WORKING PAPER 451

AN ALTERNATIVE ALGORITHM FOR Q-ANALYSIS

S.M. MACGILL AND T. SPRINGER

School of Geography
University of Leeds
Leeds LS2 9JT

September 1985

An alternative algorithm for Q-analysis

S M Macgill and T Springer

School of Geography, University of Leeds, Leeds LS2 9JT

September 1985

Introduction

Atkin's (1974) recognition of the power of exploiting the geometrical representation of an array of binary data (more properly, a binary relation between two sets) to yield insights into the data's distinctive features has been pursued by researchers in a wide range of fields - urban design, social networks, medical diagnostics to name three (see Springer and Macgill (1985) for a more comprehensive review). Geometry is a powerful thinking tool, and the geometry of the simplicial complex is a unique and complete representation of the data from which it is constructed. The properties of uniqueness and completeness along with the interpretative power of the geometrical representation have been a source of tremendous appeal to researchers using Atkin's methodology of Q-analysis, and further aspects of the approach have built directly on these properties. Since for large data sets (sets with high cardinality and many interset linkages) it is typically impossible to represent the geometry on paper, various indicators as to what the geometry might approach can be challenged on one of its often (but wrongly) assumed strengths.

In this paper the use of the alternative algorithm is demonstrated in the context of three data sets previously fed through the better known Q-analysis algorithm. The nature of the output generated by the alternative algorithm is compared with that previously generated, and conclusions are drawn on the basis of this comparison. The initial presentation of the alternative algorithm is found in Ho (1982) and a simplified account of it, with initial comment on its relevance to Q-analysis can be found in Macgill (1985). What this paper adds to the last is empirical illustration, verifying certain assertions and further demonstrating the utility of the alternative algorithm.

Example 1 Prey - predator relation Gatrell (1983)

The prey-predator relation (table 1) shows that particular predators (rows) prey on particular food sources (columns). A geometrical representation of this data is given in the simplicial complex in Figure 1, predators represented by named simplices the vertices of which are the prey on which they feed. An important point is, of course, that species have food sources in common, so that the simplices representing particular individual predators can be connected together. Gatrell draws attention to parallels between this representation and ecological concepts arising out given dimensional level according to the following rule: for any two predators in a group at level n, they must either have n prey in common, or there must be further predators in the group through which successive connections of dimension n can be found to "chain" the original predators together at that dimension. Which prey are involved is not explicitly given. Such groupings are searched out for all possible values of n, working down from the highest possible number. Thus the algorithm produces a hierarchical clustering (partitioning) of predators as a basis for interpreting significant features of the prey-predator relation.

The alternative algorithm (Ho 1982, Macgill 1985), the output of which is given in figure 2, implements a procedure designed to identify how many prey each predator consumes, how many prey particular predators consume in common with other predators, and which prey are consumed in each case. Again the "output" is organised in succesive dimensional levels. It can be compared with that given previously (table 2), as follows.

At the highest dimensional level is Rana pipiens; the Q-analysis listing (table 2) indicates that it feeds on four (un-named) species, the alternative listing (figure 2) specifies that these four are Pontania, Disyonicha, Rana pipiens and snails. Galerucella is specified at the next level, the Q-analysis listing indicates that it feeds on three species, the alternative listing specifies these three as Salix discolor, Salix petiolaris and Salix longifolia. At the next level there are three matrix or q-nearness graphs (figure 3). The alternative lattice output is again highly informative about such features. We see straight away the significance of the herbivore grouping: Galerucella and Pontania petiolaridis both feed on Salix petiolaris: Galerucella and Disyonicha quinquevitata both feed on Salix longifolia and Galerucella provides the common link because this feeds on the lot. This group is completely distinct from (has no prey in common with) the other group, the carnivores. The new information about this group is that Redwinged blackbird and Maryland yellowthroat each prey on spiders: Maryland yellowthroat, spiders and Rana pipiens each prey on Disyonicha and Rana pipiens and garter snake each prey on Rana pipiens. The four species appearing for the first time at this level, namely, redwinged blackbird, Pontania petiolaridis, Disyonichia quinquevitata and garter snake, all prey on a single species which, in turn, is also the prey of some other species. This can all be disentangled by following the edges on the lattice back to earlier levels.

