The static and dynamic analysis of network data using information theory

Loet Leydesdorff

Department of Science Dynamics, Nieuwe Achtergracht 166, 1018 WV Amsterdam, Netherlands

Information theory provides us with methods for both the static and dynamic analysis of network data. Since the models are derived within one framework, the results of the multivariate analysis and the time-series analysis can be made relevant for one another. Additionally, using the static model, one can create an exact dendrogram, and determine the precise number of clusters. The algorithm is generalizable to clique analysis. Using the dynamic model, developments can be revealed which were not suggested by the comparison of results of various forms of multivariate analysis for each year separately. The question of using these methods to design research about structure/action relations is discussed.

1. Introduction

In addition to unsolved problems relating to inferences among different levels of aggregation (Langbein and Lichtman 1978; Van den Eeden and Huttner 1982), the structural analysis of social action, as contained in network data, has to solve the methodological problem of how to account for change over time. On the one hand, in most common forms of multivariate analysis, this dynamic aspect is handled by "cutting" the three-mode matrices into two-mode "slices", and then by comparing the results for each "slice" separately (see, e.g., Tijssen and De Leeuw 1988: 727). The disadvantage of this method is that "change with time" is not accounted for systematically, but only treated as a difference.

On the other hand, in time-series analysis, a time series is defined as a set of observations obtained by measuring a *single* variable regularly over a period of time (see, e.g., McClearly and Hay 1980). In the multivariate case, one has to be able to specify the dependent

variable. [1] Data that have been collected for the purpose of network analysis, usually will not readily permit extension to this kind of modelling, or only at the expense of a rapid increase in the methodological complexities of the analysis (see, e.g., Tuma and Hannan 1984).

The further elaboration of time-series analysis to the point where the theoretically interested researcher is able to combine the interpretation of results from it with results from network analysis seems a longer term methodological objective; at present by using these methods one obviously runs the risk of becoming distracted from theoretical questions to methodological problems. However, this problem is even more general: the proliferation of methods for statistical analysis makes the choice of parameters (like similarity criteria, algorithms, cut-off levels, etc.) a matter of increasing methodological sophistication. In most cases these initial, seemingly technical (but yet analytic) choices heavily influence statistical results, and hence, interpretation.

In social network analysis, more particularly, a "relational approach" and a "positional approach" have been distinguished (Burt 1982). While the former focusses on diadic relations, cliques and hierarchies using graph-analytic notions, the latter focusses on "structural equivalence" using factor-analytic notions. However, the question of how to relate the two approaches analytically has not yet been systematically addressed.

Additionally, the intelligibility of relations among results may be affected when the measurement scales of the variables differ – as is often the case when we want to infer among levels of aggregation. For example, co-occurrences which can be measured as frequencies, i.e., at interval scale in the overall data set, may either occur or not occur dichotomously at lower levels. Consequently, we may have to combine results from non-metric and metric forms of analysis. In terms of methodologies, this often implies the choice of other parameters, and therefore again an increase in the incomparability of results, and consequently another obstacle to the drawing of inferences.

---

[1] In a design based on cross-lagged panel correlations, one may also use multivariate distributions, e.g., matrices, if one is able to compute a *single* (cross-) correlation. However, thereafter, the systematic decomposition of the results in terms of the different variables is impossible. See also: Krempel (1989); Cook and Campbell (1979).

The results of each statistical analysis should preferably be directly relatable to the specified theoretical problems without causing additional problems of interpretation, and the results of one type of analysis should add to our understanding of the results of another analysis in an intelligible way. We need methods to study problems, and not problems to study methods.

In this study, I will show that we can use measures derived from information theory – also known as statistical decomposition analysis (Theil 1972) – as *one* conceptual framework in order to study the most common problems of multivariate analysis, both in a static and in a dynamic mode. In addition to the integration of results from these two perspectives, statistical decomposition analysis allows for a precise study of the effects of aggregation and disaggregation. Despite the fact that it requires the variables to be measured only at the nominal scale, it preserves any additional information contained in more refined measurement (see, e.g., Krippendorff 1986).

In a final section, I will discuss the potential of using these methodologies to design research questions about structure / action relations. The dynamic extension makes it necessary to further specify the difference between a "structuralist theory of action" (Burt 1982) and a theory of "structure"/"action" contingencies: dynamic change in *structure* should not *ex ante* be equated with change in *eigenstructure* among data matrices, which are measured as aggregates of actions at different points in time.

As data, I will use the matrix of aggregated citations among 13 major chemistry journals as a typical set of network data. (See also Table 1.) [2] Aggregated journal–journal citations are often used as a high-level sociometric structure in scientometric studies in order to create "maps of science" (Price 1965). In this context, this type of matrices has been thoroughly analyzed in recent decades using various forms of multivariate analysis (Carpenter and Narin 1973; Narin 1976: 185–190; Doreian and Fararo 1985; Leydesdorff 1986, 1987; Doreian 1986; Tijssen *et al.* 1987). The data are readily available from the

---

Journal Citation Reports of the *Science Citation Index*. [3] In the static analysis, I will focus on 1984 data only (the data matrix is given in Table 2), and then include the corresponding data for other years in order to develop the dynamic model.

2. Information theory: a brief introduction

If we define h as the information content of the message that an event occurred, then the expected information content of the distribution of a variable with relative frequency p_i can be written as:

$$H = \Sigma_i p_i * h_i \tag{1}$$

By using Shannon (1948)'s classical function for information ($h_i = {}^2\log(1/p_i)$), [4] we may write:

$$H = -\Sigma_i p_i {}^2\log p_i \tag{2}$$

and for the multivariate case:

$$H = -\Sigma_i \Sigma_j \Sigma_k p_{ijk} {}^2\log p_{ijk} \tag{3}$$

Analogously to chi-square, H can be used as a measure of the association among variables. The overall uncertainty for two variables x and y, $H(x, y)$, is equal to $H(y)$ plus the amount of uncertainty which x adds to it, given the uncertainty in y, i.e., $H(x \mid y)$. Therefore:

$$H(x, y) = H(y) + H(x \mid y) \tag{4}$$

---

[3] However, in the printed edition, the tails of the citation distributions are summed under "all others". As a rule of thumb, values equal to or lower than five are omitted. (See also Garfield 1972.) In addition to these thresholds, Rice *et al.* (1989) noted that there are statistical and systematic errors caused by the lack of standardization in the spelling of journal names. However, as these authors note, for the type of journals which we will use here (with large numbers of total citations included in the *JCR*), these effects are probably insignificant.

[4] I will use the binary base of the logarithm throughout this study, and therefore express the information in bits.

Table 1
Journals used for the construction of a journal–journal citation network

Journal title	Variable name	Sequence number
Chemical Physics	ChemPhys	1
Chemical Physics Letters	ChemPhLt	2
Inorganic Chemistry	InorgCh	3
J. of the American Chemical Society	JACS	4
J. of Chemical Physics	JChemPh	5
J. of the Chemical Society – Dalton T	JChemSc	6
J. of Organic Chemistry	JOrgChem	7
J. of Organometallic Chemistry	JOrgmetC	8
J. of Physical Chemistry	JPhChUS	9
Molecular Physics	MolPhys	10
Physical Review A	PhysRevA	11
Tetrahedron	Tetrahe	12
Tetrahedron Letters	TrahLt	13

The *mutual information* or *transmission* between x and y is consequently defined as:

$$T(x, y) = H(x) - H(x \mid y) \tag{5}$$

This is the reduction in the uncertainty of the prediction of x, given knowledge about the distribution of y. Figure 1 gives a visual representation of the various relations (cf. Attneave 1959).

While "delta chi-square" does not have a clear interpretation, [5] the decomposition of H (and I, below) in terms of the contribution to the uncertainty of each of the component cells (or subsets) is straightforward. [6] Therefore, additionally the following formula can be derived for disaggregation of H into g groups (Theil 1972):

$$H = H_0 + \Sigma_g P_g * H_g \tag{6}$$

H_0 is a measure of the uncertainty among the groups g, or in other words, a measure of the specificity of the distribution of the relevant variables within the groups.

[5] Strictly speaking, chi-square tests only independence; it provides little information about the strength or form of the association between the two variables.
[6] The so-called likelihood ratio chi-square ($I^2 = 2\Sigma_i \Sigma_j F_{ij} \ln(F_{ij} / \hat{F}_{ij})$) is equally decomposable into interpretable parts that add up to the total. This measure is in essence an information theoretical formulation of the chi-square. See also Krippendorff (1986).

On the basis of the above definition of information, it can be shown that if we have a system of mutually exclusive events, E_i, with prior probabilities p_i, then the expected information content I of the message which transforms the *prior* probabilities p_i into the *posterior* probabilities q_i is given by the following expression: [7]

$$I = \Sigma_i q_i * {}^2\log(q_i/p_i) \tag{7}$$

Correspondingly, for the multivariate case, the expected information content of the message transforming the prior probability distribution p_{ijk} of events into the posterior probability distribution q_{ijk}, is equal to:

$$I = \Sigma_i \Sigma_j \Sigma_k q_{ijk} * {}^2\log(q_{ijk}/p_{ijk}) \tag{8}$$

Although overall $I \geq 0$, [8] ΔI can become negative if $q < p$. [9] (Of course, ΔH is always ≥ 0.)

In the dynamic perspective we do not just take "time" as another variate which may then co-vary with other variables, but we compare among events in a sequence. In our case, aggregated data for one year will constitute the "unit of event. [10] Again, one may either consider the multivariate distribution as one event, or compare among values in corresponding cells as events, and relate the two (after proper normal-

---

[7] The expected information content of the message which transforms prior probability p into posterior q, if event E ultimately occurs, is equal to:

$$i(q:p) = h(p) - h(q) = \log(q/p)$$

However, the probability that the event will occur is only equal to the posterior probability q, and hence:

$$I(q:p) = \Sigma_i q_i * \log(q_i/p_i)$$

[8] See for the proof Theil (1972: 59f).
[9] In the case of $q = p$, no information is added or lost; since the $\log(1) = 0$, I vanishes. Note that a zero in the prior distribution would make a non-zero value in the posterior distribution a complete surprise, and therefore, $I \to \infty$. In order to minimize the amount of expected information content originated by missing values in otherwise similar distributions, I replaced all missing values in this study with the value of five, since this is the cut-off level of the printed edition of the *Journal Citation Reports* from which the data were obtained. See for a discussion of the calculation of I in the case of "emergence" Leydesdorff (1990a).
[10] However, one may think of "units of event" as disaggregatable (months, weeks, days, etc.) or as aggregatable (decades, centuries, etc.)

