Graph Theory as a Mathematical Model in Social Science*

Frank Harary and Robert Z. Norman

Research Center for Group Dynamics, University of Michigan

1. Introduction

In recent years many psychologists and sociologists have concerned themselves with problems of structures resulting from relations between various entities. These have included the structure of a communication network of people in a psychological group, relations of dominance and submission within a group, influence or power of some people over others and relations between different aspects of a person's psychological field or his personality. The structures have been discussed under the names of “configurations”, "networks”, "patterns", etc., all representing the same abstract concept. This abstract concept has already been intensively studied in mathematics in the discipline known as the theory of graphs. We contend that the theory of graphs is an exceedingly appropriate mathematical model for several domains of psychology and sociology.

Why should a social scientist be interested in separating the formal aspects of the subject from its concrete sociological or psychological setting? Why should mathematical models be used in the social sciences at all ? The advantages are many, as Kaplan (10) has eloquently pointed out. The following brief discussion of mathematical models is of course familiar to anyone acquainted with the foundations of mathematics.

A mathematical model is a set of unproved statements called postulates or axioms, a set of undefined terms called primitives , and the collection of all theorems deducible from these postulates and the laws of logic. It is this definition which motivated Bertrand Russell's famous quotation* "Pure mathematics is the discipline In which we don't know what we're talking about nor whether what we say is true. 0 The power of this abstract approach is that the theorems of a mathematical model give information about each and every interpretation , i.e. concrete system or realization , satisfying its postulates. If the postulates are sufficiently rich , it is possible to deduce many consequences about the concrete system of which one may not have been aware. These consequences are obtainable as simple translations of the theorems of the abstract mathematical model. Of course, the richer the postulates, the smaller the number of concrete realizations which will satisfy the postulates. A further advantage of the abstract approach is the ease of manipulating postulates and definitions instead of actual situations and entities.

There are many other advantages to be gained by using abstract models in the social sciences. Among these are the facilitation of applying methods of deductive logic , the gain in logical self-

---

* The writing of this paper was supported by a grant from the Rockefeller Foundation. The authors wish to thank Dorwin Cartwright, Leon Festinger, John R.P. French Jr., and Abraham Kaplan for their constructive criticisms

* B. Russell, *Principles of Mathematics*, Cambridge: The University Press, 1903. p.4

consistency in the concrete system itself , the creation of a source of new hypotheses which can be experimentally tested, and a framework for testing the appropriateness of old concepts and suggesting new ones. This note is written (a.) to provide an introductory exposition of some of the mathematical concepts and results in graph theory, (b) to furnish references to the mathematical literature , (c) to encourage the use of the relatively uniform language of graph theory by social scientists , and (d) to indicate some possible uses of graph theory as an abstract model.

In the next section we shall present many of the classical definitions of graph theory, essentially following König (13) , together with some illustrations of the suitability of these notions to psychology, one of which occurs in Wertheimer (26). The succeeding section discusses further illustrations from the psychological literature, showing what concepts that are included in the theory of graphs have already been employed. Lewin (16) for example, sought a geometrical system in which to express his notions of “life space” which did not have the familiar and inappropriate properties of Euclidean space. He was greatly impressed with notions of topology presentedin Veblen (24), Lewin’s representation of his “life space” is essentially the same as that of graph theory. We shall show, however, that there are some situations which he would be unable to handle but which can be represented by means of graphs. More recently Bavelas (l) introduced a model for 'group structures" in which he condensed much of Lewin’s work, essentially preserving Lewin’s terminology, and presented the usual mathematical depiction of graphs as collections of points and lines rather than planar regions.

Section 4 will exhibit the intimate interrelationships between graphs, matrices, and relations while Section 5 will present some theorems regarding graphs. We then introduce some new concepts into the theory of graphs and discuss their psychological interpretations. A program of possible future uses of graph theory in social psychology is outlined briefly. For easy reference to the definitions of all the mathematical terms included here, we conclude with a glossary*

2. Some Concepts from Graph Theory

The most comprehensive reference to date on the theory of graphs is König (13). All of the mathematical definitions and theorems discussed in this section are based on this book. In order to provide rationale for the use of graph theory in psychology, we have included along with the exposition, some elementary psychological illustrations. Consider a finite collection of points P1, P2,... ,Pn and the set of all lines joining pairs of these points. A graph of n points consists of the n points together with a subset of this set of lines. The subset may contain none of the lines , all of them, or some intermediate number. If all of the lines are present, the graph is called complete. For example, the complete graph of 4 points may be pictured:

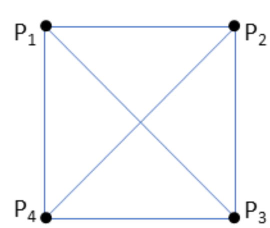


Figure 1.

(It is understood that the intersection of the lines P1P3 and P2P4 is not a point of the graph.)

Two points P and Q of a graph are called *adjacent* if the line PQ is one of the lines of G. Two graphs G and G’, each of *n* points, are called *isomorphic* if there exists a one-to-one correspondence between the points of G and those of G' which preserves adjacency. That is, G and G’ are isomorphic if it is possible to label the points of G by i $P_1,...,P_n$ and those of. G' correspondingly by P'_1>...P'_n> in such a way that a generic line P_iP_j is in G if and only if the correspondingline P'_i P'_j is in G’. Such a 1-1 correspondence is called an *isomorphism*.

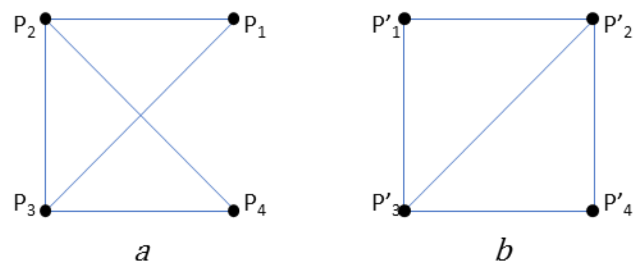


Figure 2.

Two graphs are called *different* if they are not isomorphic. An *automorphism* of a graph is an isomorphism of the graph with itself. Two points of a graph are called similar if there exists an automorphism sending one point into the other. For example, in the graph of Figure 2a, P_1 and P_4 are similar, as are P_2 and P_3. In the graph of Figure 1, all points are similar. Two individuals in a communication network who have indistinguishable positions would also be represented by similar points. The concepts of “role” and "status" have been loosely usedin the literature to refer to position in a group. Similar points of a graph correspond in this context to individuals having the same status.

A *path* from point P to point Q is a collection of lines of the form PA, AB, ..., CQ, where all the points P, A, B, ..., C, Q are different from each other. A *cycle of a graph* is a collection of lines of the form PA, AB, ..., CP, where again all the points are different from each other. The *length* of a path or a cycle is the number of lines in it. To illustrate the concepts of path, cycle, and length, consider the following graph:

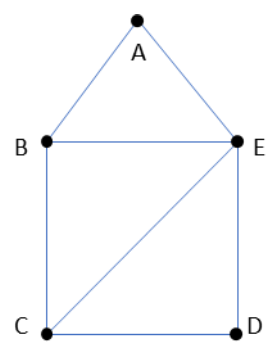


Figure 3.

In this graph, there are 4 paths from A to E, namely: (1) AE; (2) AB, BE; (3) AB, BC, CE; (4) AB, BC, CD, DE; and their lengths are 1,2,3, and 4 respectively. It has three 3-cycles, two 4-cycles and one

5-cycle. A graph is connected if there exists a path between every pair of its points. Thus, all of the graphs illustrated above are connected, while the following two graphs are not.

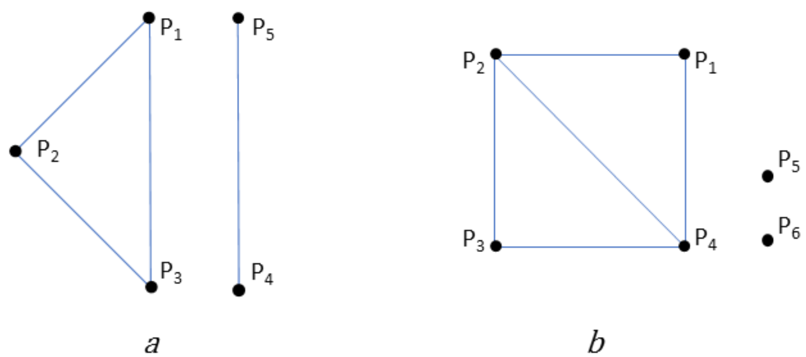


Figure 4.

A graph H is called a *subgraph* of a graph G if the points and lines of H are also in G. A *component* of a graph G is a maximal connected subgraph; that is, it is a connected subgraph which is not a subgraph of any larger connected subgraph of G. A connected graph, therefore, has one component, while the graphs of Figures 4a and 4b have 2 and 3 components respectively.

The *degree* of a point of a graph is the number of lines of the graph on which the point lies. An *endpoint* is a point of degree one. In the following five-point graph, the degree of each point is indicated.

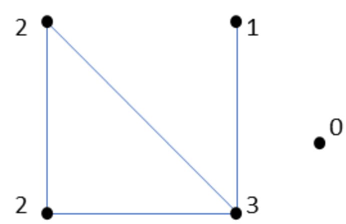


Figure 5.

A connected graph may represent, for example, the communication pattern of a group in which information held by any member may be transmitted to every person in the group. The degree of a point describes the number of people with whom the corresponding group member may communicate directly. In a cycle, information originating with a certain person may return to him with each person transmitting the information exactly once.

A point P will be called an *articulation point* of a graph G if it is possible to divide the points of G into two sets U and V having only P in common, such that every path from a point of U to a point of V includes P. Equivalently, an articulation point of a connected graph may be defined as a point whose removal separates the graph into disjoint components, where by the removal of a point is meant the deletion of the point and all the lines on which it lies.

Let P be an articulation point of a graph G. An articulation component or branch of G at P is a subgraph which contains P, does not have P as an articulation point, and is maximal. A connected graph which has no articulation points is called a *star* (or following Whitney (28), a non-separable graph.) A *bridge* is a line of a connected graph whose removal separates the graph into two components each of which has more than one point. Itis obvious, then, that the endpoints of a bridge are articulation points. Bridges and articulation points may correspond to important and specialized roles in psychological settings.

In the following illustration,

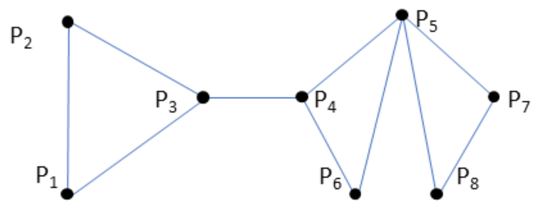


Figure 6.

the line P_3P_4 is a bridge, and of course and P_3 and P_4 are articulation points. Clearly, P5 is also an articulation point. The two articulation components at P3, for example, consist of the points P_3, P_4, P_5, P_6, P_7, and P_8 together with the lines of the graph joining them, and the points and lines of the triangle $P_1P_2P_3$. The second of these two is a star; the first is not.

A graph would be a suitable representation for the communication pattern among delegates to a truce convention attended by two rival groups using different languages. We would represent an interpreter by an articulation point. The interpreter, together with the members of either team, being able to communicate with each other, would form articulation components.

Often, instead of one interpreter, several may be necessary, suggesting a generalization of the notion of articulation point. We might say that if the points of G can be divided into three disjoint sets such that every path from a point of the first to a point of the second contains a point of the third , then the points of the third , together with the lines of G that join them, constitute an articulation subgraph of G.

A clear example of a bridge in the representation of communication among individuals is that of communication between radio operators of two small ships of a naval task force. Inside each ship communication is relatively unrestricted, but any message from an individual of one ship to an individual of the other ship must pass from one radio operator to the other.

A *tree* is a connected graph in which no cycles occur. Several properties of trees follow readily from the definition. The number of points of a tree is one more than the number of lines. If one additional line is added between points of a tree, the resulting graph is no longer a tree. Between each pair of points of a tree there is exactly one path. Conversely, if there is exactly one path between each pair of points of a graph, then the graph must be a tree. The *distance* between two points of a connected graph is the length of any shortest path joining them. Of course, such a shortest path need not be unique. The *diameter* of a connected graph is the maximum of the distances between any two of its points. By the *associated number* of a point of a connected graph, we mean the maximum of the distances from this point to each of the other points. In the following two illustrations of trees, each point is labeled with its associated number.

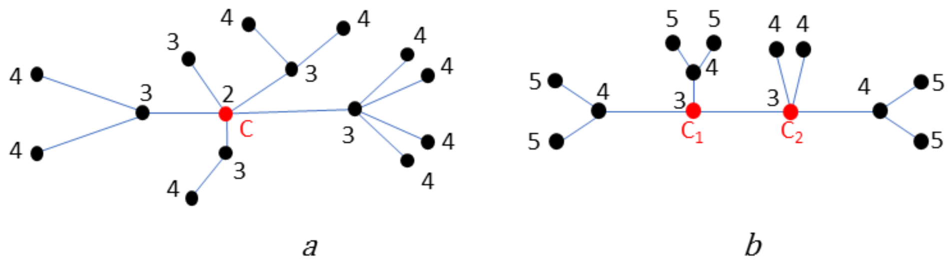


Figure 7.

Consider all points of a tree whose associated number is a minimum. It is proved in König (13) that any tree has either one or two such points. If it has one, that point is called the *center* of the tree; if it has two, they are called the *bicenters*. Thus C is the center of the tree of Fig. 7a; C-ÃÇ, C2 the bicenters of the tree of Fig. 7b.

Similarly, a *central point* of a connected graph is a point of minimal associated number. This generalization of the idea of center to any connected graph was formulated in Lewin (17) and Bavelas (1). A *rooted graph* is a graph in which one of the points is singled out. This distinguished point is called the *root* of the rooted graph, and it may be pictorially differentiated from the other points by drawing a small circle around it. For example,

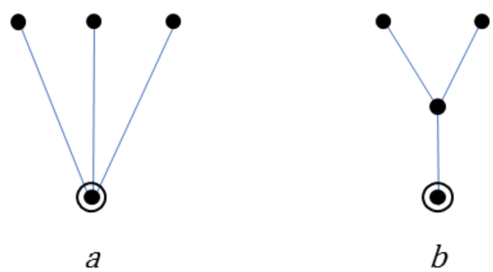
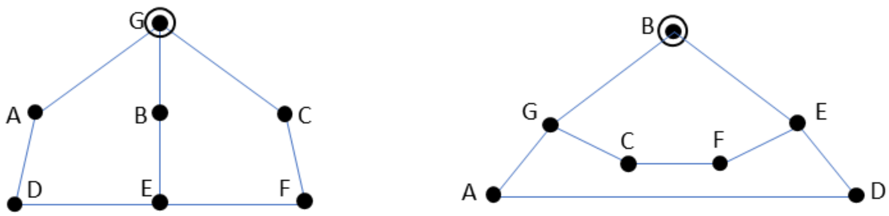


Figure 8.

Figures 8a and 8b illustrate rooted trees. Two rooted graphs are called isomorphic if they are isomorphic as graphs, with the stipulation that the isomorphism send the root point of one into the root of the other. Thus, the above two rooted graphs, although isomorphic as graphs, are not isomorphic as rooted graphs.

The root point of a rooted graph may correspond structurally to the leader of the group, or perhaps, for other purposes, to the person having the smallest attraction to the group, etc. Rooted trees are often used as organization charts with the root representing the head of the organization. The distance from the root then varies in an inverse manner with the status of the position within the organization.

Wertheimer (26) discusses a woman G's [*characterization updated – ed.*] description of her office, and his own attempts to determine the chain of command within the office by questioning her. Although the graph representing the communication pattern is the same for the woman's description as for the chain of command, these representations are quite different as rooted graphs. The woman's egocentric orientation gave her a distinguished position as in Figure 9a, but from the organizational standpoint, individual B held the leading role as in Figure 9b.



a *b*

Figure 9.

An *Euler graph* is one in which each point is of even degree, i.e. has an even number rather than an odd number of lines at each point. Obviously no tree is an Euler graph, for a tree must have at least one end point. A *Hamilton line* of a graph is a cycle passing through all the points of the graph. In the following graph, a Hamilton line is emphasized by a solid line. Clearly no tree has a Hamilton line.

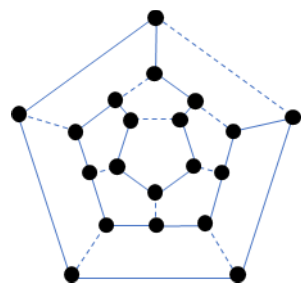


Figure 10.

In the communication structure of an organization, a route for circulating memoranda so that each individual sees them exactly once before they are returned to the sender is a Hamilton line.

One of the well known unsolved problems of the theory of graphs is that of finding a criterion for determining whether or not a given graph has a Hamilton line. It is certainly necessary that the graph have no articulation points, *i.e.* that it be a star. But this condition is not sufficient to characterize all graphs which have Hamilton lines, as the star of Figure 11 shows.

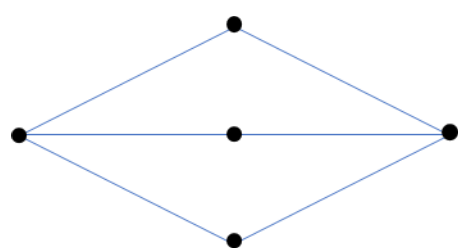


Figure 11.

One approach to this problem is discussed by Harary and Ross (6).

A *directed graph* (or *digraph*) consists of *n* points together with a subset of the set of all directed lines between each pair of points. (A graph [*such as has been discussed up to now – ed.*] is sometimes called an *ordinary graph* to distinguish it from a directed one.) In a directed graph, there may be two directed lines between each pair of points, for one may occur in each direction. Directed graphs may be pictured as follows:

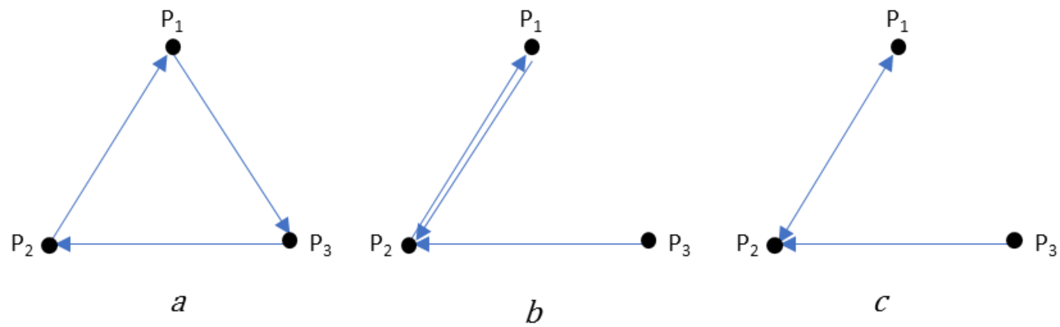


Figure 12.

Here, Figures 12b, 12c are alternate ways of depicting the same directed graph.

An ordinary graph may be regarded as a directed graph by replacing each line PQ of the graph by both of the directed lines between points P and Q. Thus directed graph is a more general structural concept than ordinary graph in that it admits paths in one or two directions rather than two-directional paths only.

A *directed path* from P to Q in a directed graph is a collection of directed lines of the form PA, AB, CQ where the points P, A, B, C, ..., Q are all different from each other. Other concepts of ordinary graphs can also be generalized to directed graphs.

It is possible to utilize the concept of a directed graph in learning experiments involving mazes. If a maze permits re-tracing of paths, then it corresponds to an ordinary graph. However, the use of barriers to prevent re-tracing in a maze is analogous to one-directional line segments in a graph. Such mazes therefore correspond to directed graphs.

3. Graphs as treated in the psychological literature

Probably the first psychologist to attempt to describe psychological phenomena in topological terms was Kurt Lewin (15). As we mentioned in the first section, he was interested in developing a representation of the “life space” of a person in which distance could be defined more suitably than in ordinary Euclidean space. By a shortest path between two regions in this representation, he meant a path which crossed the smallest number of "boundaries” of regions. His representation corresponds to a linear graph in which the regions correspond to points of the graph, and passable "boundaries” between regions correspond to lines joining these points. In the following example, R_1 to R_7 denote regions and P_1 to P_7 denote points of the corresponding graph. Two points of the graph are joined by a line whenever the corresponding regions touch each other.

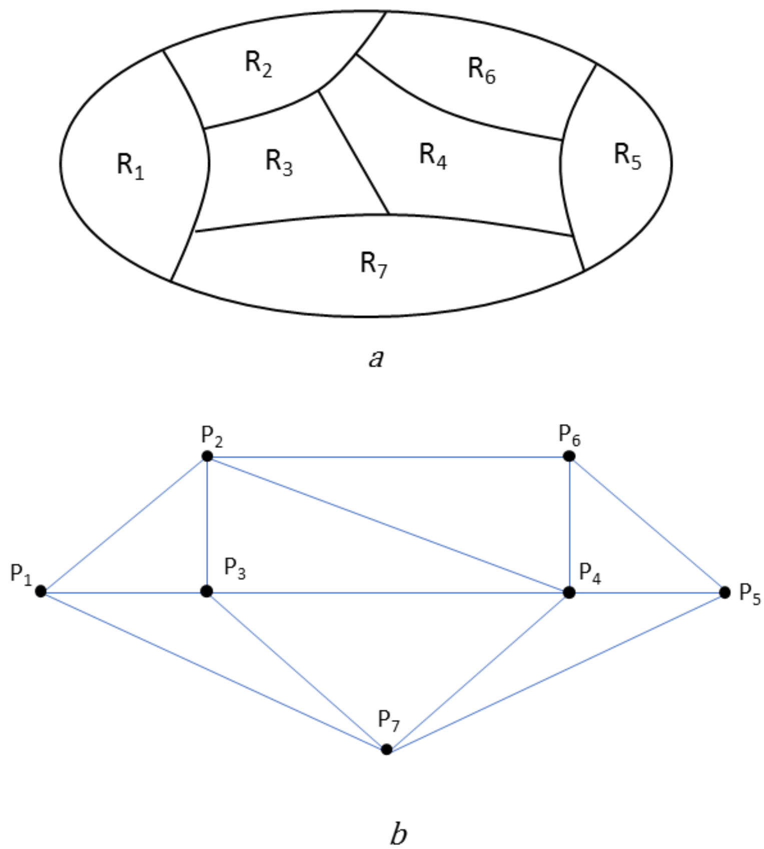


Figure 13.

Although Lewin's diagrams depict two dimensional planar maps, they are actually one dimensional in character, for he is interested solely in regions and boundaries and not in individual points in these regions. By using his particular method of representation by planar maps he is not able to represent all configurations which might arise. It is not clear through his published works whether the diagram of Figure 14a corresponds to the graph of Figure 14b or to that of Figure 14c,

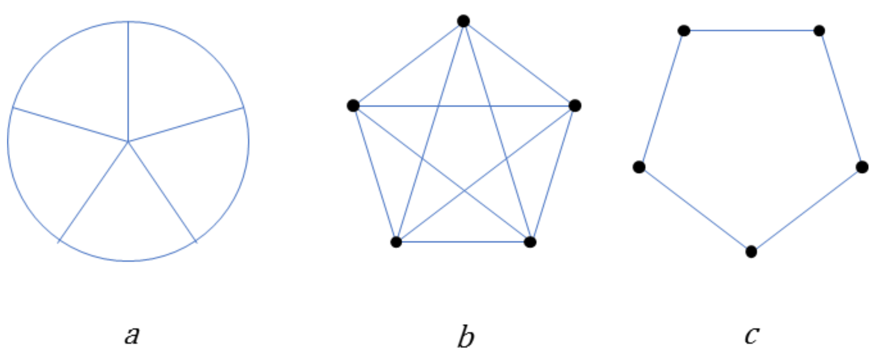


Figure 14.

for Lewin does not state whether a point is sufficient to form a boundary between regions. If Figure 14a corresponds to Figure 14b, then it is impossible to give a Lewinian representation of Figure 14c ; similarly , if Figure 14a corresponds to Figure 14c, then Figure 14b is not representable.

A graph which can be drawn in such a way that the intersection of any two lines is a point of the graph is called a *planar graph*. Using graphs, one can represent all the configurations that Lewin would wish to represent, for one can use both planar and non-planar graphs. Furthermore, graphical representation preserves all the advantages of Lewinian diagrams.

Lewin is concerned with special regions or points called inner and outer points. Let us consider a connected graph H as a subgraph of a larger connected graph G. By an outer point of H, relative to G, is meant a point of H whose distance from the set of {all points of G not in H} is 1. An inner point of H is one whose distance from the set of points of G not in H is greatest. For example, in the following graph G, if H is the subgraph of G consisting of the points P_1, P_2, P_3, P_4 and the lines P_1P_3, P_2P_3, P_3P_4, then P_4 is the only outer point of H while P_1 and P_2 are both inner points.

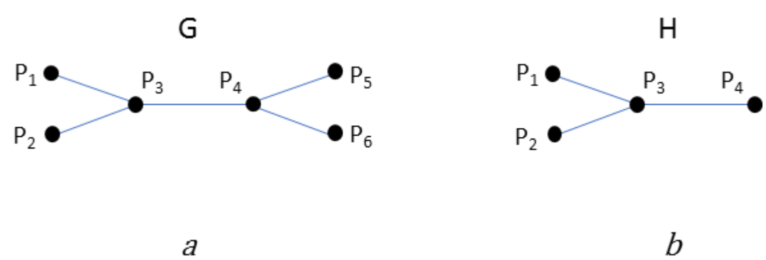


Figure 15.

Lewin (15, 16) also considered peripheral and central points. As in Section 2, central points were those points of least associated number. Those points whose associated number is greatest Lewin called *peripheral*. Centrality and peripherality , therefore, are intrinsic terms which relate to the graph itself , whereas the concept of inner and outer points is concerned with the way in which a subgraph is imbedded in a larger graph.

In Figure 16 we have placed the associated number of each point in parentheses,

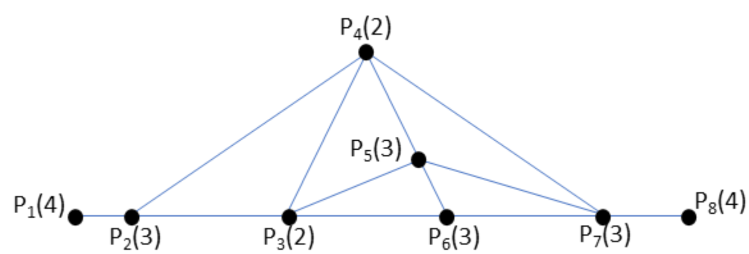


Figure 16.

Thus P_1 and P_8 are peripheral points, while P_3 and P_4 are central points. Lewin (15, p.180) gives the following example: "In conversation, the way to the peripheral regions of the person is almost always open. But it is difficult to touch the real core of the person." It is clear from his later works (such as 17) in which precise definitions are given differentiating the concepts inner and outer from central and peripheral, that it is the outer regions which are open to contact in ordinary conversation, and the inner regions which are more difficult to reach. While one would ordinarily expect the central regions to be inner regions, Lewin shows by an illustration (16, p. 124) that this need not be the case.

Bavelas (1) was interested in getting a model to represent interpersonal relations in groups as well as the “life space” of a person. He organized Lewin’s assumptions and definitions in a concise way, using neater terminology. One of his definitions, namely that of peripheral "cell" was essentially different from that due to Lewin. A *peripheral point*, following Bavelas, is one whose distance from the set of all central points is a maximum. He cites an example to show that these two definitions are different. In the following example, the points P and Q are the only peripheral points according to Bavelas, because they are the only points of maximal distance (namely 2) from the central points, whereas all six points whose associated number is 3 are peripheral according to the definition given in Lewin (16). It is easy to show that each point peripheral according to Bavelas is also is peripheral according to Lewin, but not necessarily conversely. Which of these two definitions is the more satisfactory is a question which can be settled only by analysis in the sociological or psychological setting, the main question being whether or not the definition of peripheral should depend on centrality.

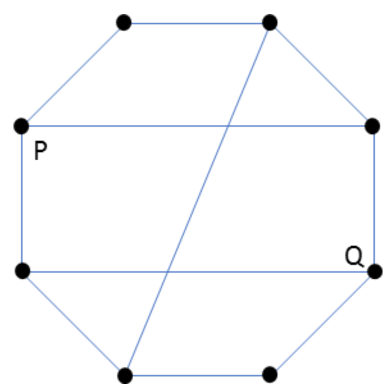


Figure 17.

Leavitt (14) experimented with several five-person groups each having a different communication “pattern”. These patterns are sets of admissible two-way communication paths between pairs of individuals, and clearly correspond to lines of five-point ordinary graphs. He considered the following graphs:

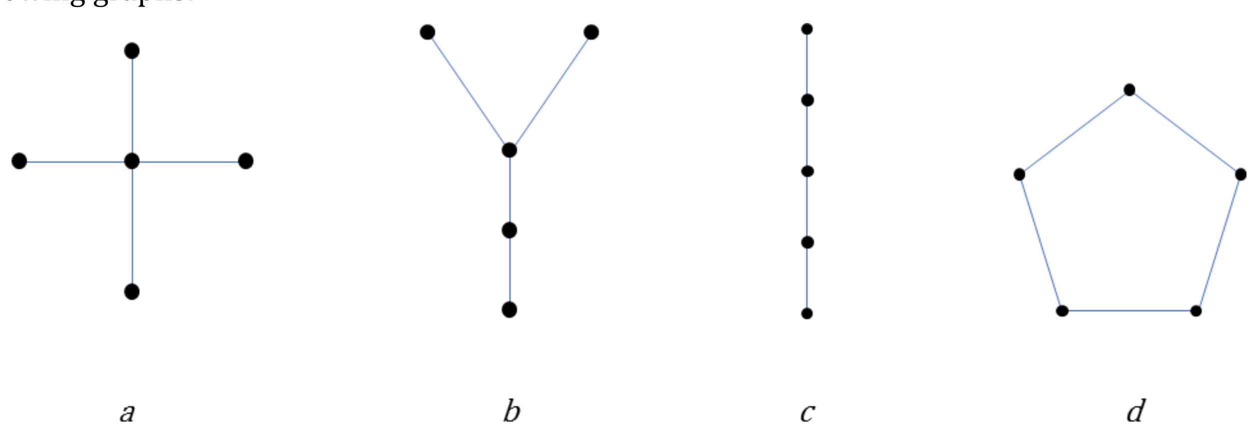


Figure 18.

He was concerned principally with the efficiency of communication as measured by productivity of the different psychological groups as a function of the group structure. He was also concerned with the manner in which the roles of the individuals involved is seen to change as experience points out to them the nature of the communication structure imposed on their group.

It is not surprising that efficiency decreases as the diameter increases and as the average associated number increases in going from the graph of Figure 18a to 18b to 18c. We note, however, that the graph of Figure 18d is more like that of 18a in diameter and in average associated number, yet it is the least efficient. The principal factor in determining efficiency is the development of a leader in the group. In each of the first three graphs there is at least one point which is different from all others in the graph, a point which corresponds to a person with a specialized role. We would say that such a point is a natural root, one that is singled out as different from the others by the nature of the graph itself. Bavelas (2) observes that "Leavitt's findings considerably strengthen the hypothesis that a recognized leader will most probably emerge at the position of highest centrality."

In those groups where the individuals were nearest to a readily recognizable central person, the role of a leader developed fastest, Leavitt's work implies that the presence of a leader in a group is much more important for the purpose of efficiency than the average degree of centrality.

Leavitt did not begin to consider all five-point connected graphs, and there are many of these which should be analyzed in the manner he treated the four he actually used. It is rather unfortunate that one finds, in the psychological literature , references to these four graphs as if they were the only ones, or at least the only important ones. Some of the other equally important 5 point connected graphs are:

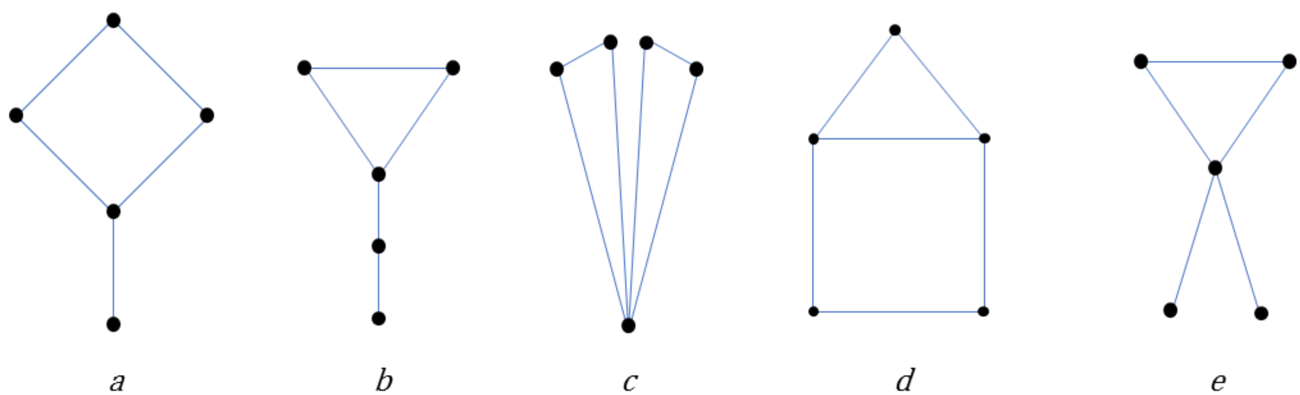


Figure 19.

Heise and Miller (8) studied, among other things, the efficiency of communication in "nets" represented by several directed three-point graphs. In their set-up a directed line runs from A to B if A can speak to B. The particular three-person groups which interested them were the ones in which information given to any of the three individuals could eventually be communicated to the other individuals in the group. They considered, therefore, all three-point directed graphs in which each pair of points is connected by a path in each direction, namely the following:

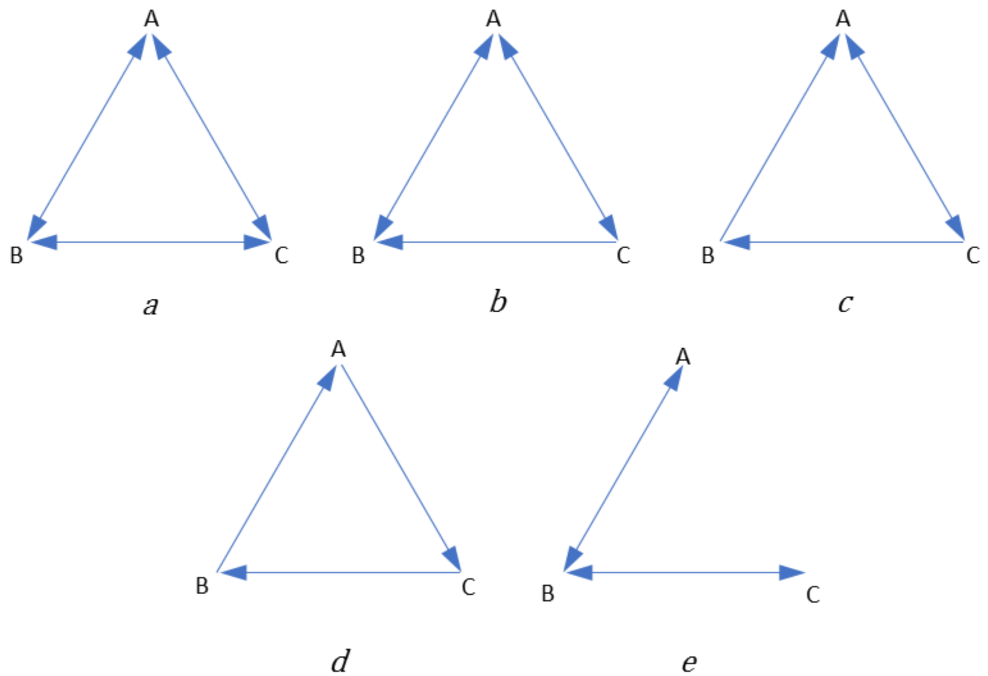


Figure 20.

4. Graphs, matrices and relations

The expressibility of graphs, matrices and (binary) relations in terms of each other is mentioned in König (13). Each of these three mathematical tools has certain operational advantages. For example, operations corresponding to matrix multiplication and inversion are not as conveniently performed on graphs or relations. An advantage of graphs is the easy pictorial identification of some of the properties, especially for small graphs, such as the presence of bridges or articulation points and disconnected components. In addition, graphs have the structure of one-dimensional complexes in combinatorial topology. There is a well-developed theory of simplicial complexes of n dimensions (25) of which graph theory is the special case n=1. This more complicated theory may eventually also be appropriate as a mathematical model for social science. Binary relations are appropriate in studies concerned with the operations of relation algebra such as converse, relative product, union, intersection, complement, etc. For a very clear and detailed exposition of the correspondence between the operations on relations and matrices, we recommend Copilowish (3).

Several recent papers have considered the analysis of sociometric data from the point of view of relations, matrices, or graphs according to the nature of the problem being studied. Before discussing some of these papers, we shall briefly review the interrelationships between these three notions.

Let G be a directed graph of n points P_1 to P_n. The matrix of graph G is the n by n matrix which has the number 1 in its i,j position if the directed line from P_i to P_j is in G, and has the number 0 as its i,j entry otherwise. The matrix of G thus depends on the particular way in which the points of G are labelled P_1 to P_n. In any such matrix the main diagonal consists entirely of zeros since a directed line from a point to itself is not admissible. By an ordinary graph we mean one in which the lines between pairs of points are not directed. Any ordinary graph may be regarded as a directed graph having two directed lines in opposite directions between each pair of points of the ordinary graph joined by a line. We see that a directed graph is ordinary if and only if its matrix is symmetric. If the directed graph G is given, then its matrix can be written by inspection. Conversely any matrix consisting entirely of zeros and ones, whose main diagonal contains only zeros, uniquely determines a directed graph.

By definition a relation is a set of ordered pairs of elements. An *ordered pair* of elements, usually written (*a,b*), is a set consisting of two elements in which *a* is designated as the first element and *b* as the second element. The directed graph G may be regarded as the relation consisting of the set of all ordered pairs of points (P_i,P_j) such that the directedline from P_i to P_j is in G. A relation is called irreflexive if it does not contain any ordered couples of the form (*a,a*). We illustrate the one-to-one correspondence between directed graphs, matrices consisting entirely of 0's and 1's with only 0's on the main diagonal, and irreflexive relations. The following example of sociometric data is taken from a 5-person group, each of whom made two choices within the group. The sociometric matrix in which the rows denote choices made and the columns choices received is:

	P_1	P_2	P_3	P_4	P_5
P_1	0	1	0	1	0
P_2	1	0	1	0	0
P_3	1	0	0	1	0
P_4	0	0	1	0	1
P_5	0	0	1	1	0

The relation corresponding to this matrix is the set of ordered pairs:

(P_1, P_2), (P_1,P_4), (P_2,P_1),(P_2,P_3), (P_3,P_1),(P_3,P_4), (P_4,P_3),(P_4,P_5), (P_5,P_3),(P_5,P_4)

The directed graph corresponding to this matrix may be pictured:

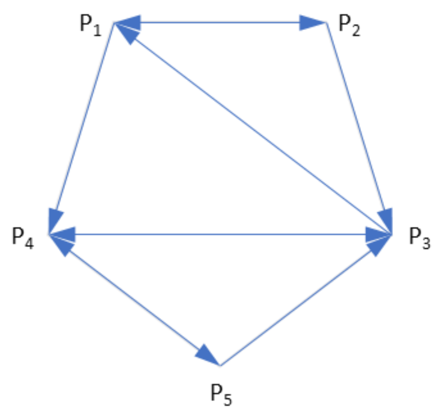


Figure 21.

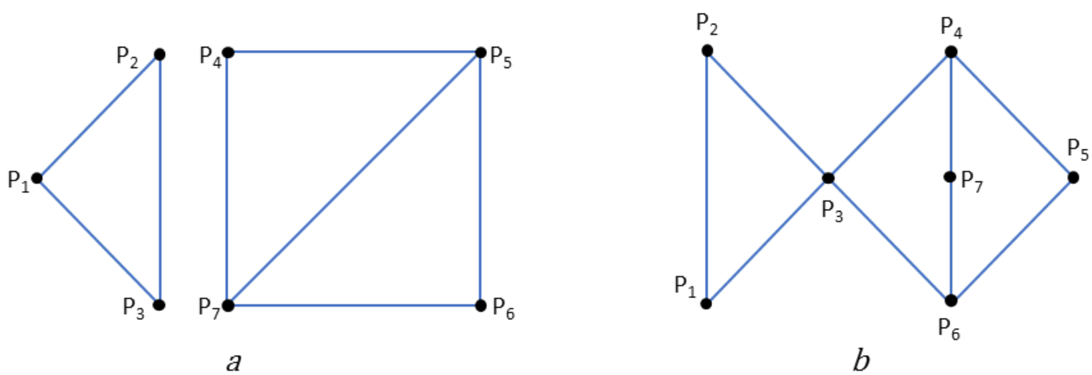
In view of the logical equivalence between graphs, matrices, and relations, each property of any one of them gives information about the other two. To illustrate the form of matrices of graphs which

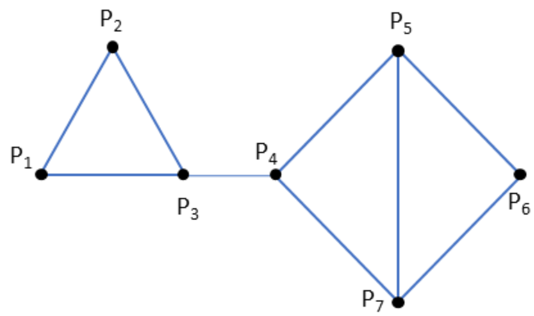
(1) are non-connected and have two components,

(2) have one articulation point with two branches,

(3) have one bridge,

we consider the following graphs:





C

Figure 22.

The matrices of these graphs are respectively:

$$
\begin{array}{c|ccccccc|}
 & P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \\
\hline
P_1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
P_2 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
P_3 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
P_4 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
P_5 & 0 & 0 & 0 & 1 & 0 & 1 & 1 \\
P_6 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
P_7 & 0 & 0 & 0 & 1 & 1 & 1 & 0
\end{array}
\quad
\begin{array}{|ccccccc|}
P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \\
\hline
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 0
\end{array}
\quad
\begin{array}{|ccccccc|}
P_1 & P_2 & P_3 & P_4 & P_5 & P_6 & P_7 \\
\hline
0 & 1 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 1 & 0
\end{array}
$$

We conclude this section with a brief discussion of some recent papers applying relation algebra and matrix theory to sociometric problems. Using relation algebra, Rosenblatt (22) solved the following problem in organization theory by finding a recursion equation: *What is the 'configuration' state of ideas held or known by members of the organization (social group) at any given time? [emphasis added –ed.]*

Introductory expositions of the utility of matrices for the handling of sociometric data was given in Katz (11) and Festinger (4). The sociometric matrices considered had entries of 1,0,-1 to represent "choice, indifference, or rejection" respectively. Continuing this approach, Katz (12) discussed in matricial terms "leaders", "selection of teams", "rumors", "isolates", and "types of structure." In this paper (12), Katz also presents a formula for redundant chains of 4 steps. The formula for redundant 3-step chains had already been deduced by Luce and Perry (17).

The problem of the determination of the number of redundant chains was essentially settled by Ross and Harary (23), who gave explicit matrix formulas for redundancies of 3,4,5 and 6 steps and a general procedure for any given number of steps. Using their result on the number of redundant chains, Harary and Ross (6) found an elementary matrix formula for the number of complete cycles in a communication network, i.e., the number of Hamilton lines in a graph.

The determination of cliques in a social group was discussed by Festinger (4) and Luce (18). Festinger indicated that the principal diagonal of the cube of the given sociometric matrix yields information regarding the cliques within the group. Luce combined concepts involving directed graphs, relation algebra, and matrices to investigate rigorously “connectivity and generalized cliques in sociometric group structure".

Weiss (25) presented a fairly systematic method of reducing a sociometric matrix of reasonable size to a canonical form designed to display the subgroups of the whole group. He defined "liaison person” and "subgroup" to stand for the concepts of graph theory called articulation point and articulation component. These ideas were further discussed by Jacobson and Seashore (9).

5. Some theorems on graphs

We have introduced many definitions of graph theory, but have not discussed any of the theorems as yet. As indicated in the introduction, any theorem on graphs states a result which is valid for any concrete setting in which the hypothesis is true. There are many theorems on graphs which when translated into a psychological setting, are trivial and inconsequential. On the other hand, some theorems are potentially applicable, but require further investigation to determine their appropriateness. Finally, there are theorems which are fairly immediately applicable. We shall mention a few such theorems in this section. In addition, to give some important aspects of graph theory, we are including brief discussions of the four color problem and the Euler characteristic.

In Section 3 the representation of maps, and especially planar maps, by graphs was discussed. In this connection it seems appropriate to discuss one of the famous unsolved problems of mathematics, the "four-color problem”. Can one color every map drawn on a sphere or a plane with four colors in such a way that no two regions, or countries, having a common boundary arc have the same color? It can be proved fairly easily that five colors are enough to color any such map, and that there are maps requiring four colors, the simplest of which is shown in Figure 23 below. It is believed that four colors are always sufficient, but this problem has baffled mathematicians for centuries. [*Solved in 1976 – ed.*]

Strangely enough, if one states the problem for a torus, i.e., a doughnut shaped surface, asking how many colors are required to color a map on a torus, the answer is quite simple. Every map on a torus can be colored with seven or fewer colors, and there exist maps requiring seven colors.

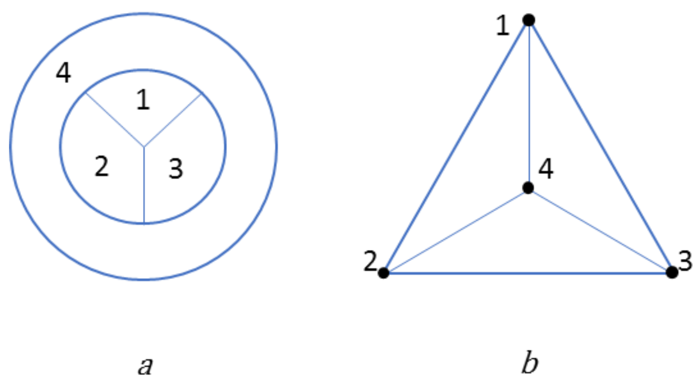


Figure 23.

A theorem with, perhaps, more immediate psychological implication is taken up in König (13, p.30). Consider a directed graph in which every pair of points is connected with a directed line in at least one direction. Then there exists a path in the graph containing all points of the graph. To illustrate this theorem, the directed graph of Figure 24 below has at least one directed line joining each pair of points and therefore there is a path containing all the points, namely the path ACDB.

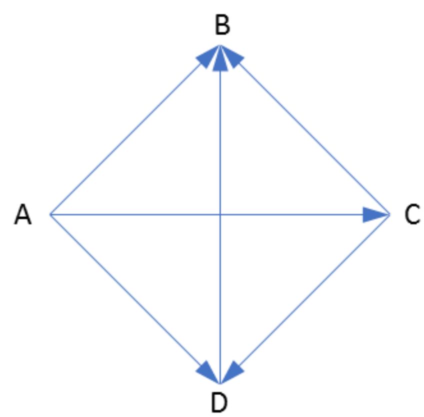


Figure 24.

Translating this theorem into the language of communication, we see that in a group in which at least one of each pair of individuals communicates [*transmits – ed.*] to the other, it is possible for a piece of information originated by some individual to be transmitted to all members of the group.

Another theorem (6) states that in any connected graph, all central points are contained in a subgraph which is a star. It is, of course, not necessary that the central points themselves be connected, as is shown in Figure 25 below.

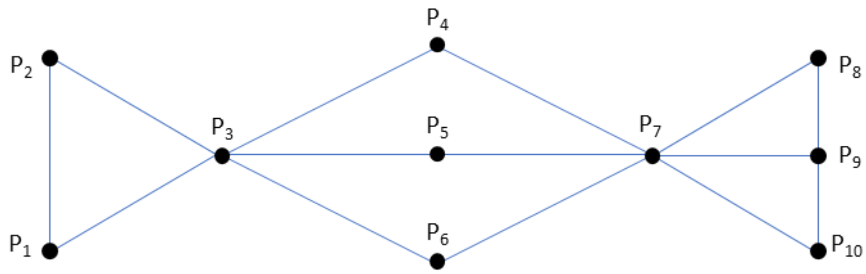


Figure 25.

Here P4, P5, and P6 are the central points, and the central star consists of them plus P3 and P7 together with the lines of the graph joining them. Such a central star corresponds to the nucleus of a psychological group. Leavitt (14) has commented that the central individual (if any) tends to become the leader of the group. Here in place of a central individual, the individuals represented by the central star might be expected to become the dominant set in the psychological group.

A classic result due to Euler states that for any polyhedron, the number of vertices less the number of edges plus the number of faces is the number two. This result can easily be translated to maps on the sphere, for we may distort such a map by making all boundaries between regions into straight lines without changing the number of boundaries, vertices, or regions. Thus the number of regions less the number of boundaries plus the number of vertices is two for any spherical map. Such a spherical map can be changed into a planar map by poking a hole inside any of the regions and stretching the erstwhile sphere into a plane. The Euler characteristic of a planar graph is thus one, because we have lost one region.

A refinement of the Euler characteristic of a tree was given by Otter (19) as a step in counting the number of trees. If we call a line whose points are similar a symmetry line, and we call two lines similar if the two points lying on one line are respectively similar to those on the other line, then we can state Otter's theorem: “In any tree the number of nonsimilar vertices minus the number of nonsimilar lines (symmetry line excepted) is the number one." If all points are dissimilar, this is the Euler relation above. For a tree is a planar graph and has no faces. This result of Otter has been extended by Harary and Norman (5) to Husimi trees, *i.e.*, graphs in which each line lies in at most one cycle.

6. Some new concepts in graph theory.

Motivated by psychological considerations, we introduce three generalizations of graphs which we call graphs of *strength s*, graphs of *type t* , and *multiply-rooted* graphs. For ease of comprehension, we shall restrict the following descriptive discussion to graphs of strength 2, graphs of type 2, and doubly-rooted graphs. It will be clear that no essential loss of generality is entailed by this restriction.

A graph of *strength 2* is one in which there may be 0, 1, or 2 lines joining a pair of points, and these lines are not distinguishable from each other.

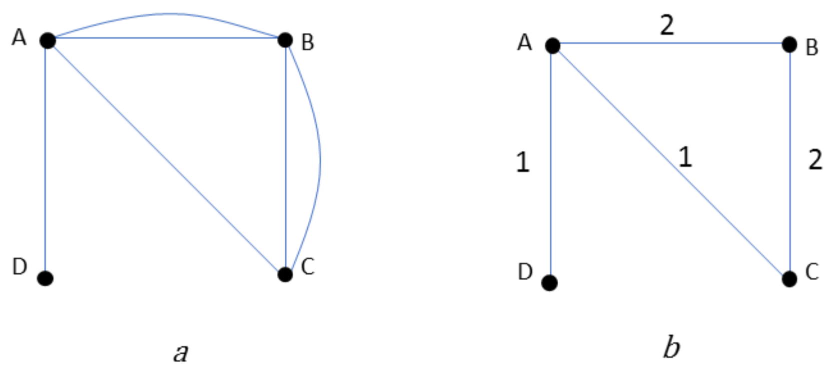


Figure 26.

Two pictorial representations of the same graph of strength 2 are shown in Figures 26a and 26b. From the point of view of probability, the above graph of strength 2 may be regarded as representing a relation between individuals together with the probability that this relation holds. Thus the probability of mutual choice, for example, between B and C is 1; between A and C is ½ ; and between C and D is 0. An advantage of the method of representation of graphs of a certain strength as in Figure 26b is that the probabilities need not be restricted to ratios of small integers. Another interpret ation of the strength of a graph may be illustratedin terms of communication; namely, the strength is the maximum possible number of units of mutual attraction between two members of a social group.

A graph of *type 2* is one in which there are lines of 2 different kinds available to characterize, by superposition, 2 different relations. A pair of points may be joined by no lines , one line of either kind, or 2 lines consisting of one of each kind. Pictorially, a graph of type 2 may have the kinds of lines distinguished by drawing a solid line or a dashed line. The three graphs

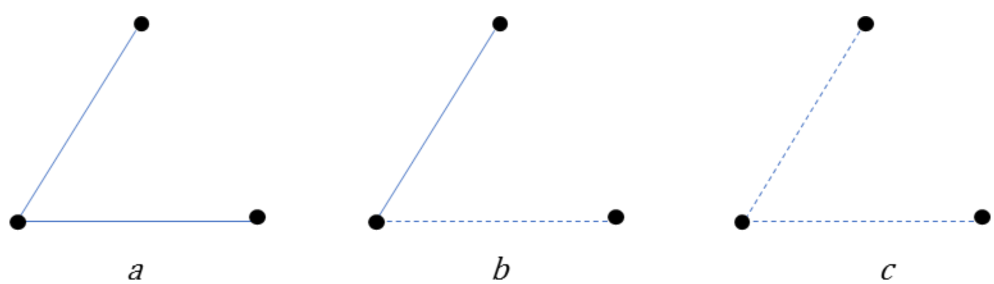


Figure 27.

of type 2 are regarded as different. Graphs of type 2 are appropriate to represent communication networks in which there are 2 different means of communication which can occur between any two members of the group. But the particular interest of this generalization of graphs lies in the interpretation of the 2 kinds of lines as 2 relations on the same group of persons. Some examples of 2 such relations are choice and communication, choice and influence, social choice and choice of a working partner, etc. Some of the possible conditions which under certain circumstances will obtain between the 2 relations are:

(1) the 2 relations are independent in that the occurrence of one of them does not necessarily imply anything about the other,

(2) one relation may imply the presence of the other,

(3) one relation may deny the other, i.e., whenever the first relation occurs, the second does not,

(4) a more complicated dependence between the 2 relations than either of the preceding two conditions.

From the above definitions and illustrations, we observe that strength and type correspond to quantitative and qualitative information about the group, respectively. The concept of a directed graph of strength *s* or of type *t* is readily defined and similarly interpreted with one-way communications or attractions being admitted as well as two-way communications or mutual choices. It is clear that a graph of strength 1 or of type 1 is an ordinary graph.

A rooted graph has been defined in Section 2 as an ordinary graph in which one point is distinguished or singled out. Two extensions of the concept of a rooted graph to that of a multiply rooted graph will next be considered.

A doubly rooted graph having *like roots* is one in which 2 of the points of the graph are distinguished from the remaining points but not from each other. A doubly rooted graph in which there are 2 *unlike roots* is one in which 2 of the points of the graph are distinguished from the remaining points and also from each other. Two like roots may be regarded as leadership vested in the hands of two persons. Two unlike roots are useful in representing two positions both of which are singled out, but in different ways, e.g., Lewinian diagrams in which one region is labelled "present status" and another "desired goal". One pictorial device which may be used when the number of unlike roots is smallis the placing of a circle around one root, a square around another root, etc. Thus a graph of 4 points with 2 unlike roots may be pictured as in Figure 28.

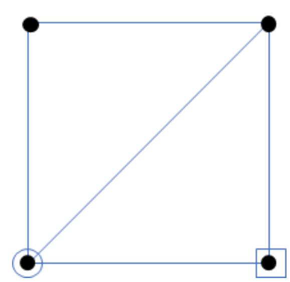


Figure 28.

There is a certain duality between graphs of type 2 and doubly rooted graphs in that the former distinguishes between the kinds of lines whereas the latter recognizes different kinds of points.

Of course all sorts of mixtures of ordinary and directed graphs of strength *s* and type *t* having several like as well as unlike roots may be defined precisely and interpreted appropriately, but we shall not discuss them here.

The determination of the number of different graphs having a given number of points and a given number of lines is a problem of considerable interest in several fields of study. In each field, graphs satisfying conditions appropriate to the situation at hand are considered.

The importance of the number of graphs for psychology is that all of the networks with a given property will be included in experimental designs as well as in theoretical investigations. Organic chemists have counted the number of trees which characterize certain hydrocarbon molecular structures. In statistical mechanics, the counting problem for a generalization of a tree (called a Husimi tree) is being studied in connection with a problem in the theory of condensation (cf. Riddell (21), Harary and Uhlenbeck (7), Harary and Norman (5)). Pure mathematicians, i.e., those who are not primarily concerned with obtaining results which can be applied, are interested in all these counting problems *per se*.

The number of different trees having a designated number of points was found independently by Polya (20) and by Otter (19), both of whom utilized a previous result due to Cayley on the number of rooted trees.

The number of ordinary graphs of n points and k lines has been found by Polya (unpublished). The present authors have succeeded in counting the number of directed graphs, graphs of type t, graphs of strength s, rooted graphs, and multiply-rooted graphs. Connected graphs were counted independently by Riddell (21) and Polya (unpublished). It would not be appropriate to reproduce these formulas here.

7. Possible further uses of graph theory in psychology

In the foregoing sections we have given a brief exposition of some concepts from the theory of graphs which are suited to analyzing some well-defined areas of social-scientific investigations. We now mention a few less-well defined topics to which the theory of graphs may be applicable. These include:

(1) study of some concepts not yet precisely defined in a quantitative manner, such as cohesiveness, stability, and efficiency of a group, and

(2) consideration of changes in graph structure brought about, for example, by a tendency toward efficiency or stability.

Each of the terms cohesiveness, stability, and efficiency of a group may be applied to describe the corresponding graph. The usual qualitative, intuitive definition of cohesiveness is the tendency of the members of the group to stick together, i.e., to remain in the group. After a specific logical quantitative definition of cohesiveness is given (and preferably unanimously agreed to by social psychologists[*]), it will be interesting to attempt to construct a formula for the cohesiveness of a graph. Directed graphs of strengths are of particular interest here as a means of describing quantitatively different attractions of the members of the group to each other. This is exemplified in a current study at the Research Center for Group Dynamics being conducted by J.R.P. French, Jr. Each member of a large organization was asked to make 10 ordered sociometric choices. The problem is the selection of special subgroups, of maximal cohesiveness.

The notion of stability overlaps that of cohesiveness, A group is unstable if there is a tendency for it to split into separate groups. In general, a group whose graph has a bridge or an articulation

---

[*] See Gross, N. and Martin, W.E., “On group cohesiveness”, *Am.J. Sociol.* 1952, **57**, 546-551 and Schachter, 3rd Comment, *ibid* 551-562.

point is less stable than one which does not. A study of stability and cohesiveness might well concern itself strongly with articulation points, bridges, and generalizations of these. Among the generalizations, there are double bridges (a double bridge is a pair of lines, neither of which is a bridge, whose removal separates the graph), double articulation points (defined similarly) , etc. We would seek conditions for stability of a graph, and a reasonable measure of stability.

Consider several different graphs of the same number of points as representing communication patterns. Which of these has the most efficient pattern for solving certain types of problems? This is the sort of question investigated by Leavitt (14) and very clearly discussed by Bavelas (2). Certainly the efficiency of a group pattern depends on the particular group task under consideration. As mentioned in preceding sections, the central points of the graph have crucial importance in questions of this sort.

If the group is free to alter its structure so as to increase its efficiency, in what way will these changes take place? A preliminary study in this field by R.Z. Norman and H. Gerard is in progress at the Research Center for Group Dynamics and is concerned with the efficiency of stars of 5 points from the point of view of the number of communications required for information to reach all members of a group.

In particular we might make the assumption that the group can remove or add one communication link at a time. We call a graph relatively efficient in this framework if adding or deleting exactly one communication link yields a less efficient structure. Then will the most efficient structure be attained by all groups, or will some groups tend to be satisfied with a relatively-efficient structure?

8. Summary

We have discussed the advantages of using a mathematical model in the social sciences, and in particular of using graphs to represent interpersonal relations in groups of individuals and the "life space" of a person. Since graphs have been discussed by many social scientists using varying terminology, we have appealed to them to use a uniform and precise language, that of the theory of graphs. To this end, an introductory exposition of the mathematical theory of graphs has been given, supplemented by relevant sociological and psychological illustrations. The interrelations among graphs, matrices, and relations and the advantages of each have been pointed out. We hope that the readers of this journal will find the concepts of graph theory useful in the formulation of ideas and in the solution of problems, not only in the fields which we have mentioned, but in other branches of the social sciences as well.

9. Glossary

Adjacent	Two points are adjacent if there is a line joining them.
Articulation component	An articulation component of a connected graph G at an articulation point P is a subgraph which contains P, does not have P as an articulation point, and is maximal.
Articulation point	An articulation point of a connected graph is a point whose removal separates the graph.
Associated number	The associated number of a point of a connected graph is the maximum of the distances from this point to each of the other points.
Automorphism	An automorphism is an isomorphism of a graph with itself.
Branch	A branch is the same thing as an articulation component. Originally, a branch was defined only for a tree, but the definition is extended to any connected graph.
Bridge	A bridge is a line of a connected graph whose removal separates the graph into two components each of which contains more than one point.
Center	The center of a graph is the collection of its central points.
Central point	Of a connected graph, a central point is a point whose associated number is a minimum.
Complete graph	A graph is complete if all pairs of points in it are joined by lines.
Component	A component of a graph is a maximal connected subgraph.
Connected graph	A graph is connected if there exists a path between every pair of its points.
Cycle	A cycle of a graph is a collection of lines of the form PA, AB,..., CP where all the points P, A, B,...,C, are different from each other.
Degree	The degree of a point of a graph is the number of lines of the graph on which the point lies.
Diameter	The diameter of a connected graph is the maximum of the distances between any two of its points.
Different graphs	Two graphs are different if they are not isomorphic.
Directed graph	A directed graph of n points is a collection of n points together with a subset of the set of all directed lines between pairs of the points.
Distance	The distance between two points of a connected graph is the length of any shortest path joining them.
End point	1. An end point of a graph is a point of degree one. 2. An end point of a line is a point of the line.
Euler graph	An Euler graph is a graph in which every point is of even degree.

Graph of *n* points	A graph of *n* points consists of these *n* points together with a subset of the set of all lines joining pairs of these points.
Hamilton line	A Hamilton line of a graph is a cycle which contains all points of the graph.
Inner point	An inner point of a graph H which is a subgraph of a graph G is a point whose distance from a point of G not in H is maximal.
Isomorphic graphs	Two graphs are isomorphic if there is an isomorphism between them.
Isomorphic rooted graphs	Two rooted graphs are isomorphic if they are isomorphic as ordinary graphs and the root point of one is sent by the isomorphism into the root point of the other.
Isomorphism	An Isomorphism is a one-to-one transformation of the set of points of one graph with that of another which preserves adjacency.
Length of a path or a cycle	The length of a path or cycle is the number of lines in it.
Maximal	A subgraph is maximal with respect to a certain property if it has the property and is not contained in any larger subgraph with this property.
Multiply rooted graph	A multiply rooted graph is one in which the set of points is partitioned into distinguished subclasses*
Ordinary graph	See graph of *n* points.
Outer point	An outer point of a graph H which is a subgraph of a graph G is a point which is adjacent to a point of G not in H.
Path	A path is a collection of lines of the form AB, BC, ..., DE where all the points A, B, C, ..., D, E are different from each other. A directed path would be, similarly, a collection of directed lines $\underset{\rightarrow}{AB}, \underset{\rightarrow}{BC}, \ldots, \underset{\rightarrow}{DE}$
Peripheral point	1. According to Lewin, a point of a connected graph is peripheral if its associated number is greatest. 2. According to Bavelas, a point of a connected graph is peripheral if its distance from the center of the graph is maximal.
Relation	A relation is a set of ordered pairs of elements.
Root	A root is a distinguished point of a graph.
Rooted graph	A rooted graph is one in which there is a distinguished point.
Similar	1. Two points are similar if there is an automorphism of the graph sending one point into the other. 2. Two lines are similar if their endpoints are similar.
Star	A graph is a star if it is connected and contains no articulation point. Subgraph. H is a subgraph of a graph G if H is a graph and every point and line of H is contained in G. Tree. A tree Is a connected graph which has no cycles.

10. References

1. Bavelas, A. “A mathematical model for group structures”, *Appl. Anthrop*. 1948, **7**, 16-30.

2. Bavelas, A. “Communication patterns in task-oriented groups”, in Lerner, D. and Lasswell, H.D. eds. *The Policy Sciences,* Stanford; Stanford Univ. Press, 1951. 193-202.

3. Copilowish, I.M. “Matrix development of the calculus of relations”, *J. Symbol. Log*. 1948, **13**, 193-203.

4. Festinger, L. “The analysis of sociograms using matrix algebra”, *Hum. Relat.* 1919, **2**, 153-158.

5. Harary, F. and Norman. R.Z. “The dissimilarity characteristic of Husimi trees”, to appear in *Annals of Math*, 1953.

6. Harary, F. and Ross, I.C. “The number of complete cycles in a communication network”. (In process of publication).

7. Harary, F. and Uhlenbeck, G.E. “On some generalizations of rooted trees”, Abstract 131, *Bull. Amer. Math. Soc*. 1952, **58**, 168.

8. Heise, G. A. and Miller, G.A. “Problem solving by small groups using various communication nets”, *J. Abnorm. Soc*. *Psychol*. 1951, **46**, 327-335*

9. Jacobson, E. and Seashore, S. “Communication practices in complex organizations”, *J. of Social Issues*. 1951, **7**, 28-40.

10. Kaplan, A. “Sociology learns the language of mathematics”, *Commentary.* 1952, **14**, 274-284.

11. Katz, L. “On the matric analysis of sociometric data”, *Sociometry* 1947, **10**, 233-240.

12. Katz, L. “An application of matrix algebra to the study of human relations within organizations”, *Inst. of Math. Stat*. Univ. of N. Carolina mimeo series, 1950.

13. König, D. *Theorie der endlichen und unendlichen Graphen*, Leipzig, 1936, reprinted New York: Chelsea, 1950.

14. Leavitt, H.J. “Some effects of certain communication patterns on group performance”, *J. Abnorm. Soc. Psychol.* 1951, **46**, 38-50.

15. Lewin, K. *Principles of topological psychology*, New York: McGraw-Hill , 1936.

16. Lewin, K. “Analysis of the concepts whole, differentiation, and unity”, Univ. of Iowa Studies in Child Welfare 1940, **18**, 226-261(reprinted as the appendix of Lewin, K. *Field theory in Social Science*, New York: Harper, 195l).

17. Luce, R.D. and Perry, A.D. “A method of matrix analysis of group structure”, *Psychometrika* 1949, **14**, 95-116.

18. Luce, R.D. “Connectivity and generalized cliques in sociometric group structure”, *Psychometrika* 1950, **15**, 169-190.

19. Otter, R. “The number of trees”, *Annals of Math*. 1948, **49**, 583-599.

20. Polya, Q. “Kombinatorische Anzahlbestimmungen fur Gruppen, Graphen, und chemische Verbindungen”, *Acta Math*. 1937, **68**, 145-254.

21. Riddell, R.J. “Contributions to the theory of condensation”, Doctoral dissertation, Dept.of Physics, Univ. of Michigan, 195l.

22. Rosenblatt, D. “A note on communication in organizations”, to appear as a chapter in a forthcoming book: *Symbolic methods in the study of organizations*.

23. Ross, I.C. and Harary, F. “On the determination of redundancies in sociometric chains”, *Psychometrika*, 1952, **17**, 195-208.

24. Veblen, O. *Analysis Situs*, second edition, New York 1931 (esp. Chap. 1).

25. Weiss, R.S. An investigation into organizational structure using a sociometric technique, Masters dissertation, Dept. of Sociology, Univ. of Michigan 1952.

26. Wertheimer, M. *Productive thinking*, New York: Harper, 1945.

27. Whitney, H. “Non-separable and planar graphs”, *Trans. Amer. Math. Soc.* 1932, **34**, 339-362.