PSYCHOMETRIKA—VOL. 14, NO. 1
MARCH, 1949

A METHOD OF MATRIX ANALYSIS OF GROUP STRUCTURE

R. DUNCAN LUCE AND ALBERT D. PERRY
GRADUATE STUDENTS, DEPARTMENT OF MATHEMATICS
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Matrix methods may be applied to the analysis of experimental data concerning group structure when these data indicate relationships which can be depicted by line diagrams such as sociograms. One may introduce two concepts, n-chain and clique, which have simple relationships to the powers of certain matrices. Using them it is possible to determine the group structure by methods which are both faster and more certain than less systematic methods. This paper describes such a matrix method and applies it to the analysis of practical examples. At several points some unsolved problems in this field are indicated.

1. *Introduction*

In a number of branches of the social sciences one encounters problems of the analysis of relationships between the elements of a group. Frequently the results of these investigations may be presented in diagrammatic form as sociograms, organization charts, flow charts, and the like. When the data to be analyzed are such that a diagram of this type may be drawn, the analysis and presentation of the results may be greatly expedited by using matrix algebra. This paper presents some of the results of an investigation of this application of matrices. Initial trials in the determination of group structures indicate that the matrix method is not only faster but also less prone to error than manual investigation.*

The second section of this paper presents certain concepts used in the analysis and associates matrices with the group in question. The third states the results obtained and the fourth gives illustrations of their application. Finally, section five contains a mathematical formulation of the theory and derivation of the results presented in section three.

2. *Definitions*

2.01. The types of relationships which this method will handle are: man a chooses man b as a friend, man a commands man b, a sends messages to b, and so forth. Since in a given problem we concern

*Some of these examples have been worked out by the Research Center for Group Dynamics, Massachusetts Institute of Technology, in conjunction with some of its research.

ourselves with one sort of relation, no confusion arises from replacing the description of the relationship by a symbol $=>$. Thus, instead of "man i chooses man j as a friend," we write "$i => j$." If, on the other hand, man i had not chosen man j, we would have written "$i \neq> j$," using the symbol $\neq>$ to indicate the negation of the relationship denoted by $=>$.

2.02. Situations such as mutual choice of friends or two-way communication would thus be indicated by $i => j$ and $j => i$, or briefly, $i <=> j$. We describe such situations by saying that a *symmetry* exists between i and j.

2.03. When the choice is not mutual, that is $i => j$ or $j => i$ but not both, we say an *antimetry* exists between i and j.

2.04.* The data to be analyzed are presented in a matrix X as follows: the i,j entry (x_{ij}) has the value of 1 if $i => j$ and the value 0 if $i \neq> j$. For convenience we place the main diagonal terms equal to zero, i.e., $x_{ii} = 0$ for all i. This convention, $i \neq> i$, does not restrict the applicability of the method, since there is little significance in such statements as "Jones chooses himself as a friend."

Suppose, for example, that we had a group of four members with the following relationships: $a => b$, $b => a$, $b => d$, $d => b$, $c => a$, $c => b$, $d => a$, and $d => c$. All other possible combinations of $a,b,c,$ and d are related by the symbol $\neq>$. The X matrix associated with this group is:

$$\begin{array}{c} \\ a \\ b \\ c \\ d \end{array} \begin{array}{c} \begin{array}{cccc} a & b & c & d \end{array} \\ \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 \end{bmatrix} \end{array}$$

2.05. From the X matrix we extract a symmetric matrix S having entries s_{ij} determined by $s_{ij} = s_{ji} = 1$ if $x_{ij} = x_{ji} = 1$, and otherwise $s_{ij} = s_{ji} = 0$. All the symmetries in the group are indicated in the matrix S. The S matrix associated with the above X matrix is:

$$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}$$

*In the course of the present work it was brought to our attention that in "A matrix approach to the analysis of sociometric data," *Sociometry*, 1946, 9, 340-347, Elaine Forsyth and Leo Katz have used matrices to represent sociometric relations. They considered a three-valued logic rather than the present two-valued one, and the operations on the matrices are different from the ones discussed in this paper.