Table 2
Original data matrix for 1984. Aggregated journal–journal citations among 13 major chemistry journals

	(citing →)												
	1	2	3	4	5	6	7	8	9	10	11	12	13
ChemPhys	984	724	51	189	1136	–	–	–	459	142	74	–	–
ChemPhLt	963	2387	206	810	2816	31	40	–	1660	331	250	53	–
InorgCh	35	157	5480	1912	138	1242	111	1228	319	14	–	28	29
JACS	344	1102	4873	15521	1185	1214	6952	2448	3240	126	–	3045	3694
JChemPh	2732	4622	715	2240	15069	166	157	163	5199	1575	1134	117	30
JChemSc	–	–	946	452	–	1443	28	830	52	–	–	–	26
JOrgChem	–	29	157	2264	–	62	5024	484	74	–	–	1617	2259
JOrgmetC	–	32	713	958	–	641	307	3765	–	–	–	106	211
JPhChUS	257	845	511	1208	1538	87	191	45	4315	122	41	51	56
MolPhys	330	455	84	220	1195	–	–	–	395	1082	113	26	–
PhysRevA	162	327	–	–	1115	–	–	–	170	183	3977	–	–
Tetrahe	13	29	49	831	–	24	891	131	49	–	–	806	724
TrahLt	–	32	84	1918	–	37	2802	548	61	–	–	1819	3385
(cited)													

$H(x,y)$

$H(x)$ $\quad$ $H(y)$
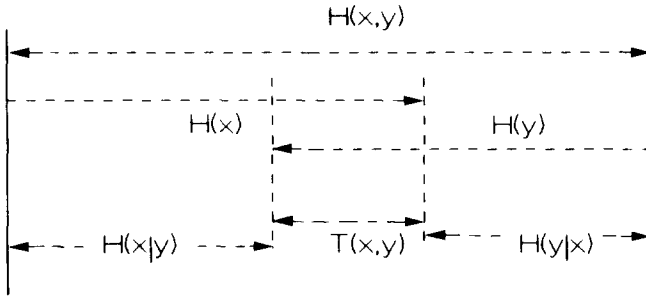
$H(x|y)$ $\qquad$ $T(x,y)$ $\qquad$ $H(y|x)$

Fig. 1. Relations of expected information contents, mutual information, and conditional entropies between two variables x and y (Attnaeve 1959).

ization) because of the additive properties of I. However, in the dynamic case the calculation rules among levels of aggregation are more complex than in the static case (see Theil 1972).

In principle, the two formulas, i.e., for H and I, provide us with a complete framework for the development of a set of methodologies equivalent to multivariate analysis and time-series analysis, respectively. [11]

3. Asymmetry: the comparison of two distributions

Let me begin my explication of the possibilities to apply information theory to social networks by using asymmetry as a simple example.

The matrix (in Table 2) is asymmetrical. It contains citing patterns and cited patterns, and the relations between the two, since each cell a_{ij} contains the number of citations journal i gives to journal j, and cell a_{ji} the number of citations journal j gives to journal i. In most forms of multivariate analysis a separate analysis is necessary for citing and cited patterns (by transposing the matrix); [12] in multi-dimensional scaling one sometimes combines the two analyses by using,

[11] For the further elaboration of relations among statistical decomposition analysis and regression analysis and to Markov chain analysis, the reader is referred to Theil (1972).
[12] In some forms of (quasi-)correspondence analysis one may be able to represent the analysis of asymmetry and the mapping of structure in one picture. However, the interpretation of these maps with respect to the attribution of cases to groups remains open, as with multi-dimensional scaling. (See also Tijssen et al. 1987.)

for example, the mean of a_{ij} and a_{ji}, while discarding the diagonal values. [13]

By using information theory, the relations between citing and cited patterns can be addressed overall as a static problem in terms of the relation between the two axes of the matrix as variables (i.e., *citing* and *cited*) or as a "dynamic" problem, i.e., as a comparison between two distributions, taking either a_{ij} as the *a priori* values and a_{ji} as the *a posteriori* (or vice versa), and by subsequent computation of $I(citing:cited)$ and $I(cited:citing)$. The advantage of this latter approach is the direct interpretability of its decomposition into each of the constituting journals (and even cells), while the static analysis only teaches us about the relation between *citing* and *cited* as generalized variables in the matrix. Therefore, I will let the discussion of the static analysis follow from the discussion of this "dynamic" perspective.

Each cell value a_{ij} contributes to the asymmetry when compared with its counterpart a_{ji}. The expected information content of the message of the difference between the two values is relative to the overall matrix. Therefore, by using formula (8) above, the total information content of the asymmetry is equal to:

$$I = \Sigma_i \Sigma_j \left(f_{ij}/N \right) * \log \left\{ \frac{\left(f_{ij}/N \right)}{\left(f_{ji}/N \right)} \right\}$$

Since all the cells of the same matrix are involved in both the computation of $I(citing:cited)$ and $I(cited:citing)$, the two sum totals for ΔIs have to be identical. For the 1984 matrix, this I is equal to 0.290 bits.

However, we may use the I and its component ΔIs as straightforward measures of the asymmetry of the matrix, and of its rows and columns, respectively. The respective sums of ΔIs for rows and columns will usually differ. Since the sign of each contribution to I is dependent on the increase or decrease of the relative frequencies ($\Delta I = (f_{ij}/N) * \log(f_{ij}/f_{ji})$), the $\Sigma \Delta I$ for each subset gives us a direct measure of whether the subset is a relative "sink" or a "source" of citations. Table 3 lists the decomposition of I in terms of rows and

---

[13] See for a discussion of the effect of diagonal values in the case of journal–journal citation matrices: Price 1981; Noma 1982; Todorov and Glänzel 1988.

Table 3
Deltas I (in bits) for cited versus citing patterns, and vice versa

	(cited : citing)	(citing : cited)
JACS	0.230	−0.073
JCHEMPH	0.148	−0.052
TRAHLT	0.016	0.010
MOLPHYS	0.007	0.000
PHYSREVA	0.005	−0.002
CHEMPHLT	−0.003	0.019
JCHEMSC	−0.005	0.020
CHEMPHYS	−0.011	0.029
INORGCH	−0.012	0.052
JORGMETC	−0.016	0.042
JORCHEM	−0.017	0.077
TETRAHE	−0.020	0.064
JPHCHUS	−0.031	0.104
	(0.290)	(0.290)

columns for the citation patterns of the 13 journals under study here; Fig. 2 plots these values as coordinates of a two-dimensional map.

Both the shape of the curve and the position of the points in Figure 2 inform us about the citing/cited relations in the network: journals
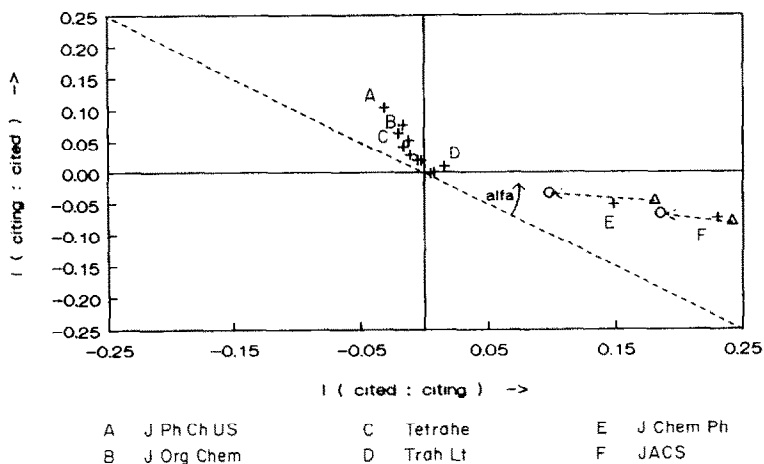

Fig. 2. "Asymmetry"; expected information values in bits of information for cited versus citing patterns of 13 major chemistry journals. +1984; △ 1981; ○ 1987; — — —→ indicates developments 1981 → 1987.

in the second quadrant are relative "sources" of citations, while journals in the fourth quadrant are relative "sinks". The closer to the origin, the more neutral the journal is, in this respect. Obviously, the two major journals in the field (*JACS* and *Journal of Chemical Physics*) are also the two major sources of citations, not only in absolute but also in relative terms. At the other extreme, the *Journal of Physical Chemistry–US* is the least cited, and in relative terms consequently the major "sink" of citations.

Because of the mathematics of *I*, points below and to the left of the diagonal ($y = -x$) are not possible. [14] The various points in the map can also be considered as vectors (from the origin) with an angle of α to the diagonal. The (co)tangent of this angle gives a measure of the relative asymmetry of each of the rows and columns (k), since it can be shown [15] that:

$$\operatorname{cotg}(\alpha) = \frac{\Sigma_k(q_k - p_k)}{\Sigma_k(q_k + p_k)} \tag{8}$$

This indicator therefore expresses the relative weight of the differences, i.e., the asymmetry, between respective ps and qs for each network element k.

In summary, the position of each point on the map informs us about a corresponding journal's position in the network (as a "sink" or "source" of citations). The distance from the origin is a measure of the size of the asymmetry effect, and the angle of the vector with the indicated diagonal is a measure of the relative weight of the asymmetry in terms of the given element.

4. "Citing" and "cited" as variables in the static analysis

By using the static measures H, expected information contents for citing and cited patterns can be computed for the journal–journal citation matrix, and the results can be evaluated in terms of their mutual information, and in terms of their constituents by subsequent

[14] It follows from considerations about the relations among the I_{matrix}, ΔI, and the $I_{subgroup}$, i.e. in this case the $I_{journal}$, that the sum of the two ΔIs for row and column for each element k of a square matrix must be larger than or equal to zero, and that therefore the points will be positioned in the right upper triangle of the graph only.

Proof:

Is have to be positive (cf. Theil 1972: 59f.) both for groups and for subgroups. ΔIs can be negative as an effect of normalization. However, $I_{journal}$ can also be obtained from the ΔIs for row and columns by appropriate normalization.

Let n_q and n_p be the margin totals for row k and column k, and N be the grand sum of the matrix; q and p are relative frequencies of the cells belonging to the respective row and column in terms of the grand sum of the matrix. Normalization relative to the margin totals for the respective row and column is achieved by multiplication of q by (N/n_q) and of p by (N/n_p). Therefore:

$$I_{journal} = \Sigma\left\{q*(N/n_q)\right\} \log \frac{q*(N/n_q)}{p*(N/n_p)}$$

$$= \Sigma\left\{q*(N/n_q)\right\}\left\{\log(q/p) - \log(n_q/n_p)\right\}$$

Since $I_{journal} \geq 0$:

$$\Sigma\left\{\log(q/p) - \log(n_q/n_p)\right\} \geq 0$$

$$\Sigma \log(q/p) \geq \log(n_q/n_p)$$

$$\Sigma q \log(q/p) \geq n_q \log(n_q/n_p)$$

However:

$$\Sigma q \log(q/p) = \Delta I_{(q:p)}$$

and therefore:

$$\Delta I_{(q:p)} \geq n_q \log(n_q/n_p)$$

Analogously:

$$\Delta I_{(p:q)} \geq n_p \log(n_p/n_q)$$

and therefore:

$$\Delta I_{(q:p)} + \Delta I_{(p:q)} \geq n_q \log(n_q/n_p) + n_p \log(n_p/n_q)$$

$$\geq (n_q - n_p) \log(n_q/n_p)$$

For $n_q > n_p$, this is a product of two positive factors; hence, > 0; for $n_q = n_p$, this product is zero; for $n_q < n_p$, this is a product of two negative factors; hence, > 0. *Q.e.d.*

decomposition. If we do so for the 1984 matrix, we find the following values:

$H(citing, cited) = 5.667$ bits
$H(citing)$ $= 3.457$
$H(cited)$ $= 3.173$

and therefore:

$H(citing \mid cited) = 2.493$
$H(cited \mid citing) = 2.209$
$T(citing, cited) = 0.964$

[15] By using Pythagoras, the (co)tangent can be calculated from the derivation of the distance (d) of point k from the origin and its distance (c) to the diagonal:

$$d = \Sigma_k \log(q_k / p_k) * \sqrt{(q_k^2 + p_k^2)}$$

$$c = \Sigma_k \log(q_k / p_k) * (q_k + p_k)/\sqrt{2}$$

and therefore:

$$\cot g(\alpha) = \frac{\sqrt{(d^2 - c^2)}}{c} = \frac{\Sigma_k (q_k - p_k)}{\Sigma_k (q_k + p_k)}$$

The advantage of using the cotg() instead of the tg() is its continuous decline over the relevant range (0 to 180 degrees), while the tg(90°) = ∞. For the 13 journals considered here, the following values were found:

	$tg(\alpha)$	$\cot g(\alpha)$
JACS	4.749	0.211
JChemPh	5.992	0.167
MolPhys	23.599	0.043
PhysRevA	33.603	0.030
TrahLt	77.721	0.013
ChemPhLt	−17.076	−0.059
InorgCh	−7.737	−0.129
JChemSc	−7.592	−0.132
JOrgChem	−6.296	−0.159
JOrgmetC	−5.654	−0.177
ChemPhys	−4.681	−0.214
JPhChUS	−3.754	−0.266
Tetrahe	−2.729	−0.366

$$\frac{H(citing) - H(citing \mid cited)}{H(citing)} = 27.9\%$$

$$\frac{H(cited) - H(cited \mid citing)}{H(cited)} = 30.4\%$$

This means that the citing pattern is a 10 percent better predictor of the cited pattern than vice versa. However, the mutual information between "cited" and 'citing" is remarkably low, at first sight (0.964 bits): one would have expected the two patterns to determine one another to a much larger degree. However, the mutual information is only of the order of a 30 percent mutual reduction of the uncertainty in the prediction, since there are discrete groupings in the data, each with high mutual information within the subsets, but hardly any "in between group" co-variance.

This result therefore raises the question of whether we can also infer the grouping of the journals using statistical decomposition analysis. Of course, this would bring the major questions back on stage, which have led to the development of various forms of multi-variate analysis. If we want to cluster, what will we regard as more or less similar, i.e., what will be the similarity criterion? Are we to use agglomerative clustering techniques or divisive ones? Do we want to take all variance within the matrix into account, or begin the cluster-ing by focusing only on diadic relations between cases which exhibit strong graphs?

5. Agglomerative clustering

A simple agglomerative clustering algorithm would be to merge as a cluster those two distributions (rows, columns or both) which are most identical, i.e., mutually have the lowest I associated with the message which converts the one distribution into the other, and subsequently to iterate the procedure. In each step this algorithm seeks the strongest graphs. [16]

Figure 3 shows the results, in the form of a dendrogram, for the citing patterns of these 13 chemistry journals in 1984. Since the initial step does not imply the construction of a symmetrical (dis)similarity
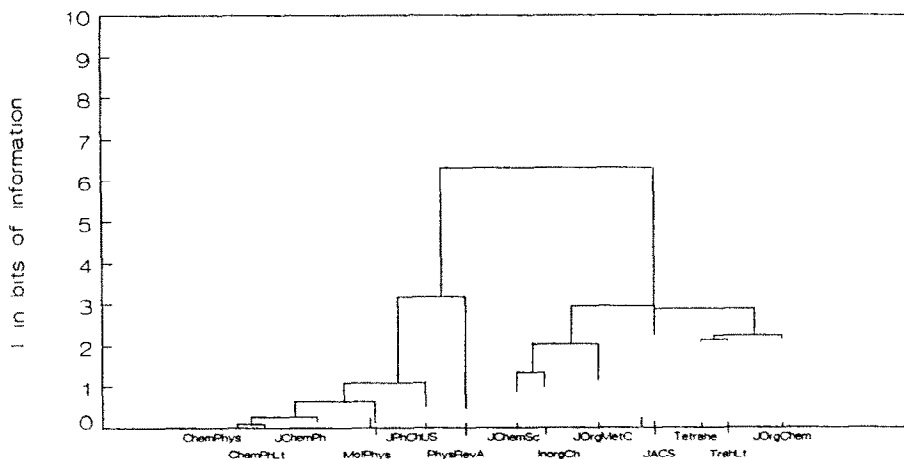
Fig. 3. Agglomerative clustering. Dendrograms of 13 major chemistry journals using measures from information theory.

matrix, the leaves of the dendrogram do not have to be equal, as is the case in most clustering programmes. The vertical distances, therefore, meaningfully represent the expected information values of the clustering in each case.

If we compare the results with a dendrogram in Figure 4 which is based on using Wards' mode of analysis for clustering the Pearson correlation as similarity measure, [17] we notice that the qualitative order is the same for the "physical chemistry" cluster, but that the complexity of the position of *JACS* is visible with extreme precision in this graph: *JACS* is only marginally more linked to the "organic chemistry" cluster than to the the "inorganic chemistry" group.

[16] However, I_{ij} will in general be unequal to I_{ji}, and therefore, it may occur (as actually happens with *JACS* in relation to the "organic" and "inorganic chemistry" clusters below) that despite the fact that two cases form a strong graph (in the sense of mutually having the lowest I associated with transformation into each other), a third case can combine with one of the two values with an in-between value for I. If that happens, we can merge this third case (as a "weaker graph") in the same step of the clustering procedure.

[17] Cluster analysis is well-known for its proliferation of options, caused by the possibilities of choosing among similarity criteria and clustering algorithms. The results are very different, accordingly. I usually found the best graphic representation of factor analytic results by using this combination of Wards' mode of analysis on a Pearson correlation matrix, although for formal reasons this combination is not allowed (cf. Leydesdorff 1987; Leydesdorff and Zaal 1988). Analyses are based on using CLUSTAN 2A. (See for further details, e.g., Everitt 1974.)
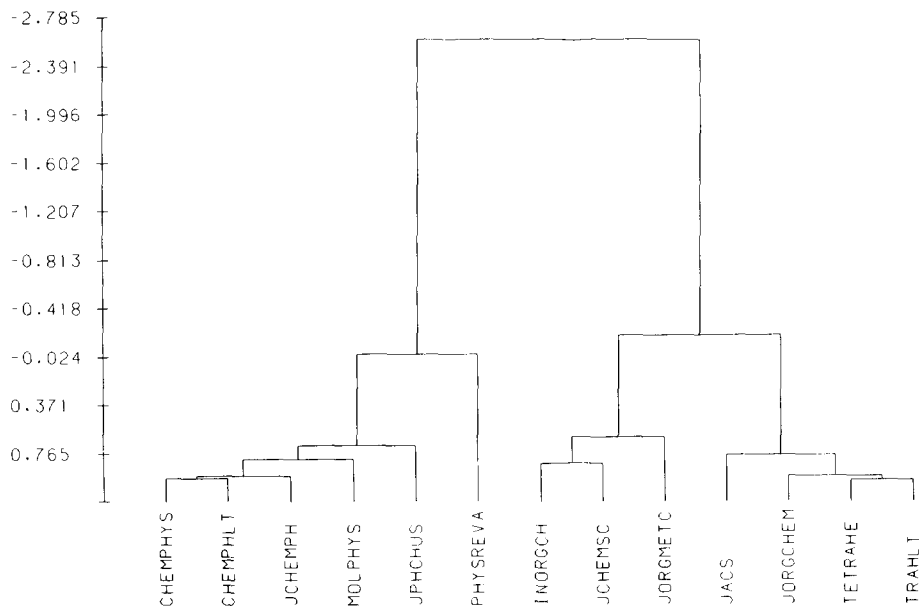
Fig. 4. Cluster analysis of citation patterns for 13 major chemistry journals; Wards' mode of analysis, Pearson correlation coefficient.

6. Divisive clustering

However, it is impossible to decide on the basis of this agglomerative procedure how many groups should be distinguished, since the agglomerative steps are formally equivalent. Therefore, we may again raise the question of whether *JACS* is to be counted as part of the "organic chemistry" group or whether it should rather be handled as a separate case? Were we to choose the latter option, then what about *Physical Review A*, since this journal bifurcates from the "physical chemistry" cluster at even a larger distance (in terms of bits) than *JACS* from the "organic chemistry" cluster core? Could we not find a more strict criterion for division into groups using the rules of statistical decomposition for divisive clustering into groups directly?

The problem of how to choose the number of clusters, factors, groups, dimensions, etc. is a pervasive one in multivariate analysis. If there are no *a priori* theoretical reasons – as it is usually the case in exploratory uses of the techniques – such decisions tend to remain somewhat arbitrary. In factor analysis, methods such as visual inspec-

tion of the scree plot or a cut-off at certain eigenvalues are common practice. In cluster analysis and multidimensional scaling, decisions based upon visual inspection of the results are also common. In bibliometrics, Small and Sweeney (1985) have proposed "variable level clustering", i.e., in essence the adaptation of the clustering level to the density of the cluster involved; the search for a formal criterion is thus replaced by a procedural one.

Indeed, statistical decomposition analysis offers a straightforward answer to this problem, since disaggregation is highly formalizable. As noted in a previous section, disaggregation of a set in g groups can be described with the following formula:

$$H = H_0 + \Sigma_g P_g H_g \tag{9}$$

in which H is the expected information content (entropy) of the aggregated distribution, and P_g the probability of each of the groups which as a subset has an uncertainty equal to the respective H_gs. The "in-between group entropy" H_0 is a measure of the specificity that prevails at the level of the subsets.

The right-hand term of the above equation is equal to the entropy of a variable (n) under the condition of a nominal variable (m) which can be attached to the grouping ($H(n/m)$). (E.g., the grouping variable m may be attributed on the basis of the factorial structure.) The left-hand term, H_0, is therefore equal to $H(n) - H(n/m)$, which is the uncertainty in n that is *not* accountable to the uncertainty within the groups, or in other words the *transmission* (*mutual information*) of the grouping variable m itself to n. The larger this transmission, the more reduction of uncertainty there will be among the groups, and therefore the better the groups will be in terms of the homogeneity of their distributions. However, by definition:

$$H(n \mid m) = H(n, m) - H(m) \tag{10}$$

and therefore:

$$H_0 = H(n) + H(m) - H(n, m) \tag{11}$$

This means that the increase of H_0 if we add another group (cluster, factor, etc.) is composed of a part that is dependent only on the grouping variable ($H(m)$), and a part which is dependent on the

interaction between the variables n and m ($H(n, m)$). The question thus becomes: for which value of m does the function $\{H(m) - H(n, m)\}$, and consequently H_0, reach a maximum? Although this problem may be addressed using differential calculus ($\delta\{H(m) - H(n, m)\}/\delta m = 0$), let me address the problem here with a more intuitive approach.

If we divide one group into two subgroups i and j, using $H_{ij} = H_0 + P_i H_i + P_j H_j$, the aggregated H_{ij} may be larger than both H_i and H_j, or larger than one of them and smaller than the other. (The two groups cannot be both larger than H_{ij}, since the "in-between group" H_0 is necessarily larger than or equal to zero.) The case of $H_i \leq H_{ij} \leq H_j$ corresponds to the removal of the more than average heterogeneous case(s) into a separate subgroup: therefore, this new subgroup has a higher uncertainty, and the remaining subgroup becomes more homogeneous than the original group. This is always possible, but it is not yet clustering, which entails the notion of reducing uncertainty in *both* subgroups. Therefore, we may define "divisive clustering" as the case where both new subgroups have a lower expected information content than the undivided group.

