RESIDUAL ANALYSIS OF THE QUASI-SYMMETRY MODEL
Applications in the citation-based assessment of journal influences

R. J. W. Tijssen,* J. de Leeuw and A. F. J. van Raan***

**Science Studies Unit, LISBON-Institute, University of Leiden*
Stationsplein 242, 2312 AR Leiden, The Netherlands
*** Department of Data Theory, University of Leiden*
Middelstegracht 4, 2312 TW Leiden, The Netherlands

Abstract - A quasi-symmetry model is applied to a matrix of cross-citations between scientific journals. The difference between the observed citation values and the model-based expected values for the citation frequencies, are used to derive a measure of interjournal-influence and an accompanying journal-influence ranking. In addition, a residual analysis method is presented to gain spatial insight in the total interjournal-influence structure. Both methods will be illustrated in two applications.

INTRODUCTION

Scientometrics can be described as a research approach in studies of science in which primarily quantifiable aspects of science are utilized. Scientometric research methods can thereby yield measures which one can use to measure and categorize specific characteristics of science and its contributors. The usefulness of these science measures for evaluating and monitoring has proved a major factor in the increasing popularity of a quantitative oriented research within studies of science and technology, for both science policy makers as well as scholars. The measures have been particularly fruitful in investigation of the (inter-) relationships between professional scientific journals, and scientific fields were scientific journals are a vital channel in the communication system. Journals reflect the interests, activities and communication processes within a scientific community. Various journal characteristics with respect to publication-, dissemination- and evaluation of scientific knowledge contain quantitative information for an objective comparison of journals (cf. Osborne, 1984). Journal-to-journal relations can be assessed by means of so-called *bibliometric* measures, i.e. based on quantified characteristics of scientific literature itself. Bibliometric data is often derived from the lists of references (citations) from publications in journals. Generally, a reference from and - vice versa - citation to a publication can be seen as an formal acknowledgement of scientific work (in the sequel we adopt common practice of refering to references as citations).

The number of citations to a publication is generally recognized as an indicator of the influence of a piece of published work on the scientific community. However, many authors questioned the validity of citations as genuine tokens of cognitive debts, suggesting additional or alternative descriptions (cf. e.g., Gilbert, 1977). The validity of individual citations remains a matter of controversy as a result of 'distrubing factors' such as the existence of negative, irrelevant, and self-citations as well as existing field-dependent citation practices (cf. Edge, 1979 ; Moed et al., 1985) which both random as well as systematic sources of bias when comparing citations as a measure of communication. In addition, MacRoberts & MacRoberts (1986) and Brook (1986) have found evidence that citations only partly cover the area intended. Nevertheless, several empirical investigations have established that that citation counts can serve as an useful measure of the certain properties of scientific documents, for example: positive correlations have been found with several objective and subjective measures of scientific prestige (Clark, 1957 ; Bayer & Folger, 1966 ; Simonton, 1984). Cole &Cole (1972) have found evidence that high citation counts relate positively with honoric awards and Nobel laureates. Moed et al. (1983) is the first large-scale study in which bibliometric measures of scientific performance are investigated within the context of several scientific disciplines. The results have been found to correlate positively with judgements of the scientists themselves and peers. In general, it is clear that high-quality work or distinguished scientists are cited more than average, whereas no evidence is found that scientists or work of less than average-quality attract relatively large numbers of citations.

The existing error at the level of single article-to-article citations is less problematic when using aggregrates of citations, such as the total number of citations between scientific journals, because random error will tend to be canceled out at higher aggregation levels and will thus only have a relatively minor effect on citation relationships between journals. Therefore it seems sufficiently justified to use citations counts in the assessment of formal communication (i.e., citation transactions) between scientific journals, as well as providing measures of the 'influence' journals have on each other. The differences in publication- and citation practices within and, in particular, between scientific fields can be dealt with by applying an appropriate normalisation-conditions, e.g. resulting in 'size'-independent citation counts (e.g., Pinski&Narin, 1976 ; Tijssen et al., 1987). Considering the points mentioned above, it will assumed in the following that citation counts yield a relatively unobtrusive indication of the merit and utility of the cited journal in the scientific community, and citation transactions can thus provide meaningful information for assessing (inter-)relationship between scientific journals.

During the last decades, journals have attracted much attention in the sociology of science as an suitable point-of-entry in assessing the cognitive structures. Citation-based quantitative measures evaluating relations between scientific journals have proved to be a valuable asset in evaluating a cognitive structure of different fields of science (e.g., Narin, Carpenter & Berlt, 1972 ; Small & Griffith, 1974).

Evaluating interjournal-citation structures has also turned out to be a suitable approach for providing journal classifications and rankings indicating the level of influence of individual journals on other journals. Citation-based rankings of journals are often derived for selection or evaluative reasons of scientific performance of scientists (cf. Garfield, 1983).

In this paper we apply a probabilistic model to the structure of the citation-transaction structure and use the residuals to derive citation-based measures indicating the influence from a journal on one or more other journals. In doing so, we realise that basic terms such as 'influence' or 'prestige' are not conceptually defined. Our operationalisation of the concept influence is based on citations only, although we are aware of the fact that empirical investigations have shown that journal influence measures include extraneous disturbing factors, i.e. factors not related to instrinsic properties of journals : e.g., the age of a journal, its type of contents (e.g. review journals), or an affiliation with a professional organization.

CITATION-BASED ASSESSMENT OF INFLUENCES BETWEEN JOURNALS

Many computational procedures for the assessment of citation-based influences (also referred to as 'impacts') share the same basis : they operate on a journal-to-journal citation data. The data is generally displayed as a square or rectangular citation transaction-matrix, i.e. an array of cross citation frequencies $C = \{c_{ij}\}$ in which the same set of journals is assigned to the rows i (i=1,...,I) and to the columns j (j=1,...,J) of the matrix. The rows classifying the cited-mode of the journal (the citations received) and the columns the citing-mode (the citations issued). The elements c_{ij} of such a citation matrix contain citation frequencies, indicating the level of citation-transaction between journals for both modes based on the number of citable (source) items s in a specified time period t. The values c_{ii} in the main diagonal represent self-citations. The (relative) self-citation magnitude can be considered an indicator of the 'centrality' of a journal within a (sub)discipline ; a journal will dominate a (sub)field or specialty over more peripheral journals in the sense that it attracts authors (with citations to the same journal). Adjustment or elimination of these intra-journal citations in an interjournal-oriented citation analysis is an important methodological point of consideration. Self-citations very often have relatively high values, as they are partly the result of the qualitatively different and empirically well-known referencing process in which the authors tend to utilize a limited number of journals for both reading and publishing. The references to publications in a journal is enhanced due to the fact that cited - as well as citing publications are relatively often published in the same journal. Of course, such a phenomenon is quite understandable since publications within specialized fields will often be concentrated in a (very) limited number of journals. However, if one is primarily interested in inter-journal relations, these intra-journal relations are of less importance. Besides, in many cases incorporating self-citations means computational problems, simply because the relatively high citation-frequencies will dominate the analysis results and thus obscure information on the actual structure of interrelations.

Gross and Gross (1927) were pioneers in citation-based assessment of scientific journals and their grading with respect to their relative importance within the set of journals. Applications of their methodology, from the thirties onward, are listed in Sengupta (1986). In the last decades, more sophisticated methods for determining journals rankings were introduced (e.g., Yanovsky, 1981 ; Todorov, 1984 ; Liebowitz & Palmer, 1984 ; Doreian, 1985, 1987). Several review articles have discussed properties of journal citation measures and their rankings (e.g., Singleton, 1976 ; Weishart&Regoli, 1984 ; Todorov & Glänzel, 1987) We discuss three methods in more detail. Garfield (1972) introduced the *Impact Factor* (IF), which has probably become the most commonly applied analytic aid in the citation-based ranking of journals.

$$IF_t = (c_{t-1} + c_{t-2}) / (s_{t-1} + s_{t-2}) \tag{1}$$

The IF incorporates citations received from all journals covered by the Institute for Scientific Information (ISI). A (sub)disciplinary assessment is possible with the *Disciplinary Impact Factor* (DIF) operating on an iteratively determined set of core journals (Hirst, 1978).

$$DIF_t = (c_{t-1} + ... + c_{t-n}) / (s_{t-1} +... + s_{t-n}) \tag{2}$$

with $n \geq 1$.

