THE STRUCTURE OF A RANDOM RELATION WITH AN APPLICATION TO A NOMINATION NETWORK *

Peter J. LENK **

University of Michigan

Stochastic relations frequently occur in the social sciences; however, their statistical properties are not well-known. A random relation can be used to model noise or observational error. This paper characterizes the structure of a random relation by deriving the distribution of the structural vector from Q-analysis and by simulating random relations. We propose a measure of structure and apply these concepts to the nominations of esteemed peers by mathematicians.

1. Introduction

Atkin (1974) developed a powerful technique, called Q-analysis or polyhedral dynamics, to analyze the structure of a relation between two sets. An implicit assumption in the analysis, as used by most researchers, is that the relation is deterministic. However, relational data in the social sciences is frequently stochastic in nature, and care must be used in drawing inferences from a Q-analysis since its statistical properties are not well-known.

Stochastic relations arise in several contexts. They occur when a large population is sampled with the objective of estimating population parameters. In this case, a subset of the entire relation is observed. For example, Granovetter (1976) samples a large network in order to estimate average acquaintance volume. Frank (1971, 1978) extensively discusses sampling from networks. Stochastic relations also occur when an entire relation is observed, but the observed relation is a sample from a population of relations. For instance, Rapoport and Horvath (1961) observe the friendship networks in two junior high schools. They treat the two data sets as replications and draw inferences about the

population of friendship networks in junior high schools. Relational data can also be considered as a realization of a stochastic relation. Gould and Gatrell (1980) analyze a soccer match. They treat the relation as being deterministic and analyze the structure of a particular game by using Q-analysis. However, to draw more general inferences about the structural play of the two teams, the data should be viewed as a realization of a stochastic relation since a re-match would produce a different data set. Observational error is another source of randomness in relational data that should not be ignored.

In order to treat stochastic relations, probability models with parameters that express the essential features of the relation and techniques to estimate these parameters need to be developed. Rapoport (1979) and Holland and Leinhardt (1981) have worked in this direction. However, probability models are not widely used for relations. Typically, researchers combine their expertise of a field with the data to draw inferences and develop insights. Since they are unable to specify their prior knowledge and the likelihood of the data due to the lack of probability models, they are unable to separate their prior information and the data in the posterior analysis as suggested by Bayesian inference. In this situation, the researcher runs the risk of drawing conclusions that are not warranted or supported by the data.

Random relations can be used to model observational error or noise in model building. This paper characterizes the structure of a random relation by deriving the distributions of the structural vector from Q-analysis and by simulating random relations. This characterization is applicable to model building and to the analysis of residuals from fitting a model to the data. It is also important to recognize this structure, for inference based on this sort of data may be primarily the product of the prior information without the researcher being cognizant of it. Researchers should be careful not to fall into the position of mistaking the structure of a random relation as being significantly different from random or, to use a colorful analogy, of seeing patterns in the clouds.

Section 2 provides a characterization of the structural vector from a Q-analysis for a random relation. We propose a measure of random structure that is based on concepts from Q-analysis. This statistic seems to be useful in measuring the deviance of the structure of a relation from the structure of a random relation. The reader who is unfamiliar with Atkin's work may find Gould's (1980) introduction to Q-analysis

insightful. In section 3 we demonstrate the concepts of section 2 with an example of nominations of esteemed peers in two fields of mathematics. The measure of random structure indicates an asymmetry in the structure of giving and receiving nominations.

2. The structural vector for a random relation

A relation between two sets identifies elements of the first set with those of the second. A canonical example in Atkin (1974) has the first set as people in a community and the second set as activities. The relation identifies the people that participate in particular activities. A directed network is also a relation where the elements of the two sets are nodes, and the relation identified which nodes are linked together. The nominations of esteemed peers in section 3 is an example of a directed network where the first set contains the nominators, and the second lists the nominees. The relation identifies a nominator with his or her nominees.

A random relation identifies each pair of elements from the two sets with a constant probability, independently of the other pairs of elements. A random relation can be considered as modeling observational error or random noise. An interpretation of a random relation is that it is non-informative about the phenomenon under consideration. However, a random relation also exhibits a structure. This section characterizes that structure.

Subsection 2.1 introduces the concepts of Q-analysis, in particular the structural vector, that are necessary for the rest of the paper. Subsection 2.2 derives the distribution of the structural vector for a random relation and qualitatively describes its most likely behavior. Since the distributions are complex, subsection 2.3 reports the results of a simulation. The simulation is useful in exploring the operational characteristics of the structural vector and indicates that its distributions can be approximated by standard distributions. Subsection 2.4 proposes a measure of random structure and derives its asymptotic distribution for random relations.

2.1. Definitions and notation

This section is a brief summary of the concepts in Appendix B of Atkin (1974) that are necessary for the derivation of the distribution of the structural vector. We will not give a complete introduction or explore

the philosophical implications of Q-analysis but merely fix ideas and notation.

Let $X = \{xi\}$, $i = 1,\ldots,m$ and $Y = \{yj\}$, $j = 1,\ldots,n$ be two, finite sets. A relation λ from X to Y as a subset of the cartesian product of X and Y. If x is related to y by λ, we write $x\lambda y$. The relation can be represented as an incidence matrix $\Lambda = (\Lambda(ij))$ with m rows and n columns where

$$\Lambda(ij) = \begin{cases} 1 & \text{if } xi\lambda yj \\ 0 & \text{if not} \end{cases}$$

The set of points in Y that are related to a fixed x is denoted by $\sigma(x, Y)$. Similarly, the set of points in X that are related to a fixed y is denoted by $\sigma(y, X)$. These sets are called simplices. If $\sigma(x, Y)$ contains $p + 1$ elements, it is called a p-simplex, is said to have dimension p, and is denoted by $\sigma_p(x, Y)$ where the subscript makes the dimension explicit. If σ_q is a proper subset of σ_p, then q is less than p, and we write $\sigma_q < \sigma_p$ which is read as "σ_q is a face of σ_p". If σ_q and σ_p have the same elements, write $\sigma_q = \sigma_p$.

A collection K of simplices is called a simplical complex if:
1. Each 0-simplex is a member of K;
2. If σ_p is in K and if $\sigma_q < \sigma_p$, then σ_q is in K.
For the relation λ define

$$K(X, Y) = \{\sigma : \sigma \leq \sigma(x, Y) \text{ for some } x \text{ in } X\} \cup \{\{y\} : y \text{ in } Y\},$$

$$K^*(Y, X) = \{\sigma : \sigma \leq \sigma(y, X) \text{ for some } y \text{ in } Y\} \cup \{\{x\} : x \text{ in } X\}.$$

K and K^* are simplicial complices. K^* is known as the dual complex of K. The largest integer N such that a N-simplex is in K is called the dimension of K and written as dim K. The face relation $<$ is a partial ordering on K and K^*.

The structure of a simplicial complex is determined by the manner in which the simplices intersect each other. Suppose that σ_r and σ_s are elements of K. If $\sigma_r \cap \sigma_s = \sigma_q$, then $\sigma_q < \sigma_r$ and $\sigma_q < \sigma_s$. Two simplices σ_r and σ_s are q-connected if there exists a sequence of simplices $\sigma_{a_1},\ldots,\sigma_{a_h}$ such that:

1. $\sigma_{a_1} < \sigma_r$,

2. $\sigma_{a_h} < \sigma_s$,   for $i = 1,\ldots,h-1$,

3. $\sigma_{a_i} \cap \sigma_{a_{i+1}} = \sigma_{b_i}$

and q is the minimum of the dimensions a_1, b_1, b_2, ..., b_{h-1}, a_h. If two simplices are q-connected, then they are also t-connected for $t = 0, ..., q - 1$.