Note that the above justification of the division is based on the right-hand term of the formula for disaggregation only ($\Sigma_g P_g H_g$). The value of the left-hand term (H_0) is sensitive to both the number of groups – since each further division adds to H_0 unless the two groups have similar Hs – and to the quality of the attribution of cases to groups given a certain number of groups. However, the two questions – (1) concerning the number of groups, and (2) concerning the attribution of cases to groups – are obviously independent, given the two parts of the above noted equation.

The possible number of attributions of n cases to m groups ($m < n$) increases so rapidly with the number of cases and the number of groups, that systematic comparison of all possible combinations can imply heavy computation. However, in practice, this type of repetitive approach to the data, which is characteristic of information theory (Krippendorff 1986), can be programmed in DO WHILE-loops: first, we investigate whether the setting apart of any of the cases leads to two subgroups, both of which have lower Hs than the overall H. If so, we begin with the one which leads to the highest H_0, and systematically evaluate whether the addition of other cases to this one subgroup leads to a further increase of H_0, etc. Once we have so investigated all
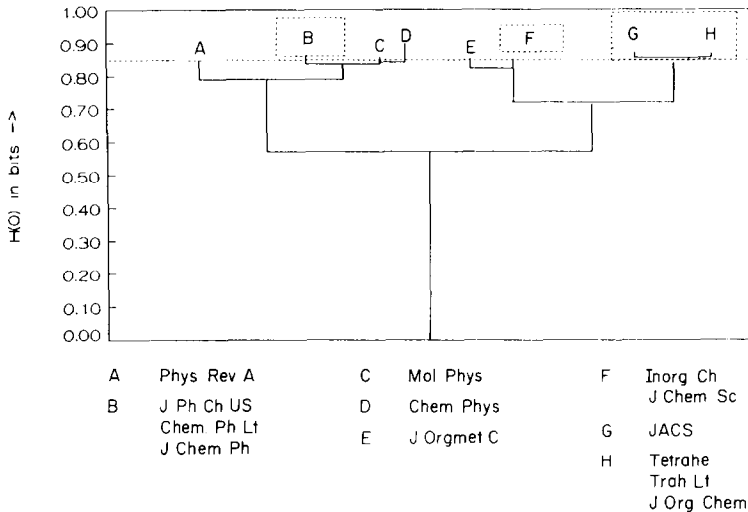
Fig. 5. Divisive clustering of *citing* patterns of 13 major chemistry journals.

the possibilities and decided upon the best division into two sub-
groups, the analysis can be repeated for the two subgroups respec-
tively. After normalization of H_0 in terms of the grand sum of the
matrix, a dendrogram can be constructed, which is exact both in terms
of the vertical distances between the nodes and in terms of where to
draw the line above which further division leads to subgroups which
are not both lower in their entropy than their respective aggregates.
This level in the graph corresponds to a maximum for H_0.

Figures 5 and 6 show the dendrograms for citing and cited patterns,
respectively. The dotted line represents the level above which further
division becomes counterproductive (i.e., would lead to a decrease of
total H_0). From these figures, we may conclude that there is a
relevant subdivision of the cluster which we considered as the one
"chemical physics" cluster above; however, this subdivision is some-
what different as "cited" and "citing". Both "cited" and "citing," the
Journal of Organometallic Chemistry has to be considered as an
isolate, and "cited" it seems that the whole cluster of "inorganic
chemistry" falls apart. However, along both axes, *JACS* firmly belongs
now to the "organic chemistry" cluster. (As was noted above, the
major limitation of divisive clustering is that each case has to be
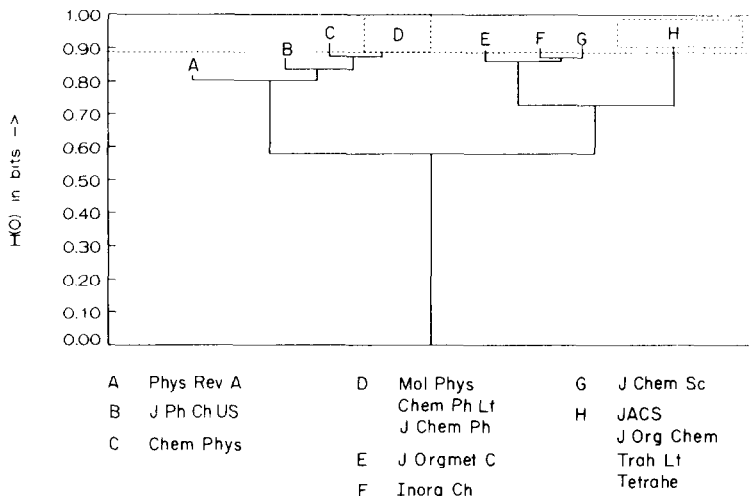attributed to one group only, and that details about inter-group

Fig. 6. Divisive clustering of cited patterns of 13 major chemistry journals.

positions, as shown by using agglomerative clustering and factor analysis, are not revealed.)

In summary, we may conclude that in terms of dividedness, the exact clustering of these two patterns leads us to the distinction of seven and eight subgroups, respectively. The "cited" pattern is more
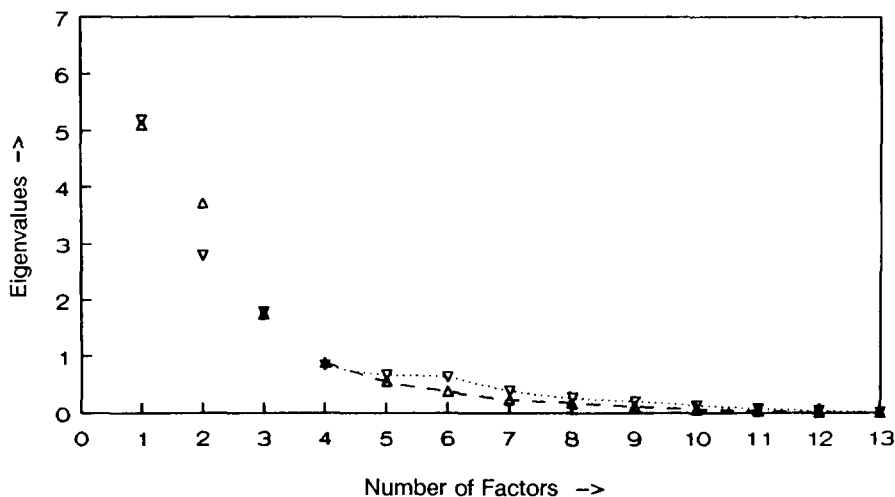

Fig. 7. Scree plots eigenvalues citation patterns for 13 major chemistry journals 1984; △ citing, ▽ cited.

Table 4
Factor analysis of citing pattern of 13 "chemistry journals"; four factors forced. Principal–component analysis, Varimax rotation – Kaiser normalization

Factor	Eigenvalue	Pct of Var	Cum Pct
1	5.10338	39.3	39.3
2	3.71624	28.6	67.8
3	1.74990	13.5	81.3
4	0.88657	6.8	88.1

Varimax converged in 5 iterations
Rotated factor matrix:

	Factor 1	Factor 2	Factor 3	Factor 4
CHEMPHLT	0.97034	− 0.04210	− 0.05748	0.04397
JCHEMPH	0.95862	− 0.09472	− 0.07715	0.10414
CHEMPHYS	0.94609	− 0.12497	− 0.13481	0.02694
JPHCHUS	0.84504	0.16857	0.11524	− 0.02427
MOLPHYS	0.83678	− 0.17293	− 0.19056	0.05371
TETRAHE	− 0.08978	0.97921	0.12886	− 0.07722
JORGCHEM	− 0.08007	0.95670	0.16671	− 0.06099
TRAHLT	− 0.15185	0.95405	0.04416	− 0.09632
JACS	0.19066	0.78732	0.52907	0.01318
JCHEMSC	− 0.12914	0.01697	0.92257	− 0.09178
INORGCH	0.04224	0.24207	0.87231	− 0.00175
JORGMETC	− 0.19436	0.19302	0.70483	− 0.14136
PHYSREVA	0.09225	− 0.13880	− 0.16657	0.96895

Factor designation:

	Factor 1	Factor 2	Factor 3	Factor 4
	"chem. phys."	"org. chem."	"inorg. chem."	

divided than the "citing" pattern. It is interesting to look back at the scree plots of the respective factor analyses along those two dimensions (Fig. 7): with hindsight, we see more components in the cited patterns than in the citing patterns, but these differences are within the "scree" of the graph, and would therefore be considered as irrelevant for the factor analysis. (The default factor analysis in SPSS generates a three factor solution.)

However, in factor analysis we aim primarily at the reduction of complexity in terms of the number of *relevant* factors, while in divisive cluster analysis we study the dividedness as such. If we force a four factor solution in the citing patterns of this matrix, *PhysRevA* loads highly significantly on this factor only (Table 4). *JACS* retains its interdisciplinary position. Subsequently, in the case of five factors the

JOrgmetC loads primarily on this factor. However, with more than four factors, factorial complexity and factor pattern correlations in the oblique solution also increase.

Therefore, we may now conclude that *PhysRevA* constitutes a fourth element in the structure of the matrix along this axis. However, this was not obviously made visible by the factor analysis from SPSS or the two dendrograms in the previous section.

7. Confirmational usage

In a confirmational design, we may now also investigate other questions with respect to this matrix. For example, let us further investigate the "interdisciplinary" postion of *JACS* in the citing pattern of this matrix. We must therefore analyze in more detail the right leaf of the dendrogram in Figure 3, which contains citing patterns for both "inorganic" and "organic chemistry" journals. Questions can be raised as to whether *JACS* should be considered as an isolate or as part of the "organic chemistry" cluster; and so on.

Table 5 gives an overview of the associated values for H for these various options. It can be concluded that by the criterion of lower Hs for subgroups than for the aggregate, we should consider this set as two subgroups of journals only: the overall H ($= 2.7088$ bits) is larger than the value for the "organic chemistry" group (2.4137 bits) or for "inorganic chemistry" (2.5534 bits). Further division of the "organic chemistry" group into a separate group containing *JACS* and a group containing *Journal of Organic Chemistry, Tetrahedron* and *Tetrahedron Letters* leads to an H for *JACS* of 2.4304, which is above the previously found value of the "organic chemistry" group as a whole (2.4137). This means that the uncertainty has increased in this sub-

Table 5
Values of H associated with the right leaf of the dendrogram

one group			2.7088		
	┌— "inorganic"			**2.5534**	
	└— "organic" ┐			**2.4137**	
		┌— "organic" – JACS			2.1101
		└— JACS			2.4304

group; therefore, in agreement with the above result from the explorative analysis, this hypothesis should be rejected.

