Using MDS to Infer Relative Status From Dominance Matrices

Stephen P. Borgatti
University of South Carolina

***Techniques** is a regular column devoted to techniques of data construction, management, and analysis. Contributions are appreciated.*

Background. Suppose we have an anti-symmetric matrix in which x_{ij} records the percentage of times that *i* "beat" *j* in a contest, and x_{ji} records the proportion of times that *j* beat *i*. By definition, $x_{ij}+x_{ji}$ equals a constant. We could be talking about sports teams playing games, animals in dominance encounters, or products in consumer taste tests. For example, in an experiment reported by Guilford (1954), respondents were shown pairs of vegetables and asked to choose the one they preferred. The result was a vegetable-by-vegetable matrix, shown in Table 1, in which x_{ij} gives the percentage of respondents who chose vegetable *i* over vegetable *j*. Making the assumption that there exists a common preference ordering across all respondents (i.e. they are all drawn from the same "culture"), the question is whether we can uncover that latent ordering by examining the pattern of frequencies in the table. There are many approaches to this question. In this article, I will give one very simple approach based on ordinary multidimensional scaling.

We begin by considering the process that generates the observed table of frequencies. What would we expect to happen if two vegetables were equally preferred? Since respondents are forced to make a choice, we would expect that in the long run, about half would choose one and half would choose the other. Consequently, x_{ij}-x_{ji} should be about zero. What if two vegetables are at opposite ends of the preference spectrum? Then one of them will always beat the other, and so the absolute value of x_{ij}-x_{ji} will approach its maximum value, which is $x_{ij}+x_{ji}$. This means, then, that the absolute difference $|x_{ij}-x_{ji}|$ may be used as an indicator of the interval between the vegetables on the latent preference scale.

Table 1. Vegetable preferences.

	Tu	Ca	Be	As	Ca	Sp	St	Pe	Co
Turnip		82	77	81	88	89	90	90	93
Cabbage	18		60	72	74	74	81	85	86
Beets	23	40		56	74	68	85	80	82
Asparagus	19	28	44		56	59	68	60	73
Carrot	12	26	26	44		49	57	71	76
Spinach	11	26	32	41	51		63	68	63
St. Bean	10	19	16	32	43	37		53	64
Peas	11	16	20	40	29	32	47		63
Corn	7	14	18	27	24	37	36	37	

Note, however, that if preference truly is a unidimensional scale, we do not have just one estimate of each interval. For example, consider the ordering of objects (labeled "A" through "E") in the line below. The interval between items B and C may be initially estimated from $|x_{BC}-x_{CB}|$ but it can also be calculated as the difference between the A-C and A-B intervals, as well as the difference between the B-D and C-D intervals, and so on.

A----------B------C-----------D---------------E

This redundancy can be put to use. For one thing, if there are missing cells in our data matrix, it seems likely that we can still recover the underlying scale. For another, if our data are subject to random error, we can use this implicit system of equations to average out the errors and get a better estimate of the underlying scale positions

In other words, we would like to find a set of points (corresponding to our items) in 1-dimensional euclidean space such that distances between the items correspond as closely as possible to a set of input proximities defined by $|x_{ij}-x_{ji}|$ for all items. This is precisely the kind of problem that multidimensional scaling is designed to solve.

How to do it. The first step is to construct a symmetric difference matrix P from a raw frequency matrix X by setting $p_{ij} = |x_{ij}-x_{ji}|/(x_{ij}+x_{ji})$. Then submit P to a standard multidimensional scaling program such as found in SPSS or UCINET (Borgatti *et al*. 1992). When running the program it is important to remember to specify that a 1-dimensional solution is desired, and that the input matrix P is a distance matrix rather than a similarity matrix. The program will then output the coordinates of the points in 1-space; these coordinates are then interpreted as preference scale scores. For the vegetable data, the scale scores using non-metric scaling are given in Table 2.