Define a relation γ_q on K or K^* by: two simplices are related if they are q-connected. The relation is an equivalence relation. Then q-connectedness determines the equivalence classes of the quotient set K/γ_q, which partitions all simplices of dimension greater than or equal to q. Consider two simplices in K such that their dimension is greater or equal to q. They belong to the same equivalence class in K/γ_q if and only if they are q-connected. If they are in the same equivalence class at dimension q, then they are in the same one at all dimensions less than q.

The cardinality of K/γ_q is denoted by Qq. Hence Qq is the number of q-connected components or equivalence classes in the quotient set. The vector $Q = (Q(N), Q(N-1), ..., Q(0))$ where N is the dimension of the complex K is called the structural vector, and the procedure of finding Q is called, aptly enough, Q-analysis. For each relation from X to Y there are structural vectors for $K(X, Y)$ and $K^*(Y, X)$. Commonly, the results of a Q-analysis is presented in three columns: the first gives the dimension q, the second gives the number of equivalence classes Qq, and the third lists the equivalence classes.

Before giving an example, we briefly comment on the geometrical interpretation of the simplices and the simplicial complex. A p-simplex has $p + 1$ elements and can be considered as the convex hull of $p + 1$ points in a p-dimensional Euclidian space such that no $q + 1$ of the points lie in a subspace of dimension less than q. In other words, the p-simplex determines a p-dimensional polyhedron with the elements of the simplex at the vertices of the polyhedron. For example, $\sigma_2(x1, Y) = (y1, y2, y4)$ is a triangle in two dimensions, and $\sigma_3(x2, Y) = (y1, y2, y3, y4)$ is a tetrahedron in three dimensions. Also σ_2 is a face of σ_3. Clearly, they are 2-connected A geometrical representation of a simplicial complex can be constructed by joining the simplices at their shared faces.

Consider the relation gives in Table 1. The simplices in $K(X, Y)$ that are determined by the relation are

$$\sigma_2(x1, Y) = (y1, y2, y4),$$

$$\sigma_3(x2, Y) = (y1, y2, y3, y4),$$

$$\sigma_2(x3, Y) = (y1, y2, y5).$$

Table 1
The Q-analysis for a relation Λ

	y1	y2	y3	y4	y5
$\Lambda = x1$	1	1	0	1	0
x2	1	1	1	1	0
x3	1	1	0	0	1

Q-analysis for $K(X, Y)$

$\Lambda \Lambda' - J$ J is a matrix of 1's.

Q-analysis for $K^*(Y, X)$

$\Lambda' \Lambda - J$

	x1	x2	x3
x1	2	2	1
x2		3	1
x3			2

	y1	y2	y3	y4	y5
y1	2	2	0	1	0
y2		2	0	1	0
y3			0	0	-1
y4				1	-1
y5					0

q	Q	components
3	1	$(x2)$
2	2	$(x1, x2)$ $(x3)$
1	1	$(x1, x2, x3)$

q	Q	components
2	1	$(y1, y2)$
1	1	$(y1, y2, y4)$
0	1	$(y1, y2, y3, y4, y5)$

and the simplices in the dual $K^*(Y, X)$ that are determined by the relation are

$$\sigma_2(y1, X) = (x1, x2, x3),$$

$$\sigma_2(y2, X) = (x1, x2, x3),$$

$$\sigma_0(y3, X) = (x2),$$

$$\sigma_1(y4, X) = (x1, x2),$$

$$\sigma_0(y5, X) = (x3).$$

The Q-analysis for $K(X, Y)$ and $K^*(Y, X)$ can be obtained from the matrices $\Lambda \Lambda' - J$ and $\Lambda' \Lambda - J$, where "A'" indicates the transpose of a matrix, and J is a matrix of ones of the appropriate order. The matrices are symmetric, so only the upper triangular elements are listed. The diagonal elements give the dimensions of the simplices. The off-diagonal elements give the dimension of the faces shared by each pair of simplices; a ' -1 ' indicates that the simplices are disjoint, a '0' indicates that they share one vertex, a '1' indicates that they are connected by an edge, and so on. The Q-analysis for the above example is given in Table 1.

2.2. Distribution of the structural vector

Let λ be a relation from X to Y given by the incidence matrix $\Lambda = (\Lambda(ij))$, $i = 1,\ldots,m$ and $j = 1,\ldots,n$. Suppose that $\Lambda(ij)$'s are independent, identically distributed random variables such that

$$P(\Lambda(ij) = 1) = p,$$

$$P(\Lambda(ij) = 0) = 1 - p.$$

Such a relation is called random. In this section we derive the marginal distributions of the structural vector for $K(X, Y)$ and qualitatively describe the most likely behavior of the distributions as the dimension q varies. The results of this section also hold for the structural vector of the dual complex with the obvious change in the parameters.

Define a m by m matrix $A = (A(ij)) = \Lambda\Lambda'$. Then

$$A(ii) = \#\sigma(xi, Y)$$

$$A(ij) = \#[\sigma(xi, Y) \cap \sigma(xj, Y)]$$

where '#' indicates the cardinality of a set. Note that $A(ii) = \dim \sigma(xi, Y) + 1$.

Since the $\Lambda(ij)$ are independent and identically distributed, the $A(ii)$, $i = 1,\ldots,m$ are independent, and the following results are easily obtained. The probability that a y is in the following sets is given by:

y is in $\sigma(x, Y)$ with probability p,

y is in $\sigma(xi, Y) \cap \sigma(xj, Y)$ with probability p^2,

y is in $\sigma(xi, Y) - \sigma(xj, Y)$ with probability $p(1 - p)$.

The joint distribution of $A(ii) - A(ij)$, $A(jj) - A(Aij)$, and $A(ij)$ is multinomial and given by

$$\binom{n}{A(ii) - A(ij), A(jj) - A(ij), A(ij)}$$

$$\times (p(1 - p))^{A(ii) - A(ij)} (p(1 - p))^{A(jj) - A(ii)}$$

$$\times p^{2A(ij)} (1 - p)^{2(n - A(ii) - A(jj) + A(ij))}$$

for $A(ii)$, $A(jj) = 0$, $1,\ldots,n$ and $A(ij) = 0$, $1,\ldots,\min(A(ii), A(jj))$ where the first factor is the multinomial coefficient. The marginal distribution of the $A(i)$'s are independent binomial $B(n, p)$, and the $A(ij)$'s are $B(n, p^2)$. The conditional probability of $A(ij)$ given $A(ii)$ and $A(jj)$ is

$$\frac{\left(\begin{array}{c} n \\ A(ii) - A(ij), A(jj) - A(ij), A(ij) \end{array} \right)}{\left(\begin{array}{c} n \\ A(ii) \end{array} \right) * \left(\begin{array}{c} n \\ A(jj) \end{array} \right)}$$

for $a(ij) = 0$, $1,\ldots,\min(A(ii), A(jj))$.

Next define $Uq = \#\{x: \dim \sigma(x, Y) = q\} = \#\{A(ii) = q + 1\}$ for $q = -1, 0,\ldots,n - 1$. Each of the m elements of X is counted once and only once. The probability that $\dim \sigma(x, Y) = q$ is given by

$$P(A(ii) = q + 1) = \left(\begin{array}{c} n \\ q + 1 \end{array} \right) p^{q+1} (1 - p)^{n - q - 1}.$$

Denote $P(A(ii) = q + 1)$ by p_q. The vector $U = (U(n - 1),\ldots,U(-1))$ is multinomial with parameter m and probabilities $p = (p_{n-1},\ldots,p_{-1})$. Define $Vq = Uq + U(q + 1) + \ldots + U(n - 1)$. Vq is the number of simplices that are of dimension q or more. Since Vq is the sum of the components of a multinomial distribution, it is $B(m, Pq)$ where $Pq = p_q + p_{q+1} + \ldots + p_{n-1}$.

Define the matrix $Aq = (Aq(ij))$ by

$$Aq(ij) = \begin{cases} 1 & \text{if } A(ij) \geq q + 1 \\ 0 & \text{if } A(ij) < q + 1. \end{cases}$$

$Aq(ii)$ indicates those elements whose dimension is greater or equal to q, and $Aq(ij)$, i not equal to j, indicates the pair of elements that share at least a q-dimensional face. If $Aq(ii) = 0$, then $Aq(ij) = 0$ for all j. If $Aq(ij) = 1$, then $Ar(ij) = 1$ for all r less than q. Similarly, if $Aq(ij) = 0$, then $As(ij) = 0$ for all s greater than q. Therefore, the structural vector Q has the Markov property in that

$$P(Qq \mid Q(q + 1),\ldots,Q(n - 1)) = P(Qq \mid Q(q + 1)),$$

and

$$P(Qq \mid Q(-1),\ldots,Q(q - 1)) = P(Qq \mid Q(q - 1)).$$

Either set of these conditional probabilities and the marginal distribution of $Q(-1)$ or $Q(n-1)$ specify the joint distribution of $\mathbf{Q}$. Unfortunately, these probabilities are extremely complicated because of the large number of ways that $Q(q+1)$ equivalence classes at dimension $q+1$ can change to Qq components at dimension q.

A more tractable problem is deriving the marginal distribution of $\mathbf{Q}$. Even these are complex and not easily computed. Fix a dimension q, and suppose that Vq is not zero. There are Vq ones and $m - Vq$ zeros in the main diagonal of Aq. Suppose that $Qq = k$ and that the k equivalence classes are of size $n_1, n_2, \ldots, n_k$. Permute the rows and columns of Aq so that the n_1 elements of the first component occupy the first n_1 rows and columns, the n_2 elements of the second component occupy the $n_1 + 1$ to $n_1 + n_2$ rows and columns, and so on with the last $m - Vq$ rows and columns being zeros.

First we compute the joint probability of $Qq = k, n_1, \ldots, n_k$ given Vq. In order that the jth component has n_j elements, at least $n_j - 1$ ones must appear tin the block above the main diagonal in the area defined by the intersection of $n_1 + \ldots + n_{j-1}$ to $n_1 + \ldots + n_j$ rows and columns. The remaining entries in these segments can be either zeros or ones. The probability that an off-diagonal entry is one is given by the probability that $A(ij)$ is greater or equal to $q+1$ given that $A(ii)$ and $A(jj)$ are greater or equal to $q+1$. In order that there are k components, the remaining entries are zero. If one of these entries were one, then two elements in different equivalence classes would share a q-face, which is a contradiction. The number of these entries is

$$g(n_1, \ldots, n_k) = n_1 n_2 + (n_1 + n_2) n_3 + \ldots + (n_1 + \ldots + n_{k-1}) n_k$$

$$= \sum_{i=2}^{k} \sum_{j=1}^{i-1} n_i n_j.$$

For $Vq > 0$

$$P(Qq = k, n_1, \ldots, n_k | Vq = v) = f(n_1, \ldots, n_k) d^{v-k} (1-d)^{g(n_1, \ldots, n_k)}$$

where

$$d = P(Aq(ij) = 1)$$

$$= P(A(ij) \geq q+1 \mid A(ii) \geq q+1, A(jj) \geq q+1),$$

$$d = \frac{\sum\limits_{r=q+1}^{n} \sum\limits_{s=q+1}^{n} \sum\limits_{t=q+1}^{\min(r,s)} P(A(ii) = r, A(jj) = s, A(ij) = t)}{P_q^2}$$

and

$$f(n_1, \ldots, n_k) = \binom{v}{n_1, \ldots, n_k}.$$

The probability of Qq given Vq is

$$P(Qq = k \mid Vq) = \sum P(Qq = k, n_1, \ldots, n_k \mid Vq) \text{ for } k = 1, \ldots, Vq$$

where the summation is over all combinations of $n_1, \ldots, n_k$ such that the n_j's are greater than zero, and they sum to Vq. If $Vq = 0$, then

$$P(Qq = k \mid Vq = 0) = \begin{cases} 1 & \text{if } k = 0 \\ 0 & \text{otherwise} \end{cases}.$$

Finally,

$$P(Qq = k) = \sum_{v=0}^{m} P(Qq = k \mid Vq = v) * P(Vq = v)$$

where Vq is a binomial $B(m, Pq)$ distribution.

Despite the complexity of the marginal distributions, it is possible to make qualitative observations about the most likely behavior of the structural vector. First we list some facts about the maximum and the minimum of f and g.

1. Fix v and treat $k, n_1, \ldots, n_k$ as variables in f.
    a. f is minimized by $k = 1$ and $n_1 = v$ with $f = 1$.
    b. f is maximized by $k = v$ and $n_1 = \ldots = n_k = 1$ with $f = v!$.
2. Fix v and k. Since f is symmetric, suppose that $n_1 \geq \ldots \geq n_k$. Decrease the ith entry by 1 and increase the jth entry by 1 for $i < j$. Observe that

$$f(n_1 \ldots n_i - 1 \ldots n_j + 1 \ldots n_k) \geq f(n_1 \ldots n_i \ldots n_j \ldots n_k)$$

if and only if $n_i \geq n_j + 1$.

a. f is minimized by $n_1 = v - k + 1$ and $n_2 = \ldots = n_k = 1$.

b. f is maximized when all of the components are approximately equal. Namely,

$$n_1 = [v/k], \quad n_2 = [(v - n_1)/(k - 1)], \quad n_3 = [(v - n_1 - n_2)/(k - 2)],$$

$$\ldots n_k = v - n_1 - n_2 - \ldots - n_{k-1}$$

where '$[a]$' denotes the greatest integer less than a.

3. Fix v and k in g. The reasoning here is identical to (2), and the minimizer of q is the same as in (2a), and the maximizer is the same as in (2b).

4. Fix v and let k be variable in g.

a. By (3) the minimum of g with k and v fixed is

$$g(v - k + 1, 1 \ldots, 1) = (v - k + 1)(k - 1) + \max((k - 1)(k - 2)/2, 0)$$

so g is minimized by $k = 1$ and $n_1 = v$.

Table 2
Q-analysis of $K(X, Y)$ for a random relation card $X = 30$, card $Y = 30$, and $p = 0.25$.

q	Q	components
14	1	1 singleton
13	1	1 singleton
12	1	1 singleton
11	3	3 singletons
10	4	4 singletons
9	5	5 singletons
8	12	12 singletons
7	19	19 singletons
6	20	1 component with 2 elements
		19 singletons
5	21	1 component with 5 elements
		1 component with 2 elements
		19 singletons
4	13	1 component with 14 elements
		1 component with 2 elements
		11 singletons
3	4	1 component with 27 elements
		3 singletons
2	1	1 component with 30 elements
1	1	1 component with 30 elements
0	1	1 component with 30 elements

b. Using (3) the maximum of g with k and v fixed is $[(k-1)v/2k]$. Treating k as variable gives the maximum at $k = v$.

A typical result of a simulation with $m = n = 30$ and $p = 0.25$ appears in Table 2. The reader may find it helpful to compare the simulated structural vector to the following description of its most likely behavior based on the above results.

For small values of q ($q = 0$ to 2 in the simulation) d is large, and $1 - d$ is small. Then $(1 - d)^g$ dominates f, and $(1 - d)^g$ is maximum when g is minimum which occurs when there is a single, large equivalence class by (4a). For large dimensions ($q = 7$ to 14 in the simulation) d is small so that the dominant factor is f. The most likely behavior is to have the Vq elements partitioned into singleton equivalence classes, i.e. $k = Vq$ as seen from (1b). As the dimension decreases $(1 - d)^g$ begins to dominate f. The transition from a large number of singleton components to a single equivalence class with Vq elements is not abrupt but goes through a transitional stage where a relatively few components have more than one element, and the rest are singletons. This behavior is seen from treating k as fixed and minimizing g as in (3). (see $q = 3$ to 6 of the simulation.)