Alternatively, the attribution of *JACS* to the "inorganic chemistry" group, under the assumption of three subgroups, leads to only a slightly lower value for H_0 (0.7144 bits) than attribution to the "organic chemistry" group ($H_0 = 0.7170$ bits). Therefore, *JACS* is more closely related to "organic chemistry" than to "inorganic chemistry" also in this analysis. However, the difference is only marginal. This complex position of *JACS* between "organic" and "inorganic chemistry," as evident from agglomerative clustering (Fig. 3) and as factorial complexity from factor analysis (Table 4) was obscured in the results from divisive clustering, since each case then has to be attributed to a specific grouping.

8. Cliques

The above decomposition in clusters was based on either citing patterns or cited patterns. In terms of the two-dimensional matrix H_{ij}, we grouped along one dimension only, i.e., i and j, respectively. Grouping over i or j led to two different analyses, and correspondingly, to two sets of results. As we have seen above, the asymmetry of the matrix in this case was low, but in comparison to the citing patterns all values of H_0 for the cited patterns were slightly higher, suggesting a more pronounced dividedness than in the former case.

However, in the case of a bivariate matrix we can also study $H(citing, cited)$ and the effects of grouping in both dimensions in one design. Analogously to the above analysis in each of the dimensions, we can study the transmission $H(citing \mid a, cited \mid b)$, in which now both a and b are grouping variables. Grouping bivariate arrays in this way also gives us the possibility to integrate relational approaches (from graph and clique analysis) with approaches based on structural equivalence (i.e., eigenvectors) into one conceptual framework. However, in order to keep the discussion as little mathematical as possible, I will address the problem again in terms of disaggregation in analogy to the discussion above.

Let me use the simple matrix depicted in Fig. 8 to explain my point. Principal component analysis, and therefore also factor and cluster analysis, is based on grouping over rows or columns. However, were

11	12	13	14
21	22	23	24
31	32	33	34
41	42	43	44

Fig. 8. Illustrative, exemplary matrix.

we for example to hypothesize that 1 and 2 form a clique, we would expect cells $11 + 12 + 21 + 22$ to form a strong "cluster" when compared with the sum of all the other cells. For both groups – the one supposedly forming a clique and the remainder – we can calculate an H_{ij}, and a P_{ij}, and after appropriate weighting an H_0 can be calculated (by substraction from H_{ij} for the whole set). Again, H_0 is a straightforward measure of the specificity of the subsets given the number of clusters (in this case two). However, the cluster may consist of any part of the matrix, and not necessarily a set of columns or rows: we can attribute each cell value to the hypothetical subset and to the grand sum respectively, and consequently calculate an H_0. As the reader should now understand for the same reasons as above, we are able to distinguish between the effects on H_0 of increasing the number of clusters ("cliques") and the effects of better grouping given a certain number of clusters.

Note additionally that it is not necessary to determine only groups which are symmetrical with respect to the diagonal, since with our formulas we can vary along both dimensions independently. We may also decide not to include the diagonal elements themselves into the clusters, or to treat them as a separate group.

Table 6
Clique analysis of the "chemical physics" and *PhysRevA* group in terms of H_0 in bits of information

	H_0
5 "chemical physics" journals	1.9982
PhysRevA	
...(various combinations with in-between values)	
ChemPhys + ChemPhLt + JChemPh	**2.1084**
PhysRevA + MolPhys + JPhChUS	
ChemPhLt + JChemPh	2.1077
PhysRevA + MolPhys + JPhChUS + ChemPhys	
ChemPhys + ChemPhLt	2.0358
PhysRevA + MolPhys + JPhChUS + JChemPh	

The use of asymmetries in i and j gives room also for the distinction between "weak" and "strong" graphs in the analysis. Strong cliques correspond to mutual diadic relations, and therefore they have to be operationalized in this framework in terms of the two variables i and j ranging over the same values only (i.e., $n < i < m$ AND $n < j < m$). Note also that while the former analysis implied the notion of "structural equivalence" as fundamental to the idea of factor and principal component analysis, this analysis addresses questions concerning relations in graphs, as they are distinguished nowadays from the former analysis in network analysis (Freeman 1978/79; Burt 1982).

With respect to the analysis of the 1984 matrix, I will now limit the discussion to the question of whether the four major groups among the 13 journals which we identified above in terms of structural equivalence also correspond to four cliques in this network. (However, in searching for four cliques in the matrix, we must allow for five groups, since the off-diagonal elements form a remainder-group.)

In terms of graph analysis, the "chemical physics"/*PhysRevA*-group separates into two cliques, which are different from the analysis in terms of structural equivalence. The attribution with highest H_0 is boldfaced in Table 6: in this analysis the "chemical physics" group is composed of two groups with the highest "within group" densities, consisting of *ChemPhys, ChemPhLt*, and *JChemPh* on the one hand, and *PhysRevA, MolPhys.* and *JPhChUS*, on the other. Note the

differences in the citing and cited patterns: as a clique, the group of three journals with "chemical physics" in their title are one, while their citing and cited patterns are much more interactive with *Mol-Phys* and *JPhChUS*. As a clique the latter two form one graph with *PhysRevA*.

The values for H_0 in the Table 6 are based on the initial assumption from the above analysis about the attribution of *JACS* to the "organic chemistry" group of journals in the other part of the matrix. However, if in terms of clique analysis *JACS* is attributed to the "inorganic chemistry" group, H_0 further increases by 0.0268 bits to 2.1352. Obviously in this analysis, the latter grouping is better than the inclusion of *JACS* as an "organic chemistry" journal. Since as we noted, we may vary the attribution over the two dimensions, we may also group *JACS* asymmetrically, for its citing and cited patterns. (However, we have then to correct for the diagonal values. Omission of the diagonal values can be argued for and against on substantive grounds, anyhow.) It can be shown that the attribution of *JACS'* citing pattern to the "inorganic" chemistry group, and of its cited pattern to "organic" chemistry, leads to a further increase of H_0 of 0.0304 bits, as against a further increase of only 0.0255 bits for the attribution of cited to "inorganic" and citing to "organic chemistry."

In summary, the grouping in four cliques shows *JACS* to be at an asymmetrical crossroads where the "inorganic" and "organic chemistry" groups meet; and the emergence of a specialty group of journals which have "chemical physics" as such in their title. Note that neither of these results is exhibited by applying cohesion analysis in STRUCTURE (Burt 1987; see Fig. 9).

9. The dynamic analysis

We now extend the analysis to similar matrices for 1981 and 1987, in addition to the one of 1984 used above. In essence, we can think of the three matrices as forming a cube, in which all kinds of questions about conditional entropies can be raised: e.g., to what extent is the uncertainty of the prediction of development of citation patterns over the years for one set of journals, such as "organic chemistry," reduced if we know the overall patterns of development for the aggregate?

```
CONSERVATIVE CLUSTERING OF THE DISTANCES
(based upon the diameter or maximum algorithm)
-- high distances within clusters are preserved

The criterion distance, D, is the largest distance
between any pair within a cluster.


     Values
      of D          0 0 0 0 0 0 1 0 1 1 0 0 1
-------------       3 6 8 1 2 9 0 5 1 2 4 7 3
      .000          . . . . . . . . . . . XXX
      .000          . . . . . . . . . . XXXXX
      .000          . . . . . . . XXX . XXXXX
      .053          XXX . . . . . XXX . XXXXX
      .154          XXX . . XXX . XXX . XXXXX
      .275          XXX . . XXX . XXX XXXXXXX
      .358          XXXXX . XXX . XXX XXXXXXX
      .743          XXXXX XXXXX . XXX XXXXXXX
      .878          XXXXX XXXXXXX XXX XXXXXXX
      .959          XXXXX XXXXXXXXXXX XXXXXXX
      .995          XXXXX XXXXXXXXXXXXXXXXXXX
      .995          XXXXXXXXXXXXXXXXXXXXXXXXX

                    I J J C C J M J P T J J T
                    n C O h h P o C h e A O r
                    o h r e e h l h y t C r a
                    r e g m m C P e s r S g h
                    g m M P P h h m R a   C e
                    C S e h h U y P e h   h L
                    h c t y L S s h v e   e t
                        C s t       A     m
```

Fig. 9. Cohesion analysis of citing patterns using STRUCTURE (Burt 1987).

However, this would still be a "static" analysis: in a dynamic perspective we do not take "time" or "sequence numbers" as another variate which co-varies with other variables. (This would lead to problems of auto-correlation in the data; see below.) Instead, we compare among (sets of) events. The events are in this case the bivariate probability distributions of the matrices for each of the three years, i.e., 1981, 1984 and 1987.

I will limit myself to the following exemplary questions among the many which could be raised:

(1) What is the overall pattern of change?
(2) How can we decompose this overall pattern in terms of citing and cited patterns and in terms of cliques?

(3) How can we best analyze these patterns? Is multivariate change or univariate change predominant? Is the structure changing, or only the composing elements?

10. Changing patterns

Since in each of the matrices the mutual aggregated citation rates of the same 13 journals are organized in a similar way, [18] as bivariate events of distributions with $13^2 = 169$ values each, between each two of which we can calculate values for ΔI_{ij}, which subsequently add up to an overall I, by using the formula:

$$I_{ij} = \Sigma_i \Sigma_j q_{ij}{}^2 \log(q_{ij}/p_{ij}) \tag{12}$$

the difference between two corresponding cells f_{ij} in two matrices can be straightforwardly transformed into one ΔI_{ij} after normalization as relative frequencies, q_{ij} and p_{ij} The overall I_{ij} between each two matrices is fully decomposable in terms of these Δs. However, as long as we normalize in terms of the grand sums of each of the matrices, the sums of any corresponding subsets can also be compared straightforwardly. The "information matrix" containing the Δs is a transformation of the differences between the two matrices only.

A very informative criterion for comparing subsets is the sign of a contribution to I_{ij}, i.e., the sum of the Δs for a subset, to the overall I_{ij}. As noted above, the log becomes negative if q is smaller than the corresponding p, and positive if q is larger than p; and therefore the sign of each of the cells, and of each subset, is a direct indicator of the relative increase or decrease of weight of that cell or that subset. (Note that I_{ij} for the whole group must be positive or zero (Theil 1972: 59f).) This means that we can directly measure the dynamics among subsets using respective sigmas. For example, if we want to compare the dynamics of citing patterns, we may sum the Δs over columns; if we want to compare cited patterns, we may sum the Δs over rows, and if we want to compare cliques as defined in the previous section, we may sum over respective rows and columns only, and compare the contributions to the overall I_{ij}. In addition, we may

[18] See for a discussion of "emergence" of new categories Leydesdorff 1990c.

Table 7
Expected information value for matrices of different years (mbits)

a posteriori a priori	1981	1984	1987
1981	* * *.*	54.3	81.7
1984	38.7	* * *.*	18.0
1987	62.7	18.0	* * *.*

also analyze the difference between off-diagonal elements and diagonal elements for each subset, using the substract of the two sigmas: for example, in terms of the model matrix above (see Fig. 8), we may sum over all deltas contained in subset 22 and compare that with the sum for the row (21 + 22 + 23 + 24) in order to give a straightforward answer to the question of how large a percentage of the increase or decrease in weight of that row is due to only the diagonal element. [19]

In summary, the matrix of values for ΔI which can be calculated between any two matrices representing the network elements can be used to answer a multitude of questions concerning the dynamics of clusters and graphs in various dimensions. Comparison of the three matrices under study here leads to the creation of three such matrices of Δs, notably one for the comparison of 1987 with 1984 data, one for the: comparison of 1984 with 1981 data, and one for the comparison of 1987 with 1981 data.