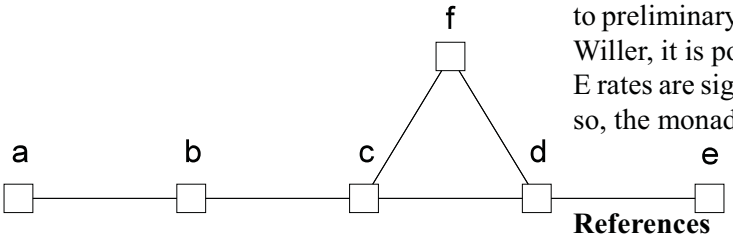
Table 2. Scale scores for vegetables.

Turnip: . . . . . -2.19	Spinach: 29
Cabbage: . . . . . -.78	Str. Bean: 82
Beets: . . . . . -.59	Peas: 87
Asparagus: . . . . -.04	Corn: 1.29
Carrots: 32	

Note that if the stress obtained by the program is excessive, we cannot interpret the 1-dimensional coordinates as preference scores. High stress means that the data are not consistent with a single underlying dimension. In our vegetable example, the stress was high but acceptable, but if not the reason could have been that the seemingly simple construct "preference for vegetables" was not a unidimensional concept, but rather had multiple aspects (e.g., taste, color, texture, cost, nutritional value) which were differentially evoked by different pairs of vegetables.

Exchange networks. Borgatti and Everett (1992:290) deplore that fact that experimental exchange theorists assume power to be a monadic attribute of network positions rather than a dyadic attribute of pairs of nodes. Emerson (1962) had made it a point to note that power is a relation between pairs of actors, not a generalized attribute of the actor as a whole. Yet in some studies (Cook et al 1983; Markovsky et al 1988) the points earned by actors are averaged across all exchanges to yield a single value for each actor. If Emerson was right, this averaging is inappropriate.

We can use the scaling method described above to investigate whether the pairwise exchange rates obtained in the experiments are consistent with an underlying ordering of actors along a unidimensional power continuum, as assumed by modern researchers. Consider the following 6-node graph:



John Skvoretz and David Willer have tested this network in the laboratory.* Preliminary results are given in Table 3.

Table 3. Avg. points received by row actor.

	A	B	C	D	E	F
A		10.31				
B	13.69		11.54			
C		12.46		12.00		12.19
D			12.00		14.20	13.01
E				9.80		
F			11.81	10.99		

The table contains missing values for all pairs of actors that, by experimental design, were not allowed to exchange (e.g., A and C). Due the pattern of missing values (e.g., A and E have only data point), a unique scaling of all points is not possible, although the core nodes C, D and F will be uniquely ordered.

Using metric MDS, the best 1-dimensional scaling has Kruskal stress equal to 0.185, which is high for so few data. This suggests that, contrary to the assumptions of Cook et al. and Markovsky et al., power cannot be reduced to a nodal attribute. However, the conclusion is not unequivocal since it is unclear whether the small differences in observed exchange rates can be attributed to sampling and measurement error. According to preliminary statistical tests by Skvoretz and Willer, it is possible that only the A-B and D-E rates are significantly different from zero. If so, the monadic assumption is supported.

*I am grateful to Skvoretz and Willer for sharing these yet-to-be-published data. Important note: these are preliminary results based on a small sample.

References

Borgatti S.P., M.G. Everett & L.C. Freeman 1992. *UCINET IV 1.00.* Columbia: Analytic Technologies.

Borgatti S.P. & M.G. Everett 1992. "Graph Colorings and Power in Experimental Exchange Networks." *Social Networks* 14:287-308.

Cook K.S., R.M. Emerson, M.R. Gillmore & T. Yamagishi. 1983. "The Distribution of Power in Exchange Networks: Theory and Experimental Results." *AJS* 89:275-305.

Emerson, R.M. 1962. "Power-Dependence Relations." *ASR* 27:31-40.

Guilford J.P. 1954. *Psychometric Methods*. McGraw-Hill.

Markovsky B., D. Willer & T. Patton. 1988. "Power Relations in Exchange Networks." *ASR* 53:220-36.

Skvoretz J. & D. Willer 1993 "Exclusion and Power: A Test of Four Theories of Power Exchange Networks." *ASR* 58:801-818.