To indicate the i,j entry of the matrix X^n, which is the n^{th} power of X, we shall employ the symbol $x_{ij}^{(n)}$. Similarly, the i,j entry of S^n is $s_{ij}^{(n)}$.

2.06. In the group considered above, we had $a => b$, $b => d$, and $d => c$ as three of the relations. If the symbol $=>$ indicates the relationship "sends messages to," it appears that a can send a message to c in three steps, via b and d. We call this three-step path a 3-chain from a to c. Rather than writing out the above sequence of relations, we may omit the symbol $=>$ and simply write the 3-chain as a,b,d,c.

In a group involving more elements one might have the 5-chains: a,e,c,b,d,f and a,d,b,c,d,e. We notice that the first sequence involves five steps between six elements of the group. The second sequence also involves five steps but only five elements of the group, since the element d appears as both the second and fifth member of the sequence. Thus, although these two five-step sequences contain different numbers of elements of the group, they both have six members. Using this concept of membership in a sequence, an n-step sequence has $n+1$ members.

These examples of 3-chains and 5-chains suggest a general definition for a property within the group: an ordered sequence with $n+1$ members, $i, a, b, \cdots, p, q, j$, is an *n-chain* from i to j if and only if

$$i => a, a => b, \cdots, p => q, q => j.$$

2.07. When two n-chains have the same elements in the same order, i.e., the same members, then they are said to be *equal,* and otherwise they are *distinct.* It is important in this definition of equality that it be recognized that both the elements of the group and their order in the sequence are considered. The two chains i,j,k,l,p and i,p,k,j,l are distinct though they contain the same five elements.

2.08. When the same element occurs more than once in an n-chain, the n-chain is said to be *redundant.* (Thus, in a group of m elements any n-chain with n greater than m is redundant). The chains a,b,e,d,b,c and a,c,a,b,d,c,e are, for example, both redundant, for the element b occurs twice in the former and the elements a and c both occur twice in the latter. An example of a non-redundant 5-chain is a,d,p,b,q,j.

2.09. A subset of the group forms a *clique* provided that it consists of three or more members each in the symmetric relation to each other member of the subset, and provided further that there can be found no element outside the subset that is in the symmetric relation to each of the elements of the subset. The application of this definition to the concept of friendship is immediate: it states that a set of more than two people form a clique if they are all mutual friends of one another. In addition, the definition specifies that subsets of cliques are not cliques, so that in a clique of five friends we shall not say that any three form a clique. Although the word "clique" immediately suggests friendship, the definition is useful in the study of other relationships.

2.10. This definition of clique has two possible weaknesses: first, if each element of the group is related by $=>$ to no more than c other elements of the group, then we can detect only cliques with at most $c + 1$ members; and second, there may exist within the group certain tightly knit subgroups which by the omission of a few symmetries fail to satisfy the definition of a clique but which nonetheless would be termed, non-technically, "cliques." It may be possible to alleviate these difficulties by the introduction of so called "n-cliques" which comprise the set of n elements which form two distinct n-chains from each element of the set to itself. This requires that the n-chains be redundant with the only recurring element being the end-point and also that all the relations in the n-chains be symmetric.

This definition means that the four elements $a,b,c,$ and d form a 4-clique if the 4-chains (for example) a,b,c,d,a and $a,d,c,b,a,$ both exist. These by the definition of n-chain require that the relations

$$a <=> b \, , \; b <=> c \, , \; c <=> d \, , \; d <=> a$$

exist, but nothing is said about the relations between a and c, and b and d. The original definition requires, in addition, that

$$a <=> c \quad \text{and} \quad b <=> d$$

for $a,b,c,$ and d to form a clique of four members. Thus we see that the definition of n-clique considers "circles" of symmetries, but it fails to consider the symmetric "cross" terms that exist between the members of the n-clique. These cross terms will be investigated, however, by determining whether any m of these n-elements form an m-clique.

The usefulness of the definition of n-clique can be judged only after experience has been gained in its application. This is not conveniently possible at present, unfortunately, because the problem of the general determination of redundant n-chains has not been solved (see §5.09).

The most general definition of a clique-like structure including antimetries will not be discussed, for it is believed that this will not be amenable to a concise mathematical formulation.