Table 7 gives the overall values for I_{ij} for the various years. Remember that this is a summary indicator of change in citing *and* cited patterns, since both these patterns are contained in each original matrix: the table gives only grand sums of each of the matrices of Δs which can be derived between each two of the original matrices.

It teaches us that the overall dynamics of the citation patterns between 1981 and 1984 were three times as large (54.3 mbits) as those between 1984 and 1987 (18.0 mbits). The values in the upper triangle are larger both in terms of cells and for its sum (153.0 mbit) than for the lower triangle ($\Sigma = 119.4$), which makes it clear that overall we gain more information by comparing with the time axis than by comparing against the time axis. The difference is due to the 1981–1984 period exclusively, and not to the 1984–1987 period: it indicates

---

[19] See also note 13.

Table 8
Decomposition of the expected information value of the change in journal citation patterns for
the period 1981–1987

Citing		Cited	
JPhChUS	74.2 mbits	JPhChUS	47.1
JChemPh	47.0	PhysRevA	38.1
PhysRevA	30.9	MolPhys	30.0
TrahLt	13.8	ChemPhLt	15.0
ChemPhLt	9.7	TrahLt	11.1
JChemS	− 0.4	ChemPhys	6.1
MolPhys	− 7.0	Tetrahe	2.1
InorgCh	− 8.0	JOrgChem	1.4
JOrgChem	− 8.2	JChemPh	1.4
ChemPhys	− 11.2	JChemSc	0.4
Tetrahe	− 12.1	InorgCH	− 2.4
JOrgmetC	− 15.6	JOrgmetC	− 14.1
JACS	− 31.4	JACS	− 54.5
	(81.7)		(81.7)

a shift away from randomness in the matrix, i.e., of more pronounced
specification and differentiation [20] among the composing elements of
the matrices during the 1981–1984 period, which came to a halt in
later years.

Where should we look for the sources of these changes? Let us now
focus on the dynamics along the time axis only, comparing 1981 with
1987 as *a priori* and *a posteriori* respectively.

11. The dynamics of relations among the 13 journals

The summation of the row and/or column elements for each journal
in the matrices containing the Δs to the overall I_{ij} can straightfor-
wardly be used as an indicator of the contribution of that journal to
the overall change of the citation patterns in the matrix, since each ΔI
is normalized in terms of the grand sums of the underlying data
matrices. From Table 7 (above) we know that the overall change over
the period 1981–1987 is equivalent to a message of 81.7 mbits of

[20] In the case of equiprobability in the *a priori* distribution, I becomes equal to the amount by
which the entropy is reduced below the original (maximum) value by the *a posteriori* distribution
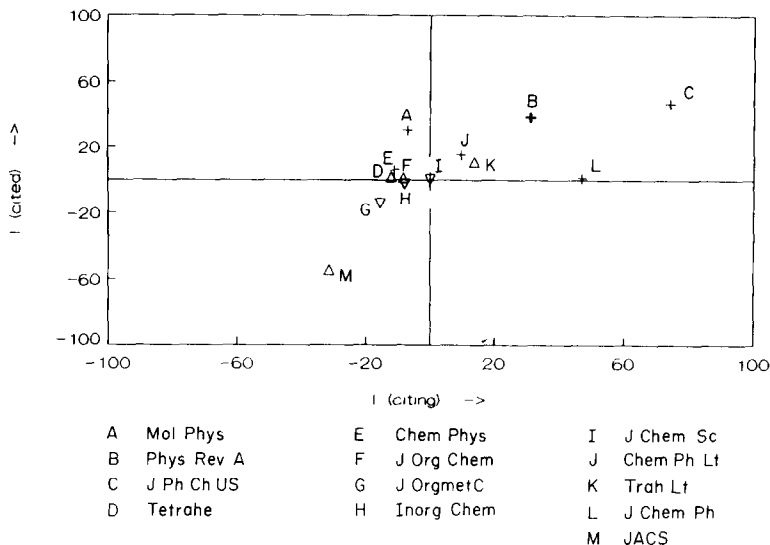(Theil 1972).

Fig. 10. Dynamic analysis of citation patterns of 13 major chemistry journals, 1981–1987. Expected information values in mbits. + chemical physics; △ organic chemistry; ▽ inorganic chemistry; + Phys. Rev. A.

information. The decomposition of these 81.7 mbits in terms of citing and cited patterns for each of the journals is given in Table 8.

The two values for each journal can also be used as coordinates on a map showing the dynamics of the citation patterns (Fig. 10). Zero gain (or loss) in terms of expected information content in either dimension means that the pattern of that journal has been stable over the period under consideration; a positive value means a gain in relative contribution to the respective pattern, and a negative value a loss. Therefore, journals represented by points in the first quadrant gain in importance in both dimensions ("citing" and "cited"), and journals represented by points in the third quadrant lose, relative to this journal set. Note that these values are multivariate and dynamic, in contrast to the "impact factors" (Garfield 1979) and "influence weights" (Narin 1976) of journals and other such indicators (Todorov and Glänzel 1988), which are time-series points based on static network analysis, i.e., computed for each year separately, and subsequently plotted against time.

I also compared the solutions for factor analysis and multidimensional scaling of two-dimensional arrays as if one were to cut the cube
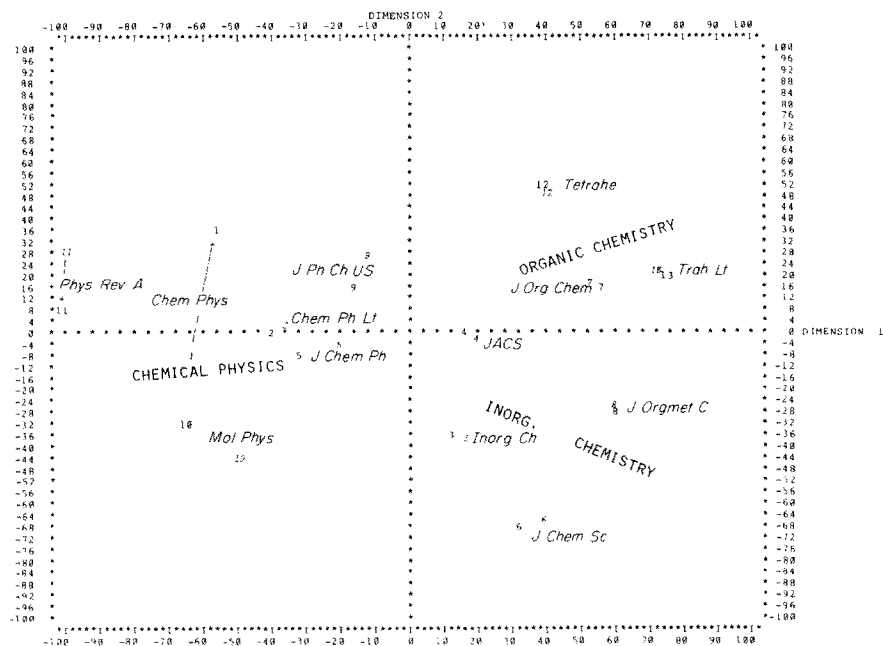
Fig. 11. MINISSA for 13 major chemistry journals – 1981 (in italics) and 1987 solutions superimposed; Dimension 2 plotted against dimension 1. 1981: dhat = 0.110, 1987: dhat = 0.088.

into slices, and then intrapolate the results. Then, I find stable patterns, particularly in the "organic" and "inorganic chemistry" parts of the map, and changes only in eigenvalues for factors. As an example, Fig. 11 shows the superposition of the multidimensional scaling solutions in two dimensions for 1981 (italicized) and 1987: the two pictures can be brought almost to coincide by rotation around the origin. However, from visual inspection of Fig. 10 it is obvious that the major effect in this matrix is the loss of relative contributions, both to the cited and the citing pattern, of *JACS* as a central journal. Note that in Fig. 11 the position of *JACS* is completely stable!

Further analysis of the matrix of values for ΔIs (see above) teaches us that more specialized journals, particularly on the "chemical physics" and "physical chemistry" part of the set, gain weight in the data matrix at the expense of *JACS*. If we focus on the "organic" and "inorganic" subsets only – excluding *JACS* – then the "inorganic" subgroup is "losing" most in relation to this set. However, recall that in the multivariate analysis above, the eigenvalue of "inorganic chem-

istry" was increasing over the same period. I will return to the explanation of this discrepancy later.

As before, we may omit the diagonal values, if we wish to compare the matrices without taking "self-citations" into account (Price 1981; Noma 1982), or we may focus on the diagonal elements in relation to off-diagonal elements if we want to elaborate on the dynamic analysis of cliques as in the previous section. Such analysis of the differences between these two matrices teaches us that the "organic chemistry" group of journals (without *JACS*) has gained influence in the matrix both as a cohesive clique and as a "being cited" unit, although it has been almost neutral with respect to its citing pattern. Additionally, I found an increase in *all* indicators (being cited, citing) for "organic chemistry" when compared with "inorganic chemistry" only.

However, in terms of citing behaviour, the expansion of citations in the "chemical physics" and "physics" part of the matrix is quantitatively more important than the relative increase of "organic chemistry" in relation to only "inorganic chemistry" and *JACS* in the other part of the matrix. More detailed analysis of this part of the matrix, i.e., the interface of chemistry with physics, teaches us that the journals with "chemical physics" in their title have lost coherence among themselves (as a clique), and hence have also become more integrated into the groups of other "physics" journals. The gains in contribution of the latter group to the overall change, both in terms of being cited patterns and citing behaviour, have been most important. [21, 22]

12. Revision of the prediction

In Table 7 we compared not only 1987 with 1981 data, but also with 1984 data. In addition to the study of relations between two matrices,

---

[21] Note that we could have used the one-dimensional data of the margin totals for subgroups, either in the citing or cited dimension, but in that case we would not only have lost the possibility of varying over the dimensions in order to investigate cliques, but we would also have lost information by reducing the dimensionality of the problem. For example, for 1981–1987, we can find in that case only 5.6 mbits of change between 1981–1987 – as against 81.7 mbits using the bivariate analysis – but the distribution over each of the journals happens to have the same rank-order for each journal, nevertheless. However, the major advantage of the direct approach of using the matrices of Δs for all cells is that we gain a more informative picture by direct summation of the contributions of any subset without the need for further assumptions.

[22] I want to note also that the above discussed analysis of citing versus cited patterns ("asymmetry") in terms of information contents of the transforming messages for the rows and columns can be developed as a special case in terms of the Δs in the dynamic analysis, however, only in the case of square matrices.

we may raise the question of whether, and if so to what extent, the prediction of the posterior matrix from the prior matrix is improved or worsened by using in-between data to revise the prediction. The importance of this technique is also that it gives us the basis for a test of whether the in-between data are just to be regarded as a case between prior and posterior cases, or as containing additional information which merits separate analysis. This may provide a particularly useful tool if one wants to reconstruct a line of actions, as, for example, is often the case in science studies (Leydesdorff 1990a).

In general, if we have a prior distribution (p), a posterior distribution (q), and a third distribution which can be regarded as a revision of the prediction (p'), the improvement of the prediction by the revision can be expressed as follows:

$$I(q:p) - I(q:p') = \Sigma q \ \log(q/p) - \Sigma q \ \log(q/p')$$

$$= \Sigma q \ \log(p'/p)$$

This improvement can be positive or negative; in the latter case the prediction is worsened.

I now define as a *critical revision* the case where the revision is not only positive, but where the following inequality also holds:

$$I(q:p) - I(q:p') > I(p':p)$$

or:

$$I(q:p) > I(q:p') + I(p':p)$$

i.e., in terms of expected information value, the pathway of the signal through the revision is more efficient than the direct transfer of the signal from the prior to the posterior event. In this case, one might see the revision as an auxiliary transmitter which boosted the signal from the original sender: there is no need for the receiver to listen to the original sender any more (cf. Leydesdorff 1990a).

If we apply this reasoning to our "1981" data as prior, and our "1987" data as posterior distributions, the "1984" data as a revision of the prediction obviously satisfies this inequality. (By using Table 7: 81.7 is greater than 54.3 + 18.0; and also in the reverse direction: 62.7 > 38.7 + 18.0.) The in-between year "1984" thus boosts the signal

from "1981" to "1987", and vice versa. In other words, when using 1984 data in the prediction of 1987 data, the 1981 data are no longer relevant. This indicates the well-known Markov property: the future behaviour of a *system* is not determined by its previous history. Therefore, it gives us a first indication that the overall development of the data set is not just the sum of the development of its components.

Note that the revision of the prediction is also disaggregatable with respect to subsets of the matrix. In some parts of the system, the revised data set may give us new information indeed, i.e., the subsystem would have changed. In other parts, the revisions may also sum to a number less than zero, indicating a change which has been compensated in other parts of the system.

In our journal–journal citation data, none of ΔIs for rows and columns changed sign for the two periods of 1981–1984 and 1984–1987, respectively. However, not in all such disaggregated cases were the in-between data for 1984 a critical revision of the prediction by the 1981 data in the sense of the inequality discussed above. Consequently, we may conclude that the data for the in-between year (1984) contains some additional information at the journal level about developments between 1981 and 1987, and therefore at that level cannot be described as only an in-between observation, but must be explained in different terms.

Here, we begin to envisage the relations between network analysis and system dynamic modelling.

13. Forecasting

In addition to the dynamic analysis of network data, the use of I as a measure of dynamic development makes it possible to make best forecasts on the basis of any time series of data, also in the multivariate case. (In Leydesdorff (1990b), this option with the dynamic measure I is discussed in more detail, and in comparison to other available statistical techniques like, for example, ARIMA.)

If we have a set of (for example, yearly) trend data for the period from year m to year n, the best prediction of the figure for year $n + 1$ could be derived from a comparison of the shares for years $m + 1, \ldots,$ $n + 1$ in terms of *a posteriori* probabilities with the shares for years $m, \ldots, n$ as *a priori* probabilities.

year →

P_m	P_{m+1}	P_{m+2}				P_{n-2}	P_{n-1}	P_n	

	Q_{m+1}	Q_{m+2}				Q_{n-2}	Q_{n-1}	Q_n	Q_{n+1}

Remember that the best prediction is the one with the lowest I. Since the data for all years are given except for the year $n + 1$, the best prediction for Q_{n+1} has to be based on the addition of $\Delta I = 0$ to the Σ which constitutes the I.

$$\Delta I = Q_{n+1} \log(Q_{n+1}/P_n) = 0 \tag{15}$$

For $Q_{n+1} > 0$, $\Delta I = 0$, only if:

$$\log(Q_{n+1}/P_n) = 0 \tag{16}$$

or:

$$Q_{n+1} = P_n \tag{17}$$

From this equation we can derive the value (F) of the variable for the year $n + 1$ as a function of the value of that same indicator in the previous years of the series: [23]

$$F_{n+1} = \left\{ \frac{(\Sigma_m^n F_i) - F_m}{(\Sigma_m^n F_i) - F_n} \right\} * F_n \tag{18}$$

[23] The probabilities of a distribution are defined in terms of frequencies F_i as follows:

$P_i = F_i / \Sigma_{i=m}^n F_i$

$Q_i = F_i / \Sigma_{i=m+1}^{n+1} F_i$

However, obviously:

$\Sigma_{i=m+1}^{n+1} F_i = (\Sigma_{i=m}^n F_i) - F_m + F_{n+1}$

Since $Q_{n+1} = P_n$:

$$\frac{F_{n+1}}{\Sigma_m^n F_i - F_m + F_{n+1}} = \frac{F_n}{\Sigma_m^n F_i}$$

From which we can calculate F_{n+1}, and then Q_{n+1} also follows.

The coefficient is the sum of the time series minus the value for the first year of the series divided by the same sum minus the value of the last year. The interpretation is simple: with no further information ($\Delta I = 0$), we may assume that the *distribution* of the time-series data remains the same for the next year with the difference of one year only. Note that this assumption is much weaker than the assumption of linearity (or of a higher-order polynomial relationship) implied in regression analysis and time series analysis. Since the measurement is non-parametric, we are not required to make any further assumptions about the character of the trend beyond the assumption that without any additional information, we have no reason to expect change in the distribution over the years except for the noted advancement of one year only.

The extension of the univariate forecast to the multivariate one is straightforward. Following the arguments presented above, we can use the following figures for the multivariate prediction:

year →

$P_{1,m}$	$P_{1,m+1}$	$P_{1,m+2}$			$P_{1,n-2}$	$P_{1,n-1}$	$P_{1,n}$
$P_{2,m}$	$P_{2,m+1}$	$P_{2,m+2}$			$P_{2,n-2}$	$P_{2,n-1}$	$P_{2,n}$
$P_{j,m}$	$P_{j,m+1}$	$P_{j,m+2}$			$P_{j,n-2}$	$P_{j,n-1}$	$P_{j,n}$

$Q_{1,m+1}$	$Q_{1,m+2}$			$Q_{1,n-2}$	$Q_{1,n-1}$	$Q_{1,n}$	$Q_{1,n+1}$
$Q_{2,m+1}$	$Q_{2,m+2}$			$Q_{2,n-2}$	$Q_{2,n-1}$	$Q_{2,n}$	$Q_{2,n+1}$
$Q_{j,m+1}$	$Q_{j,m+2}$			$Q_{j,n-2}$	$Q_{j,n-1}$	$Q_{j,n}$	$Q_{j,n+1}$

As above: for $Q_{j,n+1} > 0$, $\Delta I = 0$, only if $Q_{j,n+1} = P_{j,n}$ and therefore:

$$\frac{F_{j,n+1}}{\sum_{i=m+1}^{n+1}\sum_j F_{ij}} = \frac{F_{jn}}{\sum_{i=m}^{n}\sum_j F_{ij}}$$

However, obviously:

$$\sum_{i=m+1}^{n+1}\sum_j F_{ij} = \sum_j\left\{\sum_{i=m}^{n} F_{ij} - F_{mj} + F_{n+1,j}\right\}$$

Table 9
Comparison of univariate and multivariate predictions

a posteriori	82	83	84	85	86	87
a priori						
81	**33.4**	115.3	96.8	16.0	16.1	23.3
82		**19.1**	187.2	15.6	17.1	24.3
83			**13.8**	32.4	20.7	28.0
84				**8.1**	29.9	30.3
85					**14.9**	44.5
86						**19.8**

and therefore:

$$F_{j,n+1} = F_{jn} * \left\{ \frac{\sum_{i=m}^{n} \sum_j F_{ij} - \sum_j F_{mj} + \sum_j F_{n+1,j}}{\sum_{i=m}^{n} \sum_j F_{ij}} \right\}$$

$$F_{j,n+1} = F_{jn} * \left\{ \frac{\text{Grandsum}_{mn} - \text{Columnsum}_m + \text{Columnsum}_{n+1}}{\text{Grandsum}_{mn}} \right\}$$

Since the column sum for the year $n + 1$ is a normalization factor only, [24] the right-hand factor is a constant, and we may conclude that according to this reasoning the best prediction for next year's distribution would always be the current distribution (n). One may formulate this alternatively: as a *system* the data set has no memory of the values of individual elements in previous states. (This is also called the Markov property in systems theory.)

However, we can now make *two* best forecasts: one on the basis of the values of individual elements of the system, and another for the data set as a system, i.e., on the basis of the last year's distribution. [25] By comparing these forecasts with actual values, we are able to develop a basis for a first test in order to distinguish whether or not the elements develop as coordinates in a system.

[24] The prediction of the value of this sum may, for example, be calculated univariately on the basis of the time series for the column sums by using methods from the previous section.
[25] The Markov property suggests modelling in terms of Markov chains. However, in order to do so, we would first have to prove that the Markov chains are also regular. This is not obvious. However, this would require a separate study of the network data under discussion.

Since three matrices did not seem enough data for such a comparison, I used additionally the corresponding data for all the years in between 1981 and 1987. The results are summarized in Table 9. The columns represent the *a posteriori* distributions, the rows the a priori ones. The expected information values are noted on the basis of univariate predictions. The last figure in each column corresponds to the multivariate prediction, since it is based on the last year only.

It is clear that in none of the cases does the prediction on the basis of the univariate time series improve the prediction on the basis of the previous year only. Therefore, we may now conclude that the data does indeed change over time as a single system. [26]

Note that if we include more previous years in the univariate prediction, the factor by which F_n must be multiplied in order to estimate F_{n+1} becomes closer to one, and therefore the proportions among the various cells in the predicted distribution for the year $n + 1$ also become more similar to the ratios in year n. This agrees with our intuition that long-term variations correspond with systems variations, when these are present.

14. Summary and conclusions

By using the information theoretical measures, I first addressed the question of measuring asymmetry in the matrix, and the respective contributions of rows and columns. Then I raised the question of whether the obvious presence of structure in the matrix can be revealed using these methods. I showed that one can create, firstly, an exact dendrogram in which the length of the leaves represents (in bits of information) the asymmetrical mutual distances among the cases; and secondly, using divisive clustering one can determine the exact number of clusters (groups, factors, etc.) if the "in-between group uncertainty" H_0 has a maximum value.

The analysis of the grouping, which up to this point had addressed one dimension of the matrix only, was subsequently generalized to any

[26] However, from Table 7 we know that the expected information value of the change between 1984 and 1987 is only 18.0 mbits, whereas we have a value of 19.8 mbits for the change between 1986 and 1987. Therefore, in this case only (and for 1987 data only), the data for a previous year would be a better predictor than those for a last year.
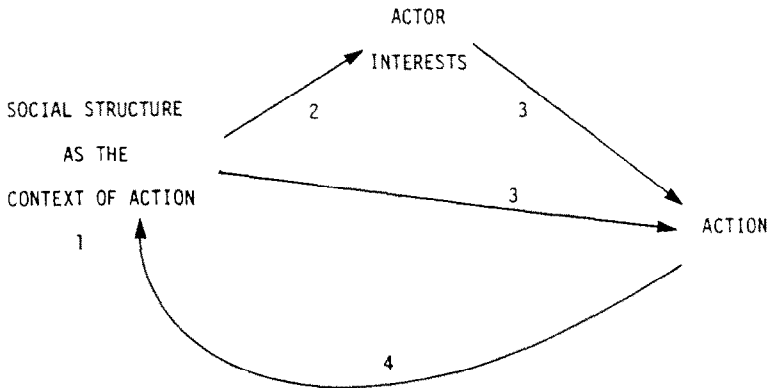
Fig. 12. Components in a structural theory of action (Burt 1982; p. 9).

grouping in terms of the two dimensions of the matrix, including clique analysis as the special case in which one part of the matrix (including diadic relations) is analysed in relation to other parts or to the remainder of it. I showed that the graph-analytic approach ("cohesion" or "relational") and the factor-analytic ("structural equivalence" or "positional") approach can be considered as special cases of a general algorithm for grouping in multivariate arrays.

The next sections of the study addressed the dynamic analysis using the corresponding data for 1981 and 1987. [27] I showed that the results give a view of the data that is completely different from the comparison of results of various forms of multivariate analysis for each year separately.

Although the non-parametric measures of information theory were primarily developed for the analysis of qualitative data, their integrating power makes it possible in principle to address issues of dynamic multivariate systems which cannot easily be analyzed in a single coherent theoretical framework using various, more common statistical tools.

15. Relevance for social network analysis

In his seminal study, Burt (1982: 9) pictured his model as in Fig. 12. Note that the arrow from "action" is the only incoming one for

[27] At the time of this analysis (May 1989), 1987 was the latest available year.
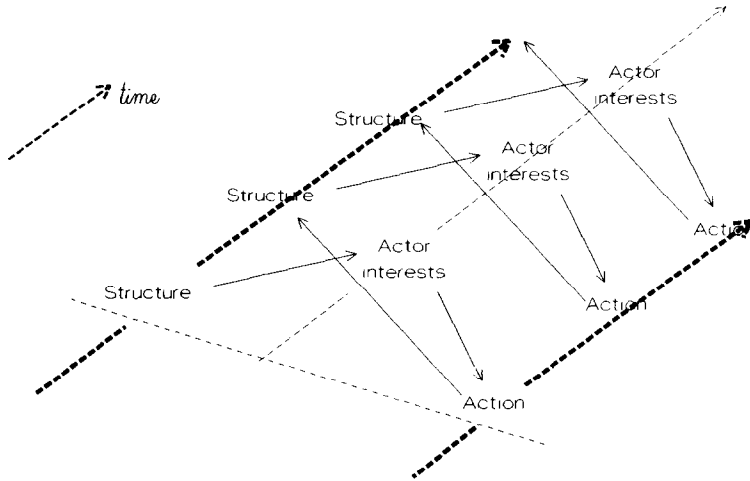
Fig. 13. A dynamic extension of the structuralist model of Fig. 12; structure conditions action; action changes structure.

"structure"; thus, structure is to be explained in terms of (aggregates and patterns of) action. Obviously, with this model one can study only the relations between various aggregates of actions, and therefore in this theory, network analysis was a special case of multivariate and multilevel analysis; but it did not yet touch the core questions concerning the *dynamics* of "structure"/"action" contingencies.

While the loop in Fig. 12 suggests a dynamic feedback, in methodological terms the model is static: it is a loop, and not a spiral! If I extend it to a spiral with *time* as a separate dimension, it takes the shape of Fig. 13. However, this is a rather different model: now, structure has an additional incoming arrow from structure at a previous moment.

The problem of structure and action in sociology is thus one step more fundamental than whether one chooses a positional or a relational approach in a static model. In either case, one still has to relate the results of the multivariate analysis to a dynamic perspective. It goes beyond the scope of this paper to fully elaborate this dynamic structure/action contingency relation (Leydesdorff 1991a and 1991b).

However, Fig. 13 shows that the underdetermination of action by structure can be conceptualized as a static conditional probability distribution, while the reproduction and change of structure by action

can never be a mere product of (static) aggregation, but must be the product of a dynamic process. Structure at t_1 may enable and constrain action at t_1 as a static relation, but *subsequently* that action may not only influence later action (at t_2) but also structure at t_2, etc. [28]

In summary, the constraints of structure upon action and the effects of action upon structure can be expressed as static and dynamic relations among conditional probability distributions. These models can be analyzed in one framework using the various methods from information theory which were introduced in this study. For example, with respect to the arrows in Fig. 13 we can raise the following questions:

(1) The constraining and enabling function of structure (s) – for example, reputation – in relation to action (a) – for example, citations – at any moment in time, may be described as the static conditional relation between the (multivariate) probability distribution of a and s, $H(a) - H(a \mid s)$: how much does knowledge about the structural conditions reduce our uncertainty of the distribution of actor behaviour? (One may wish to include structure at previous moments into the analysis by using $s_{t=n-1}$ instead of $s_{t=n}$.)

(2) Analogously, the effect of (aggregated) action on structure is a dynamic conditional probability relation, which can be expressed by formulas like $I(s_{t=2}: s_{t=1}) - I(s_{(t=2)} \mid a_{(t=1)}: s_{(t=1)} \mid a_{(t=1)})$. Despite the phenomenological complexity, this formula is rather easy in its computation.

At this stage, it seems almost unnecessary to stipulate that every relation is again decomposable down to the level of the individual case, and that at each level we may apply directly the methodologies which were elaborated in previous sections of this article, for example, in order to identify clusters and cliques. Additionally, the explicit introduction of the time axis is not only methodologically fruitful, but also theoretically meaningful. Structure is not just an aggregate of

---

[28] Of course, one may wish to complicate the analysis by adding more arrows to the scheme, such as for example, action at t_1 having an effect on structure at t_3 without influencing the in-between structure at t_2, or action being guided by perceptions of previous structures as intermediating variables, etc.

action; [29] as visible in Fig. 13, the dynamic conceptualization urges us to reformulate their relations as the interaction between *two* self-referential loops (Luhmann 1984; Leydesdorff 1991b).

References

Attneave, F.
    1959   *Applications of Information Theory to Psychology*. New York: Holt, Rinehart and
        Winston.
Burt, R.S.
    1982   *Toward a Structuralist Theory of Action*. New York: Academic Press.
    1987   *Structure, Version 3.2, Technical Report No. TR2*. New York: Columbia University.
Carpenter, M.P. and F. Narin
    1973   "Clustering of scientific journals". *Journal of the American Society of Information
        Science 24*: 425–436.
Cook, T.D. and D.T. Campbell
    1979   *Quasi-experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand
        McNally.
Cozzens, S.E. and L. Leydesdorff
    1988   *Dynamic Structures of Science: Journal – Journal Citation Methods*. Troy/Amsterdam:
        RPI.
Doreian, P.
    1986   "A revised measure of standing of journals in stratified networks". *Scientometrics* 11:
        63–72.
Doreian, P. and T.J. Fararo
    1985   "Structural equivalence in a journal network". *Journal of the American Society for
        Information Science 36*: 28–37.
Everitt, B.
    1974   *Cluster Analysis*. Social Science Research Council. London: Heinemann.
Freeman, L.C.
    1978/79   "Centrality in social networks. Conceptual clarification". *Social Networks 1*: 215–
            239.
Garfield, E.
    1972   "Citations analysis as a tool in journal evaluation". *Science 178*: 471–479.
    1979   *Citation Indexing*. New York: Wiley.
Giddens, A.
    1979   *Central Problems in Social Theory*. London: Macmillan.
Krempel, L.
    1989   "Big science and the state". Paper presented at the *European Conference on Social
        Network Analysis (EC-SNA)*, Groningen.
Krippendorff, K.
    1986   *Information Theory, Structural Models for Qualitative Data*. Beverly Hills: Sage.

---

[29] Cf. Parsons (1954: 337): "It is essential from the point of view of social science to treat the social system as a distinct and independent entity which must be studied and analyzed on its own level, not as a composite resultant of the actions of the component individuals alone."

Langbein, L.I. and A.J. Lichtman
    1978   *Ecological Inference*. Beverly Hills: Sage.
Leydesdorff, L.
    1986   "The development of frames of references". *Scientometrics 9*: 103–125.
    1987   "Various methods for the mapping of science". *Scientometrics 11*: 291–320.
    1990a   "Relation among science indicators or more generally predictions on the basis of
            anything one might know about texts. II. The dynamics of science". *Scientometrics 19*:
            271–296.
    1990b   "The prediction of science indicators using information theory". *Scientometrics 19*:
            297–324.
    1990c   "Irreversibilities in science and technology networks: An empirical and analytical
            approach". *Scientometrics* (forthcoming).
    1991a   "'Structure'/'action' contingencies and the model of parallel computing" (forthcom-
            ing).
    1991b   "A general algorithm for structure/action contingencies" (forthcoming).
Leydesdorff, L. and R. Zaal
    1988   "Co-words and citations: relations between document sets and environments", in: L.
            Egghe and R. Rousseau (eds.) *Informetrics '87 / '88*. Amsterdam: Elsevier.
Luhmann, N.
    1984   *Soziale Systeme, Grundrisz einer allgemeinen Theorie*. Frankfurt a.M.: Suhrkamp.
McClearly, R. and R.A. Hay Jr
    1980   *Applied Time Series Analysis for the Social Sciences*. Beverly Hills: Sage.
Narin, F.
    1976   *Evaluative Bibliometrics*. Cherry Hill, NJ: Computer Horizons, Inc.
Noma, E.
    1982   "An improved method for analyzing square scientometric transaction matrices".
            Scientometrics 4: 297–316.
Parsons, T.
    (1950) 1954   "Psychoanalysis and the theory of social structure". *Psycho-analytic Quarterly*,
                  July 1950; reprinted in: *Essays in Sociological Theory*. New York: Free Press.
Price, D. de Solla
    1965   "Networks of scientific papers". *Science 149*: 510–515.
    1981   "The analysis of square matrices of scientometric transactions". *Scientometrics 3*:
            55–63.
Rice, R.E., C.L. Borgman, D. Bednarski and P.J. Hart
    1989   "Journal-to-journal citation data: Issues of validity and reliability". *Scientometrics 15*:
            257–282.
Shannon, C.H.
    1948   "A mathematical theory of communication". *Bell System Technical Journal 27*: 379–
            423, and 623–656.
Small, H. and E. Sweeney
    1985   "Clustering the *Science Citation Index* using co-citations I. A comparison of methods".
            Scientometrics 7: 391–409.
Theil, H.
    1972   *Statistical Decomposition Analysis*. Amsterdam: North-Holland.
Tijssen, R., J. de Leeuw and A.F.J. van Raan
    1987   "Quasi-correspondence analysis on square scientometric transaction matrices". *Scien-
            tometrics 11*: 347–361.
Tijssen, R.J.W. and J. de Leeuw
    1988   "MVA methods in bibliometric studies". in: A.F.J. Van Raan (ed.) *Handbook of
            Quantitative Science and Technology Studies*. Amsterdam: Elsevier.

Todorov, R. and W. Glänzel
  1988 "Journal citation measures: a concise review". *Journal of Information Science 4*:
       47–56.
Tuma, N.B. and M.T. Hannan
  1984 *Social Dynamics, Models and Methods*. Orlando, FL: Academic Press.
Van den Eeden P. and H.J.M. Huttner
  1982 *Multilevel Research*. Beverly Hills: Sage.