To summarise then, up to this point, the alternative (lattice) algorithm seems to yield a more informative product - prey are identified, and the explicit nature of all links is clearly identified. Are there features that the traditional Q-analysis yields which this does not? Two such features might be thought to be "eccentricity" and "structure vector". Both these follow from the considered importance of dimensionality in the

We can observe that such features can be followed directly in the lattice diagram (though with greater precision, as prey are now named), and also that the "connectivity" that "matters" is not most forcefully represented in the q-listing because this does not distinguish between direct and indirect linkages, nor does it name which prey are involved.

A further point is to examine possible effects arising from the removal of an entire species. Removing Disyonicha, generates a new predator-prey relation, a new geometry, a new Q-analysis and a new lattice (see table 3 and figure 4). It is seen that the structure has been yet further fragmented; not only have the niches of certain species narrowed as a result of Disyonicha no longer being available as a food source, but the two bird species find themselves potentially in direct competition for the spider population. Again, more detail is given on the lattice than on the traditiional Q-analysis listing, so that the availability of an algorithm that was not of general availability at the time the Q-analysis was originally undertaken would appear to offer definite advantages.

Example 2 Q-analysis and the structure of friendship networks - Freeman (1980)

Several researchers in social network analysis (Doreian, 1979, 1980, 1982; Freeman,1980; Spooner and Batty, 1981) have found the Q-analysis two sets of results should emerge: (i) the 'before' pattern of friendship traffic should conform to this backcloth, and (ii) new friendship patterns may be created because of the computer conference. The setting in which friendships developed was determined from the relation defined on pairs of the 29 participants and whether or not they were or had been linked together as colleagues, fellow students or teacher-student pairs. This relation produced 19 cliques (simplices) or linking events, each of which drew two or more of the participants into a common setting. This generated a 29x19 incidence matrix (see table 4) defining a backcloth of interpersonal contacts between the 29 individuals and the 19 linking events. This data are represented as a simplicial complex in figure 5 and the listing of q-connected components produced from the Q-analysis algorithm is given in table 5.

Paraphrasing Freeman (p.385) as we look down the listing, we are led to look successively at chains from high to low dimension. All in all, it is a rather meagre structure. The highest order persons, P3 and P13, have been involved in only four linking events and have a dimension of 3, so that the overall complex requires no more than 3 dimensions for its structural description. No links between simplices - no shared faces in the terminology of Q-analysis - are observed at any dimension greater than 1, where p13, p19 and p21 share a pair of events, as do p2 and p28. But this linkage is not impressive - the 12 simplices that appear at q=1 are so loosely linked that they form nine separate components. It is not is reflected in the lattice in only indirect links between these individuals in the lattice, along a non-convex closed path. The structural significance of the q-hole is ambiguous in Freeman's paper, and we do not consider it appropriate to dwell on this aspect in this paper.

In the final part of the analysis, Freeman examined the degree of change in the pattern of friendship ties observed before and after the conference, drawing out the significance of the backcloth structure in interpreting such change. This aspect of the analysis in context of figure 6 rather than table 5 as backcloth is left to the reader.

Example 3 Attitudes to ECT and placebo response (Cowley 1985)

In the field of clinical psychiatry Cowley (1985) used Q-analysis to investigate the relationship between patients' attitudes towards electroconvulsive therapy (ECT) and the placebo response. Data were collected during an ECT study in Leicestershire which was a double-blind controlled trial examining the efficacy of ECT in a group of clinically selected patients. The data comprised a set of named patients on whom attitude information had been collected. A subset of the data is used in this paper, recording whether each of 41 patients scored positively or negatively on each of 8 attitudes, thus generating a 41 x 16 incidence matrix. The output of the Q-analysis algorithm is given in table 6.

respectively, +psychopathology and -psychopathology, for example, may share various other attitudes named in the component with 19 (q+1) other patients. As with the previous examples the chaining effect can make it very difficult to interpret q-connected components sensibly. The difficulty seems particularly acute in the case of opposite attitudes appearing in a single component, though the difficulty is not specific to such a case. For example there are difficulties with the component at q=22 which are discussed below.

By way of comparison, the nature of the output produced from filtering the same data through the lattice algorithm can be examined. Part of the output is given in table 7. Each level refers to the number of patients who possess the various combinations of attitudes. For example, the node +insight at level 31 indicates that 31 people have insight into their medical condition. For convenience and to focus concentration on interpretations of attitudes, the listings of the individual patients concerned for each node is not presented in table 7, though it can be readily obtained from the computer output.

Again we would suggest that the ability of the lattice representation to identify the precise groupings of attitudes (and, if required, the individuals concerned) is an important improvement over the groupings of connected attitudes given by the Q-analysis algorithm because it pinpoints exactly the occurrence of commonly shared attitudes between patients.

treatment. Finally, if we examine lower levels, we find that the traditional Q-analysis listing yields no discrimination whatever between components below level 8 (the same component is simply repeated at all levels). The lattice output reveals additional detail at all levels. For example at level 9 (q=8) instead of a single component in the Q-analysis listing, 19 separate groupings are identified from the alternative algorithm. A further 20 are identified at level 8. Thus there is consdiderable additional discrimination to be revealed where the Q-analysis listing repeats the same single component.

The interpretation of counts of patients displaying particular combinations of characteristics as "traffic" could be superimposed on the lattice listing in an analogous way as would be done for the traditional Q-analysis listing.

Further comment

The utility of Ho's lattice algorithm as a possible alternative to the conventional Q-analysis algorithm for indicating the geometry of a simplicial complex - and therefore facilitating the type of insight sought via the methodology of Q-analysis - has been examined in this paper. It has been found to provide a more detailed, more complete and more accurate representation and to yield somewhat different interpretations. One drawback is a possible swamping effect that might arise over the extent of

Ongoing work by the authors is exploring in greater depth, and in collaboration with specialists in particular fields of application, the utility of the lattice representation as a new research tool.

Acknowledgements. This research was supported by a grant (ref D00232055) from the Economic and Social Research Council. We gratefully acknowledge the co-operation of Peter Cowley in connection with example 3.

References

Atkin R H (974) Mathematical structure in human affairs, Heineman.

Cohen J E (1978) Food webs and niche space, Princeton Universiy Press, Princeton.

Cowley P (1985) An investigation of patients' attitudes to ECT using the methodology of Q-analysis, Psychological Medicine, vol 15, pages 131-139.

Doreian P (1969) A note on the detection of cliques in valued graphs, partial re-specification of Q-analysis. Environment and Planning A, vol 17, pages 1089 - 1109.

Spooner R and Batty M (1981) Networks of urban systems analysts. Environment and Planning B, vol 8, pages 449-475.

Springer T and Macgill S M (1985) Q-analysis through it applications (forthcoming working paper).

TABLE 2

Q-analysis of a prey-predtor relation

Level 4 q = 3 (Rana pipiens)

Level 3 q = 2 (Galerucella) (Rana pipiens)

Level 2 q = 1 (Galerucella) (Maryland yellowthroat) (Spiders, Rana pipiens)

Level 1 q = 0 (Galerucella, Pontania, Disyonicha) (Redwinged blackbird, Maryland yellowthroat, spiders Rana pipiens, garter snake)

Structure vector $Q = (\overset{3}{1}\ 2\ \ 3\ \overset{0}{2})$

Simplex	Eccentricity
Galerucella	2.0
Redwinged blackbird	0
Maryland yellowthroat	1.0
Spiders	0
Pontania petiolaridis	0
Disyonicha quinquevitata	0
Rana pipiens	1.0
Garter snake	0

TABLE 4

29 x 19 Incidence matrix of participants and their linking events

	L_1	L_2	L_2																L_{19}
P_1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P_2	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0
P_3	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	1	0	0	0
	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	1
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1
	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
	0	0	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	0
	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
P_{29}	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0

TABLE 6

q	Q	Components*
30	1	(+ insight)
29-28	3	(+ insight(, (+ need), (- know)
27	4	(+ insight), (+ need), (- know), (- fear)
26	5	(+ insight), (+ need), (- know), (+ attitude), (- fear)
25-24	4	(+ insight, + need), (as above)
23	5	(as above), (+ want)
22	3	(+ insight, + need, - fear, + attitude), (+ want), (- know)
21	3	(+ insight, + need, + want, - fear, + attitude), (- know), (+ trust)
20	3	(+ insight, + need, + want, - fear, + attitude, - know), (+ trust), (- psychopathology)
19	2	(+ insight, + need, + want, - fear, + attitude, - know, + trust, - psychopathology), (+ psychopathology)
18-17	2	(+ insight, + need, + want, - fear, + attitude, - know, + trust, - psychopathology, + psychopathology), (- trust)
16	2	(+ insight, + need, + want, - know, + trust, - fear, + attitude, - psychopathology, + psychopathology, - trust), (- want)
15-14	2	(+ insight, + need, + want, - fear, + attitude, - know, + trust, - psychopathology, + psychopathology, - trust), (- want)
13	1	(+ insight, + need, + want, - fear, + attitude, - know, + trust, - psychopathology, + psychopathology, - trust, - want, - attitude)
12-11	2	(as above), (+ fear)
10	2	(+ insight, +need, + want, - fear, + attitude, - know, + trust, - psychopathology, + psychopathology, - trust, - want, - attitude, + fear, + know), (- need)
9	2	(+ insight, + need, - need, - fear, + fear, + attitude, - attitude, + want, - want, + trust, - trust, + psychopathology, - psychopathology, + know, - know), (- insight)
9-0	1	(+ insight, - insight, + need, - need, + fear, - fear, + attitude, - attitude, + want, - want, + trust, - trust, + psychopathology, - psychopathology, + know, - know)

* the first appearance of an attitude is underlined

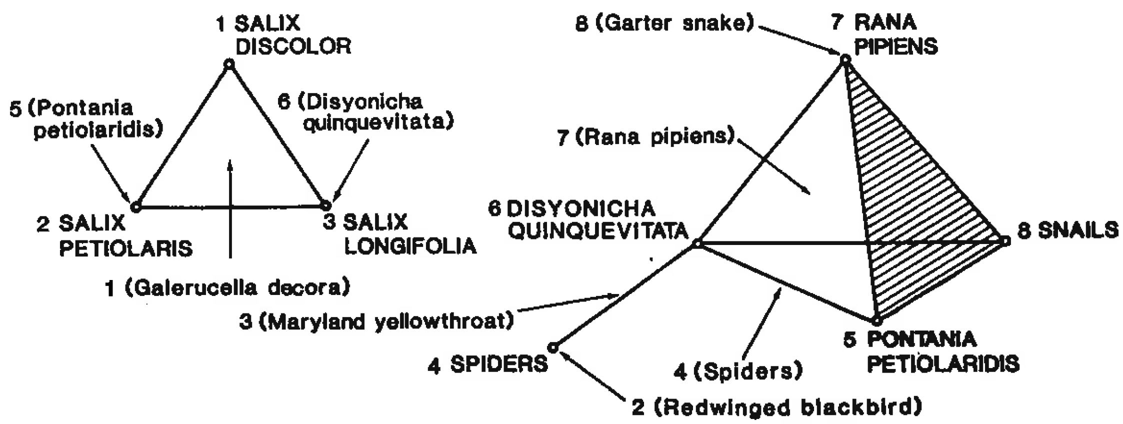

Fig. 1 : The structure of a prey–predator relation.

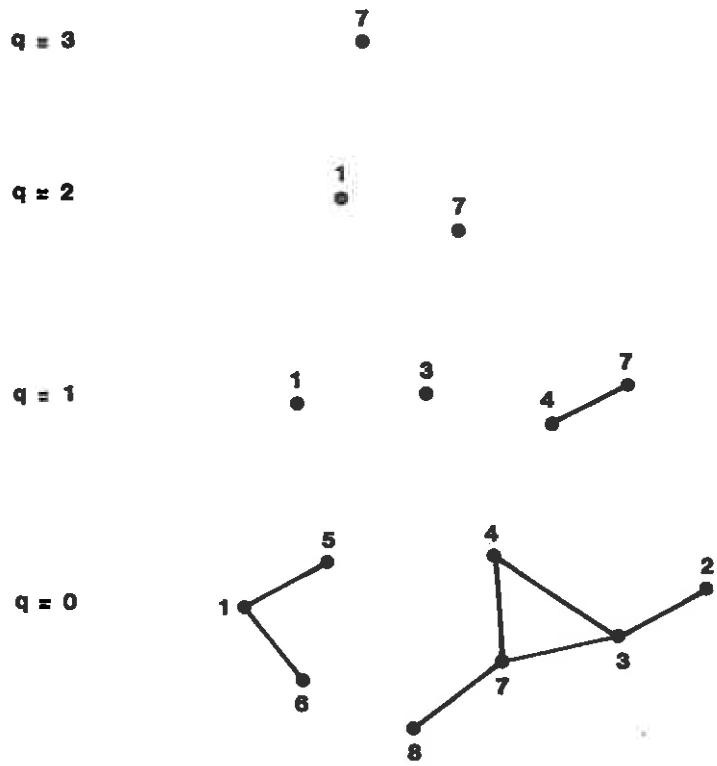

Fig. 3 : q-nearness graphs for prey-predator relation

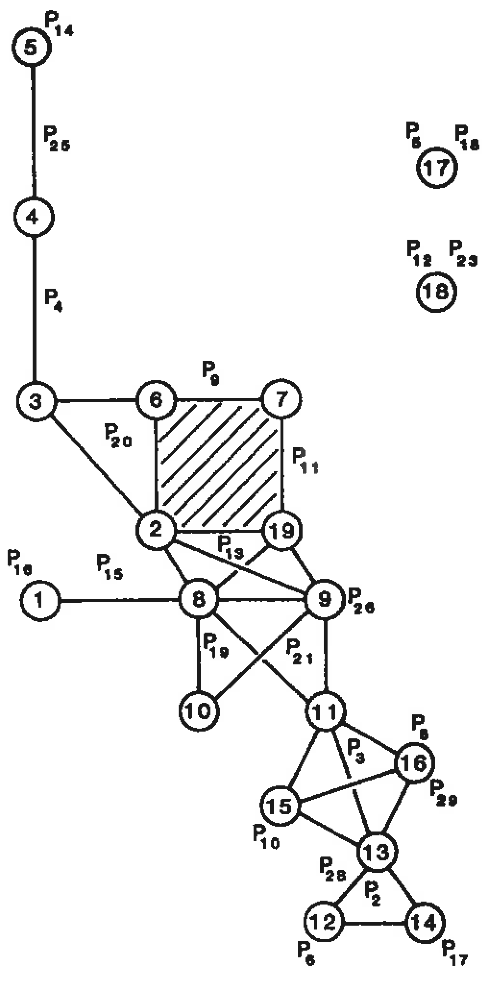

Fig. 5 : Simplicial complex, showing the pattern of links between persons (P_1–P_{29}) through shared linking events (L_1–L_{19})