2.3. Simulation

The marginal distributions of the structural vectors are mixtures of a distribution, whose functional form resembles a multinomial distribution, with a binomial distribution. The underlying parameters for the relation from X to Y are $m = $ card X, $n = $ card Y, the probability p that x is related to y, and the dimension q. The distributions of the structural vector Q are complicated functions, that are not amenable to analysis, of these parameters. However, a simulation indicates that the marginal distributions, with a few exceptions, can be approximated by standard distributions.

We studied six random relations. The first three had $m = 20$ and $n = 10$; the next three had $m = 30$ and $n = 10$. For each choice of m and n, p was 0.1, 0.2, and 0.3. A uniform random number on (0,1) was generated for each pair of x and y. If the random number was less than or equal to p, then x and y were related. For each combination of m, n, and p, we generated 1000 random relations, performed a Q-analysis for $K(X, Y)$ and $K^*(Y, X)$, estimated the marginal distributions of Q, and computed the mean and variance of the number of equivalence classes at each q.

The simulation indicates considerable regularity and dependency for the marginal distributions for different m, n, p, and q. The simulated distributions for $K^*(Y, X)$ when $m = 30$, $n = 10$, and $p = 0.2$ are given in figure 1. The dimensions are labeled on the graphs. The patterns in Figure 3 recur for the other simulations. As q increases from 0 to 3, the mode shifts to the right and then to the left as q increases from 4 to 11. The skewness is to the right for $q = 0$ and 1, almost symmetric for $q = 2$, to the left for $q = 3$, to 5, approximately symmetric at $q = 6$, and to the right for $q = 7$ to 11. The curves are suggestive of a wave starting
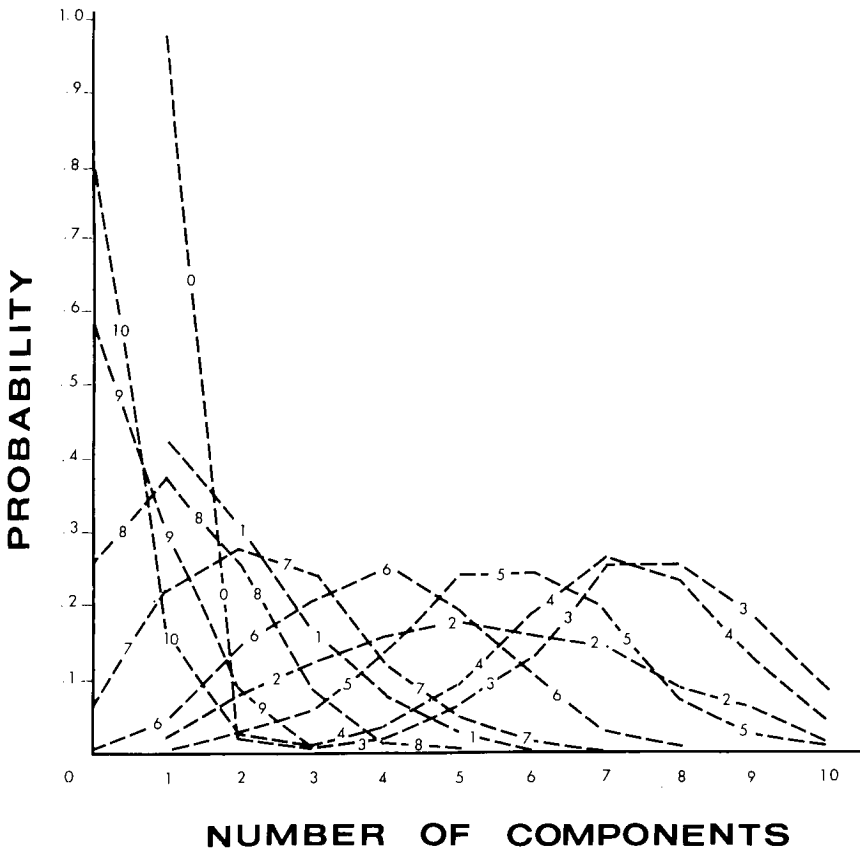

Figure 1. Simulated marginal distributions for the structural vector of $K^*(Y, X)$ for a random relation. Card $X = 30$, Card $Y = 10$, and $p = 0.2$. The numbers on the distributions indicate the corresponding dimension q of the structural vector.

at the left, traveling to the right until it hits a barrier, and returning.

The simulated distributions are approximated by one of three distributions:

1. Binomial $B(n, p)$, $n = 1, 2, \ldots$ and $0 < p < 1$.

$$P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ for } k = 0, 1, \ldots, n,$$

with mean $M = np$ and variance $V = np(1 - p)$.
2. Poisson $P(z)$, $z > 0$.

$$P(X = k) = e^{-z} z^k / k! \text{ for } k = 0, 1, \ldots,$$

with mean $M = z$ and variance $V = z$.
3. Negative binomial $NB(r, p)$, r is real valued, $0 < p < 1$.

$$P(X = k) = \binom{r + k - 1}{k} p^r (1 - p)^k \text{ for } k = 0, 1, \ldots,$$

with mean $M = r(1 - p)/p$ and variance $V = r(1 - p)/p^2$.

Denote the apparent beginning of the simulated distribution by $\min(k)$. Usually it is zero, but occasionally it is greater than zero. For example, when q is small, say 0 or 1, the distribution is 'L' shaped starting at 1. Then $\min(k) = 1$. However, $\min(k)$ is not always the first positive frequency of the simulation. When the distribution is skewed to the left, the first positive frequency in the simulation is greater than zero, but it appears that the distribution tails off to zero, in which case $\min(k) = 0$. A better approximation is obtained by fitting $X + \min(k)$. Denote the mean and variance of the simulated distributions by M and V. If $M - \min(k) > V$, fit $B(n, p)$ by estimating

$$p = (M - \min(k) - V)/(M - \min(k)),$$

$$n = (M - \min(k))/p.$$

Since n has to be an integer, round the above estimate of n to the nearest integer, and re-estimate p by

$$p = (M - \min(k))/n.$$

Use these new estimates in fitting the binomial. If $M - \min(k) < V$, fit $NB(r, p)$ by estimating

$$p = (M - \min(k))/V,$$

$$r = p^2 V/(1 - p).$$

If $M - \min(k)$ is approximately equal to V, fit $P(z)$ by estimating

$$z = M - \min(k).$$

Note that the Poisson distribution is an approximation of the binomial when n is large, p is small, and $np = z$.

Table 3 reports the mean, variance, $\min(k)$, and the fitted distribution for the various simulations. The chi-squared statistic, degrees of freedom, and the p-value for testing the hypothesis that the simulated distribution follows the fitted distribution are listed in the table. Generally, the p-values are high which indicates a good fit. However, a class of distributions cannot be fitted by the above procedure. In the dual K^* the maximum number of equivalence classes is 10. When the simulated distribution has a mode to the right and skewness to the left, sometimes the procedure for fitting the distributions will try to fit a binomial distribution with n greater than 10 and significant mass to the right of 10. The lack of fit in these instances arises from the truncation of the simulated distribution. Examples are

K^*, card $X = 20$, $p = 0.3$, $q = 3$ and 4,

K^*, card $X = 30$, $p = 0.2$, $q = 2$ and 3 (see Figure 1),

K^*, card $X = 30$, $p = 0.3$, $q = 4, 5, 6$.

An anomaly also occurred at K, card $X = 20$, $p = 0.3$, and $q = 2$. The simulated distribution was flat relative to the other distributions and had two modes, which may indicate that 1000 simulations was not sufficient to estimate the distribution accurately.

2.4. A measure of random structure

Let λ be a relation defined on X into Y with incidence matrix Λ and card $X = m$ and card $Y = n$. Define $A = \{A(ij)\}$ to be $\Lambda \Lambda'$. Subtracting

Table 3
Distributions fitted to the simulated structural vectors 1000 random relations were simulated for choices of the parameters.

Dimension	Mean	Variance	Min(k)	Fitted * distribution	df	Chi-square	p-value (>)
$K(x,y)$	Card X = 20	Card Y = 10	p = 0.1				
0	3.1010	2.1430	1	NB(109.44, 0.9804)	5	6.3480	0.25
1	4.5580	2.6393	0	B(11, 0.4144)	6	8.2945	0.1
2	1.3890	1.3131	0	P(0.3890)	4	2.6939	0.5
3	0.2490	0.2432	0	P(0.2490)	1	0.0257	0.5
$K^*(y,x)$	Card X = 20	Card Y = 10	p = 0.1				
0	3.0920	2.1556	1	NB(68.8, 0.9705)	5	6.4714	0.25
1	5.4430	2.2028	0	B(9, 0.4937)	5	5.8180	0.25
2	3.2480	1.9466	0	B(8, 0.4060)	5	3.1421	0.5
3	1.339	1.1012	0	B(8, 0.1674)	3	5.2225	0.1
4	0.451	0.4340	0	B(12, 0.0376)	1	0.4168	0.5
$K(x,y)$	Card X = 20	Card Y = 10	p = 0.2				
0	1.118	0.1422	1	NB(0.58, 0.8298)	1	0.8344	0.25
1	5.001	5.1660	0	P(5.001)	10	13.2970	0.1
2	5.390	3.3694	0	B(13, 0.3850)	8	6.5723	0.5
3	2.320	1.9836	0	B(16, 0.1450)	4	1.011	0.9
4	0.676	0.6496	0	P(0.676)	3	2.4302	0.25
$K^*(y,x)$	Card X = 20	Card Y = 10	p = 0.2				
0	1.119	0.1450	1	NB(0.54, 0.8207)	1	0.5959	0.25
1	3.500	3.2030	1	NB(8.84, 0.7805)	6	10.7212	0.05
2	6.234	2.2517	0	B(11, 0.5667)	6	15.7300	0.01
3	5.585	2.3010	0	B(9, 0.6206)	5	15.3959	0.005
4	3.644	2.3040	0	B(10, 0.3644)	6	2.1778	0.9
5	1.963	1.6030	0	B(11, 0.1785)	4	2.8714	0.5
6	0.8310	0.7912	0	B(17, 0.0489)	2	2.7105	0.25
7	0.285	0.2901	0	P(0.285)	1	0.7540	0.25

$K(x,y)$	Card $X=20$	Card $Y=10$	$p=0.3$				
1	1.961	1.4410	1	$NB(1.92, 0.6667)$	3	1.5282	0.5
2	5.472	6.2116	1	$NB(40.48, 0.8809)$	11	29.1790	0.0
3	5.521	3.8076	0	$B(18, 0.3067)$	8	8.9494	0.25
4	2.763	2.183	0	$B(13, 0.2125)$	5	4.5411	0.25
5	0.9020	0.8712	0	$P(0.9020)$	4	1.7338	0.75

$K^*(y,x)$	Card $X=20$	Card $Y=10$	$p=0.3$				
1	1.242	0.2447	1	$NB(7.61, 0.9692)$	1	0.2567	0.5
2	2.856	2.812	1	$NB(3.60, 0.6600)$	6	7.3991	0.25
3	5.419	3.8833	0	$B(19, 0.2852)$	6	14.9048	0.01
4	6.456	2.7689	0	$B(11, 0.5869)$	4	8.4125	0.05
5	5.525	2.5319	0	$B(10, 0.5525)$	7	5.0390	0.5
6	3.831	2.3368	0	$B(9, 0.4257)$	7	4.3683	0.5
7	2.221	1.7098	0	$B(10, 0.2210)$	4	5.8718	0.1
8	1.107	1.0147	0	$B(13, 0.0852)$	2	0.8969	0.5
9	0.468	0.4574	0	$P(0.468)$	2	0.1780	0.9

$K(x,y)$	Card $X=30$	Card $Y=10$	$p=0.1$				
0	2.348	1.6086	1	$NB(6.97, 0.8380)$	4	8.0241	0.05
1	6.262	3.6089	0	$B(15, 0.4175)$	8	12.1441	0.1
2	2.076	1.924	0	$P(2.076)$	5	4.1635	0.5
3	0.392	0.3607	0	$B(5, 0.0784)$	1	1.9608	0.1

$K^*(y,x)$	Card $X=30$	Card $Y=10$	$p=0.1$				
0	2.3420	1.5447	1	$NB(7.13, 0.8415)$	4	2.5276	0.5
1	6.663	2.2317	2	$B(9, 0.5368)$	5	4.0489	0.5
2	5.831	2.1325	0	$B(9, 0.6479)$	5	6.3534	0.25
3	3.503	2.1162	0	$B(9, 0.3892)$	5	5.4562	0.25
4	1.740	1.4860	0	$B(12, 0.1450)$	3	2.6027	0.25
5	0.724	0.6365	0	$B(6, 0.1207)$	1	1.3110	0.25

$K(x,y)$	Card $X=30$	Card $Y=10$	$p=0.2$				
1	4.8010	5.9673	1	$NB(6.67, 0.6380)$	9	13.2488	0.1
2	7.149	4.9018	0	$B(23, 0.3108)$	9	12.2988	0.1
3	3.406	2.7300	0	$B(17, 0.2004)$	6	3.9566	0.5
4	1.005	1.004	0	$P(1.005)$	2	0.7761	0.5

Table 3 (continued)

Dimension	Mean	Variance	Min(k)	Fitted* distribution	df	Chi-square	p-value (>)
$K^*(y,x)$	Card X = 30	Card Y = 10	p = 0.2				
1	1.999	1.3003	1	NB(3.31, 0.7683)	3	1.9054	0.5
2	5.305	4.3322	0	B(30, 0.1768)	6	16.1464	0.01
3	7.501	2.2602	0	B(11, 0.6819)	5	13.5070	0.01
4	7.161	2.0252	0	B(10, 0.7161)	5	3.9339	0.5
5	5.667	2.3624	0	B(10, 0.5667)	5	4.3464	0.5
6	3.869	2.358	0	B(10, 0.3869)	6	5.8792	0.25
7	2.339	1.8559	0	B(11, 0.2126)	4	2.8069	0.5
8	1.251	1.0851	0	B(9, 0.1390)	3	1.3567	0.5
9	0.570	0.5616	0	P(0.5700)	2	0.1280	0.95
10	0.224	0.2300	0	P(0.2240)	1	0.9509	0.25
$K(x,y)$	Card X = 30	Card Y = 10	p = 0.3				
1	1.494	0.6312	1	NB(1.78, 0.7826)	2	3.1781	0.1
2	5.535	7.7445	1	NB(6.41, 0.5856)	11	13.4884	0.25
3	7.364	5.9137	0	B(37, 0.1990)	10	4.2147	0.9
4	4.185	2.9957	0	B(15, 0.2790)	7	2.3023	0.9
5	1.388	1.2708	0	B(16, 0.0868)	3	0.7444	0.75
$K(y,x)$	Card X = 30	Card Y = 10	p = 0.3				
2	1.341	0.3972	1	NB(2.07, 0.8585)	1	1.6549	0.1
3	2.739	2.4355	1	NB(4.34, 0.7140)	5	1.1548	0.9
4	5.378	4.5199	0	B(34, 0.1582)	6	15.0072	0.025
5	7.307	3.236	0	B(13, 0.5621)	6	25.4968	0.0
6	7.741	1.9799	0	B(10, 0.7741)	4	12.9459	0.01
7	6.979	2.0366	0	B(10, 0.6979)	5	4.6415	0.25
8	5.622	2.3935	0	B(10, 0.5622)	5	3.7049	0.5
9	4.098	2.4608	0	B(10, 0.4098)	6	1.555	0.95
10	2.646	2.0869	0	B(13, 0.2035)	5	3.4518	0.5
11	1.582	1.3067	0	B(14, 0.1130)	3	6.0966	0.1
12	0.865	0.8123	0	B(14, 0.0618)	2	0.2287	0.96

* $B(n, p)$ = binomial. $NB(r, p)$ = negative binomial. $P(z)$ = Poisson.

a matrix of ones from A is the input for a Q-analysis of $K(X, Y)$, and

$$A(ii) = \#\sigma(xi, Y),$$

$$A(ij) = \#[\sigma(xi, Y) \cap \sigma(xj, Y)] \text{ for } i \text{ not equal to } j.$$

If $K(X, Y)$ is well structured, then the simplices tend to share high dimensional faces, and $A(ii) - A(ij)$ will tend to be small. On the other hand, if $K(X, Y)$ lacks structure, then the simplices tend to be disconnected or connected only at low dimensions, $A(ij)$ is close to 0, and $A(ii) - A(ij)$ will tend to be large. Therefore, a reasonable measure for the structure of a relationship should be based on $A(ii) - A(ij)$.

The measure of random structure, R, that we are proposing is

$$R = (S - ES)/(\text{Var } S)^{1/2}$$

where

$$S = \sum_{i=1}^{m} \sum_{j=1}^{m} (A(ii) - A(ij))^2,$$

and the expected value ES and the variance Var S are computed under the assumption that the relation is random with probability p, in which case $A(ii) - A(ij)$ is $B(n, p(1-p))$. Then ES is

$$ES = m(m-1)[n(n-1)p^2(1-p)^2 + np(1-p)]$$

and Var S is

$$\text{Var } S = m(m-1)C + m(m-1)(m-2)D$$

where

$$C = -4n(n-1)(2n-3)p^4(1-p)^4 + 4n(n-1)(n-4)p^3(1-p)^3$$
$$+ 2n(3n-4)p^2(1-p)^2 + np(1-p),$$

and

$$D = -8n(n-1)(2n-3)p^4(1-p)^4 + 4n(n-1)(n-4)p^3(1-p)^3$$
$$+ 2n(2n-5)p^2(1-p)^2 + np(1-p).$$

Typically p is unknown and is estimated by the maximum likelihood estimate which is

$$\sum_{i=1}^{m} A(ii)/mn.$$

The distribution of R is complex; however, the following theorem gives an asymptotic result.

Theorem. Let Λ be a random relation form X into Y with probability p, card $X = m$, and card $Y = n$. If m and n go to infinity and p goes to zero so than np converges to a constant and m/n converges to 1, then R converges to a normal distribution with mean 0 and variance 1.

Proof. Write

$$S - ES = S' - ES' + S'' - ES'' + S''' - ES''',$$

where

$$S' = (m-1) \sum_{i=1}^{m} A(ii)^2,$$

$$S'' = -2 \sum_{i \neq j}^{m} A(ii) A(ij),$$

$$S''' = \sum_{i \neq j}^{m} A(ij)^2,$$

and ES', ES'' and ES''' are their expectations under the assumption of being from a random relation. Now Var $S' = (m-1)^2 \, mF$ where

$$F = -2n(n-1)(2n-3)p^4 + 4n(n-1)(n-3)p^3$$

$$+ 4n(6n-7)p^2 + np.$$

Then Var $S'/$Var S goes to 1 as p goes to 0 and m, n tend to infinity. Since the $A(ii)$, $i = 1, \ldots, m$ are independent and identically distributed, the central limit theorem is enforced, and $(S' - ES')/(\text{Var } S')^{1/2}$ converges in distribution to a standard normal. Now Var S'' and Var S'''

are $O(m)$ while Var S is $O(m^3)$. Therefore Var $S''/$Var S and Var $S'''/$Var S converge to 0. This implies that

$$(S'' - ES'')/(Var\ S)^{1/2} \text{ and } (S''' - ES''')/(Var\ S)^{1/2}$$

converge to zero in probability. Hence

$$R = (\text{Var } S'/\text{Var } S)^{1/2}(S' - ES')/(\text{Var } S')^{1/2} + (S'' - ES'')/(\text{Var } S)^{1/2}$$

$$+ (S''' - ES''')/(\text{Var } S)^{1/2}$$

converges to a standard normal in distribution.

Remarks.

(1) If p is unknown, using the maximum likelihood estimate of p does not change the theorem since it is a consistent estimator.

(2) Basing R on $(A(ii) - A(ij))^2$ leads to a more complicated statistic than the one based on $A(ii) - A(ij)$. However, when we use the maximum likelihood estimate of p, the quantity

$$(m - 1)\sum_{i=1}^{m} A(ii) - m(m - 1)np$$

is zero, and the corresponding statistic converges to 0 in probability.

(3) R is related to the concept of "eccentricity" in Q-analysis. The eccentricity of xi in $K(X, Y)$ is defined as

$$A(ii) - \max_{i \neq j}\{A(ij)\}.$$

The distribution of the eccentricity is not obvious since the $A(ij)$'s are not independent.

(4) There is an obvious dual measure of random structure R^* for the dual complex $K^*(Y, X)$.

3. An application: Nominations among mathematicians

Kochen and Blaivas asked editors of selected journals in six fields to nominate researchers whom they esteem. The nominees were, in turn, asked to nominate other researchers so that the nominees became

respondents. This procedure, called snowball sampling by Goodman (1961), was repeated four times. In this manner, the network of esteemed colleagues was traced. for a more detailed description of the first three rounds see Blaivas *et al.* (1981) and Blaivas *et al.* (1982), Crickman *et al.* (1982), Kochen *et al.* (1982), and Lenk (1983).

Two of the fields in the study are topology and differential geometry. These fields are closely related and interact. In fact, some mathematicians work in both areas. The analysis in this section is for all four rounds of nominations for the mathematicians. Subsection 3.1 reports the results of a Q-analysis for the nominations. Subsection 3.2 introduces a graphical technique that synthesizes Q-analysis and multidimensional scaling in a mutually reinforcing manner. The measure of structure is computed in subsection 3.3 and interpreted for the nominations.

3.1. Q-analysis of the nominations

Let X be the set of respondents or nominators and Y be the set of nominees. The card X is 33, and the card Y is 163. Table 4 is the Q-analysis for $K(X, Y)$, which is the complex for the nominators, and Table 5 is that for $K^*(Y, X)$, which is the complex for the nominees. The complices K and K^* have dimensions 27 and 12 respectively. If the relation was random, the maximum likelihood estimate of p is 0.06972. A comparison of the structural vectors of a simulated random relation with the observed relation indicates that the nominations form non-singleton components at a higher dimension, and these components remain distinct to a lower dimension than in a random relation. This effect is more pronounced for K^* than for K.

Consider the complex $K(X, Y)$ for the mathematicians. A component forms around $\sigma_{19}(3)$ and $\sigma_{18}(25)$ at $q = 11$ and another around $\sigma_{27}(2)$ at $q = 8$. These two components merge at $q = 6$. A component consisting of $\sigma_{11}(11)$ and $\sigma_9(29)$ forms at $q = 4$ and remains isolated until $q = 1$. The structural vector indicates that there is a high degree of consensus about esteemed peers among respondents in the components (3, 15, 25, 26) and (2, 22, 31). On the other hand, the high degree of eccentricity and the large number of singleton components of the other respondents is similar to the structure resulting from a random relation.

Next consider the dual complex $K^*(Y, X)$ of nominees. A component forms around $\sigma_{12}(40)$ at $q = 8$, and another forms around $\sigma_{10}(2)$

Table 4
The structural vector for $K(X, Y)$, simplicial complex of the nominators. (Card $X = 33$ and card $Y = 163$)

q	Q	components
27	1	(2)
26	1	(2)
25	2	(2)(15)
24	2	(2)(15)
23	2	(2)(15)
22	2	(2)(15)
21	2	(2)(15)
20	2	(2)(15)
19	3	(2)(3)(15)
18	4	(2)(3)(15)(25)
17	5	(1)(2)(3)(15)(25)
16	5	(1)(2)(3)(15)(25)
15	6	(1)(2)(3)(15)(21)(25)
14	6	(1)(2)(3)(15)(21)(25)
13	6	(1)(2)(3)(15)(21)(25)
12	8	(1)(2)(3)(15)(21)(25)(26)(30)
11	11	(3,25)(1)(2)(11)(15)(21)(22)(26)(30)(31)(32)
10	14	(3,25)(1)(2)(6)(9)(11)(14)(15)(21)(22)(26)(30)(31)(33)
9	19	(3,25,26)(1)(2)(4)(6)(7)(9)(11)(14)(15)(16)(18)(21)(22)(24)(29)(30)(31)(33)
8	18	(2,22,31)(3,15,25,26)(1)(4)(6)(7)(9)(11)(12)(14)(16)(18)(21)(24)(29)(30)(32)(33)
7	20	(2,22,31)(3,14,15,25,26)(1)(4)(6)(7)(9)(11)(12)(13)(16)(18)(21)(24)(27)(28)(29)(30)(32)(33)
6	22	(2,3,14,15,22,25,26)(1)(4)(6)(7)(8)(9)(10)(11)(12)(13)(16)(18)(19)(21)(24)(27)(28)(29)(30)(32)(33)
5	20	(2,3,4,14,15,16,22,25,26,31)(6,33)(1)(5)(7)(8)(9)(10)(11)(12)(13)(18)(19)(21)(24)(27)(28)(29)(30)(32)
4	13	(1,2,3,4,6,12,14,15,16,19,22,24,25,26,28,31,32,33)(11,29)(5)(7)(8)(9)(10)(13)(18)(20)(21)(27)(30)
3	7	(1,2,3,4,5,6,7,10,12,14,15,16,19,20,21,22,23,24,25,26,27,28,31,32,33)(11,29)(8)(9)(13)(18)(30)
2	5	(1,2,3,4,5,6,7,8,9,10,12,14,15,16,19,20,21,22,23,24,25,26,27,28,31,32,33)(11,29)(13)(18)(30)
1	1	all but respondent 17 who nominated only 1 mathematician
0	1	all

and $\sigma_{10}(34)$ at $q = 9$. These two components do not merge until $q = 2$. This behavior indicates the existence of two sub-areas which correspond to topologists and differential geometricians.

In $K(X, Y)$ if xi and xj are in the same component at dimension q, then the probability that xi nominates yk given that xj nominates yk is

Table 5
Structural vector of $K^*(Y, X)$, simplical complex of the nominees. (Card $X = 33$ and card $Y = 163$.)

q	Q	components
12	1	(40)
11	1	(40)
10	3	(2)(34)(40)
9	2	(2,34)(40)
8	2	(2,34)(40,54)
7	5	(2,34,42)(40,54)(14)(15)(43)
6	7	(2,34,42)(14,40,54)(4)(15)(35)(43)(53)
5	5	(2,10,34,35,42,51)
		(4,14,15,40,43,52,54)
		(3)(44)(53)
4	11	(2,10,34,35,42,51)
		(4,14,15,40,43,46,52,54)
		(1)(3)(6)(29)(36)(44)(48)(50)(53)
3	5	(2,10,28,34,35,37,41,42,51)
		(1,3,4,6,14,15,30,36,38,39,40,43,44,45,46,47,48,50,52,54)
		(29)(49)(53)
2	7	(1,2,3,4,6,10,14,15,20,23,28,29,30,34,35,36,37,38,39,40,41,42,43,44,45,46,
		47,48,49,50,51,52,53,54,92,106)(156,161)
		(7)(13)(94)(100)(134)
1	5	1 component with 63 elements
		4 singletons
0	1	all

large and tends to increase as q increases. The large amount of randomness in $K(X, Y)$, except for the two exceptional components, indicates that, in general, this probability is small. In $K^*(Y, X)$ if yi and yj are in the same component at dimension q, then the probability that xk nominates yi given that xk nominates yj is large and increases with increasing q. If they are in different components, this probability tends to be small and decreases with increasing q. The structure of $K^*(Y, X)$ indicates that if yi and yj work in the same sub-area, they tend to be nominated together; whereas, if they work in different sub-areas, they are not. This effect becomes more pronounced at higher dimensions.

3.2. Graphical representation

This author (1983) uses a common technique to map the nominees so that if the respondents collectively view two nominees as being 'close'

Figure 2. Multi-dimensional scaling for the nominators. The nominators are the circled numbers. The numbers on the lines are the dimensions of the faces that are shared by the nominators.

together, they are mapped close together in two dimensions. Let Λ be an incidence matrix for a relation, and define $A = \Lambda\Lambda'$. Treat the columns of A as vectors. To each element of X there corresponds a vector which is called its 'profile'. Compute the correlations between each pair of profiles, and use these correlations as a measure of similarity in a multidimensional scaling program. The program attempts to map X into a low dimensional, usually 2, space so that the similarities are consistently maintained. If two x's have a large positive correlation, they are mapped close together. If they are negatively correlated, they are mapped far apart. In this sense the scaling gives a convenient representation of the correlation matrix of the profiles. A similar procedure can be performed for Y. The connection to Q-analysis is that $\Lambda\Lambda'$ and $\Lambda'\Lambda$ are the input to the scaling procedure and the Q-analysis. The two procedures are consistent with each other and mutually supportive.

Figure 2 is the multi-dimensional scaling of the respondents, and Figure 3 is for the 32 nominees with 4 or more nominations. A scaling
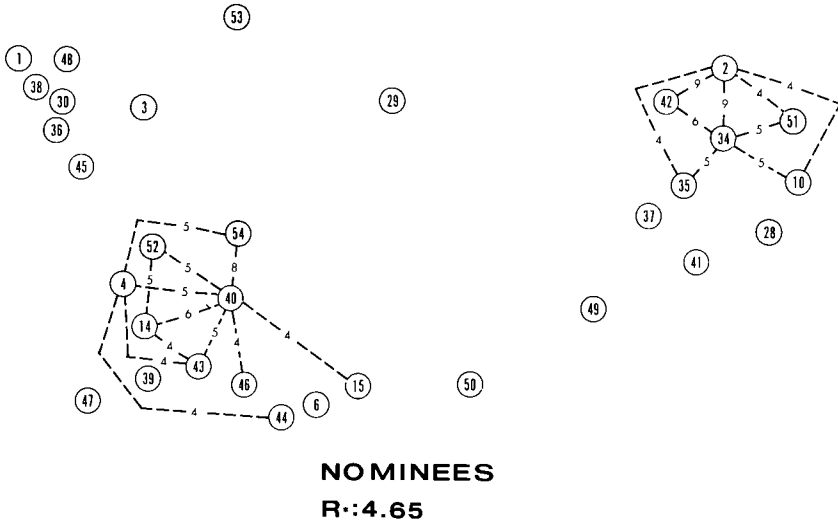
NOMINEES

R·:4.65

Figure 3. Multi-dimensional scaling for the nominees. The nominees are the circled numbers. The numbers on the lines are the dimensions of the faces that are shared by the nominees.

of all 163 nominations becomes too chaotic. Figure 2 should be viewed with the Q-analysis of $K(X, Y)$ in Table 4, and Figure 3 with the analysis of $K^*(Y, X)$ in Table 5. The numbers in the circles represent the respondents or the nominees. If a line connects two mathematicians, then their simplices share a face with the dimension given by the number on the line. For example, in Figure 2 nominators 26 and 15 share a 7-dimensional face. In practical terms, they both gave nominations to the same eight people. Likewise, in Figure 3 the simplices for nominees 40 and 15 share a 4-dimensional face, *i.e.* five respondents nominated both 40 and 15. Note also that if the same number appears in both figures, it identifies the same person. The figures only show connections of dimension 4 or more since including the lower order dimensions produces an illegible graphic.

Consider Table 4 and Figure 2 for $K(X, Y)$ which represents the nominators. The components (3, 14, 15, 25, 26) and (2, 22, 31) at $q = 7$ give $K(X, Y)$ most of its structure. In Figure 2, (3, 15, 25, 26) and the second component are completely connected networks at $q = 7$. The component (11, 29) is isolated. An outstanding feature is the central role of respondent 15. He is a cut point when only simplices of dimension 4 or more are considered. If 15 is removed, the left and right

halves become two separate components. Note also that many of the nominators on the left side are q-connected where 15 is the intermediate node. We infer that 15 probably is important in facilitating the transmission of ideas in the invisible college of topologists and differential geometricians. Inspection of the configuration formed by 15, 6, 33, and 24 seems to indicate the existence of a Q-hole at dimension 4.

The right side (right of 15) of Figure 2 consists of differential geometricians. Eight of the eleven indicated that they work with measure theory. Mathematicians 1, 3, 7, 11, 13, 16, 25, 26, 29, and 30 in the upper left quadrant are topologists, and 4, 6, 9, 14, 15, 19, 21, 24, 27, 32, and 33 in the lower left quadrant are differential geometricians. None of these indicated an interest in measure theory.

Next consider Table 5 and Figure 3 for $K^*(Y, X)$ which represent the nominees. The two large components at $q = 3$ in the table are separated in the scaling. The component (2, 10, 28, 34, 35, 37, 41, 42, 52) consists of differential geometricians, and the other component consists of topologists and mathematicians who were identified as working in both areas. At dimension 4 the two graphs of these components are disconnected and have an extensive internal structure. At $q = 3$, the left and right sides remain separate, but each half becomes well integrated except for 29, 49, and 53 which remain isolated until $q = 2$. Nominees 2 and 34 are central to the network of differential geometricians, and 40 is the center of the topologists. These three researchers also have high dimensional simplices.

3.3. The measure of structure for the nomination relation

Using the maximum likelihood estimate 0.06972 of p, the measure of random structure R for $K(X, Y)$ the respondents, is 0.87. The p-value for testing the null hypothesis that the structure of $K(X, Y)$ corresponds to a random structure is 0.3844, *i.e.* p-value $= P(|Z| > 0.87)$ where Z has a standard normal distribution. Therefore, the structure arising from the process of giving nominations cannot be distinguished from the structure resulting from a random relation when using the statistic R. It is important to note that the conclusion does not say that respondents randomly nominate but that the resulting structure is similar to that of a random relation. The conclusion supports the observations in subsections 3.1 and 3.2.

The measure of random structure R^* for $K^*(Y, X)$, the nominees, is

4.65. The p-values for testing the null hypothesis that the structure of $K^*(Y, X)$ corresponds to a random structure is approximately 0. This result indicates that the manner in which nominations are received by nominees is not random. This conclusion is supported by the remarks in subsections 3.1 and 3.2.

The asymmetry in the structures for giving and receiving nominations needs a more detailed investigation. We infer that a respondent nominates at least two types of researchers. The first type is a highly visible, well established mathematician whose work is applicable to a large number of problem areas in a sub-field. These mathematicians appear in Figure 3. The second group consists of mathematicians with less visibility than the first. His work is relevant to the problem areas of the nominator, but perhaps not to the entire sub-field. The mathematicians in the first group are consistently nominated by the respondents in their sub-field, while the ones in the second group receive only one or two nominations. The number in the first group is small compared to the second. In fact, the distribution of nominations follow Zipf's law (Blaivas *et al.* 1983).

The result of having these two groups vying for nominations is that there is little overlap between two sets of nominations for different nominators. In other words, $A(ii) - A(ij)$ is close to that of a random relation. Hence, $K(X, Y)$ appears to be random, and R is close to zero. For the dual define $B = \Lambda'\Lambda$. If the relation is random, the average number of nominations that a nominee receives is 2.3. If a nominee yi is in the highly visible group, yi is frequently nominated, and $B(ii)$ is much larger than 2.3. Generally, $B(ii) - B(ij)$ is larger than that for a random relation – especially if yj is in a different area since then $B(ij)$ is close to zero. Therefore, the R^* statistic becomes inflated for the members of the first group. If yi is in the second group, they yi receives few nominations, and $B(ii) - B(ij)$ is approximately the same as a random relation. On the whole, the R^* statistic becomes significantly larger than what would be probable if the structure were random.

4. Conclusion

The characterization of the structure of a random relation is a first step in constructing models for stochastic relations. A random relation can be used to model the noise component. Further work is needed to

develop probability models and methods of inference for stochastic relations. The characterization is also an aid in avoiding the error of making inferences that are not supported by the data when the data is a realization from a random relation. The measure of random structure is useful in recognizing random structure.

An application of these concepts to the nomination of esteemed peers by mathematicians indicates that the structure of giving nominations has an unexpectedly high degree of randomness; whereas, the process of receiving nominations is highly structured. This result suggests the hypothesis that two groups are being nominated: a small group of highly visible mathematicians and a large group of less well-known researchers.

References

Atkin, R.
    1974 *Mathematical Structure in Human Affairs*. London: Heinemann.
Blaivas, A., R. Brumbraugh, R. Crickman and M. Kochen
    1982 "Consensuality of peer nominations among scientists". *Knowledge 4*: 252–270.
Blaivas, A., M. Kochen and R. Crickman
    1981 "Geographical patterns of choice among peers". *Social Science Information Studies 1*: 283–295.
Crickman, R., C. Barr and M. Kochen
    1982 "Qualitative interpretation and sampling validity of peer recognition studies". *Social Science Information Studies 2*.
Frand, Ove
    1971 *Statistical Inference in Graphs*. Stockholm: Research Institute of National Defence
    1978 "Sampling and estimation in large social networks". *Social Networks 2*: 155–163.
Goodman, Leo
    1961 "Snowball Sampling". *Annals of Mathematical Statistics 32*: 148–170.
Gould, Peter
    1980 "Q-analysis or a language of structure: An introduction for social scientists, geographer, and planners". *International Journal of Man-Machine Studies 13*: 169–199.
Gould, Peter and Anthony Gatrell
    1980 "A structural analysis of a game: the Liverpool v Manchester United cup final of 1977". *Social Networks 2*: 253–273.
Granovetter, Mark
    1976 "Network sampling: some first steps". *American Journal of Sociology 81*: 1287–1303.
Holland, Paul and Samuel Leinhardt
    1981 "An exponential family of probability distributions for directed graphs". *Journal of the American Statistical Association 76*: 33–50.
Kochen, M., R. Crickman and A. Blaivas
    1982 "Distribution of scientific experts as recognized by peer consensus". *Scientometrics 4*: 45–56.

Lenk, Peter
   1983 "Mappings of fields based on nominations". *Journal of the American Society for Information Science 34*: 115–122.
Rapoport, Anatol
   1979 "A probabilistic approach to networks". *Social Networks 2*: 1–18.
Rapoport, Anatol and William Horvath
   1961 "A study of a large sociogram". *Behavioral Science 6:* 279–291.