Computation of IF and DIF thus involves the whole citation flow between journals, without weighting the citations received. However, as mentioned in the previous section, more recent empirical investigations have given evidence that several types of citations exist, carrying unequal 'weights' as a valid basic measure of influence. Also citing practices differ between scientific disciplines. A discplinary-oriented assessment method remedying citation-equalities was given by Pinski & Narin (1976). They developed an iterative method operating on the journal-to-journal citation matrix. A journal-size independent influence measure - the *Influence Weight* (IW) - is computed for an assessment of influences within a given set of journals, in which a higher amount of citations received by a (more prestigious) journal yields a higher weight for the citations given by that journal. The self-citations are excluded from the computations. For each citing journal they compute

$$IW_j = \Sigma_i \, IW_i \, (c_{ij} / S_j) \tag{3}$$

with $S_j = \Sigma_i c_{ij}$, normalised as $\Sigma_j IW_j S_j / S_j = 1$.

Pinski & Narin's method also treats the the citation flows between journals as a unity. However, analogously to the variety of reasons for citing individual publications, one may assume that not only influence determines the strenght and direction of citation flows. At a journal-level one can think of such as the range of subjects, type of contents, or extraneous factors like the age of a journal or an affiliation with a professional organization. Such factors can be translated into the following - broadly described - terms : 'size', 'contents-similarity' and 'visibility'. Of course, such journal properties are generally interdependent.

In the following section we introduce an alternative approach which also operates on the journal-to-journal citation matrix, enabling a computation of influence measures within a (disciplinary-bound) set of journals. It is based on a probalistic model of the network of citation flows between journals. The model adjusts for the abovementioned 'disturbing' factors, by dividing the citation flow into a number of subflows. Each subflow is incorporated in the model with a separate parameter. In addition, the self-citations are discarded. The journal 'influence' is operationalized as a model-based (normalized) residual number of citations between journals.

MODELING THE CITATION TRANSACTION MATRIX

The so-called *independence model* can be seen as the basic probabilistic model to approximate citation relationships between journals within a journal-to-journal citation matrix. In this model one assumes that cited journals i and citing journals j have an independent contribution to the transaction values, i.e. the probability p_{ij} of a citation transaction from journal i to journal j is equal to the probability of a citation from j, multiplied by the probability of receiving citations in row i. In formula :

$$p_{ij} = p_{i+} \; p_{+j} \tag{4}$$

were '+' denotes the summation over the omitted index.

Multiplying by the grand sum of the citation transactions - $N = c_{++}$ - yields the expected number of citation transactions from journal i to j

$$e_{ij} = N \, p_{i+} \; p_{+j} \tag{5}$$

The independence model is a member of a family of modeling techniques described as *loglinear analysis*. In general, loglinear analysis investigates the structural relationships between units contained in a matrix, by modeling the data as a linear function of a number of parameters (cf. Bishop, Fienberg & Holland, 1975). By taking the logarithms in Eq.(5), we can define the following loglinear independence model for the probabilities of transactions, with parameters u, expressed in analogy of analysis-of-variance :

$$\log e_{ij} = u + u_{1(i)} + u_{2(j)} \tag{6}$$

with the parameter u for the overall mean $u = 1/4\ \Sigma_i\ \Sigma_j \log e_{ij}$, and two parameters for the so-called first-order interactions ; (a) the effect of row i, i.e. $u_{1(i)} = \Sigma_j \log e_{ij} - u$, (b) the effect of the column j, i.e. $u_{2(j)} = \Sigma_i \log e_{ij} - u$.

The following constraints are imposed on Eq. (!) : $\Sigma_i\ u_{1(i)} = \Sigma_j\ u_{2(j)} = 0$.

In loglinear analysis one computes the *maximum likelihood estimates*[1] - $\hat{e}_{ij}$ - of the expected values. In case of the independence model $\hat{e}_{ij}$ is identical to e_{ij}, thus

$$\log e_{ij} = \log \hat{e}_{ij} = \hat{u} + \hat{u}_{1(i)} + \hat{u}_{2(j)} \tag{7}$$

Loglinear analysis also enables one to evaluate the difference between the observed citation values and the estimates of expected values derived from a specified loglinear model, with aid of the *Pearson chi-square statistic*

$$\chi^2 = \Sigma_i\ \Sigma_j\ (c_{ij} - \hat{e}_{ij})^2 / \hat{e}_{ij} \tag{8}$$

The χ^2-statistic has an asymptotic chi-square distribution with df=(I-1) (J-1) degrees of freedom, when the specified model is true. If the specified model is untenable, the statistic provides quite strong evidence yielding a high χ^2-value, with a probability (p) near zero, according to the chi-square distribution.

This model will only incorporate the extent to which citation exchanges are due to 'visibility' and 'size' of journals (i.e, the row and column sums of the matrix, respectively). When dealing with journal-citation matrices the expected values under the independence model will generally deviate from the observed values. A more sophisticated model will yield a better approximation. As a next step one can extend the model with a second-order interaction parameter - $u_{12(ij)}$. The estimated value of this parameter represent the symmetric interaction-effect between journal i and journal j [i.e., $u_{12(ij)}=u_{21(ji)}$]. Analogous to Eq. (7) this so-called *symmetry model* is defined as

$$\log e_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \tag{9}$$

with constraints : $u_{1(i)} = u_{2(j)}$; $u_{12(ij)} = u_{21(ji)}$; $\Sigma_i u_{1(i)} = \Sigma_j u_{2(j)} = \Sigma_i u_{12(ij)} = \Sigma_j u_{21(ji)} = 0$

Fitting this model to the transaction-matrix means that both the first order ('size')-effects and the second-order ('similarity') effect of the journals will be accounted for.

In view of the foregoing, one would thus like to fit a model like the symmetry model, and at the same time eliminate self-citations in order to obtain a suitable matrix for constructing influence measures of journals. These conditions prevent the use of the symmetry model because the suitable model must provide a separate and adequate modeling of the self-citations. Moreover, the symmetry model assumes equal row and column sums which is generally not the case in journal to journal citation-matrices. The model in which these conditions are fulfilled is referred to as the *quasi-symmetry model.*

In general, this model enables one to investigate if significant asymmetry exists in a matrix, which is not due to the difference in the row- and/or column sums. In this model it is assumed that :

$$\log e_{ij} = u + u_{1(i)} + u_{2(j)} + u_{12(ij)} \qquad (10)$$

with constraints : $u_{12(ij)} = u_{21(ji)}$; $\Sigma_i u_{1(i)} = \Sigma_j u_{2(j)} = \Sigma_i u_{12(ij)} = \Sigma_j u_{21(ji)} = 0$.

The maximum likelihood estimates of the expected transaction-values must satisfy the following likelihood equation :

$$\log \hat{e}_{ij} = \hat{u} + \hat{u}_{1(i)} + \hat{u}_{2(j)} + \hat{u}_{12(ij)} \qquad (11)$$

with restrictions : $\Sigma_i \Sigma_j \hat{e}_{ij} = c_{++}$; $\Sigma_j \hat{e}_{ij} = c_{i+}$; $\Sigma_i \hat{e}_{ij} = c_{+j}$; $\hat{e}_{ij} + \hat{e}_{ji} = c_{ij} + c_{ji}$.

These equations imply that $\hat{e}_{ii} = c_{ii}$, which is the required property to eliminate the self-citation values, as will become clear in the sequel.

This set of maximum likelihood equations can not be solved in a closed form, i.e. it is not possible to compute a direct estimation of the expected values from the observed values, as in the independence model. Instead, the equations are solved by a so-called iterative maximum likelihood algorithm of the '*iterative proportional fitting*'-type (cf. Deming & Stephan, 1940). The principle of this algorithm is to compute values $\hat{e}_{ij}$, satisfying Eq.(11), through a cycle in which the abovementioned restrictions are subsequently carried out. This cycle is repeated until all $\hat{e}_{ij}$'s converge to a stable value with an acceptable degree of accuracy. The goodnessof-fit of the quasi-symmetry model can then be computed by means of the χ^2 -statistic [cf. Eq. (8)], with (I-1)(J-1)/2 degrees of freedom.