3. *Statement of Results*

3.01. In X^n the entry $x_{ij}^{(n)} = c$ if and only if there are c distinct n-chains from i to j (for proof see §5.04). Thus, if in the fifth power of a matrix of data X we find that the number 9 occurs in the third row of the seventh column, we may conclude that there are 9 distinct 5-chains from element 3 to element 7.

3.02. In X^2 the i^{th} main diagonal entry has the value m if and only if i is in the symmetric relation with m elements of the group (§5.05). Since by the definition of a clique each element i in a clique of t members must be in the symmetric relation to each of the t—1 other elements, it is necessary that $x_{ii}^{(2)} \geqslant t-1$ for i to be in a clique of t members. We may not, however, conclude from the fact that $x_{ii}^{(2)} \geqslant t-1$ that i is necessarily contained in a clique of t members.

3.03. An element i is contained in a clique if and only if the i^{th} entry of the main diagonal of S^3 is positive (§5.06). The main diagonal terms of S^3 will be either 0 or even positive numbers in all cases, and when the value of the entry is 0 the associated element is not in a clique.

3.04. If, in S^3, t entries of the main diagonal have the value $(t-2)(t-1)$ and all other entries of the main diagonal are zero, then these t elements form a clique of t members (§5.08). It also follows from the next statement (§3.05) that if there is only one clique of t members then these t elements will have a main diagonal value in S^3 of $(t-2)(t-1)$. The former statement is, however, the more significant in analysis, for it is the aim to go from the matrix representation to the group structure. There is no difficulty in going from the structure to the matrices.

3.05. Since by statement 3.03 the main diagonal values of S^3 are dependent only on the clique structure of the group, it is to be expected that a formula relating these values and the clique structure is possible. If an element i is contained in m different cliques each having t_ν members, and if there are d_k elements common to the k^{th} clique and all the preceding ones, then

$$s_{ii}^{(3)} = \sum_{\nu=1}^{m} \{ (t_\nu - 2)(t_\nu - 1) - (d_\nu - 2)(d_\nu - 1) \} + 2$$

(§5.07). Thus, if we have three cliques: (5,7,9,10), (1,4,9), and (1,2,5,9,11), then $d_1 = 0$, for there are no preceding cliques; $d_2 = 1$, for only element 9 is common to the second and first cliques; and $d_3 = 3$, for clique three has the elements 1,5, and 9 common with the first two cliques. Substituting $t_1 = 4$, $t_2 = 3$, $t_3 = 5$, $d_1 = 0$, $d_2 = 1$, and $d_3 = 3$ and evaluating the formula for element 9, which is the only one common to all three cliques, we obtain

$$\begin{aligned} s_{99}{}^{(3)} &= [(4-2)(4-1)-(0-2)(0-1)] \\ &+ [(3-2)(3-1)-(1-2)(1-1)] \\ &+ [(5-2)(5-1)-(3-2)(3-1)] + 2 \\ &= 18 . \end{aligned}$$

In the evaluation of this formula it is immaterial how the cliques are numbered initially; however, it is essential once the numbering is chosen that we be consistent.

3.06. The redundant 2-chains of a matrix X are the main diagonal entries of X^2 (§5.09). Thus for a matrix

$$X = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

with the square

$$X^2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 2 & 0 & 0 & 0 & 2 \\ 1 & 2 & 0 & 1 & 0 \\ 1 & 2 & 0 & 1 & 0 \\ 0 & 1 & 1 & 2 & 0 \end{bmatrix},$$

the matrix of redundant 2-chains is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

To obtain the matrix of redundant 3-chains we compute the following matrix, in which the symbol $R^{(2)}$ stands for the matrix of redundant 2-chains:

$$XR^{(2)} + R^{(2)}X - S .$$

Deleting in this sum the main diagonal and replacing it by the main diagonal of X^3 gives the matrix of redundant 3-chains (§5.09). If the main diagonal of $XR^{(2)} + R^{(2)}X - S$ is denoted by $Y^{(3)}$ and the main diagonal of $X^{(3)}$ by $Z^{(3)}$, then let $E^{(3)} = Z^{(3)} - Y^{(3)}$ and thus the matrix of redundant 3-chains, $R^{(3)}$, is given by

$$R^{(3)} = XR^{(2)} + R^{(2)}X + E^{(3)} - S .$$

It has not yet been possible to develop formulas which will give the matrix of redundant n-chains for n larger than 3. What work that has been done in this direction is presented in §5.09.

