stichting
mathematisch
centrum

MC

AFDELING MATHEMATISCHE BESLISKUNDE (DEPARTMENT OF OPERATIONS RESEARCH)

BW 88/78

MEI

D.S. BRÉE, B.J. LAGEWEG, J.K. LENSTRA,
A.H.G. RINNOOY KAN, G. VAN BEZOUWEN

A HIERARCHICAL CLUSTERING SCHEME FOR ASYMMETRIC MATRICES

Preprint

2e boerhaavestraat 49 amsterdam

Printed at the Mathematical Centre, 49, 2e Boerhaavestraat, Amsterdam.

The Mathematical Centre, founded the 11-th of February 1946, is a non-profit institution aiming at the promotion of pure mathematics and its applications. It is sponsored by the Netherlands Government through the Netherlands Organization for the Advancement of Pure Research (Z.W.O).

AMS (MOS) subject classification scheme (1970): 15A18, 62P15, 90C10, 92A25

A HIERARCHICAL CLUSTERING SCHEME FOR ASYMMETRIC MATRICES

D.S. BRÉE
Graduate School of Management, Delft

B.J. LAGEWEG

J.K. LENSTRA

A.H.G. RINNOOY KAN
Erasmus University, Rotterdam

G. VAN BEZOUWEN
Graduate School of Management, Delft

ABSTRACT

This paper is concerned with the problem of grouping elements given asymmetric relationships between them. Two criteria for distinguishing between arrangements are developed: the distortion, based on the eigenvalues of the interaction matrix, and the composition, based on the amount of information lost by combining groups. Our goal is to locate those arrangements in the space of the two criteria that lie close to the Pareto-optimal points. A heuristic search procedure is proposed, which first allocates each element to a separate group and then proceeds by combining groups until the final group containing all the elements is reached. The method is applied to illustrative examples and compared to some other approaches.

KEY WORDS & PHRASES: *clustering, asymmetric interaction matrix, distortion, eigenvalues, linear assignment, composition, entropy, agglomerative procedure, heuristic search, Pareto-optimal points.*

NOTE: This report is not for review; it will be submitted for publication in a journal.

1. INTRODUCTION

The problem of grouping a set of elements according to some measure of their similarity has been approached from several directions. It is unlikely that a single method will ever be accepted as the best, since the use made of the results of the grouping is so varied. At most we can expect a collection of good methods for various applications. This paper is concerned with developing a method for grouping elements given asymmetric relationships between them.

We know of no technique that can cope with this problem without first destroying the asymmetric information by some averaging procedure [Sokal & Sneath, 1963; Gower, 1967; Jardine & Sibson, 1971; Anderberg, 1973; Everitt, 1974; Duran & Odell, 1975; Hartigan, 1975]. Even in Hubert's [1973] attempt to deal with asymmetric matrices first three symmetric matrices are produced, in which for each corresponding pair of entries both values are made equal to either their minimum or their maximum or their arithmetic mean, and then Johnson's [1973] hierarchical clustering techniques are applied to the three matrices.

The problem of finding suitable groupings of elements with asymmetric relationships arises in several contexts. The aggregation problem in Leontief input-output analysis is one example; the detection of highly interactive groups of individuals is another. In general the problem may arise when the elements are variables and the relationships represent the effect of the variables upon each other, or when the elements are entities and the relationships represent the frequencies of asymmetric events in which two entities are involved. The problem has many other interpretations; e.g., the relationships could represent ratings of people by each other, but the solution developed below was created with the above interpretations in mind.

There are two main aspects of the problem: the selection of criteria for choosing suitable groupings and the choice of a method for searching among the possible groupings. To develop an interpretable clustering procedure it is important to distinguish between these two aspects. As noted by Gower [1967], the failure to formulate this distinction has led to the proliferation of a large number of methods without the possibility of adequate comparison between them.

Criteria for choosing between possible groupings can be related either to properties of the individual groups or to the complete arrangement. Various linkage methods [Gower, 1967] have a criterion based on a measure for each separate group, resulting in a simple search procedure. But inevitably, a second criterion is required by which entire arrangements can be compared. Most methods thus employ two criteria, even if one of them is not stated explicitly. In the linkage methods the explicit criterion, e.g. the minimum average similarity of elements in a group, is used by the analyst to determine a suitable number of groups - his implicit criterion. It would be preferable to make both criteria explicit. The goal for the search procedure is then to locate those arrangements in the space of the two criteria that lie close to the *Pareto-optimal* points, i.e. the points for which one can only achieve improvement with respect to one criterion at the expense of the other. This will usually lead to a collection of good arrangements rather than to a single "best" arrangement. If a single solution is required, then some further assumption will be needed, such as a weighting function.

The evaluation of the criteria for all possible arrangements is prohibitive even for quite small numbers of elements, since the number of possible arrangements grows exponentially in the number of elements. Thus, some sort of heuristic search procedure is required. In many methods search is based on the form of the criteria for selection [Gower, 1967], and sometimes a simple search method is used without reference to a well-defined criterion of choice in order to yield a rough analysis of the data [King, 1967]. Some search methods have been designed to generate a suitably structured set of solutions; for example, biologists analysing various species to determine their phenetic grouping require a solution in the form of a dendogram or taxonomic tree.

In the rest of this paper we will first develop two criteria for distinguishing between arrangements: a measure of *distortion* based on the eigenvalues of the matrix of interactions and a measure of *composition* based on the amount of information lost by combining groups. We then look at the search problem and suggest an *agglomerative procedure* that, in fact, yields solutions similar to a dendogram. Finally we look at how the proposed method performs on an example, as compared to some other approaches.

2. PROBLEM DEFINITION

Assume that we have a set E of n elements $e_1,\ldots,e_n$ and an n×n-matrix $P = (p_{jk})$ $(j,k) = 1,\ldots,n)$, where p_{jk} measures the effect that e_j has on e_k. A *clustering* corresponds to a partition of E into N subsets $F_1,\ldots,F_N$:

$$E = \bigcup_{\ell=1}^{N} F_\ell, \quad F_\ell \cap F_m = \emptyset \quad (\ell,m = 1,\ldots,N;\ \ell \neq m).$$

Let f_ℓ be the cardinality of F_ℓ, and assume the elements to be reindexed in such a way that

$$F_\ell = \{e_j \mid j = \sum_{h=1}^{\ell-1} f_h + 1,\ldots,\sum_{h=1}^{\ell} f_h\} \quad (\ell = 1,\ldots,N).$$

To determine the quality of a particular clustering, we shall make use of $f_\ell \times f_\ell$-matrices Q_ℓ representing the interaction within F_ℓ:

$$Q_\ell = (p_{jk}) \quad (j,k = \sum_{h=1}^{\ell-1} f_h + 1,\ldots,\sum_{h=1}^{\ell} f_h) \quad (\ell = 1,\ldots,N),$$

combined to form the completely decomposed n×n-matrix Q:

$$Q = \begin{bmatrix} Q_1 & 0 & \cdots & 0 \\ 0 & Q_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & Q_N \end{bmatrix}.$$

2.1. The measure of distortion

If a system is adequately represented by some interaction matrix, such as P, then the behaviour of that system through time often depends on powers of P. For example, if the elements correspond to individuals and p_{jk} represents the probability of the diffusion of a message from person j to person k in a unit time period, then P^t is the transformation to be applied to an original knowledge vector to establish the probability distribution of knowledge after t time periods. Another common example is when the elements are variables in a closed system and P represents the set of linear difference equations that connects these variables. Then the value of these variables at

some time t is given by an n-dimensional vector $x(t) = Px(t-1) = P^t x(0)$, where $x(0)$ represents the initial values of the variables. This suggests that a suitable criterion for choosing between different arrangements is to make the difference between P^t and the Q^t associated with the arrangement as small as possible for all t.