Quasi-symmetry models will generally suffer of a lack of fit when applied to a matrix with journal to journal citation transactions, especially journals with asymmetric citation-relationships. This was illustrated in Noma (1982) who also fitted a - a slightly different version[2] - of our quasi-symmetry model to a journal to journal citation-matrix. The emphasis in Noma's study was put on evaluation of journal similarities. In contrast to our aim, values of symmetric interaction-parameters (i.e., $u_{12}=u_{21}$) were used as input data in a multidimensional scaling technique to obtain a spatial display of journal similarity. Complementary, Noma also evaluated the values of the set of residuals by means of the Freeman-Tukey (F-T) statistic (see Bishop, Fienberg & Holland, 1975). Using this statistic, the residuals are evaluated as samples from a normal distribution. Consequently, one can classify the values of the residuals as being significantly or not significantly deviant from what would be expected on the basis of a normal distribution. One can thus specify a statistical cutoff-value, say .001, which indicates that 1% of the significant deviations may occur due to mere chance. Highly significant positive residual values indicate an additional citation-flow and thus a significant relative influence between journals. No absolute measure of influence was given. However, evaluation of the F-T values can only be done for each element of the matrix separately. In our method we attempt to complement Noma's approach, by introducing an analysis of the joint interrelations between residuals. We will therefore revisit this approach, more in particular the data used and the corresponding results of his method, in one of the following applications of our method.

The quasi-symmetry model is merely used as a baseline-model, i.e. to filter out certain information from the citation transaction matrix ; our interest is centered on the deviates from the model. Subtracting the estimates of the expected citation-values, under the quasi-symmetry model, from the observed values will yield a residual matrix $N=\{n_{ij}\}$ with values usable as an indication of the influence between journals. We now suppose that an influence of a citing journal j on a cited journal i is marked by a net citation-flow n_{ij} from journal j to i. In addition, self-citation values $n_{ii}= c_{ii} - \hat{e}_{ii} = 0$.

RESIDUAL ANALYSIS

Influence measures

Normalisations are in order to transform residual values n_{ij} into a meaningful measure of influence. A simple influence value z_{ij} can computed as the amount of citations after fitting the quasi-symmerty model, thus

$$z_{ij}^{(a)} = (n_{ij} / x_{ij}) - 1 \tag{12a}$$

A positive z_{ij} will reflect a positive net amount of citations, i.e., a positive citation-balance. In this case the journal i has a positive influence on the other journal j ; such a journal thus receives more citations from j then expected. A negative value will indicate an reverse citation-flow ; journal j has an influence on journal i.

Additional standard transformations which can be applied in this situation are :

$$z_{ij}^{(b)} = r_{ij} / \hat{e}_{ij}^{1/2} \tag{12b}$$

These values are generally referred to as the standardized residuals and are simply the square-roots of the components used in the chi-square statistic.

$$z_{ij}^{(c)} = \sqrt{c_{ij}} + \sqrt{(c_{ij}+1)} - \sqrt{(4\,\hat{e}_{ij} + 1)} \tag{12c}$$

These are the already mentioned Freeman-Tukey (F-T) residuals. When it is assumed that the citation data are accumulated according to a Poisson process, these values have an asymptotic N(0,1)-distribution

For simplicity and illustrative reasons we will use the $z_{ij}^{(a)}$ - transformation in this report. The F-T residual transformation is used for determining journal-influence rankings in Tijssen & Van Raan (1988).

Subsequently, a overall influence of a journal i on all other journals j is equal to the row sum :

$$z_i = \Sigma_j\, z_{ij} \tag{13}$$

Analogous to Noma's assessment of the residual values as a sample of the normal distribution, these influence measures need a validation as to their importance in the whole of residual citation-flows between journals. In the second stage, our interest is focussed on the analysis of this remaining citation transaction-structure. A suitable data analysis-technique is introduced, which can display the relations within the structure of residual values.

Analysis of the residuals structure

The analysis technique used is based on the so-called *singular value decomposition* (see e.g., Golub & Reisch, 1970), a well-known method to obtain an optimal approximation of the data structure in a rectangular matrix through a limited number of components. The general idea behind the use of this decomposition is to arrive at a sparse representation of data in the independent (in a geometric terms referred to as orthogonal) components which account for most information ('variance') in the matrix. It is important to note that, contrary to *Noma's* approach, the evaluation of residual values by means of this decompostion assumes no model-distribution. In non-mathematical terms, the singular value decompositon can be seen as a decompostion of a matrix in a number of pairs of mutually independent vectors for the rows and columns, each vector containing a score for each row or column, respectively. Each pair of vectors has a corresponding singular value ; the hight of the value indicates the extent in which the vectors approximate the structure of the rows and columns of a matrix. The set of vectors corresponding to the highest singular value are given first. They yield the best rank-one approximation of a matrix. Thus the singular value decomposition reorganizes the data and helps us to understand the structure of the data better. The singular value decomposition is the basis of a data-analysis technique with some attractive descriptive properties, *Correspondence Analysis* (CA hereafter), which will be first introduced in the following. In a futher stage, CA is adapted to suit the form of the residual data.

In general, CA can be seen as a modelfree technique to find an optimal multidimensional representation of the (inter-)relations between rows and columns (Greenacre, 1984). In fact, CA is basically a singular value decomposition of the residuals, albeit after fitting the independence model to a rectangular matrix. CA can also be used for the analysis of residulas from models other than the independence model (cf. Van der Heijden & De Leeuw, 1985)

CA can be defined in terms of deviations from the independence model as follows. With C as the observed citation matrix, entries c_{ij} adding up to n,.the row sums c_{i+} are contained in the diagonal matrix D_r, while D_c contains the column sums c_{+j}. The matrix with expected values E is equal to $(D_r\, t\, t'\, D_c)/n$, with elements e_{ij} based on the independence model. The accent denotes the transpose of a vector or matrix. The vector t contains elements equal to one. The matrix containing the standardized residuals, after accounting for the row and column effects, is analyzed by computing the singular value decomposition

$$D_r^{-1/2}\,(C-E)\,D_c^{-1/2} = U\Omega V' \tag{14}$$

where U'U=I and V'V=I (I denotes the identity matrix). Ω is a diagonal matrix containing positive descending singular values ω_s , where s (s=1,...,s,.,s',...,S) is the index for the independent analysis solutions or, in geometrical terms, referred to as the so-called 'dimensions'. Each row and column is quantified using the corresponding elements of the row and column vectors U and V. The resulting weighted row and column scores are :

$$X = D_r^{-1/2}\ U\ n^{1/2} \qquad (15a)$$
$$Y = D_c^{-1/2}\ V\ n^{1/2} \qquad (15b)$$

with a weighted average equal to 0 and a weighted variance equal to 1. Futhermore, $X'\ D_rX=I$, $Y'\ D_cY=I$, $t'D_rX=0$ and $t'D_cY=0$. These row and column scores can be interpreted as points and can therefore be presented in an s-dimensional space.

Define the profile of row i as the row of the values c_{ij}/c_{i+}.The row and column scores can be normalised in such a way that the Euclidean distance between a row i and row i' of $X^*=X\Omega$ is equal to the chi-square distance δ^2. In this metric, the distance between the respective row and column profiles is defined as :

$$\delta^2(i,i') = (I_i - I_{i'})'\ D_r^{-1}\ C'\ D_c^{-1}\ C'\ D_r^{-1}\ (I_i - I_{i'})' = (r_i\ - r_{i'})'\ \Omega^2\ (r_i - r_{i'}) = (r^*_i - r^*_{i'})'(r^*_i - r^*_{i'}) \qquad (16)$$

I_i and I_j are unit vectors from the identity matrix I. Approximations of the chi-square distances are found by considering only the columns of X^* corresponding to the s largest singular values ω_s (s'«S) of Ω. A similar approximation of the columns can be given by the Euclidean distances between the columns of the matrix $Y^*=Y\Omega$.

To facilitate interpretation one can integrate separate plots of row and column scores into a single plot. If one interprets relations between row scores (or column scores) using chi-square distances, one must bear in mind that rows or columns with similar profiles will have small distances between them. The profiles of weighted row and column sums of C are always located in the origin of the plot. Row and column scores with profiles which are (very) deviant from these mean-profiles, thus adding significantly to the chi-square total [cf. Eq. (8)], are found in the periphery of the plot. The distance between a row i and a column j is small if x_{ij}»e_{ij}, they are far apart if c_{ij}«e_{ij}.

The Pearson-statistic χ^2 can be defined in terms of total variance in the data found by the CA-solution:

$$\mathrm{tr}\,\Omega^2 = \Sigma_s\ \omega^2_s = \chi^2/n \qquad \text{(tr=trace : the sum of the diagonal elements)} \tag{17}$$

The relative importance of a dimension can now be interpreted as the ratio of the inertia in a dimension and the total inertia $\omega^2_s/\mathrm{tr}\ \Omega^2$, or more simply, as that part of χ^2, which is decomposed (the amount of variance within the data 'accounted for') in dimension s.

Secondly, we must adjust our data analysis-technique to the particular type of data we obtain after fitting a quasi-symmetry model. As mentioned before, the observed citation transaction matrix C is divided in a quasi-symmetric matrix E, with the maximum likelihood-estimates of the citation-values based on the quasi-symmetry model, and the matrix N with the residuals. This matrix N will always have a so-called skew-symmetrical form, i.e. $n_{ij} \neq -n_{ji}$, for $i \neq j$. Gower (1977) has shown that the singular value decomposition of N can be formulated as

$$N = U\Omega V' = U\Omega JU' \tag{18}$$

where J is a block-diagonal matrix with 2x2 blocks $\begin{vmatrix} 0 & 1 \\ -1 & 0 \end{vmatrix}$.

Singular values ω_s are ordered pairwise (i.e., $\omega_1=\omega_2$, $\omega_3=\omega_4$,...). The matrices U and V now contain elements which are related in pairwise fashion ; $v_{1i} = u_{i2}$, $v_{i2} = -u_{i1}$, $v_{i3} = u_{i4}$, $v_{i4} = -u_{i3}$, etc. Analogous to Eq. (14), we define U as the row scores and JU' as the column scores. Due to the block-property, the spatial position of column scores only differs from the position of the row scores by a 90-degree rotation factor for each pair of dimensions. The elements of N may thus be expended as a sum of skew-symmetric components :

$$n_{ij} = \omega_1(u_{i1}\,u_{j2} - u_{i2}\,u_{j1}) + \omega_2(u_{i3}\,u_{j4} - u_{i4}\,u_{j3}) + \ldots + \ldots \tag{19}$$

Note that in this particular decomposition, the elements of matrix N are not standardized (i.e., weighted with row and column sums as in CA), this decomposition will thus only reflect on the actual number of residual citations. To maintain the CA-properties, and thus a more comprehensive insight of the structure of the structure, one must investigate the proportional distribution of residual citations. Applying Eq. (18) in combination with a set of appropriate weights $\mathbb{D}= \mathrm{diag}\ (D_r\ D_c)^{1/2}$, based on the observed marginal frequencies D, one can recover a weighted singular value decomposition which serves as a basis for CA on the skew-symmetric residuals, as in Eq. (14) :

$$\mathbb{D}^{-1/2}\,N\,\mathbb{D}^{-1/2} = U\,\Omega\,J\,U' \tag{20}$$

The skew-symmetric properties of matrix N thus remain unaltered. This weighting scheme thus aviods the incorporation on unjustified high residual values ; the relatively high values stemming from journals with a relatively high observed citation frequency are proportionally downgraded. Note that $\mathbb{D}$ yields similar values, whether one uses the marginal frequencies of the original matrix X or the symmetrical matrix M.

Analogous to Eqs. (15a) and (15b), row and column scores are defined as :

$$\mathbb{X} = \mathbb{D}^{-1/2}\, U \tag{21a}$$
$$\mathbb{Y} = \mathbb{D}^{-1/2}\, JU' \tag{21b}$$

As noted, a relatively strong relationship between cited journal i and citing journal j is indicated in CA by scores i and j which are located near each other in a space with a chi-square metric. Instead of using the chi-squared distances, the geometrical interpretation of the relations between scores in combination of both approaches (which we shall adopt as the *Gower/CA-decomposition*), is done by using the triangular area between two scores i and j and the origin $\mathbb{O}$ [with spatial coordinates (0,...,0)] in non-metric space. The size of this triangular area is proportional to the hight of the residual citation-value between the units i and j. Due to the block-diagonal property, small distances between row score i and column score j are equivalent to large area's between either row scores $\mathbb{X}$ or column scores $\mathbb{Y}$. To avoid redundancy and faciliate interpretation of the score-configuration, we only use row scores $\mathbb{X}$. Pairs of row scores $\mathbb{X}$ are represented as coordinates per two-dimensional non-metric space, the so-called *bimension* (Harshman, 1981), corresponding to ω-pairs. The best one-bimensional approximation of the skew-symmetric elements in N, through the scores $\mathbb{X}$, is then given by

$$n_{ij} = \omega_1\, (x_{i1}\, x_{j2} - x_{j1}\, x_{i2}) \tag{22}$$

with the corresponding points $\mathbb{X}_i=(x_{i1}\,,x_{i2})$ and $\mathbb{X}_j=(x_{j1}\,,x_{j2})$.

The first-bimensional configuration yields an approximation of the most prominent features in the structure, accounting for the largest amount of variance within the data. Analogous to the separate dimensions, each bimension can be seen as representing one component of the total pattern of skew-symmetry, although the bimensions have a different geometrical interpretation. For example, if the line $(\mathbb{X}_j\mathbb{X}_k)$ is parallel to $(\mathbb{O}\mathbb{X}_i)$, then $n_{ij}=n_{ik}$. Moreover, scores on the same line through the origin (the so-called collinearity) have no residual relationship (within the specified bimension). Skew-symmetry is reflected by the fact that the size of the triangle $(\mathbb{O}\mathbb{X}_i\mathbb{X}_j)$ equals $-(\mathbb{O}\mathbb{X}_j\mathbb{X}_i)$. Similar pairs of points, and pairs of points lying on the same line from the origin, will yield triangular area's ($\mathbb{O}\mathbb{X}_i\mathbb{X}_j$) with a zero area.

With respect to the geometrical interpretation of the triangles and the rotation property of the scores, we will adopt the rule that elements of n_{ij} are positive if we interpret the positions of the points in a clockwise-direction ; if the citation transaction-flow runs from j to i (rows i representing cited journals and columns j the citing journals), it defines an influence from i on j. Counterclockwise rotations from score i to score j indicate a negative value for n_{ij}, and an influence of i on j. Given the direction of rotation, skew-symmetric relationships are assessed by investigating triangles ($\mathbb{O}\mathbb{X}_i\mathbb{X}_j$) with an angle $\alpha(\mathbb{X}_i\mathbb{O}\mathbb{X}_j)<180°$. Interpretation of the configuration of circular triads of the scores $\mathbb{X}$ as a whole, is independent of the point-of-entry within the configuration. The CA properties remain principally the same within this framework, e.g. the row scores still represent differences in row profiles, and distances between row scores are (approximately) chi-squared distances. Row scores near the origin, having a mean profile, are of course less important in the whole.

Combining properties of CA and the Gower-decomposition thus enables us to interpret the skew-symmetrical structure of relative number of residual citations in terms of a configuration of triangular area's. Results can be displayed as a spatial presentation of rowpoints $\mathbb{X}$ or an enumeration of the sizes of areas for each bimension. In the following examples only plots of the configuration of $\mathbb{X}$ will be given.

It is clear that differences can occur between influence values computed with Eqs. (12), (13) and results of the Gower/CA decomposition. The influence values are simply based on comparing n_{ij} to c_{ij}, whereras the latter results compare all n_{ij}'s simultaneously and, instead of x_{ij}, incorporates marginal frequencies c_{i+} and c_{+j} as weights. Differences between the results of both approaches may arise as a result of, for example, a relatively large influence value which is not accountable to a relatively large amount of residual citations, but stems from a relatively small number of observed citation. This will thus not be recovered as a large influence in the configuration of scores from the Gower/CA decomposition. Of course, the opposite situation is also possible. In general, the Gower/CA decomposition detects if notable residuals are really influences of importance considering the whole influence structure.

EXAMPLES

Two examples are discussed. In both cases data sets with high numbers of inter- and self-citations were chosen between a comparatively small and somewhat arbitrary set of highly related journals. No attempt has been made to ascertain a well-bounded disciplinary domain of journals or include all relevant related journals, nor have we restricted ourselves to similar types of journals (e.g. excluding review journals). These data sets were merely selected for illustrative reasons, in order to compare the discussed method with results of two related data analysis-methods : *Noma's* approach and the application of a *quasi-independence model* in the so-called *quasi-correspondence analysis* (cf. De Leeuw & Van der Heijden, 1988). Since no adjustment of the citations (i.e., a priori weighting of citation frequencies) was carried out in either of these two methods, a similar approach was taken with respect to the analyses of the citation counts with our method, as illustrated in the following two examples. The necessary computations for the application of the quasi-symmetry model and the Gower/CA decomposition were carried out with ad hoc computer programs written in APL.

Example 1 : Comparing two methods for the assessment of residuals based on a quasi-symmetry model.

For reasons of direct comparison with *Noma's* assessment of residuals, the data consist of the same citation matrix as used by Price (1981) and subsequently by Noma (1982), i.e. citation counts between eight journals containing publications on subjects in biochemistry, biophysics or molecular biology (Table 1). The cumulative citation counts were taken from the 1977 *Journal Citation Reports*. (Garfield, 1977)

Five journals primarily concentrate on biophysical or biochemical subjects : Journal of Biological Chemistry (JBC), Biochimica et Biophysica Acta (BcBpA), Biochemistry U.S. (BcUS), Biochemical Journal (BcJ), and Biochemical and Biophysical Research Communications (BcBpRC). The other three journals emphasize on molecular biology:: Journal of Molecular Biology (JMB) and, to a lesser extent, the multidisciplinary journals : Nature (Nat) and Proceedings of the National Academy of Sciences (PNAS).

Table 1
Journal-to-journal citation-data from the 1977 *Journal Citation Reports*

Cited journal	Citing journal 1	2	3	4	5	6	7	8
1 BcBpA	7550	1120	1478	1757	2406	408	365	865
2 BcBpRC	1719	1313	564	1040	1624	241	263	695
3 BcJ	1812	528	2464	632	1183	150	201	326
4 BcUS	2591	887	653	3827	2553	601	299	1057
5 JBC	6181	2511	2335	3750	9384	719	609	2107
6 JMB	1136	367	216	1347	1109	2545	504	1251
7 Nat	1230	630	379	837	1007	603	2963	1407
8 PNAS	2184	1329	488	1964	2770	1239	1470	3995

The quasi-symmetry model does not fit the journal to journal data adequately (the chi-square value equals χ^2=200.6 ; df=21; p<0.001), indicating that a significant part of the values in the journal to journal citation-structure can be ascribed to residuals, i.e. journal-to-journal influences. The corresponding influence values [cf. Eqs. (12a) and (13)] are presented in Table 2.

Table 2
Journal-to-journal citation-data from the 1977 *Journal Citation Reports*

Cited journal	Citing journal IV 1	2	3	4	5	6	7	8	SIV	SIR
1 BcBpA	0	+.01	+.05	+.00	-.07	+.05	-.01	+.11	+.14	2
2 BcBpRC	+.00	0	-.04	+.05	-.02	+.16	-.08	-.01	+.07	3
3 BcJ	-.04	+.04	0	+.03	-.08	+.32	+.20	+.26	+.74	1
4 BcUS	+.00	-.06	-.03	0	+.05	-.06	-.14	.05	-.20	8
5 JBC	+.03	+.01	+.05	-.03	0	-.10	-.09	-.03	-.16	6
6 JMB	-.02	-.08	-.14	+.03	+.08	0	-.04	-.08	-.18	7
7 Nat	+.00	+.04	-.08	+.06	+.06	+.04	0	-.08	+.04	4
8 PNAS	-.04	+.01	-.12	-.02	+.02	+.01	+.09	0	-.06	5

IV - Influence values
SIV - Sum influence values
SIR - Summed influence value-ranking

High influences are shown as relatively high positive influence values in this table. For example, the influences BcJ on Nat, BcJ on PNAS and BcJ on JMB (with a 20%, 26% and 32% excess of observed citations to BcJ, respectively). In the ranking of the overall influence, BcJ takes the first place, mainly due to these specific influences. It is noteworthy that the relatively small number of citations from JMB to BcBpRC (see Table 1) corresponds with a relatively large influence of BcBpRC on JMB.

The analysis of the matrix with residual values yields two bimensions of interest, accounting for almost all variance within the data : 69% and 30%, respectively.[3] The configuration of both bimensions thus actually reflects the total structure-structure of residual values. The journal scores $\mathbb{X}$ in the first and second bimension are displayed in Figures 1a and 1b, respectively.

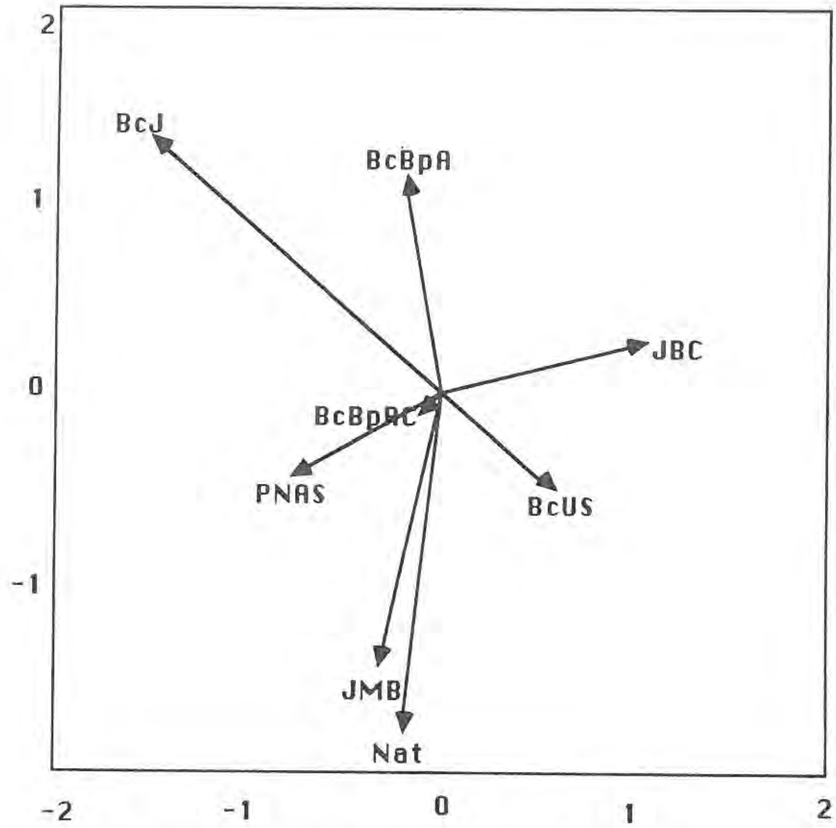

Fig. 1a : Gower/CA decomposition on the Price/Noma data ; first bimension. The labels are centered at the location of the scores of journals.

The scattered position of the row scores in the first bimension calls for an interpretation of relative journal influences, in terms of circular triads. Only journal scores with a relatively large contribution (i.e. a relatively large part the variance due to these journals is accounted for) in the analysis-solution will be examined in more detail. These are the scores corresponding to longer vectors, and thus larger areas, in Figure 1a. The first bimension is dominated by BcJ, PNAS, JMB and Nat. Intermediate positions are taken by BcBpA, BcUS and JBC. The journal BcBpRC has a location near the origin of the space, indicating a minor role in this bimension. The row points of Nat and JMB are approximately collinear, indicating a comparatively small net influence between these journals - at least as far as the first-bimensional approximation is concerned.

Use of the earlier discussed clockwise interpretation for positive residuals, means that counterclockwise relations indicate the direction of positive influences. We see that BcJ thus has a large relative influence on PNAS, and an even larger influence on Nat and JMB (i.e., these journals have positive residuals with respect to BcJ). Futhermore, following the circle, PNAS has an relative influence on Nat and JMB. The configuration also reveals that Nat and JMB have a relative influence on BcUS and, more in particular, on JBC. The journal JBC has an influence on BcBpA and an even larger influence on BcJ. In a nutshell, this is also the basic structure between the journals in Table 2. In fact, all abovementioned results correspond with the (row-wise) structure of the positive influence values.

The second bimension is also of interest, for it accounts for nearly one-third of the variance. When interpreting the journal positions in the second bimension (Fig. 1b), one must keep in mind that this configuration is independent from the configuration in the first bimension. The journals Nat, PNAS and BcBpRC are now the most important journals, followed by BcUS, JMB and JBC. The journals BcJ and BcBpA are relatively unimportant in this bimension ; apparently their relations to the other journals are adequately displayed in the first bimension. Conversely to the first bimension, Nat is now separated from JMB, and placed in a position indicating a relative influence of JMB on Nat (a result which is not found in Table 2). PNAS has an additional influence on JMB.Also, Nat now has a small influence on PNAS, thus partly undoing the relative influence of PNAS on Nat in the first bimension. In some aspects the configuration thus yields a correction on the results in the first bimension.

As mentioned, the influence value of BcBpRC on JMB is relatively large influence, but in the Gower/CA decomposition this is only partly reflected in Fig. 1b and not found at all in Fig. 1a. Apparently this residual value was mainly due to the relatively small number of citations, and thus only partly caused by the residual citation numbers if one compares these residuals within the total residual structure.

The two-bimensional configurations displays independent components of the structure of the skew-symmetrical data. The structure of the data dictates the division into the bimensions. Sometimes it is no quite clear which labelling can be given of the separate configurations. In such cases detailed expert-opinion with respect to the data and the results is required. Anyway, it is obvious that a primary advantage of this type of representation of data is found in the explorative features of this method, i.e. providing a better insight in a complex structure of interrelations.

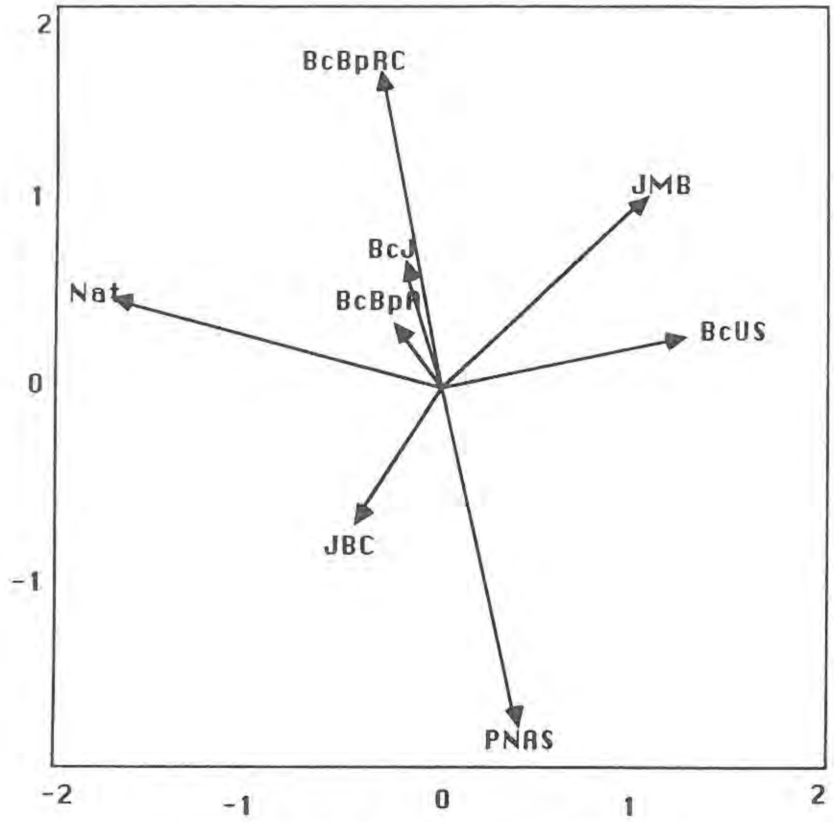

Fig. 1b : Gower/CA decomposition on the Price/Noma data ; second bimension. The labels are centered at the location of the scores of the journals.

As could be expected, the configurations of structure of influences (Figs. 1a and 1b) and values of the computed influences (Table 2) are largely in accordance with the values of the Freeman-Tukey statistic presented in Noma (1982, p. 300) ; large influences correspond with high positive values of the F-T statistic for the corresponding pair of journals, whereas small influences coincide with non-significant F-T values.

Example 2 : Comparing residuals from the quasi-independence and quasi-symmetry model.

In Tijssen et al. (1987) the quasi-independence model was fitted to a journal-to-journal citation-matrix. The residuals after fitting this particular model were analyzed with a modification of correspondence analysis, *quasi-correspondence analysis* (henceforth abbreviated to QCA). In QCA, the effect of the self-citations on the analysis results was also eliminated, but now in combination with the more restrictive quasi-independence model, i.e. the baseline model for this approach only accounts for the 'size'-effects [cf. Eq. (6)}. The residuals after fitting the quasi-independence model will therefore contain both a symmetric interaction-component as well as the (asymmetric) influence component. Analogous to the Gower/CA decomposition, QCA decomposes the structure of residuals in a number of independent components (dimensions, in this case). Differences between the analysis results of both methods on the same data are thus caused by the similarity-component which remains in the residuals after fitting the quasi-independence model. Comparing QCA-results to results from the Gower/CA decomposition thus enables one to gain insight as to whether residual citation-relations between journals are due to either the symmetric-component ('contents'), or the asymmetric-component ('influence'), or a combination of both.

The data-set presented in Tijssen et al. (1987) consists of citation counts between seven highly cited U.S. and European journals, specialized in the fields of astronomy and astrophysics. The cumulative total of citations up until 1983 were collected by manual search from the 1983 *Journal Citation Reports* (Garfield, 1983). The resulting citation data-matrix is given in Table 3.

Table 3
Journal-to-journal citation-data from the 1983 *Journal Citation Reports*

Cited journal		Citing journal 1	2	3	4	5	6	7
1	AA	2714	269	2009	297	159	867	129
2	AN	454	651	965	100	58	433	126
3	AP	4506	1358	16079	975	1186	3383	895
4	APSS	424	191	978	464	33	261	94
5	ARAA	237	64	601	74	45	140	63
6	MNRAS	1163	268	2327	315	141	1959	175
7	PASP	282	170	576	69	32	208	183

Two US journals, i.e., the Astronomical Journal (AN) and Publications of the Astronomical Society of the Pacific (PASP), and the British journal Monthly Notices of the Royal Astronomical Society (MNRAS) emphasize astronomical subjects, while the US journal Astrophysical Journal (AP) concentrates on astrophysical topics. The European journal Astronomy and Astrophysics (AA) covers both fields. In addition to publications on astrophysics, the US journal Astrophysics and Space Science (APSS) also contains publications on space physics and related topics on the solar system. The Annual Review of Astronomy and Astrophysics (ARAA) contains papers in which an overview is given of past and current developments in various subfields of astronomy and astrophysics; these papers generally contain a large amount of references. All journa's, except ARAA, have a self citation-frequency (far) beyond the expected frequency based on the independence model.

Although the quasi-symmetry model naturally gives a better approximate fit of the citation structure compared to the quasi-independence model (the model has the additional symmetric-parameters), it does not fit these journal to journal data adequately either (the chi-square value equals χ^2=139.6 ; df=15 ; p<0.001). Apparently, a significant part of the citation-transactions is due to asymmetric relations between journals, thus indicating notable influences.