3.07. The several theorems on cliques give a method that to some extent determines the clique structure independent of the rest of the group structure. It would be desirable to find a simple scheme that determines the clique structure directly. Since a certain amount of knowledge in this direction can be obtained from S^3, it is conjectured that possibly there is a simple formula relating clique structure to the numbers in S^3. As yet no such formula has been developed.

In a consideration of this problem, it was questioned whether certain aspects of the structure would be lost in the multiplication, which, if true, might make the discovery of the desired formula impossible. The following theorem shows that neither the clique structure nor any of the properties of S are lost in the matrix S^3: Any real symmetric matrix has one and only one real symmetric n^{th} root if n is a positive odd integer (§5.12). This theorem is somewhat more general than was required, since it does not restrict the entries in the n^{th} root to 0 and 1, and since it is true for any odd root rather than just the cube root. (In general the real symmetric even roots are not unique.)

This theorem suggests a further problem to be solved: to find a symmetric group structure which will insure the presence of certain prescribed minimum n-chain conditions for odd n. To carry this out it will probably prove necessary to discover a theorem that uses not only the realness and symmetry of the S matrix and its powers, but in addition the fact that only the numbers 1 and 0 may be entries in S.

4. *Examples*

4.01. As the first example, let us compare and analyze the friendship structure in the two following hypothetical groups. The matrices are (where a blank entry indicates a zero):

I

	1	2	3	4	5	6	7	8	9	10
1		1		1			1	1		1
2	1			1	1		1	1		1
3	1							1		
4	1	1					1			1
5						1				
6					1					
7	1	1		1				1		
8	1	1		1			1		1	
9										
10	1	1		1	1					

II

	1	2	3	4	5	6	7	8	9	10
1			1	1	1					1
2						1			1	
3	1			1						
4	1		1			1			1	1
5	1									
6		1					1			
7		1				1				
8					1				1	
9				1	1			1		
10	1		1				1			

The associated S matrices are:

	1	2	3	4	5	6	7	8	9	10
1		1		1			1	1		1
2	1			1			1	1		1
3										
4	1	1					1			1
5						1				
6					1					
7	1	1		1				1		
8	1	1					1			
9										
10	1	1		1						

	1	2	3	4	5	6	7	8	9	10
1			1	1	1					1
2						1				
3	1			1						
4	1		1						1	
5	1									
6		1					1			
7						1				
8									1	
9				1				1		
10	1									

The S^2 matrices are:

I

	1	2	3	4	5	6	7	8	9	10
1	5	4		3			3	2		2
2	4	5		3			3	2		2
3										
4	3	3		4			2	3		2
5					1					
6						1				
7	3	3		2			4	2		3
8	2	2		3			2	3		2
9										
10	2	2		2			3	2		3

II

	1	2	3	4	5	6	7	8	9	10
1	4		1	1					1	
2		1					1			
3	1		2	1	1				1	1
4	1		1	3	1			1		1
5			1	1	1					1
6						2				
7		1					1			
8				1				1		
9	1		1						2	
10			1	1	1					1

Here the differences between the groups are becoming evident. In group I, men 3 and 9 have no mutual friends, since $s_{33}^{(2)} = s_{99}^{(2)} = 0$

(§3.02). Thus, as far as symmetric relationships are concerned, these men are isolated from the group. In the same way we determine that 5 and 6 each have just one symmetric friendship relation ($s_{55}^{(2)} = s_{66}^{(2)} = 1$, §3.02) which we determine to be 5 <=> 6 from the S matrix. The remaining elements in S^2 form a rather dense set of quite large numbers, which means, roughly, a tightly knit group.