The validity of this approach when considering the behaviour in time of nearly decomposable systems has been demonstrated by Simon and Ando [1961]. They show that it is sufficient in the short run to consider the behaviour of each subsystem separately, and that each subsystem can be treated as a single composite variable when studying the behaviour of the entire system in the long run. A more intuitive approach to this question has been put forward by Simon [1969].

Assume for the moment that the *eigenvalues* of P and Q are all distinct, so that there exist nonsingular n×n-matrices Y and Z such that

$$Y^{-1}PY = \Lambda = (\lambda_j \delta_{jk}) \quad \text{for } j,k = 1,2,\ldots,n,$$
$$Z^{-1}QZ = M = (\mu_j \delta_{jk}) \quad \text{for } j,k = 1,2,\ldots,n,$$

where $\delta_{jk} = 1$ if $j = k$ and $\delta_{jk} = 0$ otherwise, and λ_j and μ_j are the eigenvalues of P and Q respectively. Y and Z are only defined modulo column permutations and scalar multiplications of each column. Now

$$P^t = (Y\Lambda Y^{-1})^t = Y\Lambda^t Y^{-1},$$
$$Q^t = (ZMZ^{-1})^t = ZM^t Z^{-1}.$$

Since Y and Z depend on Λ and M respectively, this in turn suggests as a suitable criterion for deciding between arrangements the *total squared difference between corresponding eigenvalues of* P *and* Q.

What are to be taken as the corresponding eigenvalues in the two cases? With small deviations between Λ and M it will be clear by inspection which eigenvalues of Q are distortions of the eigenvalues of P. However, when Q is a poor approximation of P, this correspondence will not be clear. We propose to find a pairing of the λ_j with the μ_k that yields the minimum value Δ of the sum of squares of their differences, i.e., a pairing that minimizes

$$\sum_{j=1}^{n}\sum_{k=1}^{n} |\lambda_j - \mu_k|^2 x_{jk}$$

subject to

$$\sum_{j=1}^{n} x_{jk} = 1 \quad (k = 1,\ldots,n),$$

$$\sum_{k=1}^{n} x_{jk} = 1 \quad (j = 1,\ldots,n),$$

$$x_{jk} \in \{0,1\} \quad (j,k = 1,\ldots,n);$$

λ_j is paired with μ_k if $x_{jk} = 1$ in the optimal solution. This is a *linear assignment problem*, for which several efficient algorithms are available [Wagner, 1975, p.183; Lawler, 1976, p.129; Dorhout, 1977].

Decomposing the eigenvalues into real and imaginary parts:

$$\lambda_j = \lambda_j' + \lambda_j''i, \quad \mu_j = \mu_j' + \mu_j''i \quad (j = 1,\ldots,n),$$

we can rewrite the cost coefficient of x_{jk} as

$$\begin{aligned} |\lambda_j-\mu_k|^2 &= (\lambda_j'-\mu_k')^2 + (\lambda_j''-\mu_k'')^2 \\ &= \lambda_j'^2 + \lambda_j''^2 + \mu_k'^2 + \mu_k''^2 - 2(\lambda_j'\mu_k'+\lambda_j''\mu_k'') \\ &= |\lambda_j|^2 + |\mu_k|^2 - 2(\lambda_j'\mu_k'+\lambda_j''\mu_k'). \end{aligned}$$

Substitution of this expression in the objective function yields

$$\sum_{j=1}^{n}\sum_{k=1}^{n}|\lambda_j-\mu_k|^2 x_{jk} = \sum_{j=1}^{n}(|\lambda_j|^2+|\mu_j|^2) - 2\sum_{j=1}^{n}\sum_{k=1}^{n}(\lambda_j'\mu_k'+\lambda_j''\mu_k'')x_{jk}.$$

When the eigenvalues of either P or Q are all real, it is not necessary to use a general linear assignment subroutine, since the problem can be solved in a much simpler way. The objective function attains its minimum value Δ when the real parts λ_j' and μ_k' are both arranged in descending order of magnitude and the eigenvalues in the corresponding rank order are paired (cf. [Lawler, 1976, p.211]). That this procedure indeed yields an optimum pairing can be proved by considering the situation in which

$$\lambda_j' \geq \lambda_k', \quad \mu_j' \geq \mu_k', \quad \mu_j'' = \mu_k'' = 0.$$

If $x_{jj} = x_{kk} = 1$, the contribution of λ_j, λ_k, μ_j and μ_k to the variable part of the objective function is given by

$$-2(\lambda_j'\mu_j'+\lambda_k'\mu_k') = -2(\lambda_j'\mu_k'+\lambda_k'\mu_j') + 2(\lambda_j'-\lambda_k')(\mu_k'-\mu_j').$$

In the rearranged expression, the first term denotes the contribution if $x_{jk} = x_{kj} = 1$. Since the second term is nonpositive, it follows that the latter pairing is no better than the former one.

We will need to make comparisons between Δ values for different groupings. Moreover it is desirable to have some measure of the distortion between matrices of different size. We therefore choose to normalize Δ so that the distortion is zero when all the elements are combined in one group of size n, and that it is unity when there are n groups each containing one element. In the former case we have $Q = P$, so that $\mu_j = \lambda_j$ $(j = 1,\ldots,n)$ and $\Delta = 0$; in the latter case we have $Q_\ell = (p_{\ell\ell})$, $\mu'_\ell = p_{\ell\ell}$ and $\mu''_\ell = 0$ $(\ell = 1,\ldots,n)$, so that $\Delta = \Delta^*$ where Δ^* denotes the value of the objective function when the λ'_j and the $p_{\ell\ell}$ are paired in descending order of magnitude. The *measure of distortion* D is now defined to be

$$D = \Delta/\Delta^*.$$

2.2. The measure of composition

The measure of distortion is by itself not a sufficient criterion for choosing suitable arrangements: the arrangement with minimum distortion is a single group containing all the elements. Therefore we also need some measure of composition in an arrangement.

This measure will be dependent only upon the format of the grouping, as defined by the number of groups and their cardinalities, and not upon the particular arrangement within the format. Although the number of groups is a simple and often-used criterion, we will not use it. To see why, consider an arrangement with two groups. It seems probable that D will be smaller if the groups contain 1 and n-1 elements than if they each contain about half the elements. Yet, in some intuitive sense, the latter format gives us more information.

The usual *measure of entropy* provides a good indication of the amount of composition achieved by a particular format. It has been used as such in the case of symmetric matrices. In general the entropy of a partition $E = \bigcup_{\ell=1}^{N} F_\ell$ is defined by

$$\frac{1}{n} \sum_{\ell=1}^{N} f_\ell \log f_\ell.$$

To get a measure of composition we wish the maximum composition to be unity and to occur when all the elements are combined in one group of size n. Since in that case the entropy is equal to log n, the *measure of composition* C is defined to be

$$C = \frac{1}{n \log n} \sum_{\ell=1}^{N} f_\ell \log f_\ell .$$

Note that the minimum composition is zero and is attained when there are n groups each containing one element.

In summary, we now have two criteria for each arrangement: the measure of distortion D and the measure of composition C. For the arrangement in which all the elements are lumped together in a single group, we have D = 0 and C = 1. For the arrangement in which there are n single-element groups, we have D = 1 and C = 0. Somewhere between these two extremes there are arrangements which are good in the sense that they are both low in distortion and composition. Our problem will be to find arrangements that are close to the Pareto-optimal ones, as defined in Section 1. The choice of which of these good arrangements is the best will not be of concern to us and is left to the individual analyst.

3. SEARCH FOR PARETO-OPTIMAL SOLUTIONS

Even for quite a small number of elements complete enumeration of all the possible arrangements is not feasible. The number of different arrangements for n elements into N groups is given by the so-called *Stirling number of the second kind* [Liu, 1968, pp.38-39, p.101; Wells, 1971, p.157, p.235]:

$$S(n,N) = S(n-1,N-1) + NS(n-1,N) = \frac{1}{N!}\sum_{\ell=0}^{N}(-1)^{\ell}\binom{N}{\ell}(N-\ell)^{n}.$$

The total number of arrangements increases exponentially with n. For n = 19 it is already of the order $4 \cdot 10^{12}$ [Fortier & Solomon, 1966].

Jensen [1969] developed a dynamic programming approach, which refrains from considering certain unsuitable arrangements. By this method he managed to reduce the number of calculations from about 10^{19} to 10^{12} for finding arrangements of 25 elements into 10 groups. But even such a method is prohibitive for situations with more than 25 elements with present computing equipment.

Fortier and Solomon [1966] proposed a sampling approach with a criterion similar to Tryon's [1939] B coefficient. They were not able to find such a good solution as the one provided by Tryon's search method.

We thus find ourselves forced by computational considerations to employ a *heuristic search procedure*. Such heuristics, applied in cluster analysis, can often be characterized as being either

- *subdivisive*: divide the set of elements into two groups and then consider each group separately, or
- *agglomerative*: allocate each element to a separate group and then proceed by combining groups.

Although the subdivisive method is sometimes considered superior [Gower, 1967, p.635], it is not practicable as the number of arrangements to be considered at the first step is already 2^{n-1}. For the agglomerative method the number of arrangements to be considered at the first step is n(n-1)/2, and the total number of arrangements that will be considered is

$$\sum_{N=2}^{n} N(N-1)/2 = n(n^2-1)/6 = O(n^3),$$

which is within the realms of possibility. The agglomerative method is the basis of the search procedure that we will employ.

3.1. The search procedure

The method chosen is a simple hierarchical scheme. The initial arrangement is that in which there are n groups each containing one element. At each step the current arrangement is taken as starting point. All possible arrangements that can be formed by combining two groups of the current one are considered. For each of these new arrangements the distortion and the composition are calculated. The lines in the composition-distortion space joining the current arrangement to each of the new ones are compared, and the line having the most negative slope is selected to determine the next current arrangement. The result is a hierarchy or dendogram in which the elements are brought together into groups and groups into larger groups until finally one group of size n is formed. The procedure is illustrated in Section 3.3.

This search heuristic does not guarantee that a Pareto-optimal solution will be found. By backing up in the search procedure and selecting one of the rejected arrangements as an alternative, better arrangements might be generated. We have chosen, however, not to embed our approach into a branch-and-bound scheme, since the computational requirements of such a scheme would probably be excessive.

3.2. Programming considerations

In calculating the distortion for an arrangement the eigenvalues for this arrangement need to be available. The matrix Q is completely decomposed and the eigenvalues of each of the diagonal submatrices Q_ℓ ($\ell = 1,\ldots,N$) can be calculated separately. As a possible new arrangement differs from the current one only in that two groups are being combined into one, it is only the eigenvalues for this new group that are needed; the remaining ones are available from previous calculations. If both groups that are being brought together had already been formed before the previous step, then they will have been brought together in an earlier phase as a possible but rejected grouping; the eigenvalues for the new group will have already been calculated and can be retrieved from memory. However, if one of the two groups being combined is the group that was formed at the previous step, then the eigenvalues for the new group will have to be calculated. Thus, in reducing the number of

groups from N to N-1, with N < n, eigenvalues will have to be calculated only for the possible new groups formed by combining the single group just formed with each of the other N-1 ones.

At the very first step of the search procedure, the eigenvalues of n(n-1)/2 2×2-submatrices will need to be calculated, which is a trivial task. A call for an algorithm for finding eigenvalues is only required from then on, i.e. (n-1)(n-2)/2 times. The subroutine used for this purpose is fairly fast and renders this approach computationally feasible up to n = 100.

Similar considerations surround the assignment algorithm used to pair the eigenvalues of the original matrix P with those of the proposed new matrix Q. For each of the $O(n^3)$ arrangements that will be considered, an assignment problem has to be solved. In each of these problems, the eigenvalues of P are the same, and the eigenvalues of a new matrix Q differ from those of the current one only as far as the new group is concerned. It follows that the optimal solution to the current assignment problem usually provides a good initial solution to the new assignment problem. Thus, we should use a highly efficient assignment algorithm that benefits from a good initial solution. The method developed by Dorhout [1977] turned out to be suitable for this purpose.

3.3. A numerical example

To illustrate the search procedure, we consider the 4×4-matrix used for similar purposes by Simon and Ando [1961] and reproduced in Fig. 1.

The initial arrangement has four groups each containing one of the elements a,b,c,d. All pairwise combinations of elements are now considered. Only two of these alter the eigenvalues of the initial arrangement, given by the main diagonal of the original matrix. These are the combinations a with b and c with d, which reduce the distortion from 1 to 0.87 and 0.26 respectively; in both cases the composition is increased from 0 to 0.25. The slope of the lines joining the initial arrangement to these two possible new arrangements in the composition-distortion space is -0.53 and -2.95 respectively (cf. Fig. 2). As the latter slope is the smaller one, c and d are combined into group 2 and the second arrangement results. Next, there are three possible combinations: a with b, a with group 2, and b with group 2. Since only

```
INTERACTION MATRIX
##################

          A           B           C           D
A       0.9700      0.0295      0.0005      0.0
B       0.0200      0.9800      0.0         0.0
C       0.0         0.0         0.9600      0.0400
D       0.0002      0.0002      0.0396      0.9600

ORIGINAL EIGENVALUES
R:  0.1000E 01  0.9996E 00  0.9502E 00  0.9202E 00
I:  0.0         0.0         0.0         0.0

SUM OF SQUARES OF DIFFERENCES BETWEEN
EIGENVALUES AND MAIN DIAGONAL OF MATRIX
IS DENOMINATOR FOR DISTORTION  = 0.2956E-02


SEARCH FOR NEXT PAIR OF GROUPS TO COMBINE
#########################################

     COMBINING      GIVES SLOPE   CHOSEN

     A         B     -.5300E 00
     C         D     -.2950E 01  GROUP  2

     A         B     -.1050E 01  GROUP  3

 GROUP  2  GROUP  3  -.5188E-04  GROUP  4


          SEQUENCE OF ARRANGMENTS
          #######################

GROUP    FORMED FROM          COMP.     DISTORTION

    1                         0.0       0.1000E 01
    2    C          D         0.2500    0.2626E 00
    3    A          B         0.5000    0.2596E-04
    4  GROUP  2  GROUP  3     1.0000    0.1808E-07


EIGENVALUES OF ARRANGMENTS
##########################

                1           2           3           4
R:  0.9800E 00 *0.9998E 00 *0.9998E 00**0.1000E 01
I:  0.0         0.0         0.0         0.0
R:  0.9700E 00  0.9800E 00* 0.9998E 00**0.9996E 00
I:  0.0         0.0         0.0         0.0
R:  0.9600E 00* 0.9700E 00**0.9502E 00**0.9502E 00
I:  0.0         0.0         0.0         0.0
R:  0.9600E 00**0.9202E 00  0.9202E 00**0.9202E 00
I:  0.0         0.0         0.0         0.0
```

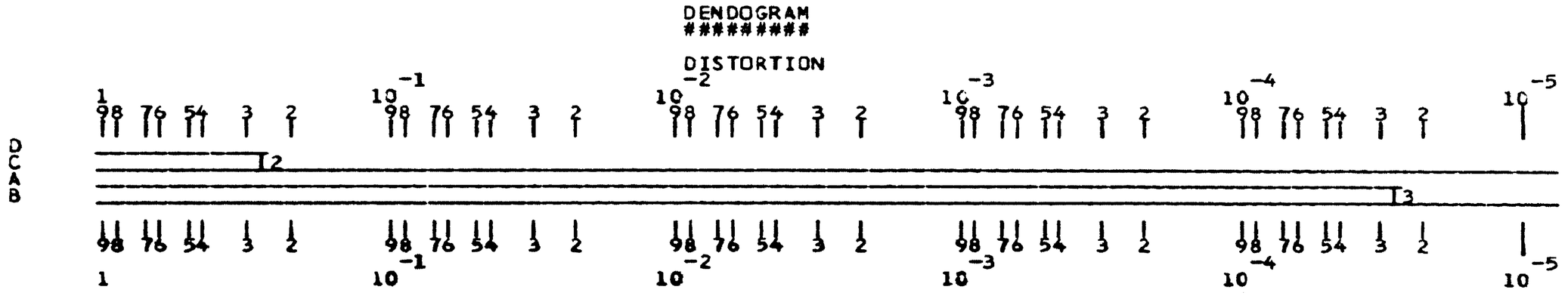

Figure 1 Data and results for the 4×4-example.

the former reduces the distortion, it is selected as group 3 to form the third arrangement. Finally, there is only one combination, namely of groups 2 and 3, which gives the completely integrated original matrix with no distortion.

The results of this search are given in Fig. 1, together with a dendogram illustrating the distortion reduction. In the dendogram, the procedure has reordered the elements so that elements in one group are always adjacent, with earlier formed groups preceding later formed ones. The horizontal axis

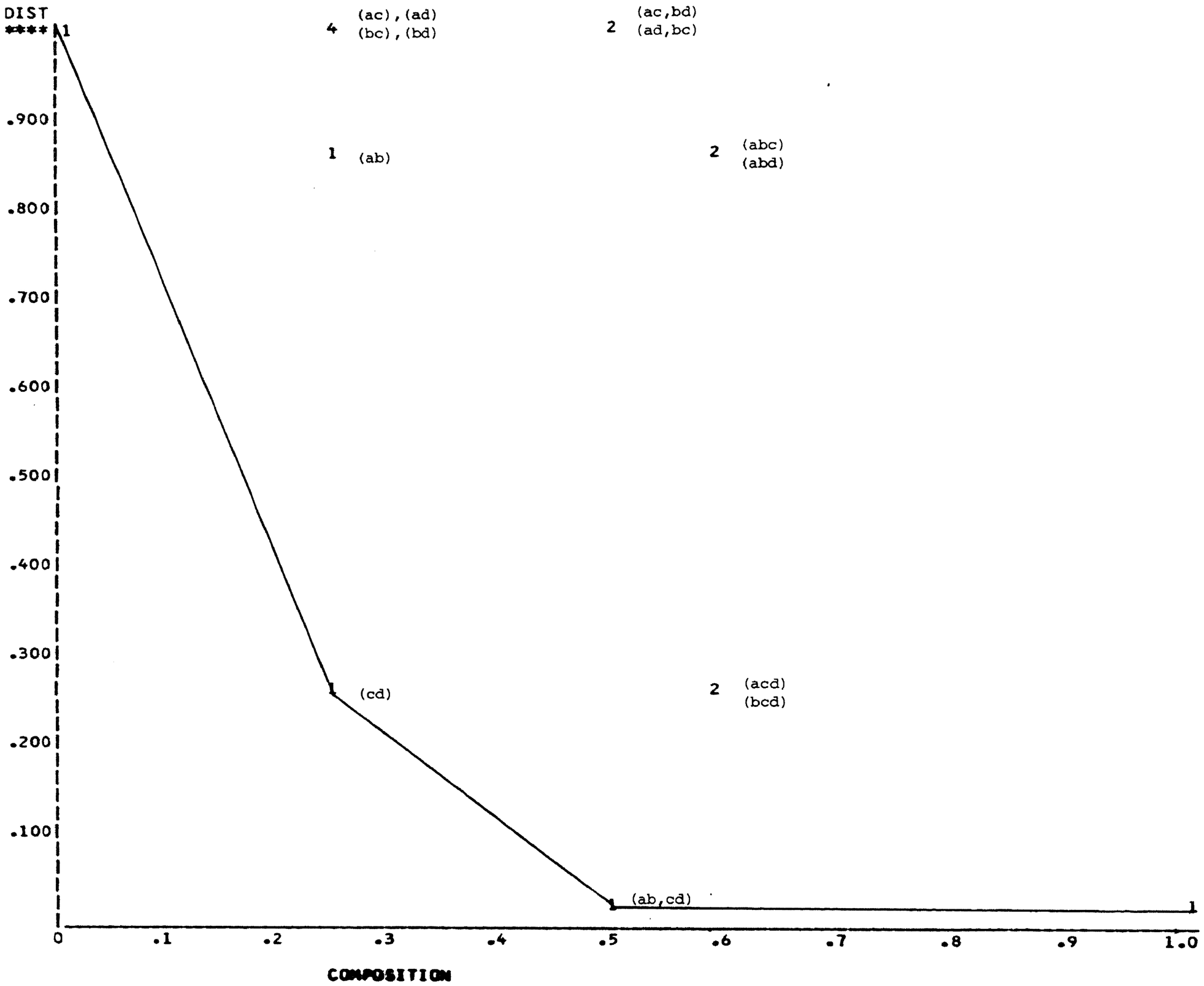

Figure 2 All arrangements for the 4×4-example.

gives the distortion on a logarithmic scale in such a way that the penultimate, but not necessarily the final, arrangement can be plotted. The horizontal lines represent the groups. When two groups are combined, the corresponding lines are joined by a vertical line, and the new group is represented by a new horizontal line. The number of the new group appears in the corner between the vertical line and the new horizontal line.

In this case it is an easy task to obtain a complete listing of all possible arrangements. For each arrangement, the composition is trivial to calculate, and since the eigenvalues of the original matrix are all real, the distortion can be calculated in the simple way outlined in Section 2.1. The

positions of all the arrangements in the composition-distortion space are shown in Fig. 2. A digit in this graph denotes the number of arrangements with the indicated distortion (± 0.01) and composition (± 0.005). The line in the graph joins the arrangements generated by our heuristic procedure. Note that these are exactly the Pareto-optimal points.

4. AN APPLICATION TO A SMALL MATRIX

To compare our asymmetric clustering procedure to some other approaches, we consider the 8×8 citation matrix used for illustrative purposes by Coombs, Dawes and Tversky [1970]. In this case, the elements are eight journals of the American Psychological Association. Each entry in the matrix gives the number of citations to the column journal occurring in the row journal in the year 1964. The matrix is reproduced in Fig. 3.

4.1. Results for our search procedure

It is instructive to follow the decisions made during the search. Those arrangements that are selected or close to being selected are given in Fig. 3, together with the eigenvalues of the selected arrangements.

After setting up the initial arrangement with one journal in each group, the procedure looks for the second arrangement. Three pairings are close rivals for the first agglomeration, all involving JExP, namely with JCPP, with JASP or with AJP in that order of preference. It is not surprising to find these three pairs being principal candidates: the two interaction entries for each pair are high.

However, not all the pairs with two high entries (e.g., JASP with JCP) come in for serious consideration. The reason is seen by comparing the eigenvalues of the original matrix, given in Fig. 3, with those for the initial arrangement, given by the self citations on the main diagonal of the original matrix. The major differences between corresponding pairs of eigenvalues occur for the largest and third largest pair. These are partially caused by the self citations of JExP and JASP, and indicate why this pair comes in for serious consideration. However, this pair is not chosen and instead JExP is combined with JCPP, which has the second largest number of self citations. This combination reduces the eigenvalue corresponding to JCPP to below that of JASP and brings it close to the third largest of the original eigenvalues, while JASP comes close to the second largest one, thus killing two birds with one stone. The pair with the highest interaction entries in the original matrix, JExP with AJP, does not get combined, because the self citation of AJP, being the fifth largest, is close to the fifth largest of the resulting

eigenvalues and so there is not much to be gained from their combination. It appears that self citations have some influence. If one finds that this feature is undesirable, then the main diagonal entries should be set to zero, as is frequently done is cluster analysis methods.

Having combined JExP with JCPP, the procedure looks for the third arrangement. Both JASP and AJP, the contenders for joining JExP as group 2, are considered for being joined to group 2 but rejected in favour of combining JASP with JCP to form group 3. AJP has to wait until the fourth step to be joined to group 2 to form group 4. At the fifth step groups 3 and 4 come together to form group 5, with no contenders. For the remaining three journals, JAP, Pka and JEdP, there is little cross citation but still sufficient for their combination in the sixth and seventh step of the search. Finally, in the eighth step the two groups of journals are brought together.

The result of this search are presented diagrammatically in the form of a dendogram in Fig. 3 (cf. Section 3.3). The citation matrix reordered into the same order as used in the dendogram is also given in Fig. 3.

4.2. Comparison with the complete listing

With eight elements to be grouped it is still possible to obtain a complete listing of all possible arrangements. The remarks made in the final paragraph of Section 3.3 apply here as well. The result is shown in Fig. 4; an asterisk denotes that there are ten or more arrangements falling within the indicated composition-distortion area. The arrangements with distortion at most 0.02 have been given again at the bottom of the graph, where the distortion scale is multiplied by ten; this is repeated for the arrangements with distortion at most 0.002.

As can be seen most of the arrangements are of no interest at all. Moreover the solution found by our procedure includes only Pareto-optimal points. Two arrangements lie very close to this lower bound. The first one is JCPP, JExP and AJP in one group and the other journals in separate groups. This arrangement has D = 0.0926 and C = 0.1981, and was considered but rejected at the third step. In fact, JCPP and JExP are joined by AJP at the fourth step. The second arrangement very close to the line is JAP and JEdP in one group and the other journals in a second group. This has D = 0.0003 and

INTERACTION MATRIX

	AJP	JASP	JAP	JCPP	JCP	JEDP	JEXP	PKA
AJP	119.0000	8.0000	4.0000	21.0000	0.0	1.0000	85.0000	2.0000
JASP	32.0000	510.0000	16.0000	11.0000	73.0000	9.0000	119.0000	4.0000
JAP	2.0000	8.0000	84.0000	1.0000	7.0000	8.0000	16.0000	10.0000
JCPP	35.0000	8.0000	0.0	533.0000	0.0	1.0000	126.0000	1.0000
JCP	6.0000	116.0000	11.0000	1.0000	225.0000	7.0000	12.0000	7.0000
JEDP	4.0000	9.0000	7.0000	0.0	3.0000	52.0000	27.0000	5.0000
JEXP	125.0000	19.0000	6.0000	70.0000	0.0	0.0	586.0000	15.0000
PKA	2.0000	5.0000	5.0000	0.0	13.0000	2.0000	13.0000	58.0000

ORIGINAL EIGENVALUES

R:	0.6898E 03	0.5291E 03	0.4606E 03	0.1984E 03	0.9635E 02	0.8694E 02	0.5579E 02	0.5010E 02
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

SUM OF SQUARES OF DIFFERENCES BETWEEN
EIGENVALUES AND MAIN DIAGONAL OF MATRIX
IS DENOMINATOR FOR DISTORTION = 0.1447E 05

SEARCH FOR NEXT PAIR OF GROUPS TO COMBINE

COMBINING				GIVES SLOPE	CHOSEN	
AJP		JEXP		-.3775E 01		
JASP		JEXP		-.4944E 01		
JCPP		JEXP		-.9780E 01	GROUP	2
JASP		JCP		-.8381E 00	GROUP	3
AJP	GROUP	2		-.8034E 00		
JASP	GROUP	2		-.2683E 00		
AJP	GROUP	2		-.8034E 00	GROUP	4
GROUP	2	GROUP	3	-.3209E 00		
GROUP	3	GROUP	4	-.1066E 00	GROUP	5
JAP		JEDP		-.8765E-02		
JAP		PKA		-.9990E-02	GROUP	6
PKA	GROUP	5		-.1798E-02		
JEDP	GROUP	5		-.4055E-03		
JEDP	GROUP	6		-.2552E-02	GROUP	7
GROUP	5	GROUP	7	-.6592E-03	GROUP	8

SEQUENCE OF ARRANGMENTS

GROUP	FORMED FROM				COMP.	DISTORTION
1					0.0	0.1000E 01
2	JCPP		JEXP		0.0833	0.1858E 00
3	JASP		JCP		0.1667	0.1151E 00
4	AJP		GROUP	2	0.2815	0.2291E-01
5	GROUP	3	GROUP	4	0.4837	0.1335E-02
6	JAP		PKA		0.5671	0.5027E-03
7	JEDP		GROUP	6	0.6819	0.2097E-03
8	GROUP	5	GROUP	7	1.0000	0.2955E-08

EIGENVALUES OF ARRANGMENTS

	1	2	3	4	5	6	7	8
R:	0.5860E 03	**0.6571E 03	0.6571E 03	**0.6749E 03	**0.6891E 03	0.6891E 03	0.6891E 03	**0.6898E 03
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.5330E 03	* 0.5100E 03	**0.5371E 03	0.5371E 03	**0.5286E 03	0.5286E 03	0.5286E 03	**0.5291E 03
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.5100E 03	*0.4619E 03	0.4619E 03	**0.4659E 03	**0.4606E 03	0.4606E 03	0.4606E 03	**0.4606E 03
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.2250E 03	0.2250E 03	**0.1979E 03	0.1979E 03	**0.1975E 03	0.1975E 03	0.1975E 03	**0.1983E 03
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.1190E 03	0.1190E 03	0.1190E 03	**0.9720E 02	**0.9720E 02	0.9720E 02	0.9720E 02	**0.9635E 02
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.8400E 02	0.8400E 02	0.8400E 02	0.8400E 02	0.8400E 02	**0.8580E 02	**0.8762E 02	**0.8694E 02
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.5800E 02	0.5800E 02	0.5800E 02	0.5800E 02	0.5800E 02	**0.5620E 02	**0.5640E 02	**0.5579E 02
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
R:	0.5200E 02	0.5200E 02	0.5200E 02	0.5200E 02	0.5200E 02	0.5200E 02	**0.4998E 02	**0.5010E 02
I:	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

REORDERED INTERACTION MATRIX

	JEXP	JCPP	AJP	JCP	JASP	JAP	PKA	JEDP
JEXP	586.0000	70.0000	125.0000	0.0	19.0000	6.0000	15.0000	0.0
JCPP	126.0000	533.0000	35.0000	0.0	8.0000	0.0	1.0000	1.0000
AJP	85.0000	21.0000	119.0000	0.0	8.0000	4.0000	2.0000	1.0000
JCP	12.0000	1.0000	6.0000	225.0000	116.0000	11.0000	7.0000	7.0000
JASP	119.0000	11.0000	32.0000	73.0000	510.0000	16.0000	4.0000	9.0000
JAP	16.0000	1.0000	2.0000	7.0000	8.0000	84.0000	10.0000	8.0000
PKA	13.0000	0.0	2.0000	13.0000	5.0000	5.0000	58.0000	2.0000
JEDP	27.0000	0.0	4.0000	3.0000	9.0000	7.0000	5.0000	52.0000

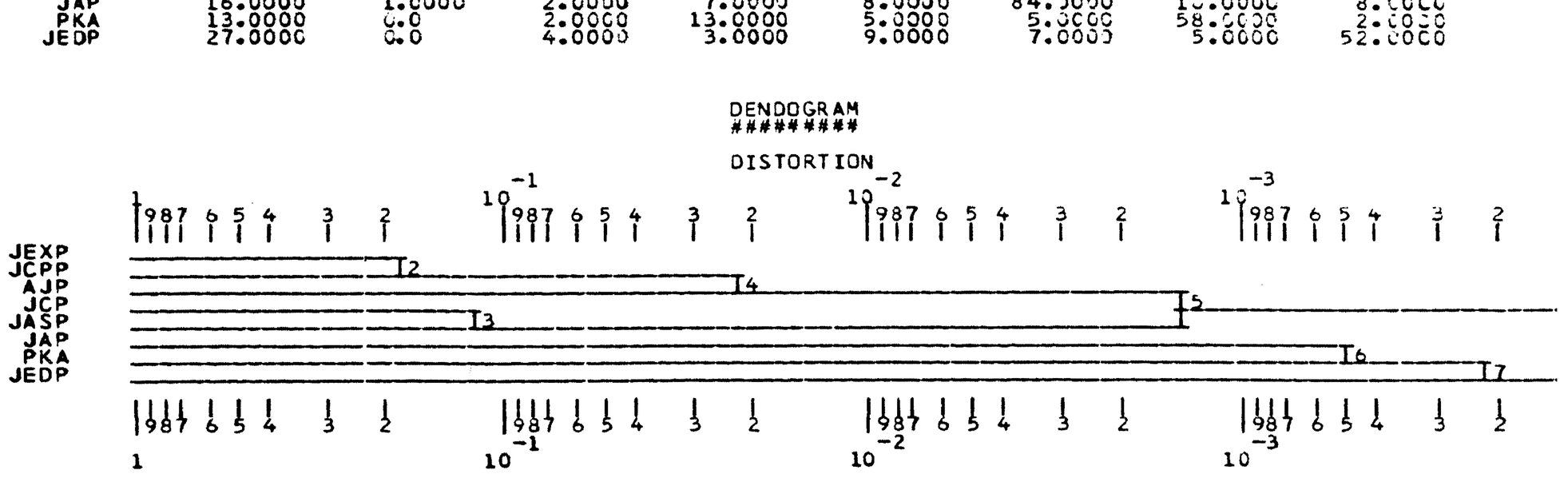

Figure 3 Data and results for the 8×8-example.

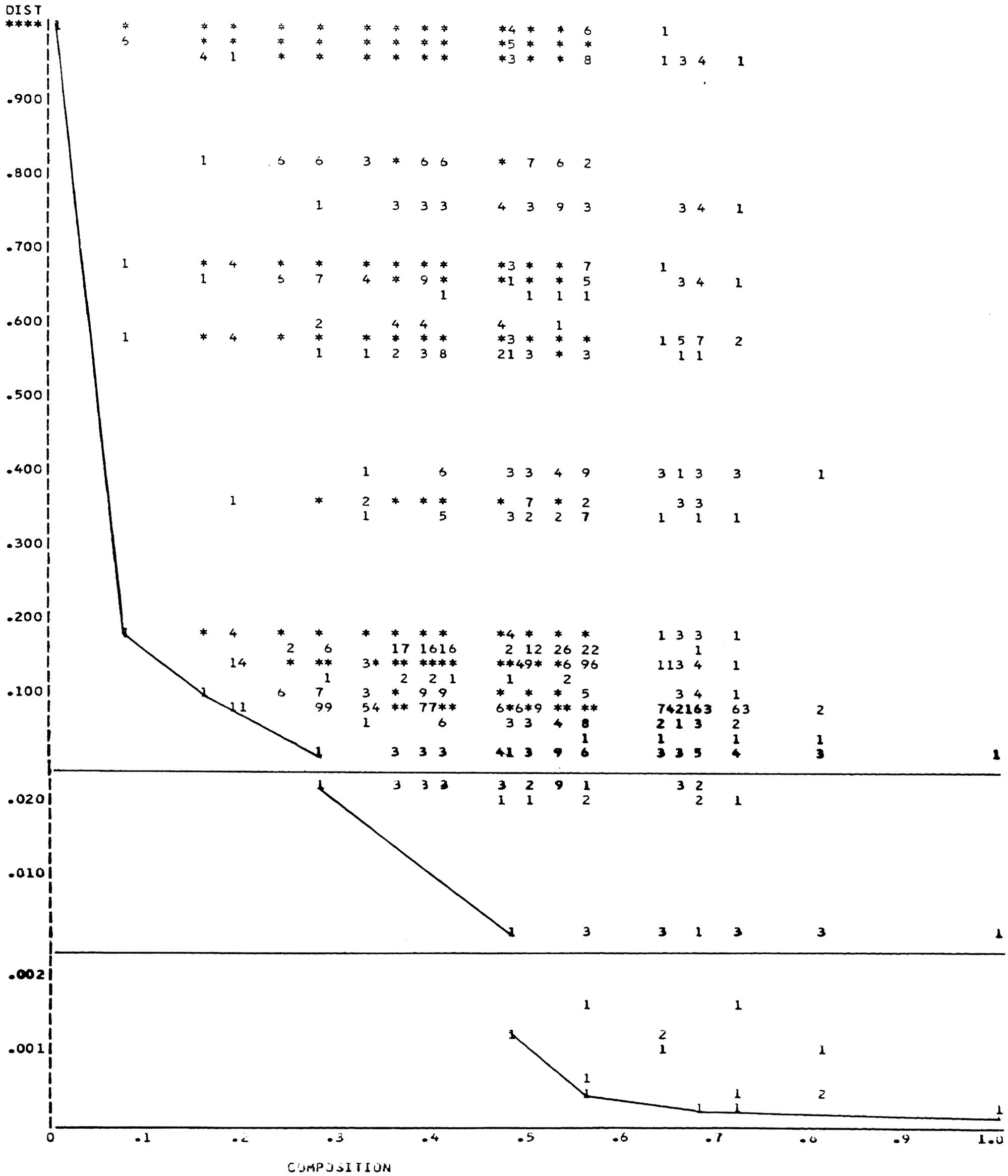

Figure 4 All arrangements for the 8×8-example.

C = 0.7296, and was never considered. It is slightly inferior to the arrangement found at the seventh step in which JAP and Pka are joined by JEdP. So in this instance the solution found by our heuristic is the optimum one, although there is no guarantee that this will always be the case.

4.3. Comparison to multidimensional scaling methods

Multidimensional scaling methods start by constructing a so-called I scale for each element j, i.e. an ordering of all elements k according to non-increasing row entries p_{jk}. They then proceed, usually in an iterative way, to create configurations of points, corresponding to the elements, in metric spaces of descending dimension, such that for each element j the ordering of the elements as defined by their distances to j is close to the I scale.

Coombs, Dawes and Tversky [1970] have analysed the 8×8 citation matrix using three multidimensional scaling methods. In an attempt to remove effects due to the total number of citations appearing in each journal, they subtracted row and column means from each entry. When locating the journals in two-dimensional Euclidean space, the three methods produced similar results. The solution obtained by the Guttman-Lingoes method is shown in Fig. 5.

How does this configuration compare to the dendogram given by our search procedure? For the multidimensional scaling method the most closely related journals are JAP, Pka and JEdP. In our procedure these journals only come together at the end of the search; they are related but not closely related. On the other hand, with us the most closely related journals are JExP with JCPP and JCP with JASP; together with ASP they form a group distinct from JAP, Pka and JEdP. For Coombs et al. JExP is close to JCPP and JCP to JASP, but not quite as close as JAP, Pka and JEdP; this group is closely related to AJP and JCP, and JExP, JCPP and JASP are outliers. The two representations of the data lead to very different interpretations!

Why it is that the Guttman-Lingoes method (and the other multidimensional scaling methods) bring together JAP, Pka and JEdP as a close group whereas our procedure, while bringing them together, only does so late indicating a weak interaction? The reason for their proximity with respect to multidimen-ional scaling methods is their small size, so that the number of citations rom any journal to these journal is also small. So already without the

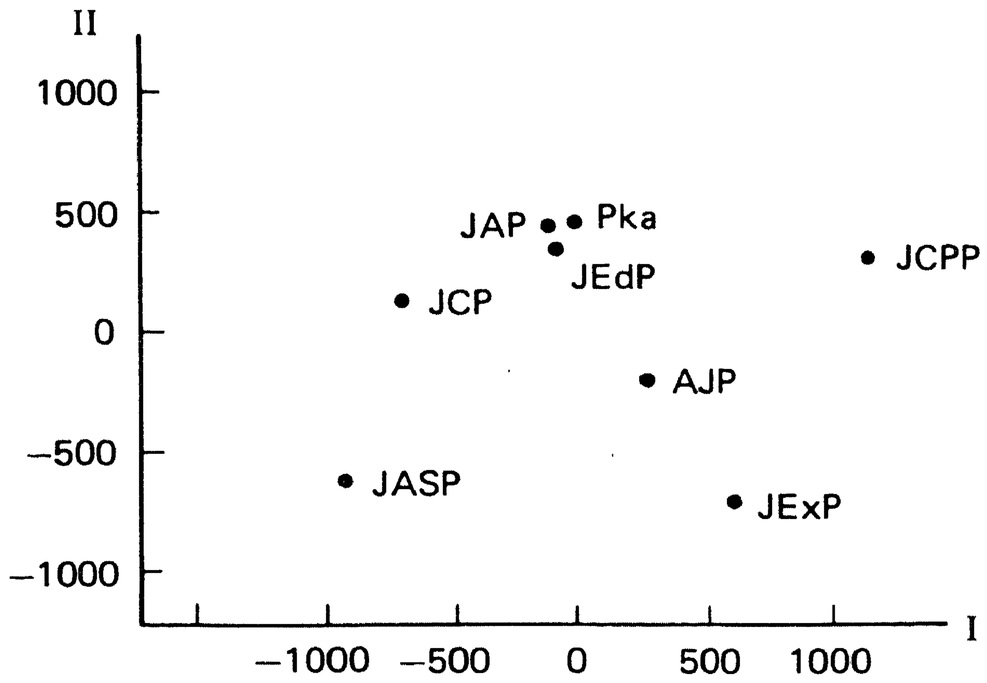

Figure 5 Guttman-Lingoes configuration for the 8×8-example.

subtraction of row and column means, the I scale for each journal would have listed them at the end. Subtracting row means does not change the I scales, but subtracting column means makes the situation worse. As there are few citations to JAP, Pka and JEdP, their column means are small (16.63, 12.75 and 10.00 respectively); none of their cross citations exceeds their respective means so that subtracting column means results in small negative entries. For other journals with higher column means the original small entries become large and negative. From the point of view of each of the five large journals, the three small journals will appear about equally distant. From the point of view of each of the three small journals, the five large journals will have large negative entries and the other two small journals will have small negative entries. Thus small journals are bound to come together, and subtracting row and column means does not remove the "bulk effects" caused by differences in journal size as was intended.

With our procedure no attempt is made to remove bulk effects, nor should it be. As the off-diagonal entries for JAP, Pka and JEdP are all small, their inclusion in a group rigthly has little effect on the eigenvalues. Once the five large journals have been combined it is not necessary that the three little journals will come together. That they do so is due to their cross citations; moreover that Pka joins the other two small journals rather than the five large ones is a marginal decision (see Section 4.2).

4.4. Comparison with Johnson's hierarchical clustering techniques

Two well-known hierarchical clustering techniques for symmetric matrices have been developed by Johnson [1967]. Initially, each element is in a separate group, and the similarity between two groups is given by the (symmetric) interaction between the corresponding elements. At each step, both methods combine the two groups with maximum similarity into a new group. The methods set the similarity between the new group and each of the remaining old groups to either the maximum (the single linkage method) or the minimum (the complete linkage method) of the similarities between the two groups being combined and the old group.

These methods can be applied to an asymmetric matrix provided that it is made symmetric, an approach taken by Hubert [1973]. He uses three different schemes for determining the similarity s_{jk} between elements j and k with interactions p_{jk} and p_{kj}:

(a) $s_{jk} = \min\{p_{jk}, p_{kj}\}$;

(b) $s_{jk} = \max\{p_{jk}, p_{kj}\}$;

(c) $s_{jk} = (p_{jk}+p_{kj})/2$.

When the single linkage method is applied to the 8×8 citation matrix, after it has been symmetrized by any of the above schemes, then the result is a string, i.e., two journals are combined and the remaining journals are added to this group one by one. The results of applying the complete linkage method depend on the scheme chosen for symmetrizing the matrix. However, schemes (a) and (c) give the same clustering and scheme (b) is very similar. The solution based on scheme (c) is presented in the form of a dendogram in Fig. 6. This is similar to the solution by our search procedure in that groups 3 and 4 are the same. Instead of their being combined to form group 5, the three remaining "little" journals are added to group 3. While these journals are certainly cited more often by the journals in group 3 than by those in group 4, they themselves cite JExP, which is in group 4, most frequently.

The main disadvantage of applying Johnson's methods lies, however, in the arbitrariness of the scheme needed to symmetrize the original matrix.

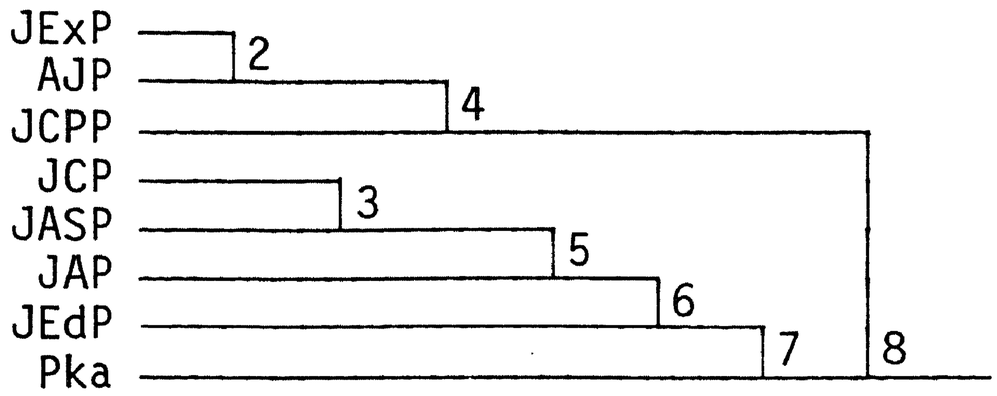

Figure 6 Complete linkage method for the 8×8-example.

4.5. Comparison with a travelling salesman approach

We know of only one other clustering technique that can handle asymmetric data without symmetrizing them. This method, originally proposed by McCormick, Schweitzer and White [1972] is not a hierarchical one and produces only one solution, which is not a clustering in our sense of the word. The objective of their approach is to permute the rows and columns of the interaction matrix so that high entries are brought together and a visually pleasing pattern emerges. The resulting clusters of high entries need not occur around the main diagonal, as in our search procedure.

As a criterion that has to be maximized over all row and column permutations, McCormick et al. chose the sum of all products of horizontally or vertically adjacent entries in the matrix. Thus, for an m×n-matrix $A = (a_{jk})$ $(j = 1,...,m;\ k = 1,...,n)$, their *measure of effectiveness* E is defined to be

$$E = \sum_{j=1}^{m}\sum_{k=1}^{n-1} a_{jk}a_{j,k+1} + \sum_{k=1}^{n}\sum_{j=1}^{m-1} a_{jk}a_{j+1,k}.$$

Maximizing E has been shown to be equivalent to solving two *travelling salesman problems* [Lenstra, 1974]. This is a standard combinatorial optimization problem, for which various algorithms have been developed; see, e.g., [Christofides, 1975, p.236; Lenstra, 1977, p.63]. If the matrix is square (i.e., m = n) and we restrict our attention to identical row and column permutations, as in the case of our example, then only one travelling salesman problem has to be solved [Lenstra & Rinnooy Kan, 1975].

Applying this approach to the 8×8-example, we have chosen to define a_{jk} as the number of significant digits of p_{jk} (i.e., $a_{jk} = 0$ if $p_{jk} = 0$, $a_{jk} = 1$ if $1 \leq p_{jk} \leq 9$, etc.). The results are shown in Fig. 7. E increases from 238

for the original matrix to 301 for the permuted matrix. Inspection reveals that the clusters of high entries correspond to the principal submatrices formed by our search procedure.

The reordered interaction matrix given in Fig. 3 has E = 269. However, the underlying reordering is quite arbitrary. Our dendogram could, for example, have been drawn on an ordering with E = 296, namely the one obtained by interchanging JCPP and AJP in the optimal permutation shown in Fig. 7(b).

AJP	3	1	1	2	0	1	2	1
JASP	2	3	2	2	2	1	3	1
JAP	1	1	2	1	1	1	2	2
JCPP	2	1	0	3	0	1	3	1
JCP	1	3	2	1	3	1	2	1
JEdP	1	1	1	0	1	2	2	1
JExP	3	2	1	2	0	0	3	2
Pka	1	1	1	0	2	1	2	2

(a) Interaction matrix; E = 238.

JCPP	3	2	3	1	0	0	1	1
AJP	2	3	2	1	0	1	1	1
JExP	2	3	3	2	0	1	2	0
JASP	2	2	3	3	2	2	1	1
JCP	1	1	2	3	3	2	1	1
JAP	1	1	2	1	1	2	2	1
Pka	0	1	2	1	2	1	2	1
JEdP	0	1	2	1	1	1	1	2

(b) Permuted interaction matrix; E = 301.

Figure 7 Travelling salesman problem for the 8×8-example.

5. CONCLUDING REMARKS

We have discussed the application of various methods to the 8×8 citation matrix to give a feel for how our method works and compares with others. Our method has also been applied to larger problems.

The search procedure is coded in FORTRAN IV and has been run on an IBM 370/158. Solution of one of the 16×16 consonant confusion matrices from Miller and Nicely [1955] required 8 seconds, and solution of a 37×37-matrix of information transfers between workers in an R&D laboratory from Whitley, Bitz and McAlpine [1975] required 110 seconds. However, a 96×96-matrix could not be solved within 40 minutes. We suspect that a 50×50-matrix is about the largest that can be handled by the present version of the program within reasonable time.

Possibilities for improving the efficiency exist within the search scheme described in Section 3. First, the eigenvalues for each new matrix are calculated from scratch using a standard subroutine. However, since the eigenvalues of the principal submatrices are known, they could be used as starting values for a tailor-made subroutine. Secondly, data manipulation for each arrangement is presently $O(n^2)$. This may be reduced to $O(n)$. Finally, each linear assignment problem is solved completely. Instead, with Dorhout's [1977] algorithm it is possible to stop the computations as soon as a given lower bound is reached, which may lead to early elimination of the arrangement considered. We plan to make some improvements along these lines.

Even with these improvements, the consideration of $O(n^3)$ arrangements leads to a time-consuming search procedure. A revision of the heuristic would be required to handle problems with more than 100 elements.

Based on these preliminary experiences, how does our approach compare with similar methods? It is closest in aim to Johnson's [1967] well-known hierarchical clustering schemes for symmetric matrices. Their advantage is clarity and computational simplicity; their disadvantage is that a decision must be made on how to symmetrize an asymmetric matrix [Hubert, 1973]. The multidimensional scaling methods provide a picture of the elements in a space from which a clustering can be deduced, but they put elements close together because their interaction totals are small rather than because they are highly interactive. The travelling salesman approach provides only a reordering

of the matrix rather than a clustering. Since it is suitable for large problems, it might be used to yield a starting point for our method to confine the agglomeration to elements and groups that are adjacent in the resulting permutation.

REFERENCES

Anderberg, M.R. *Cluster analysis for applications*. New York: Academic Press, 1973.

Christofides, N. *Graph theory: an algorithmic approach*. New York: Academic Press, 1975.

Coombs, C.M., Dawes, R.M., and Tversky, A. *Mathematical psychology: an elementary introduction*. Englewood Cliffs, New Jersey: Prentice-Hall, 1970.

Dorhout, B. Experiments with some algorithms for the linear assignment problem. Amsterdam: Mathematisch Centrum, Report BW 39, 1977.

Duran, B.S. and Odell, P.L. *Cluster analysis: a survey*. Berlin: Springer-Verlag, 1975.

Everitt, B. *Cluster analysis*. London: Heineman, 1974.

Fortier, J.J. and Solomon, H. Clustering procedures. In: Krishnaiah, P.R., ed. *Multivariate analysis: proceedings of an international symposium, Dayton, Ohio, 1965*. New York: Academic Press, 1966.

Gower, J.J. A comparison of some methods of cluster-analysis. *Biometrics*, 1967, 23, 623-637.

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

Hubert, L. Min and max hierarchical clustering using asymmetric similarity measures. *Psychometrika*, 1973, 38, 63-72.

Jardine, N. and Sibson, R. *Mathematical taxonomy*. New York: Wiley, 1971.

Jensen, R.E. A dynamic programming algorithm for cluster analysis. *Operations Research*, 1969, 17, 1034-1057.

Johnson, S.C. Hierarchical clustering schemes. *Psychometrika*, 1967, 32, 241-254.

King, B. Step-wise clustering procedures. *Journal of the American Statistical Association*, 1967, 62, 86-101.

Lawler, E.L. *Combinatorial optimization: networks and matroids*. New York: Holt, Rinehart and Winston, 1976.

Lenstra, J.K. Clustering a data array and the traveling-salesman problem. *Operations Research*, 1974, 22, 413-414.

Lenstra, J.K. *Sequencing by enumerative methods*. Amsterdam: Mathematisch Centrum. Mathematical Centre Tracts 69, 1977.

Lenstra, J.K. and Rinnooy Kan, A.H.G. Some simple applications of the travelling salesman problem. *Operational Research Quarterly*, 1975, 26, 717-733.

Liu, C.L. *Introduction to combinatorial mathematics*. New York: McGraw-Hill, 1968.

McCormick, Jr., W.T., Schweitzer, P.J., and White, T.W. Problem decomposition and data reorganization by a clustering technique. *Operations Research*, 1972, 20, 993-1009.

Miller, G.A. and Nicely, P.E. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America*, 1955, 27, 338-352.

Simon, H.A. The architecture of complexity. *Proceedings of the American Philosophical Society*, 1969, 106, 467-482.

Simon, H.A. and Ando, A. Aggregation of variables in dynamic systems. *Econometrica*, 1961, 29, 111-138.

Sokal, R.R. and Sneath, P.H.A. *Principles of numerical taxonomy*. San Francisco: Freeman, 1963.

Tryon, R.C. *Cluster analysis: correlation profile and orthometric (factor) analysis for the isolation of unities in mind and personality*. Ann Arbor, Michigan: Edward Bros, 1939.

Wagner, H.M. *Principles of operations research with applications to managerial decisions, second edition*. Englewood Cliffs, New Jersey: Prentice-Hall, 1975.

Wells, M.B. *Elements of combinatorial computing*. Oxford: Pergamon Press, 1971.

Whitley, R.A., Bitz, A., and McAlpine, A. The production, flow and use of information in research laboratories in different sciences. Manchester: Manchester Business School, Report ISSN 0306-5227, 1975.