When comparing the residual/observed citation-ratios (cf. Table 4), the review journal ARAA has top position in the influence-ranking, mainly as a result of its relatively large net influence on APSS and AN. The high rank of ARAA in this set of journals is not unexpected considering the type of publications (review articles) in this journal. The second position is held by PASP due to its influences on ARAA, AN and AA. Comparing the relations between, e.g. AP and MNRAS in Tables 3 and 4, makes it clear that a high number of observed citations does not necessarily lead to a high influence value.

Table 4
Journal-to-journal citation-data from the 1983 *Journal Citation Reports*

Cited journal	Citing journal IV 1	2	3	4	5	6	7	SIV	SIR
1 AA	0	+.01	-.02	+.10	-.02	+.04	-.17	-.07	5
2 AN	-.01	0	+.00	-.29	-.09	+.18	-.13	-.33	7
3 AP	+.01	+.00	0	-.04	+.02	-.02	+.06	+.04	3
4 APSS	-.06	+.27	+.05	0	-.43	-.15	+.14	-.18	6
5 ARAA	+.02	+.10	-.03	+.49	0	+.00	-.27	+.29	1
6 MNRAS	-.03	-.20	+.03	+.18	+.00	0	.-03	-.06	4
7 PASP	+.11	+.12	-.08	-.15	+.24	+.02	0	+.27	2

IV - Influence values
SIV - Sum influence values
SIR - Summed influence value-ranking

We first mention some of the more striking results of the QCA-analysis presented in Tijssen et al (1987), for a comphrehensive comparison between QCA and the Gower/CA-decomposition. These results indicated that the quasi-independence model was not an adequate approximation of the observed citation-matrix ; the residuals thus contain information of interest. Secondly, analysis of those residuals (incorporating both the symmetric and asymmetric relations) yielded a first dimension focussing on two groups of related journal-modes : first, the relationship between citing journal AN and cited journals PASP and APSS, and secondly, relations between cited journals ARAA, APSS and MNRAS with citing journals MNRAS and AA. The second dimension showed the following differentiation between journals : on one side, the relation between citing MNRAS and cited journals AA and AN, the other side indicated a relationship between citing PASP and citing ARAA, with cited ARAA.

The Gower/CA decomposition of residual values yields two bimensions of interest, accounting for a substantial part of the variance within the data : 73% and 22% percent, respectively. The configuration of the scores in the first and second bimension is depicted in the Figures 2a and 2b, respectively.

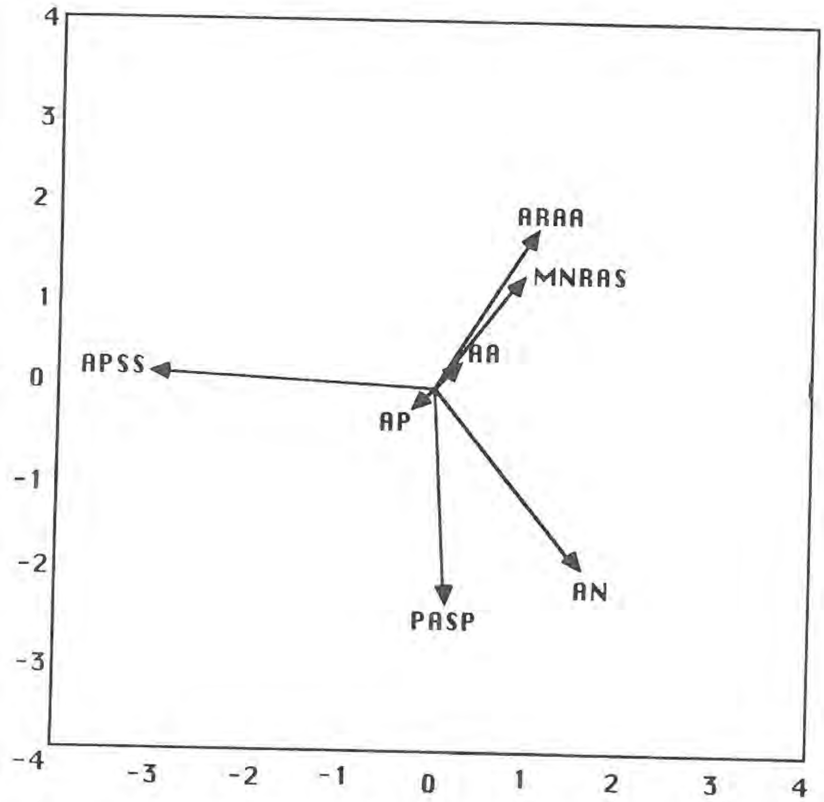

Fig. 2a : Gower/CA decomposition on the Tijssen et al. data ; first bimension. The labels are centered at the location of scores of journals.

The structure-configuration in the first bimension is clearly dominated by the journals AN, APSS, ARAA, MNRAS and PASP. Interpreting influences in a counterclockwise manner gives rise to the following influence-relationships : ARAA and MNRAS have a clear influence on APSS. In its turn, APSS has an influence on PASP and AN, whereas PASP has a smaller influence on AN. Finally, AN has an influence on both MNRAS and ARAA. However, the first-bimensional configuration does not account for the relatively important influences from AARA on APSS, ARAA on AN, PASP on AARA and PASP on AN (see Table 4).

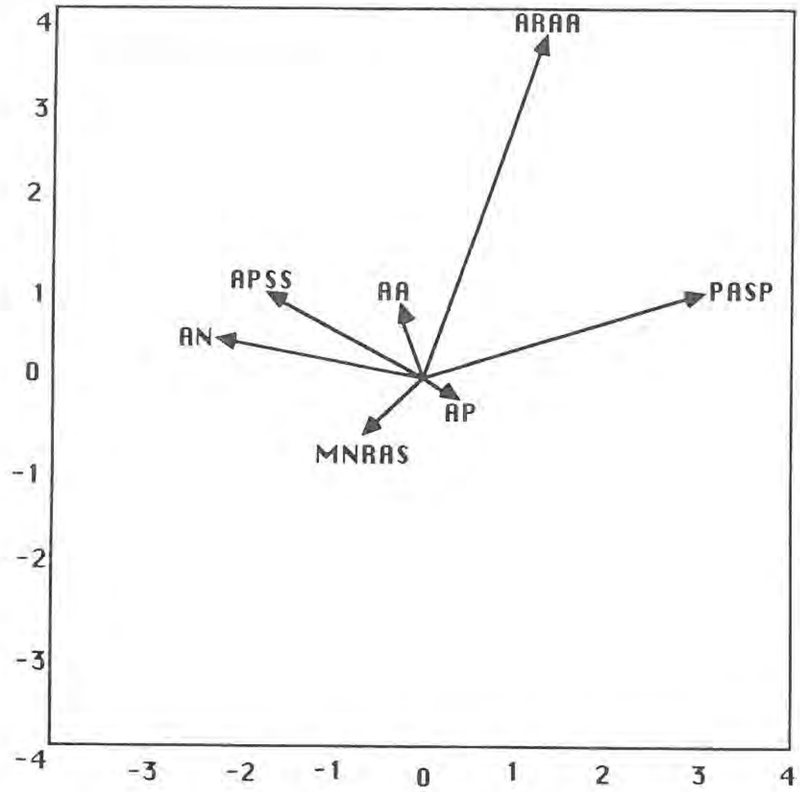

Fig. 2b : Gower/CA decomposition on the Tijssen et al. data ; second bimension. The labels are centered at the location of the scores of journals.

The additional component of the influence structure in the second bimension is mainly characterized by the journals AN, ARAA and PASP. A comparatively simple configuration of the additional information on the influence structure emerges, in which the four 'missing' influences from the first bimension are located. Also, PASP has an additional influence on both ARAA and AN, whereas ARAA now has a counter-influence on AN.

Some of the influence relations where also found in the QCA results ; for instance, the relations between cited PASP and citing AN, and citing MNRAS and cited AN. In such cases a relationship is apparantly also determined through an influence component. In other cases (e.g., cited ARAA and citing PASP) residual citation-flows were not found in the results of the Gower/CA decomposition, indicating that the relations between journals as identified in QCA-results were mainly due the similarities between the contents of the journals and not the influences between the journals. In general, a comparison of analysis results based on the residuals of both models can thus yield an insight in the decomposition of the nature of the citation-based relations between journals.

CONCLUSIONS

Most computational methods deriving citation-based influence measures operate on the total number of citations to and from journals. Sometimes journal influences are computed directly from those citation values, sometimes iteratively. The iterative method presented in this paper is based on a probalistic model. It does not incorporate all citations to determine influence-measures, but focusses on the influence-component in the citation flow. This concept is operationalized as the residual segment of the citation flow after fitting the quasi-symmetry model (i.e. filtering out two important external factors from the total citation flow ; such as the size of the journal). The residuals of this particular modelling of the citation data enables two approaches to gain information on the structure of the influences between journals : first, an influence measure is computed by as a transformation of residual citations (the part of the citations which are not accounted for by the model), secondly, a data-analysis method is developed with which one can easily obtain an additional insight in the total structure of residual citations by means of spatial display. These two complementary methods can of course also be used independently, depending on which aspect of the influences the interest of the user centres. Naturally, the information stemming our method is highly dependent upon the analyzed set of journals. This prevents a generalization of conclusions based on influence values to a set of journals beyond the scope of the assessment.

In general, the presented influence measures can provide additional and meaningful information in studies of science, especially to augment an in-depth understanding of citation-based relations between journals within a (sub)discipline. In sum, we conclude that our analysis approach based on the quasi-symmetry model proves to be a useful explorative aid in assessments of journal relationships. As a last remark it must be stressed that such a quantitative assessment of journal relations should, if possible, be combined with expert opinions and/or results of other quantitative (citation-based) methods.

NOTES

1 The maximum likelihood estimates (MLE) are based on the general estimation method of maximum likelihood. The MLE are identical under Poisson-, multinomial and product-multinomial sampling distributions of the observed data (cf, Bishop, Fienberg and Holland).

2 Noma restricts the elements on the main diagonal to be equal to the product of the overall, row and column parameters [i.e., log $e_{ii} = u + u_{1(i)} + u_{2(j)}$].

REFERENCES

Bayer, A.E. J. Folger, Some correlates of a citation measure of productivity in science, **Sociology of Education**, 39 (1966) 381-390.

Bishop, Y.M.M., S.E. Fienberg, P.W. Holland, **Discrete multivariate analysis : theory and practice**, Cambridge : MIT-Press, 1975.

Brooks, T.A., Evidence of Complex Citer Motives, **Journal of the American Society for Information Science**, 37, 34-36, 1986.

Clark, K.E., **America's Psychologists : A survey of a Growing Profession**, American Pychological Association, Washington D.C., 1957.

Cole, S., J.R. Cole., Scientific output and recognition : a study in the operation of a reward system in science. **American Sociological Review**, 32 (1972) 471-479.

De Leeuw, J., P.G.M. van der Heijden, Correspondence analysis of incomplete matrices, **Psychometrika**, (in press).

Deming, W. E., F.F. Stephan, On the least squares adjustment of a sampled frequency table when the expected marginal totals are known, **Annals of Mathematical Statististics**, 11 (1940) 427-444.

Doreian, P., A measure of standing of journals in stratified structures. **Scientometrics**, 8 (1985) 341-368.; P. Doreian. A revised measure of standing of journals in stratified structures. **Scientometrics**, 11 (1987) 71-80

Edge, D. O., Quantitative measures of communication in science : a critical review, **History of Science**, 16 (1979) 102-134.

Garfield, E. Citation analysis as a tool in journal evalution, **Science** , 178 (1972) 471-479.

Garfield (ed.), E., **SCI Journal Citation Reports**, 13, Philadelphia, PA : Institute for Scientific Information, 1977.

Garfield (ed.), E. **SCI Journal Citation Reports**, 18, Philadelphia, PA : Institute for Scientific Information, 1983.

Garfield, E., How to Use Citation Analysis for Faculty Evaluation and When is it Relevant ?, **Current Contents**, 44 (1983) 5-13 ; 45 (1983) 5-14.

Gilbert, G. N., Referencing as Persuasion, **Social Studies of Science**, 7 (1977) 113-122.

Golub, G.H., C. Reinsch, Handbook series linear algebra : Singular value decomposition and least squares solutions, **Numerical Mathematics**, 14 (1970) 403-420.

Gower, J.C., **The analysis of asymmetry and orthogonality**, Recent Developments in Statistics (eds.- J.R. Barra et al.), Amsterdam : North-Holland, 1977.

Greenacre, M, **Theory and Applications of Correspondence Analysis**, New York : Academic Press, 1984.

Gross, P. L. K. ; E.M. Gross, College libraries and chemical education, **Science**, 66 (1927) 385.

Harshman, R.A., **DEDICOM Multidimensional Analysis of Skew-Symmetric Data**, Unpublished technical memorandum, Bell Laboratories, Murray Hill, New Jersey, 1981.

Hirst, G., Discipline Impact Factors : A method for Determining Core Journal Lists, **Journal of the American Society for Information Science**, 29 (1978) 171-172.

Liebowitz, S. J., J. P. Palmer, Assessing the relative influences of economics journals, **Journal of Economic Literature**, 22 (1984) 77-88.

MacRoberts, M. H., B. R. MacRoberts, Quantitative Measures of Communication in Science : A Study of the Formal Level, **Social Studies of Science**.16, 151-172, 1986.

H.F. Moed, W.J.M. Burger, J.G. Frankfort, A.F.J. van Raan, **On the Measurement of Research Performance : the use of bibliometric measures**. Research Policy and Science Studies Unit, University of Leiden, 1983.

Moed, H. F., W. J. M. Burger, J. G. Frankfort, A. F. J. van Raan, The Application of Bibliometric Indicators : Important Field- and Time-dependent Factors to be considered, **Scientometrics**, 8, 177-203, 1985.

Narin, F., M.P. Carpenter, N.C. Berlt, Interrelations of scientific journals, **Journal of the American Society for Information Science**, 23 (1972) 323-331.

Noma, E., An improved method for analysing square scientometric transaction matrices, **Scientometrics**, 4 (1982) 297-316.

Osborne, C.B, The Place of the Journal in the Scolarly Communication System, **Library Resources and Technical Services**, 28 (1984) 315-324.

Pinski, G. ; F. Narin., Citation influence for journal aggregates of scientific publications : Theory, with an application to the literature of phyiscs. **Information Processing and Management**, 12 (1976) 297-312.

Price, D.J. The analysis of square matrices of scientometric transactions, **Scientometrics**, 3 (1981) 55-63.

Sengupta, I. N., Three new parameters in bibliometric research and their application to rerank periodicals in the field of biochemistry, **Scientometrics**, 10 (1986) 235-242.

Simonton, D.K., Scientific eminence historical and comtemporary : a measurement assessment, **Scientometrics**, 6 (1984) 169-182.

Singleton, A., Journal ranking and selection : a review of physics, **Journal of Documentation**, 32 (1976) 258-289.

Small, H., B.C. Griffith, The Structure of Scientific Literature, I - Identifying and Graphing Specialities, **Science Studies**, 4 (1974) 17-40.

Tijssen, R.J.W., J. de Leeuw, A.F.J. van Raan, Quasi-correspondence analysis on scientometric transaction matrices. **Scientometrics**, 11 (1987) 351-366.

Tijssen, R.J.W.,.A.F.J. van Raan, **Net Citation Balances - A measure of influence between scientific journals**. (submitted for publication).

Todorov, R., Determination of influence weights for scientific journals ; an example from elementary physics, **Scientometrics**, 6 (1984) 127-138.

Todorov, R., W. Glänzel, Journal Citation Measures : A concise review, **Journal of Information Science**, 13 (1987).

Van der Heijden, P.G.M., J. de Leeuw, Correspondence analysis used complementary to loglinear analysis, **Psychometrika**, 50 (1985) 429-447.

Weishart, R. A., R.M. Regoli. Ranking journals, **Scholary Publishing**, 15 (1984) 317-325.

Yanovsky, V. I., Citation analysis significance of scientific journals, **Scientometrics**, 3 (1981) 223-233.