In the second group, on the other hand, every man has a non-zero main diagonal in S^2. The men 2, 5, 7, 8, and 10 each have a single mutual friend, which we determine to be: 2 <=> 6, 5 <=> 1, 7 <=> 6, 8 <=> 9, and 10 <=> 1. Then since $s_{66}^{(2)} = 2$ and since we have just cited 6's two mutual friends, 6 need not be considered further. We note that the off-diagonal areas of this S^2 matrix are not so completely filled as group I, indicating that the group is not so tightly bound.

The S^3 matrices indicate the differences in compactness of the structures quite clearly:

I

	1	2	3	4	5	6	7	8	9	10
1	14	15		14			14	12		12
2	15	14		14			14	12		12
3										
4	14	14		10			13	8		10
5						1				
6					1					
7	14	14		13			10	10		8
8	12	12		8			10	6		7
9										
10	12	12		10			8	7		6

II

	1	2	3	4	5	6	7	8	9	10
1	2		5	6	4			1	1	4
2						2				
3	5		2	4	1			1	1	1
4	6		4	2	1				4	1
5	4		1	1					1	
6		2					2			
7						2				
8	1		1						2	
9	1		1	4	1			2		1
10	4		1	1					1	

Since the corresponding main diagonal terms are non-zero, men 1, 2, 4, 7, 8, and 10 of group I are in cliques (§3.03). These, with 3 and 9 which have no symmetries in the group and 5 and 6 which are mutual friends, account for all members of the group. The terms $s_{88}^{(3)} = s_{1010}^{(3)} = 6$ suggest a clique of four members; however, the existence of other main diagonal terms makes it impossible to apply the formula $(t-2)(t-1)$ (§3.04). Investigating in S first the elements 1, 2, and 4 because their columns have the largest values in the tenth row, we find that elements 1, 2, 4, and 10 form a clique of four members. In the eighth row the largest entries are in columns 1, 2, and 7, and an investigation reveals that 1, 2, 7, and 8 form a clique of four men, which then overlaps the first clique by the men 1 and 2. In row four the largest entries are found in columns 1, 2, and 7. We then find that 1, 2, 4, and 7 form a clique of four elements which overlaps the previous two. All the men contained in cliques have been accounted for at least once, and a check either with the formula for main diagonal entries (§3.05) or directly in the S matrix indicates that all the cliques have been discovered. This, coupled with what we discovered in S^2, completely determines the symmetric structure of the first group.

For purposes of qualitative judgment and a guide to carrying out analysis, we note that the first two rows of S^3 present an interesting summary of the clique structure. The entries $s_{12}^{(3)}$ and $s_{21}^{(3)}$ have the largest values, next largest are in columns four and seven, and then finally in columns eight and ten. Men 1 and 2 are contained in all three cliques, 4 and 7 are each contained in two cliques, and finally men 8 and 10 are each in only one clique. This indicates that the magnitude of the off-diagonal terms determines to some extent the amount and structural position of the overlap of cliques.

In group II there are only three elements with non-zero main-diagonal entries, all with the value 2. This fits the formula $(t-2)(t-1)$ with $t = 3$ (§3.04). Thus the men 1, 3, and 4 form a clique of three members. Returning to S^2, we see that there remains one unaccounted symmetry each for men 4 and 9, hence $4 < = > 9$.

In group I, the off-diagonal terms are large in magnitude and are quite dense in the array, with some rows completely empty or with single entries in the S^3 matrix. This indicates a closely knit group with certain men definitely excluded. The S^3 matrix for the second group has fewer entries of a smaller value indicating a less tightly knit structure, but it has no empty rows and only one row with a single entry; that is, it has fewer people than group I who are not accepted by the group or who do not accept it.

A consideration of the matrix $X - S$ will give all the antimetries in the groups and complete the analysis of the structures.

It is clear that this procedure gains strength as the complexity of the problem increases, for the analysis of a twenty-element group is little more difficult than that of a ten-element group.

4.02. The second example is a communication system comprising two-way links between seven stations such as might occur in a telephone or telegraph circuit. The number of channels of a given number of steps (i.e., n-chains in the general theory) between any two points and the minimum number of steps required to complete contact between two stations will be determined. Suppose the matrix of one-step contacts is: