0000000000000000000000000000010000000000
0000000000000000000010000001000000000000
0000000000000000000000010000000000000000
0000000000000000000000000000000000000000
0000000000000000000100000000000001000000
0000000000000000000100000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000010000000000000000000000000000000000
0000000100000000000000000000000000000000
0000000000000000000010000000000000000000
0000000000000000000010000000000000000000
```

(c) Teacher's structures

```
0000110000000001000000000000000000001010
0001100100000000100000000000000000000101
0000000000000000000100000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000101000000000
0000001000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000111110000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000001000000000000000000000000000000
0000010000000000000000000000000000000000
0000010000000000000000000000000000000000
0000000100000010001000000000000000000000
0000010000000010000000000000000000000000
0000000000000000011000000000000000000000
0000000000000000000001100000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000000000000000000000000000000110110000
0000000010000000000000000000000000000000
0000000000000000000000001010000000000000
0000000000000000000000000000000000000000
0000000000000000000000000100000000000000
0000000000000000000000000000010000000000
0000000000000000000110000001000000000000
0000000000000000000000010000000000000000
0000000000000000000000000000000000000000
0000000000000000000100000000000000000000
0000000000000000000100000000000010000000
0000000000000000000000000000000000000000
0000000000000000000000000000000000000000
0000010000000000000000000000000000000000
0000000100000000000000000000000000000000
0000000000000000000010000000000000000000
0000000000000000000010000000000000000000
```

Figure 2. The data structures of (a) the L student, (b) the H student, and (c) the teacher.

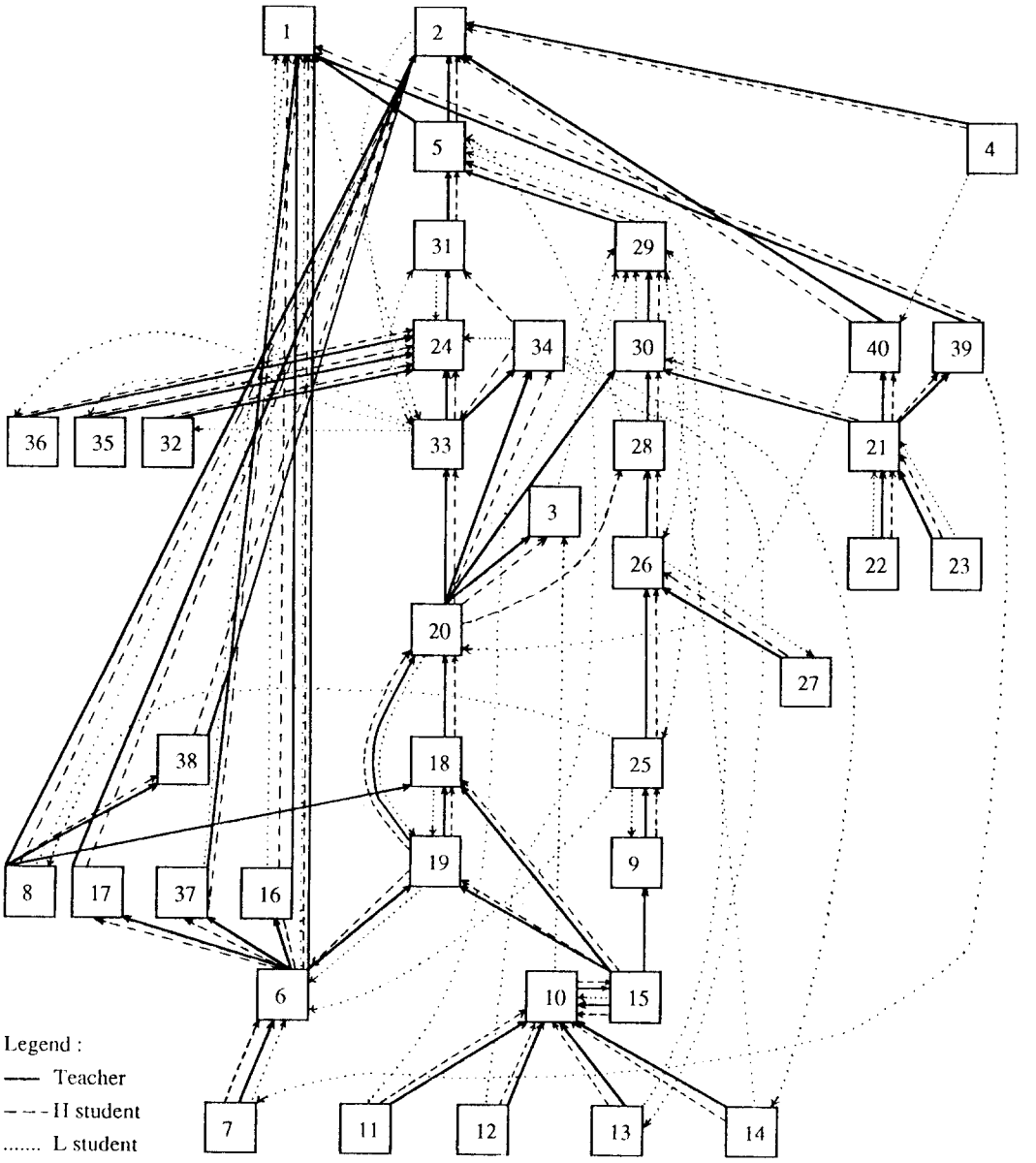

Figure 3. Knowledge structures of forty chemical concepts for the L student, the H student, and the teacher.

of a structure pattern with those of the teacher than the L student has (as shown in Figure 3), although both students have significant correlations with teacher's structures. Consequently, the L student's misconceptions can be easily spotted by mapping his/her knowledge structure to the teacher's. A particular structure between two consecutive or adjacent concepts where the deviation occurred marks the misconception. The sum of those deviations from the teacher's structure can be indexed as a measure of misconceptions. And those concepts where misconceptions occurred need to be taught again. This sparkles the possibility of make-up instructions for low ability students. Therefore, the QAP not only confirms the similarity between two proximity matrices but also provides a methodological future for cognitive psychology.

IV. Summary Comments

From the above presentation, the procedures of using QAP should be evident with no doubt. The purpose of QAP is used to compare two proximity (symmetric and asymmetric) matrices. Given the proximity measure X_{ij} and some conjecture specified in terms of a structure function Y_{ij} (i.e., both are some numerical measures of proximity between objects o_i and o_j in the same set of S), the observed correlation between X_{ij} and Y_{ij} is compared with a reference distribution generated under a hypothesis of randomness. If the obtained correlation is at extreme percentage point, the correspondence between X_{ij} and Y_{ij} is declared significant, with the added implication that the conjecture leading to the construction of Y_{ij} may help explain some of the variation presented in the empirical proximity measures. As usual, the size of the correlation can be considered an index of the degree of correspondence or confirmation. And the whole process is what the QAP talks about.

As specified above, the QAP is suggested for applications rather than for theories. A wide variety of applications of the QAP depends on how we define the proximity measures in matrices **X** and **Y**. For example, the Γ index can be viewed as a particular measures of association between two matrices by specifying the following possible definitions of proximity (X_j is a measure on object j and X_i is a measure on object i; similar measures applied to Y_j and Y_i) (Hubert, 1985, 1987; Hubert & Arabie, 1989):

(a) Pearson's product-moment correlation: $X_{ij}=(X_j-X_i)$; $Y_{ij}=(Y_j-Y_i)$;

(b) Spearman's rank-order correlation: $X_{ij}=(p_j-p_i)$; $Y_{ij}=(q_j-q_i)$, where p_j and p_i refer to the ranks of X_j and X_i within the **X**; a similar way applied to Y_j and Y_i in **Y**;

(c) Kendall's τ: $X_{ij}=\text{sign}(X_j-X_i)$; $Y_{ij}=\text{sign}(Y_j-Y_i)$, where

$$\text{sign}(u) = \begin{cases} +1 & \text{if } u>0 \\ 0 & \text{if } u=0 \\ -1 & \text{if } u<0; \end{cases}$$

all three of above correlations can be put into a more general framework referred to as Daniels' generalized correlation coefficient (1944);

(d) Geary's spatial autocorrelation (1954):$X_{ij}=(X_j-X_i)^2$; $Y_{ij}=(Y_j-Y_i)^2$;

(e) Moran's index (1950):$X_{ij}=(X_j-\overline{X})(X_i-\overline{X})$; $(Y_j-\overline{Y})(Y_i-\overline{Y})$;

(f) Royalty, Astrachan and Sokal's index (1975): $X_{ij}=|X_j-X_i|$;$Y_{ij}=|Y_j-Y_i|$;

(g) other extentions to the one-way analysis-of-variance (Mielke, 1979; Mielke, Berry, & Johnson, 1976; Hubert, Golledge, & Costanzo, 1982), spatial autocorrelation (Hubert, Golledge, & Costanzo, 1981; Hubert Golledge, Costanzo, & Gale, 1985), and so on.

So, the quadratic assignment paradigm can be viewed as a general data-analytic strategy (Hubert & Schultz, 1976).

Another major application of the QAP is to be used as a confirmatory approach in a confirmatory research. For more illustrative examples readers are referred to Gliner, Goldman, & Hubert (1983), Hubert & Levin (1976a, 1976b, 1977) and Pellegrino & Hubert (1982).

References

Borg, I., & Lingoes, J. (1987). *Multidimensional similarity structure analysis.* New York:

Springer-Verlag.
Cohen, J. (1968a). Multiple regression as a general data-analytic system. *Psychological Bulletin*, 70, 426-443.
Cohen, J. (1968b). Weighted kappa: Nominal scale agreement with provision for scaled disagreement or partial credit. *Psychological Bulletin, 70,* 213-220.
Cohen, J. (1982). Set correlation as a general multivariate data-analytic method. *Multivariate Behavioral Research, 17,* 301-341.
Daniels, H. E. (1944). The relation between measures of correlation in the universe of sample permutations. *Biometrika, 33,* 129-135.
Fisher, R. A. (1934). Two new properties of mathematical likelihood. *Proceedings of the Royal Society, 144 (A),* 285-307.
Fliess, J. L., Cohen, J., & Everitt, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin, 72.* 323-327.
Friedman, M. (1937). The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the American Statistical Association, 32,* 675-701.
Geary, R. C. (1954). The contiguity ratio and statistical mapping, *Incorporated Statistician,* 5, 115-145.
Gliner, G., Goldman, S.R., & Hubert, L. J. (1983). A methodological study on the evaluation of learning from story narratives. *Multivariate Behavioral Research, 18,* 9-36.
Golledge, R.G., & Rayner, J.N. (Eds.) (1982). *Proximity and preference: Problems in the multidimensional analysis of large data sets.* Minneapolis, Minn.: University of Minnesota Press.
Hubert, L.J. (1977). Kappa revisited. *Psychological Bulletin,* 84, 289-297.
Hubert, L.J. (1978). A general formula for the variance of Cohen's weighted kappa. *Psychological Bulletin, 85,* 183-184.
Hubert, L.J. (1979a). Comparison of sequences. *Psychological Bulletin, 86,* 1098-1106.
Hubert, L.J. (1979b). Matching models in the analysis of cross-classifications. *Psychometrika, 44,* 21-41.
Hubert, L.J. (1985). Combinatorial data analysis: Association and partial association. *Psychometrika, 50,* 449-467.
Hubert, L.J. (1987). *Assignment methods in combinatorial data analysis.* New York: Marcel Dekker.
Hubert, L.J., & Arabie, P. (1989). Combinatorial data analysis: Confirmatory comparisons between sets of matrices. *Applied Stochastic Models and Data Analysis, 5,* 273-325.
Hubert, L.J., & Baker, F.B. (1978). Evaluating the conformity of sociometric measurements. *Psychometrika, 43,* 31-41.
Hubert, L.J., Golledge, R.G., & Costanzo, C.M. (1981). Generalized procedure for evaluating spatial autocorrelation. *Geographical Analysis, 13,* 224-233.
Hubert, L.J., Golledge, R.G., & Costanzo, C.M. (1982). Analysis of variance procedures based on proximity measure between subjects. *Psychological Bulletin, 91,* 424-430.
Hubert, L.J., Golledge, R.G., Costanzo, C.M., & Gale, N. (1985). Measuring association between spatially defined variables: An alternative procedure. *Geographical Analysis, 17,* 36-46.
Hubert, L.J., & Levin, J.R. (1976a). Evaluating object set partitions: Free-sort analysis and some generalizations. *Journal of Verbal Learning and Verbal Behavior, 15,* 459-470.
Hubert, L.J., & Levin, J.R. (1976b). A general statistical framework for assessing categorical

clustering free recall. *Psychological Bulletin, 83,* 1072-1080.

Hubert, L.J., & Levin, J.R. (1977). Inference models for categorical clustering. *Psychological Bulletin, 84,* 878-887.

Hubert, L.J., & Schultz, J. (1976). Quadratic assignment as a general data analysis strategy. *Brish Journal of Mathematical and Statistical Psychology, 29,* 190-241.

Kendall, M.G. (1970). *Rank correlation methods* (4th ed.). London: Griffin.

Koopmans, T.C., & Beckmann, M. (1957). Assignment problems and the location of economic activities. *Econometrika,* 25, 53-76.

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

Mielke, P.W. (1979). On asymptotic non-normality of null distributions of MRPP statistics. *Communications in Statistics — Theory and Methods, A8,* 1541-1550.

Mielke, P.W., Berry, K.J., & Johnson, E.S. (1976). Multi-response permutation procedures for priori classifications. *Communications in Statistics — Theory and Methods,* A5, 1409-1424.

Moran, P.A.P. (1950). A test for serial independence of residuals. *Biometrika, 37,* 178-181.

Morrison, D.F. (1990). *Multivariate statistical methods* (3rd ed.). New York: McGraw-Hill.

Page, E.B. (1963). Ordered hypotheses for multiple treatments: A significance test for linear ranks. *Journal of the American Statistical Association, 58,* 216-230.

Pellegrino, J.W., & Hubert, L.J. (1982). The analysis of organization and structure in free recall. In C.R. Puff (Ed.), *Handbook of research methods in human memory and cognition* (pp. 129-172). New York: Academic Press.

Royalty, H.H., Astrachan, E., & Sokal, R.R. (1975). Tests for patterns in geographical variation. *Geographical Analysis, 7,* 369-395.

Schultz, J.V., & Hubert, L.J. (1976). A nonparametric test for the correspondence between two proximity matrices. *Journal of Educational Statistics, 1,* 59-67.

Sato, T., Yamaji, Y., Kurata, M., Shimada, N., & Harnisch, D.L. (1990, April). *Analysis of students' concept maps and comparisons with a teacher's structure of the subject matter.* Paper presented at the 2nd Symposium on "Concept mapping of subject matter content and its applications to meaningful learning," Nagasaki, Japan.

Young, F.W. (1987). *Multidimensional scaling: History, Theory, and applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.