

## A Graph Theoretic Approach to the Investigation of System-Environment Relationships<sup>1</sup>

The purpose of this paper is to undertake a graph theoretic analysis of those empirical structures for which it is meaningful to consider a part of the structure as a system and the remainder as its environment. We shall be primarily concerned with social systems such as groups, social networks or organizations, but the results are applicable to any structure that can be appropriately represented by means of graph theory.

The term *empirical structure* is here used to refer to a set of empirical elements (such as people, roles or positions) together with a set of empirical relationships (such as liking, communication or influence) between pairs of elements. Given such a structure, its digraph  $D$  is obtained as follows. The set of points of  $D$ , denoted  $V = \{v_1, v_2, \dots, v_p\}$ , corresponds to the set of empirical elements, and there is an arc (directed line) from  $v_i$  to  $v_j$  if and only if the corresponding ordered pair of elements is in the specified empirical relation. Throughout the following discussion we shall be concerned with relationships that can be viewed as "links" of communication by which "messages" are transmitted from one element to another. In graph theory, a *walk*  $W$  of  $D$  is defined as an alternating sequence of points and arcs which begins and ends with a point and has the property that each arc is preceded by its first point and followed by its second one. Thus each walk of  $D$  represents a "chain of communication" within the structure, and it is possible for a message to reach element  $v_j$  from  $v_i$  if and only if  $D$  contains a walk from  $v_i$  to  $v_j$ .

It should be noted that the term *message* is used here to refer to anything that can be transmitted from one element to another. Thus, for example, if the elements of a structure are thought of as roles, subgroups or positions of an organization, then a message might be a person, memorandum, unit of work or some other "object" that can change its location in the organization. If, on the other hand, elements are taken to be individual people, then a message might be any of the following: (a) an item of information, an opinion or a rumor; (b) an influence attempt, such as an order, request or suggestion; (c) some ma-

<sup>1</sup>Reprinted from *Journal of Mathematical Sociology*, 5:87-111, 1977

terial object that may be given, lent or sold; or (d) some symbolic object such as a favor, approval, help or support. This latter type of interpretation has been employed in research on social networks, as reported by Mitchell (1969) and Barnes (1972), where it is assumed that a variety of such messages can be transmitted through a given network. A link that carries more than one kind of message is said by these authors to be "multistranded."

Let us assume that we have some empirical basis for identifying a particular system within a given structure so that the set  $V$  of points of its associated digraph can be partitioned into two subsets,  $S = \{s_1, s_2, \dots, s_m\}$ , corresponding to the elements of the system, and  $E = \{e_1, e_2, \dots, e_{p-m}\} = V - S$ , corresponding to the elements of its environment. Strictly speaking, a *system* and its *environment* correspond to the subgraphs  $\langle S \rangle$  and  $\langle E \rangle$  that are induced by their sets of points,<sup>2</sup> but it will be convenient to denote them more simply by  $S$  and  $E$ .

For illustrative purposes, two rather different sorts of interpretation will be employed throughout this paper. In the first, we assume that a digraph  $D$  represents the *interpersonal communication structure* of a group of people and that system  $S$  is a specific subgroup within the group. Messages are taken to be items of information which can be transmitted only via links of the structure. Thus, points correspond to individual people, arcs correspond to links of interpersonal communication and walks indicate permissible chains of communication. In the second interpretation, we assume that  $D$  represents the *career structure* of an organization and that  $S$  represents a specified department within the organization. Here, points correspond to positions (or jobs), messages are thought of as individual people, arcs indicate permissible changes of position and walks designate permissible career lines of individuals.

Figure 1 shows the digraph of a small structure containing a system,  $S = \{s_1, s_2, s_3\}$ , and its environment,  $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ . It can readily be seen that in this digraph every walk  $W$  between a point of  $S$  and a point of  $E$  must contain arcs  $s_1e_1$  or  $e_1s_1$  and hence points  $s_1$  and  $e_1$ . The corresponding communication links and elements of the structure are thus uniquely involved in all transactions between the system and its environment and would seem intuitively to be located in the "boundary" between  $S$  and  $E$ .

If this digraph is interpreted as an interpersonal communication structure, then all chains of communication between a member of subgroup  $S$  and a non-member must contain "boundary" persons  $s_1$  and  $e_1$  who are uniquely able to monitor, modify or intercept any item of information going between the subgroup and its environment. It is also reasonable to assume, in keeping with the analysis of the functions of gossip in social networks presented by Epstein (1969), that when an item of information traverses a "boundary link" of such

<sup>2</sup>The definitions of concepts of graph theory that are not given in this paper may be found in Harary, Norman and Cartwright (1965).

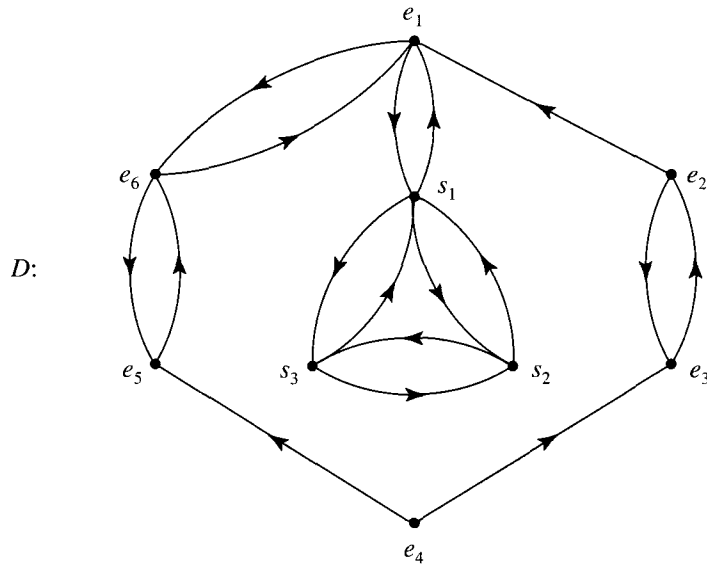


Figure 1. The digraph of a system  $S = \{s_1, s_2, s_3\}$  and its environment  $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$ .

a system, its content or meaning may be significantly altered. In a similar way, if  $D$  is interpreted as the career structure of an organization, then all career lines that include positions in both department  $S$  and its environment must contain “boundary positions”  $s_1$  and  $e_1$ . Anyone entering or leaving the department can do so only by going directly from one of these positions to the other, and when a person traverses a “boundary link” in the career structure, he changes departmental membership and consequently becomes subjected to a different set of influences and expectations.

The concept of boundary has been employed extensively by theorists (Emery, 1969; Katz and Kahn, 1966; Miller and Rice, 1967; Rice, 1963) who adopt a “systems approach” to social organizations. According to this view, those members of an organization who are directly involved in transactions between the organization and its environment are said to occupy “boundary roles” and various consequences have been identified for such individuals. Thus, for example, Kahn et al. (1964) have reported that occupants of boundary roles are more likely to experience role conflict than are members located “deep within the organization.” The explanation proposed by Kahn et al. to account for this difference can be elucidated by means of the digraph shown in Figure 1. Let us assume that  $S$  and  $E$  represent a (miniature) organization and its environment, that the points of  $D$  correspond to individuals, that role expect-

tations are communicated via the arcs of  $D$  and that these expectations differ when they come from inside as opposed to outside the organization. It then follows immediately that person  $s_1$  should experience more role conflict than  $s_2$  or  $s_3$ .

The concept of boundary has also been used implicitly by Korte and Milgram (1970) in a study of acquaintance networks connecting racial groups in the United States. The procedure involved asking a number of "source" individuals, who were white, to try to send a booklet via a sequence of acquaintances to a black "target" person in a different community. The investigators were able to record the progress of each booklet as it proceeded through the network and thus to identify a chain of communication from each source to the target. If we conceive of the white segment of an acquaintance network as system  $S$  and the black segment as environment  $E$ , then every successful chain must contain a "boundary link" of  $S$ . One striking feature of this study is that such links tended to appear only late in each chain of communication; in 23 of the 35 completed chains, the only boundary link was the last one of the chain, and in seven it was next to the last. Korte and Milgram call the white member of such a link a "gatekeeper," a term introduced by Lewin (1943:186) to refer more generally to anyone who can control the flow of objects across the boundary of a system.

In the following pages, we first develop a number of concepts which give precise meaning to the intuitive notions of the "boundary" of a system and the "degree of stratification" of a system and its environment. We then consider the properties of boundaries of systems contained in three common types of digraphs and conclude by showing how the concept of convexity can be used in the analysis of system-environment relationships.

### *Properties of Boundaries*

From the foregoing discussion it is evident that the processes at the boundary between a system  $S$  and its environment  $E$  involve (a) *links* between the elements of  $S$  and  $E$ , (b) *elements of  $S$*  that are in direct contact with  $E$  and (c) *elements of  $E$*  that are in direct contact with  $S$ . The formal definition of boundary to be presented below makes use of the corresponding graph theoretical terms (arcs of  $D$ , points of  $S$  and points of  $E$ ), but its statement requires some preliminary definitions which permit a distinction between links from  $S$  to  $E$  and those from  $E$  to  $S$ .

For any system  $S$  within a digraph  $D$ , its set of *out-liaisons*, denoted  $L_o$ , consists of those arcs  $v_i v_j$  of  $D$  from a point  $v_i$  of  $S$  to a point  $v_j$  of  $E$ , and its set  $L_i$  of *in-liaisons* consists of the arcs to a point of  $S$  from a point of  $E$ . Thus,

every out-liaison is of the form  $s_i e_j$  and every in-liaison is of the form  $e_j s_i$ . The out-liaisons and in-liaisons, taken together, constitute the set  $L$  of liaisons of  $S$ . The out-frontier  $F_o$  of  $S$  is the set of first points of all out-liaisons of  $S$ , the in-frontier  $F_i$  is the set of second points of all in-liaisons, and the frontier  $F$  is the union of  $F_o$  and  $F_i$ . The out-neighborhood  $N_o$  of  $S$  is the set of second points of all out-liaisons of  $S$ , the in-neighborhood  $N_i$  is the set of first points of all in-liaisons and the neighborhood  $N$  is the union of  $N_o$  and  $N_i$ .

With this background, we may now state a formal definition of the boundary of a system. The boundary  $B(S, D)$  of a system  $S$  within a digraph  $D$  is the subgraph of  $D$  induced by the liaisons of  $S$ . In digraph theory, a bipartite digraph has its points partitioned into two subsets,  $V_1$  and  $V_2$ , such that every arc joins a point of  $V_1$  with a point of  $V_2$ . Thus, the boundary  $B(S, D)$  is that bipartite subgraph of  $D$  in which the points of one set are all the points in the frontier of  $S$ , the points of the second set are all those in the neighborhood of  $S$  and the arcs are the liaisons of  $S$ .

In Figure 2, the boundary  $B(S, D)$ , for system  $S = \{s_1, s_2, s_3, s_4\}$ , is displayed below the digraph  $D$ . It can readily be seen that its two sets of liaisons  $L_o$  and  $L_i$ , its out-frontier  $F_o$  and its in-frontier  $F_i$ , and its out-neighborhood  $N_o$  and in-neighborhood  $N_i$  are  $L_o = \{s_3 e_1, s_4 e_2\}$ ,  $L_i = \{e_1 s_3, e_3 s_1\}$ ,  $F_o = \{s_3, s_4\}$ ,  $F_i = \{s_1, s_3\}$ ,  $N_o = \{e_1, e_2\}$ ,  $N_i = \{e_1, e_3\}$ . It will be noted that in this particular case the point  $s_3$  lies in both the out-frontier and in-frontier of  $S$  and that  $e_1$  lies in both the out-neighborhood and in-neighborhood of  $S$ .

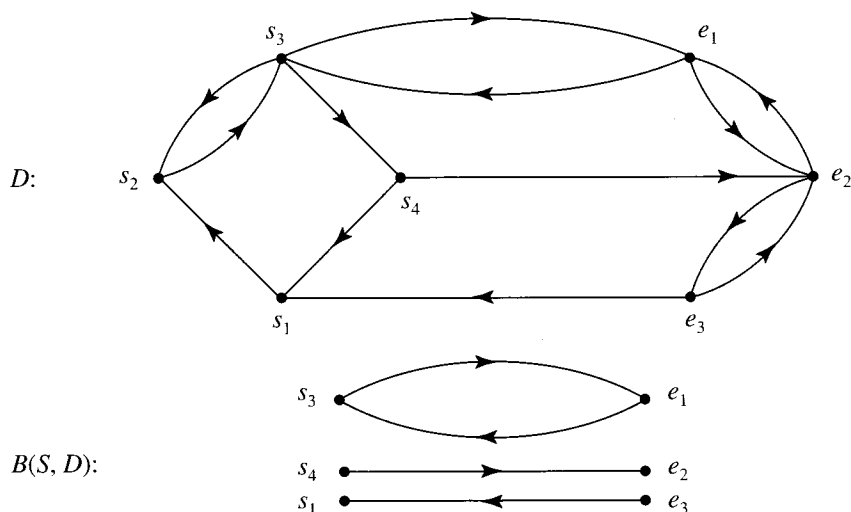


Figure 2. A system  $S$  and its boundary  $B(S, D)$ .

If the digraph of Figure 2 is interpreted as an interpersonal communication structure, then we see that the boundary of subgroup  $S$  is made up of (a) its members  $s_1$ ,  $s_3$  and  $s_4$  who are located in the frontier of  $S$ , (b) nonmembers  $e_1$ ,  $e_2$  and  $e_3$  who are in its neighborhood and (c) communication links  $s_3e_1$ ,  $s_4e_2$ ,  $e_1s_3$  and  $e_3s_1$  which join members and nonmembers. The out-frontier of the subgroup consists of all members who can communicate directly to at least one nonmember, and the in-frontier consists of those members who can receive a message directly from the outside. The out-neighborhood is composed of those people in the environment of the subgroup who can receive a message directly from a member, and its in-neighborhood is composed of those in the environment who can transmit one directly to a member. The out-liaisons consist of those communication links by which a message can leave  $S$ , whereas its in-liaisons are those by which a message can enter  $S$ .

If, on the other hand, the digraph is interpreted as the career structure of an organization, then  $B(S, D)$  represents the boundary of department  $S$ . Its out-frontier consists of those positions from which a person may leave the department, and its in-frontier consists of those positions at which a person may enter. Its out-neighborhood consists of the positions outside the department to which a member may move, and its in-neighborhood consists of those from which a person may enter the department. Finally, the out-liaisons indicate permissible changes of position that take a person out of the department, and the in-liaisons indicate those by which a person may enter.

We shall refer to the sets  $F_o, F_i, F, N_o, N_i, N, L_o, L_i$  and  $L$  as the *boundary sets* of  $B(S, D)$  and denote their cardinalities by  $f_o, f_i, f, n_o, n_i, n, l_o, l_i$  and  $l$ , respectively. The minimum cardinality of each of these sets is clearly zero, in which case the set is empty. If we denote the number of points in  $S$  and  $D$  by  $m$  and  $p$ , then for the frontier sets  $F_o, F_i$  and  $F$  their maximum cardinality is  $m$ , for the neighborhood sets  $N_o, N_i$  and  $N$  it is  $p - m$ , for the liaison sets  $L_o$  and  $L_i$  it is  $m(p - m)$  and for  $L$  it is  $2m(p - m)$ .

The *absolute size* of a boundary set is defined as its cardinality. Loosely speaking, a system may be said to be in closer contact with its environment the larger the size of its frontier, neighborhood, or liaison set. The *relative size* of  $F_o, F_i$  and  $F$  is defined as  $f_o/m, f_i/m$  and  $f/m$ . These ratios vary between 0 and 1 and indicate the proportion of elements of  $S$  that are in its out-frontier, in-frontier and frontier, respectively. The corresponding indices of the *relative size* of  $N_o, N_i$  and  $N$  are  $n_o/(p - m), n_i/(p - m)$  and  $n/(p - m)$  and have analogous meanings. The *relative size* of  $L_o, L_i$  and  $L$  is given by  $l_o/m(p - m), l_i/m(p - m)$  and  $l/2m(p - m)$ . All of these ratios equal zero when all of the boundary sets are empty and all equal one when  $S$  is in maximal contact with its environment. For the structure shown in Figure 2, we see that the relative sizes of  $F_o, F_i$  and  $F$  are  $2/4, 2/4$  and  $3/4$ , the relative sizes of  $N_o, N_i$  and  $N$  are  $2/3, 2/3$

and  $\frac{3}{3}$ , the relative sizes of  $L_o$  and  $L_i$  are each  $\frac{2}{12}$  and the relative size of  $L$  is  $\frac{4}{24}$ .

If the cardinality of each boundary set of a system is maximum, then every point of  $S$  is in both its out-frontier and in-frontier, every point of  $E$  is in both the out-neighborhood and in-neighborhood of  $S$  and there is an out-liaison from every point of  $S$  to every point of  $E$  and an in-liaison to every point of  $S$  from every point of  $E$ . In other words, every element of  $S$  can engage in direct two-way communication with every element of  $E$ .

When all of the boundary sets of  $S$  are empty, we say that  $S$  has no boundary. Such a situation obviously arises when  $S = V$ , since  $S$  then has no environment within  $D$ . And it can be easily shown that  $S$  has no boundary if and only if  $S$  is a union of weak components of  $D$ . A system with no boundary is completely self-contained with respect to a given structure either because it has no environment within the structure or because it has no contact with its environment.

If none of the boundary sets of  $S$  is empty, then the frontier  $F$  of  $S$  is as small as possible when it contains just one point  $s$ . It follows at once that both  $F_o$  and  $F_i$  must consist of the single point  $s$ . Similarly, its neighborhood  $N$  is of minimum size when it contains just one point  $e$ , which is then also the only point in  $N_o$  and  $N_i$ . When both of these conditions are met, the boundary consists of one point  $s$  in  $S$ , the point  $e$  in  $E$ , and the symmetric pair of liaisons joining them. Such a structure is shown in Figure 1.

It can be readily seen that if  $L_o$  is empty, then so are  $F_o$  and  $N_o$ . In this case, no message originating in the system can reach the environment. An example of this sort is provided in the theory of Markov chains, where points correspond to "states" and there is an arc  $v_i v_j$  in  $D$  if and only if the probability that there will be a direct transition from state  $v_i$  to  $v_j$  is greater than zero. A subset  $S$  of the points of a Markov chain is said to be "closed" if  $S$  has no out-liaisons (see Feller, 1957: 349). Thus, when a Markov chain is closed, there can be no transition from a state in  $S$  to one not in  $S$ . If such a system has any liaisons, its boundary consists of the nonempty sets  $L_i$ ,  $F_i$  and  $N_i$ . All of its liaisons thus go from  $E$  to  $S$  and we say that the boundary of  $S$  has an *inward orientation*. Figure 3 shows a system whose boundary has such an orientation.

If the boundary of a system contains out-liaisons but no in-liaisons, we say that it has an *outward orientation*. Since, in this case, all liaisons go from  $S$  to  $E$ , it follows that  $S$  can influence its environment but not be influenced by it.

We conclude this section by considering two indices for characterizing each element  $s$  of a system  $S$ . The *out-degree*  $od_D(v)$  of a point  $v$  in a digraph  $D$  is the number of arcs from  $v$ , and the *in-degree*  $id_D(v)$  is the number of arcs to  $v$ . We employ the notation  $od_s(s)$  and  $id_s(s)$  to denote the out-degree and in-degree of a point  $s$  within the subgraph  $S$ . These terms give the number of internal arcs that terminate at  $s$ . Similarly, the notation  $od_B(s)$  and  $id_B(s)$  refers to the

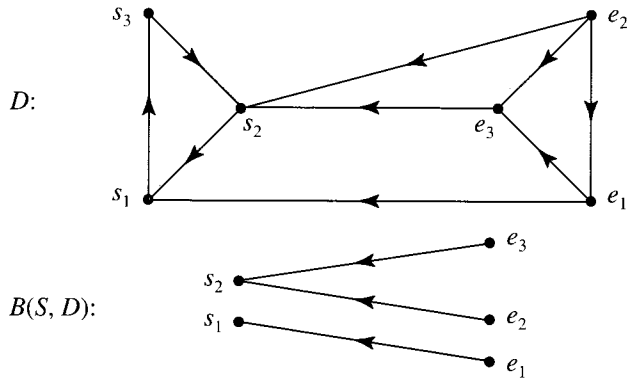


Figure 3. A boundary with an inward orientation.

boundary subgraph  $B(S, D)$  and refers to the liaisons incident with  $s$ . Since every arc of  $D$  that is incident with  $s$  is either an internal arc or a liaison, it follows that:

$$od_s(s) + od_B(s) = od_D(s) \quad \text{and}$$

$$id_s(s) + id_B(s) = id_D(s)$$

The *index of out-liaison concentration of a point  $s$*  is the ratio  $od_B(s)/od_D(s)$ . This index gives the proportion of arcs originating at  $s$  that are out-liaisons and thus the proportion of points adjacent to  $s$  that lie in the environment of  $S$ . Clearly, it is greater than zero if and only if  $s$  is in the out-frontier of  $S$ . The *index of in-liaison concentration of  $s$*  is the ratio  $id_B(s)/id_D(s)$  and has an analogous meaning. Referring back to the system shown in Figure 1, we see that the indices of out-liaison and in-liaison concentration for  $s_1$  are both  $1/3$ . Thus, this element lies in both the in-frontier and out-frontier of  $S$ , and  $1/3$  of its contacts are outside of  $S$ . On the other hand, these indices are both  $0/2$  for  $s_2$  and  $s_3$  whose contacts lie entirely within  $S$ .

### Stratification of a System and Its Environment

In the discussion thus far, a distinction has been made between two types of points within a system  $S$ : frontier points (those adjacent with a point of  $E$ ) and inner points (all other points of  $S$ ). We now develop a more general method by which the points of  $S$  can be assigned to strata according to their proximity to

*E*. To do so, we need two directionally dual definitions of point-set distance. Let  $U$  be a subset of the set of points of  $D$ . Then the *distance*  $d(v, U)$  from a point  $v$  to a set  $U$  is the minimum distance from  $v$  to a point in  $U$ , and the *distance*  $d(U, v)$  from  $U$  to  $v$  is the minimum distance from a point  $u$  in  $U$  to  $v$ .<sup>3</sup> From a point  $s$  of  $S$ , the distance  $d(s, N_o)$  is thus the minimum distance from  $s$  to any point in the out-neighborhood of  $S$ , and the distance  $d(N_i, s)$  is the minimum distance to  $s$  from any point in the in-neighborhood of  $S$ . And since these distances are clearly the shortest distances between  $s$  and any point of  $E$ , they may be thought of as the distances between  $s$  and the environment of  $S$ , explicitly,  $d(s, N_o) = d(s, E)$  and  $d(N_i, s) = d(E, s)$  for all points  $s$  of  $S$ .

A *geodesic from a point  $v$  to a set  $U$*  is a path of minimum length from  $v$  to any point  $u$  in  $U$ . Thus, if  $s$  is a point in  $S$ , then each geodesic from  $s$  to  $N_o$  is a shortest path from  $s$  to any point in its out-neighborhood. A geodesic from a set of points to a single point is defined similarly. Thus each geodesic from  $N_i$  to  $s$  is a shortest path from any point in  $N_i$  to  $s$ . Clearly,  $d(s, N_o) = k$  if and only if  $D$  contains a geodesic from  $s$  to  $N_o$  of length  $k$ , and  $d(N_i, s) = k$  if and only if  $D$  contains a geodesic from  $N_i$  to  $s$  of length  $k$ .

It will be recalled that the boundary  $B(S, D)$  is the subgraph of  $D$  that is induced by the liaisons of  $S$ . Let us now consider two analogous subgraphs of  $D$  that are induced by the geodesics from  $s$  to  $N_o$  and from  $N_i$  to  $s$ , respectively. The *out-stratification subgraph*  $Z_o(S, D)$  is the subgraph of  $D$  induced by the geodesics from  $s$  to  $N_o$ , for all points  $s$  in  $S$ , and the *in-stratification subgraph*  $Z_i(S, D)$  is the subgraph of  $D$  induced by the geodesics from  $N_i$  to the points  $s$  in  $S$ . Thus the point set of  $Z_o(S, D)$  consists of  $N_o$  and the set  $S_o$  of all points of  $S$  that can reach  $N_o$ , and its arc set consists of just those contained in a geodesic from a point of  $S_o$  to the out-neighborhood  $N_o$ . Similarly, the point set of  $Z_i(S, D)$  consists of  $N_i$  and the set  $S_i$  of all points of  $S$  that are reachable from  $N_i$ , and its arc set consists of just those contained in a geodesic from  $N_i$  to a point  $s$  of  $S_i$ . It should be noted that the boundary  $B(S, D)$  is a subgraph of the union of  $Z_o(S, D)$  and  $Z_i(S, D)$ . Figure 4 shows the stratification subgraphs for the system  $S = \{s_1, s_2, s_3, s_4, s_5\}$ .

We now define the  *$k$ 'th out-stratum of  $S$* , denoted  ${}_kS_o$ , as the subset of the points of  $S$  that can reach a point of  $N_o$  for which  $d(s, N_o) = k$ , and the  *$k$ 'th in-stratum of  $S$* ,  ${}_kS_i$ , as the subset of the points  $s$  of  $S$  that are reachable from a point of  $N_i$  for which  $d(N_i, s) = k$ . Thus, the  $k$ 'th out-stratum consists of those points of  $S$  that can reach a point of  $N_o$  by a geodesic of length  $k$  but not by any shorter path. And the  $k$ 'th in-stratum consists of those points of  $S$  that can be reached from a point of  $N_i$  by a geodesic of length  $k$  but not by any shorter

<sup>3</sup>The definition of point-set distance gives precise meaning to the concept of distance from a point to the outside, which Lewin (1941) employed to characterize the stratification of "natural wholes."

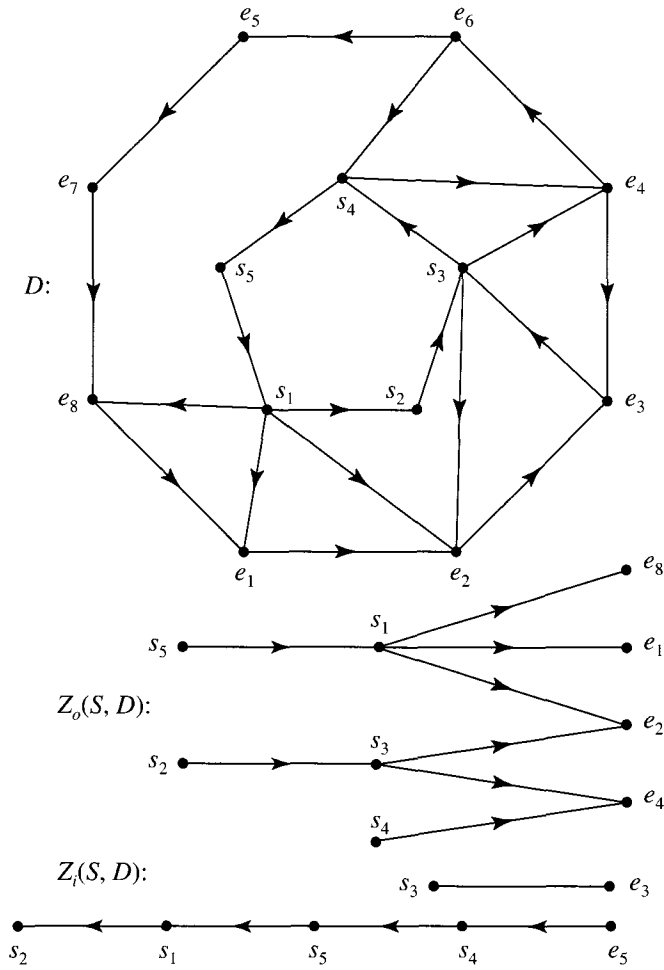


Figure 4. The stratification subgraphs  $Z_o(S, D)$  and  $Z_i(S, D)$  of a system  $S$ .

path. Clearly, the first out-stratum  ${}_1S_o$  is the out-frontier of  $S$ , and  ${}_1S_i$  is the in-frontier of  $S$ . It should be noted that no point of  $S$  lies in more than one out-stratum (or in-stratum), but there may be points that are not in any out-stratum because they cannot reach a point of  $N_o$  (or not in any in-stratum because they cannot be reached from a point of  $N_i$ ). If, however,  $S$  is strong, then both the out-strata and the in-strata do partition the points of  $S$ .

The *degree of out-stratification* of  $S$  is the number of its out-strata, and the

*degree of in-stratification of S* is the number of its in-strata. Clearly, the degree of out-stratification and the degree of in-stratification for a given system  $S$  need not be the same. If  $S$  contains  $m$  points, both of these indices may vary between 0 and  $m$ . The degree of out-stratification of  $S$  is 0 if and only if the out-frontier  $F_o$  is empty. It is 1 if and only if every point of  $S$  is in  $F_o$ . And it is  $m$  if and only if the following conditions are satisfied: (1)  $F_o$  consists of a single point  $s$ ; (2) there exists a path  $W$  to  $s$  containing all points of  $S$  and no other ones and (3) every path from a point of  $S$  to  $s$  is a subpath of  $W$ . These three conditions together ensure that each point of  $S$  constitutes an out-stratum, and conversely. Corresponding statements about the degree of in-stratification can readily be formulated.

Referring again to Figure 4, we see that  $S$  has two out-strata,  ${}_1S_o = F_o = \{s_1, s_3, s_4\}$  and  ${}_2S_o = \{s_2, s_5\}$ . Thus, its degree of out-stratification is 2. And since every point of  $S$  is contained in  $Z_o(S, D)$ , they can all reach a point of  $E$  in two steps or less along the geodesics contained in  $Z_o(S, D)$ . On the other hand, the degree of in-stratification of  $S$  is 4, and its in-strata are  ${}_1S_i = F_i = \{s_3, s_4\}$ ,  ${}_2S_i = \{s_3\}$ ,  ${}_3S_i = \{s_1\}$  and  ${}_4S_i = \{s_2\}$ . And since every point of  $S$  is contained in  $Z_i(S, D)$ , they can all be reached by a point of  $E$ , but here 4 steps are required.

Let us assume that this digraph represents an interpersonal communication structure in which one unit of time is required for a message to traverse each communication link. Then each member in the  $k$ 'th out-stratum of subgroup  $S$  can send a message to some nonmember in no more than  $k$  units of time, and everyone can do so in at most 2 units. By similar reasoning, it requires at least 4 units of time for a message to reach every member of the subgroup from the outside. If we make the further assumption that a message undergoes a unit of distortion each time it traverses a communication link, then any message originating with a member in the  $k$ 'th out-stratum will have undergone at least  $k$  units of distortion when it first reaches a nonmember. And any message originating with a nonmember will have undergone at least  $k$  units of distortion when it reaches a member in the  $k$ 'th in-stratum. Other implications of the stratification of a system can be readily derived by assigning other meanings to the terms point, arc and message, or by postulating other consequences of the transmission of messages.

It is evident that the points in the environment of a system can be stratified in a similar fashion. The  $k$ 'th out-stratum of  $E$  (relative to  $S$ ), denoted  ${}_kE_o$ , is the subset of the set of points of  $E$  that are reachable from a point of  $F_o$  for which  $d(F_o, e) = k$ , and the  $k$ 'th in-stratum of  $E$ ,  ${}_kE_i$ , is the subset of the points of  $E$  that can reach a point of  $F_i$  for which  $d(e, F_i) = k$ . The *degree of out-stratification* and the *degree of in-stratification of E* are the number of its out-strata and in-strata, respectively. For the digraph shown in Figure 4, the degree of out-stratification of  $E$  is 4 and its degree of in-stratification is 6.

### *Boundaries in Different Kinds of Digraphs*

The concepts discussed thus far are applicable to digraphs in general. If, however, it is known that  $D$  is a digraph of some particular type, certain implications may be drawn concerning the boundary properties of systems contained within  $D$ . We now consider some of these implications for three common types of structures: symmetric digraphs, transitive digraphs and signed graphs.

For this purpose, we need to introduce some additional basic concepts of graph theory. A *semipath joining*  $v_1$  with  $v_n$  is a collection of distinct points,  $v_1, v_2, \dots, v_n$ , together with  $n - 1$  arcs, one from each pair  $v_1v_2$  or  $v_2v_1, v_2v_3$  or  $v_3v_2, \dots, v_{n-1}v_n$  or  $v_nv_{n-1}$ . A digraph  $D$  is *weak* if every two points of  $D$  are joined by a semipath, and it is *strong* if every two points are mutually reachable. For any digraph  $D$ , a subgraph of  $D$  is *maximal* with respect to some property if  $D$  has no larger subgraph containing it which has this property. A *weak component* of  $D$  is a maximal weak subgraph. A *strong component* is a maximal strong subgraph. And a *clique* is a maximal complete symmetric subgraph.

#### SYMMETRIC DIGRAPHS

An arc  $uv$  of a digraph  $D$  is said to be *symmetric* if the arc  $vu$  is also in  $D$ ; otherwise, it is said to be *asymmetric*. A *symmetric digraph* is one containing only symmetric arcs. Thus, for example, a communication structure is symmetric if all of its links are two-way, and a sociometric structure is symmetric if all choices are mutual. It is evident that if two points,  $u$  and  $v$ , are joined by a semipath in a symmetric digraph, they are also joined by two paths of the same length, one from  $u$  to  $v$  and one from  $v$  to  $u$ . Thus, every weak component of a symmetric digraph is strong, and the distance  $d(u, v)$  equals  $d(v, u)$  for any two points of  $D$ .

Let us now consider a system  $S$  within a symmetric digraph  $D$ . Clearly, every out-liaison  $s_i e_j$  of  $S$  has a corresponding in-liaison  $e_j s_i$ , and conversely. Thus every liaison of  $S$  is a symmetric arc. It follows immediately that every point in the frontier  $F$  of  $S$  is in both its out-frontier  $F_o$  and in-frontier  $F_i$ , and every point in  $N$  is in both  $N_o$  and  $N_i$ . And since the distance between every pair of points is symmetric, the  $k$ 'th out-stratum  ${}_k S_o$  and the  $k$ 'th in-stratum  ${}_k S_i$  contain exactly the same set of points, so that the degrees of out-stratification and in-stratification are the same. Similar observations obviously hold for the stratification of the environment of  $S$ .

If  $D$  is a symmetric digraph and  $S$  is a strong component of  $D$ , then it can easily be shown that the boundary  $B(S, D)$  must be empty. Thus, if a system  $S$  is a strong component of a symmetric structure, then messages can go, directly

or indirectly, from each element to every other element of  $S$  but no messages can go in either direction between an element of  $S$  and an element of  $E$ .

#### TRANSITIVE DIGRAPHS

A digraph  $D$  is *transitive* if it contains an arc  $uw$  whenever arcs  $uv$  and  $vw$  are in  $D$ , for any distinct points  $u$ ,  $v$  and  $w$ . Thus, for example, an interpersonal power structure is transitive if, for any three people, person  $u$  has power over  $w$  whenever  $u$  has power over  $v$  and  $v$  has power over  $w$ . And a sociometric structure is transitive if all members follow the principle: "A friend of my friend is my friend." A useful characterization of transitive digraphs is given in the following theorem (Harary, Norman and Cartwright, 1965: 120):

**Theorem I.** *A digraph is transitive if and only if it has the following properties:*

- (1) *Every strong component  $S$  is maximal complete symmetric (that is, a clique).*
- (2) *There is an arc from every point in clique  $S_i$  to every point in  $S_j$  whenever there is a path from any point in  $S_i$  to any point in  $S_j$ .*

Some of the boundary properties of cliques in transitive digraphs may now be specified:

**Corollary 1a.** *Let  $D$  be a transitive digraph. Then*

- (1) *Every symmetric arc of  $D$  is an internal arc of some clique  $S$ , and every asymmetric arc is a liaison of some clique.*
- (2) *The out-frontier of any clique  $S$  is either empty or contains every point of  $S$ , and the same holds for its in-frontier.*
- (3) *The out-neighborhood of any clique  $S$  consists of all points in its environment that are reachable from any point of  $S$ , and the in-neighborhood consists of all points in  $E$  that can reach any point of  $S$ .*

**Corollary 1b.** *Let  $S_i$  and  $S_j$  be any two cliques of a transitive digraph. Then, in the subgraph induced by these two cliques the boundary of  $S_i$  satisfies exactly one of the following conditions: (1) it is empty; (2) it has an outward orientation and contains an out-liaison from every point of  $S_i$  to every point of  $S_j$ ; (3) it has an inward orientation and contains an in-liaison from every point of  $S_j$  to every point of  $S_i$ .*

The *condensation*  $D^*$  of a digraph  $D$  is the digraph whose points correspond to the strong components  $S_1, S_2, \dots, S_n$  of  $D$  and whose arcs are determined as

follows: there is an arc from point  $S_i$  to  $S_j$  in  $D^*$  if and only if there is at least one arc from a point of  $S_i$  to one of  $S_j$  in  $D$ . Clearly, every condensed digraph  $D^*$  is acyclic (and hence asymmetric), and each point  $S_i$  of  $D^*$  can therefore be assigned a level  $b_i$  such that  $b_i < b_j$  for each arc  $S_i S_j$  in  $D^*$  (Harary, Norman and Cartwright, 1965:268). Now, if  $D$  is transitive, then each point  $S_i$  of  $D^*$  corresponds to a clique  $S_i$  of  $D$ , and the level  $b_i$  of  $S_i$  in  $D^*$  gives the level of its associated clique in  $D$ . Thus, for any clique  $S_i$  of a transitive digraph, each of its internal arcs joins two points at the same level, each of its out-liaisons goes to a point at a higher level and each of its in-liaisons comes from a point at a lower level. And since the condensation of a transitive digraph is transitive and asymmetric, we see that a transitive structure may be characterized as a partial order (or "hierarchy") of cliques.

It has been proposed by Davis (1967), Davis and Leinhardt (1970) and Holland and Leinhardt (1970; 1971) that sociometric choices display a tendency toward transitivity. If this tendency is fully realized by a given group of people, the resulting sociometric structure has the properties discussed above. The group thus consists of subgroups in which all members mutually choose each other. These subgroups can be assigned levels (perhaps indicative of their sociometric "status") such that all choices between members of different subgroups are unreciprocated and go from a lower to a higher level. And for every pair of subgroups  $S_i$  and  $S_j$ , either every member of  $S_i$  chooses every member of  $S_j$  and no member of  $S_j$  chooses any member of  $S_i$ , or there are no choices whatsoever between members of  $S_i$  and  $S_j$ .

The properties of transitive digraphs make it clear that as the number of people in a sociometric structure increases, the attainment of transitivity may require an increasingly large number of choices on the part of some people. If there are  $m$  members of a clique  $S$ , then each member of  $S$  must choose the  $m - 1$  other members of  $S$ . And if there are  $r$  people in the environment of  $S$  who can be reached from any member of  $S$ , then every member of  $S$  must also choose  $r$  members of  $E$ . Thus the quantity  $r + m - 1$  gives the number of choices required of each member of a clique  $S$  in a transitive structure. And it is clear that this number may be quite large if  $S$  is large or is located at a lower level of the structure. We might expect, therefore, that even if there is a tendency toward transitivity in such structures, they may not always contain all of the arcs required by transitivity due to some empirical constraint on the number of choices made by a given person.

This observation suggests that it would be useful to have a characterization of digraphs that approximate, but do not necessarily fully satisfy, the property of transitivity. We shall say that digraph is *quasi-transitive* if it has at least one arc from a point of a strong component  $S_i$  to one of  $S_j$  whenever there is a path from a point of  $S_i$  to one of  $S_j$ . Clearly, every transitive digraph is quasi-transitive, but not conversely. It can be easily shown that the condensation  $D^*$  of