

5

Graph Theoretical Dimensions of Informal Organizations

David Krackhardt
Carnegie Mellon University

In his classic work on the architecture of complexity, Simon (1981) noted the uncanny presence of hierarchy in virtually all complex systems. He further argued that there was a universal function to such hierarchical forms: They are efficient and robust against disruptions that might threaten the cybernetic goals of the system. And although formal organizational charts are obviously hierarchical, he argued that informal organizations also would be found to be hierarchically structured: "If we make a chart of social interactions, of who talks to whom, the clusters of dense interaction in the chart will identify a rather well-defined hierarchic structure. The groupings in this structure may be defined operationally by some measure of frequency of interaction in this sociometric matrix" (Simon, 1981, p. 197).

This idea that informal organizations will naturally evolve into a hierarchical structure is intriguing and has intuitive appeal. The theme can be found with empirical support elsewhere. For example, Michels (1915) noted that even democratically based voluntary organizations evolve toward a ~~centralized, hierarchical structure as they grow. Guetzkow and Simon~~ (1955) discovered that small groups that are allowed unlimited choice of communication channels tend to centralize their communication flows into a hierarchical "wheel" structure.

The normative part of Simon's claim, that hierarchy exists because it allows the system to operate more efficiently and survive outside disturbances, is also appealing. From this, we may deduce a hypothesis about the structure of informal organizations and performance.

*Krackhardt, David

89

1994 "Graph Theoretical Dimensions of Informal Organizations." In Kathleen Carley & Michael Prietula (eds.), Computational Organizational Theory. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc. Pp. 89-111.

Simon's model of hierarchy raises three unresolved issues, two theoretical and one methodological. First, as Simon admitted himself (1981, p. 213), people do communicate outside the preferred boundaries defined by the formal hierarchy of the organization. That is, perfectly hierarchical informal organizations are rare. Nonetheless, he argued, these exceptional communication links are relatively limited and not consequential to the overall pattern of hierarchy in the organization.

The second theoretical problem with Simon's model is that it flies in the face of several normative theories of organizational structure that emphasize the value of communication and information flows outside the normal, hierarchical boundaries. Burns and Stalker (1961) argued that when an organization is faced with a dynamic environment, an organic, nonhierarchical, informal structure is more appropriate for organizational effectiveness and survival. Allen (1977) demonstrated that research and development organizations can enhance their effectiveness by promoting communication outside the formal, hierarchical boundaries:

Increased communication between R&D projects and other elements of the laboratory staff were in every case strongly related to project performance. Moreover, it appears that interaction outside the project is most important. On complex projects, the inner team cannot sustain itself and work effectively without constantly importing new information from the outside world. (pp. 122-123)

And Krackhardt and Stern (1988) presented experimental evidence that under some conditions organizations are better off if they maximize strong cross-departmental relationships.

The third problem with Simon's model is one of measurement and consequent testability. Given that pure hierarchies do not exist, we must somehow differentiate between structures as being more or less hierarchical. Otherwise, we have no way of confirming or disconfirming his predictions. He offered no specifics for measuring the degree of hierarchy in social systems. In fact, systems whose elements appear to have no hierarchical structure (systems he calls *flat* hierarchies), he argued, are still hierarchical with an indefinitely large "span."

The purpose of this chapter is to broaden Simon's ideas of hierarchical structures as they pertain to informal organizations. ~~First, I dispense with~~ the assertion that informal organizations are necessarily hierarchical. Instead, I argue that this is an empirical question and an appropriate object of research. Second, I propose that hierarchical forms will have implications for organizations, but that some of the implications may not necessarily enhance the organization's efficiency or ability to survive (i.e., they are not necessarily functional). Finally, I specify a method for

measuring the degree to which the informal organization is structured in a way that Simon would call hierarchical. Because the ability to answer the former research questions depends on this last measurement issue, I begin with the development of measures of informal structure.

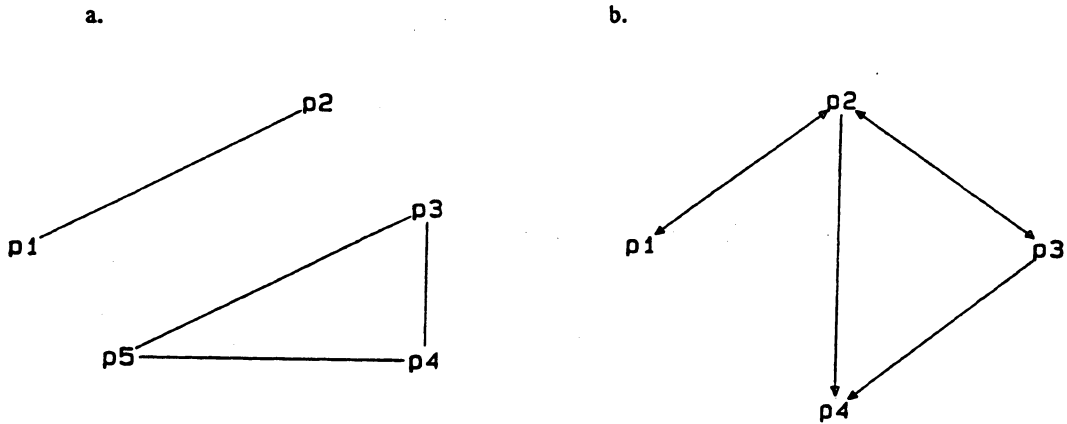
GRAPH THEORY AND MEASURES OF STRUCTURE

Graph theory (Harary, 1969) provides us with a precise language for representing structures of all forms, including the structures Simon referred to as hierarchical. Because I draw on graph theory in this chapter, I provide definitions as necessary for clarity. A *graph* (G) is defined as a set of N points $P = \{P_i\}$ and a set of unordered pairs of those points $L = \{P_i, P_j\}$; these latter elements are often referred to as lines connecting those points. In the immediate context, these points represent people in the organization, and the pairs of points represent relationships (such as interaction, communication) between those organizational members. For example, if person i interacts with person j , then the ordered pair (P_i, P_j) is included in set L that defines the relationship *interaction*.

A directed graph, or *digraph* (D), is defined as a set of points $P = \{P_i\}$ and a set of *ordered* pairs of points $L = \{P_i, P_j\}$. A digraph is used to represent relations that are potentially asymmetric, such as authority or giving advice. For example, if i is the immediate supervisor to j , and L is defined as the set of formal authority relationships, then L would contain the ordered pair (P_i, P_j) but would not contain the ordered pair (P_j, P_i) .

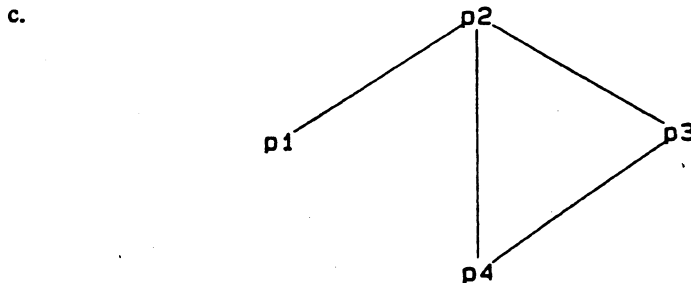
Graph theory definitions are often easier to convey by example. Figure 5.1 provides an example of two graphs and one digraph. Graphs are represented with points and lines connecting the points. Digraphs are represented by points and lines with arrowheads on them to indicate the order of the pair of points being connected. The graph in Fig. 5.1c represents a special function. It is the *underlying graph* of the digraph in Fig. 5.1b. The underlying graph of a digraph is the graph obtained by "removing the arrows" from the digraph. That is, if digraph D contains either the ordered pair (P_i, P_j) or (P_j, P_i) , then the underlying graph G will include the unordered pair (P_i, P_j) .

A point and a line are *incident* with one another if the line contains the point in its pair. In Fig. 5.1a, P_2 is incident with line (P_1, P_2) . A *walk* is an ordered sequence of alternating points and lines, starting and ending with points, such that each line is incident with the point that precedes it and with the point that follows it. A *path* is a walk with no repeating points. In Fig. 5.1c, the sequence $P_3, (P_3, P_5), P_5, (P_5, P_4), P_4$ constitutes a path from P_3 to P_4 . One point is said to be able to reach another if there exists a path that



A graph with 5 points
and 4 lines

A digraph with 4 points
and 6 lines



The underlying graph of digraph b
has 4 points and 4 lines.

FIG. 5.1. Examples of graphs and digraphs.

starts at the first point and ends at the second. All pairs of points in the graph in Fig. 5.1c are mutually reachable.

With a small restriction in the definition of incidence of points and lines, these same definitions apply to digraphs as well. In an ordered point-line pair, $[P_i, (P_j, P_k)]$, the point and line are incident with each other if $P_j = P_i$. In an ordered line-point pair, $[(P_j, P_k), P_i]$, the line and point are incident with each other if $P_k = P_i$. The definitions of paths and reachability are identical to those in graphs. In the digraph represented in Fig. 5.1b, a path exists from P_1 to P_4 but not from P_4 to P_1 . Therefore, P_1

can reach P_4 but not vice versa. In fact, P_4 cannot reach any other point, but each other point can reach P_4 .

A *connected graph* is a graph in which each point can reach every other point. Figure 5.1c is connected; Fig. 5.1a is not. A subgraph (S) of graph (G) is a graph whose points and lines are also in G. A component (C) of graph (G) is a connected subgraph of G with two characteristics: (a) All the lines in G incident to every point in C are included in C, and (b) there is no point in G not included in C that, in G, can reach a point included in C. Figure 5.1c has one only component; Fig. 5.1a has two components.

A *connected digraph* is a digraph in which each point can reach every other point in the underlying graph of the digraph. Each point in the digraph of Fig. 5.1b is reachable from every other point in the underlying graph Fig. 5.1c. Thus, the digraph in Fig. 5.1b is connected. A *component* (C) of a digraph (D) is a connected subgraph of D with the following characteristics: (a) All the lines in D incident to every point in C are included in C, and (b) there is no point in D not included in C that, in the underlying graph of D, can reach a point included in C. The digraph in Fig. 5.1b has one component only.

With these tools and definitions developed so far, it is possible to represent the informal structure of any organization. But to fully explore Simon's notions of hierarchical structure, it will be necessary to develop some additional operations. From these, one can determine the extent to which an informal structure approximates a pure hierarchical structure.

PURE HIERARCHICAL STRUCTURES: THE OUTTREE

The first task before us is to establish a pure structure as a standard against which other structures can be compared. For the purposes of this analysis, the ideal candidate for such a structure is, in graph theory terms, the *outtree*. Before a formal definition is presented, an intuition of what constitutes an outtree is provided for the reader in Fig. 5.2, which contains four examples. First, it should be noted that outtrees are digraphs. Second, every point, with the exception of the one point at the "top" of the outtree, has exactly one arrow pointing to it, although the points may have several arrows emanating from them. ~~If these arrows represented authority relationships, then we might interpret this statement as noting that each point has one and only one "boss," but each point may have any number of subordinates.~~ In fact, it should be immediately apparent to the reader that each of these figures could be examples of organizational charts—the *archetypical formal hierarchy*, as Simon termed it (1981, p. 197).

There are several reasons that the outtree serves as a reference base for

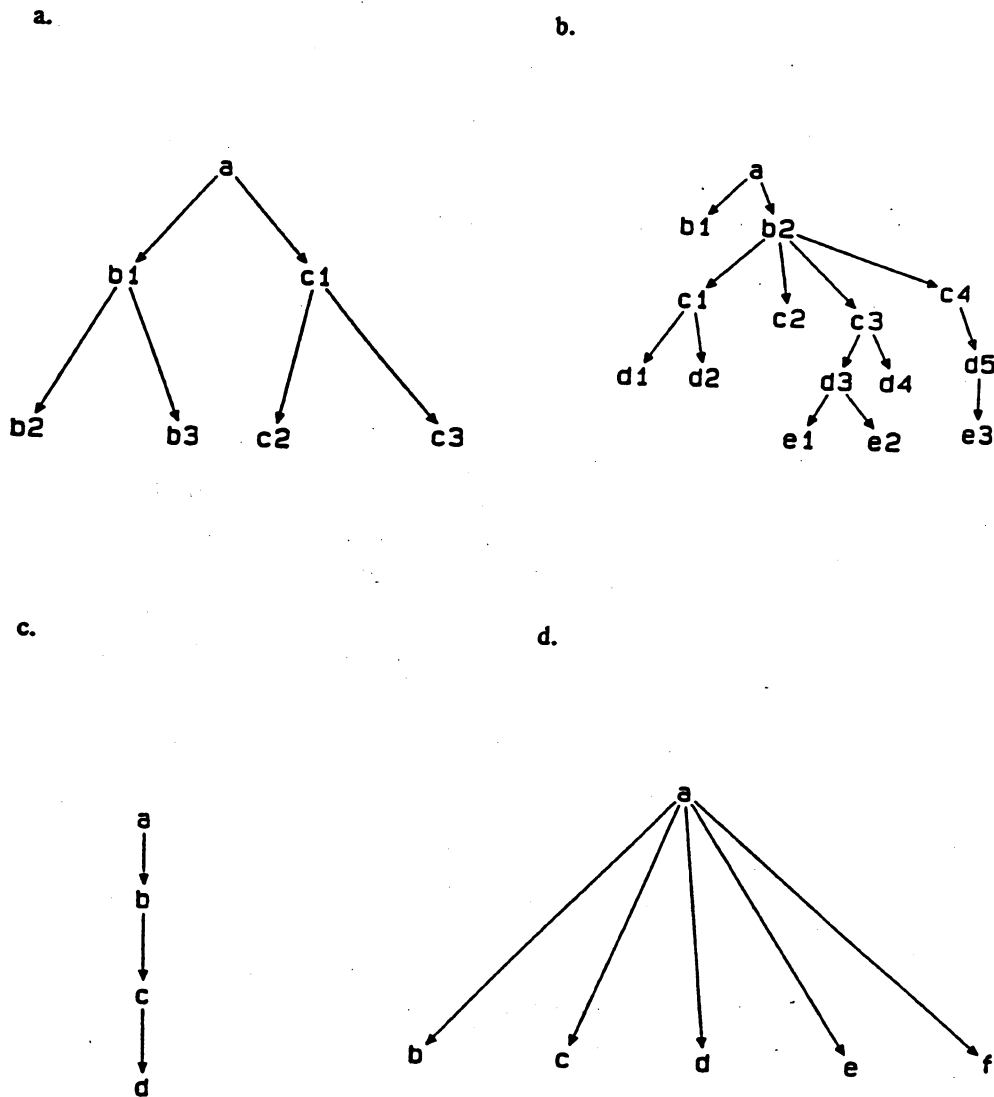


FIG. 5.2. Examples of outtrees.

the study of informal structures. First, all of Simon's hierarchical systems can be represented as an outtree.¹ Second, as mentioned earlier, they correspond to our intuition of the archetypical hierarchy, the formal organization. Third, they preserve several fundamental principles of classic organizational structure, including unity of command, unambiguous chain of command, and the scalar principle.

But this graph theory concept provides more than an archetype. It provides a basis for describing observed organizational structures and measuring their deviance from this archetype. To pursue this, it is first

¹It is worth noting at this point that most of Simon's hierarchical systems are represented by an inclusion relation, rather than the type of interpersonal relations used throughout this chapter.

necessary to formally define an outtree using more graph theory. There are many ways to so define an outtree (e.g., Wilson, 1979, p. 45). For reasons that become clear shortly, I use the following four conditions of a digraph as a definition of an outtree. These conditions are both necessary and sufficient for the digraph to be an outtree:

1. The digraph is connected.
2. The digraph is graph hierarchic.
3. The digraph is graph efficient.
4. Every pair of points in the digraph has a least upper bound.

If a graph violates any of these four conditions, then it is not an outtree. Moreover, we can count the number of violations in each of the dimensions to give us a measure of distance from the archetypal structure. Because these violations are based on independent criteria, the picture of the structure described by each dimension differs considerably. Figure 5.3 displays some examples of these differences. In the center of the figure is an outtree. Each of the other four figures surrounding the outtree represents an extreme case where one (and only one) of each of the four conditions is violated to the maximum extent possible. It is useful to refer to this figure as each of the four dimensions is defined next.

Each of the four measures of degree of structure is based on the number of outtree violations that exist in any particular structural arrangement. As such, each condition becomes a *dimension* of structure, continuously varying in value from 0 to 1. That a graph has a value of 1.0 on all four dimensions is equivalent to stating that a graph is an outtree. Also, each of the four dimensions has different implications for the organization. Each of these dimensions and its implications for the organization is next described in turn.

1. Connectedness. The definition of connectedness was already provided earlier: A digraph is connected if each point can reach every other point in the underlying graph. To say that a digraph is *disconnected* implies that there are at least two components in the digraph. The degree to which the digraph is disconnected is a function of the number of violations of the connectedness condition. A violation is defined as a point being unable to reach another point in the underlying graph. If we divide the number of violations by the maximum number of possible violations (i.e., in the case where no point can reach any other point), we have a continuum representing the degree to which the graph is disconnected. Subtracting this ratio from 1 gives us the degree of connectedness in the structure. The *degree of connectedness* is then defined as:

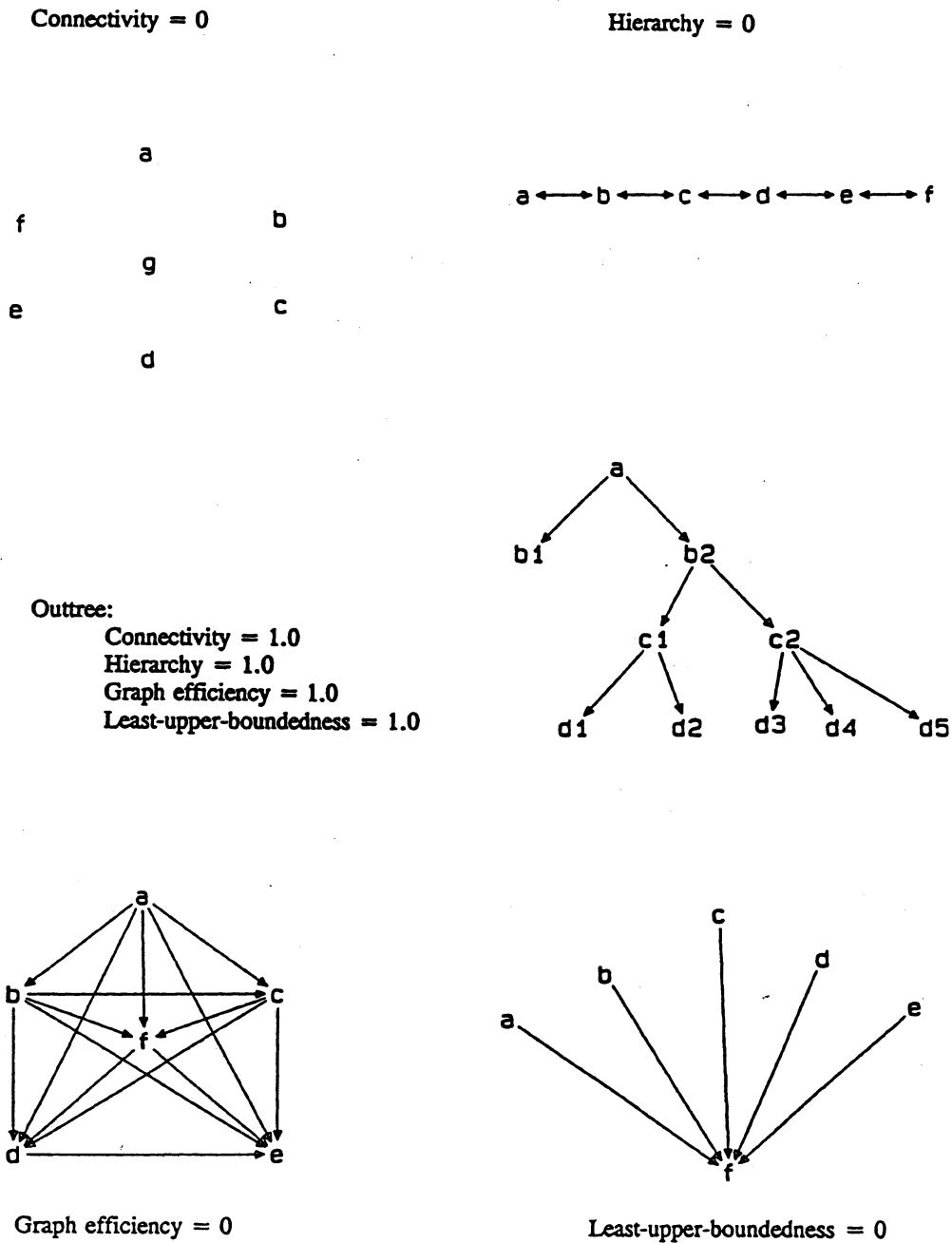


FIG. 5.3. The four dimensions of structure.

$$\text{Connectedness} = 1 - \left[\frac{V}{N(N-1)/2} \right]$$

where V is the number of pairs of points that are not mutually reachable, and the maximum number of violations is the total number of pairs of points = $N(N-1)/2$.

The degree of connectedness in a set of social relations is the simplest of the measures. At one end of the spectrum, an outtree is completely

connected. A disconnected graph represents a division in the social system. The more people are separated from each other, the more difficult it is to organize them through the network. At the extreme, no one is connected to anyone (connectedness = 0); everyone is an independent actor.

If the task facing the organization is routine, and the environment does not change, then connectedness may not be essential to the performance of the organization in its task. But if the organization has many exceptions that require consultation, a set of established communication and advice relations that incorporates all actors, at least indirectly, would be essential. Also, lack of connectedness may be a reflection of a major political division, such that one side does not talk to the other side(s).

2. *Graph Hierarchy.*² The *graph hierarchy* condition states that in a digraph D , for each pair of points where one (P_i) can reach another (P_j), the second (P_j) cannot reach the first (P_i). For example, in a formal organizational chart, a high-level employee can "reach" through the chain of command her subordinate's subordinate. If the formal organization is working properly, this lower level employee cannot simultaneously "reach" (i.e., cannot be the boss of a boss of) the higher level employee.

To measure the degree of hierarchy of digraph D , a new digraph D_r must be created. D_r is defined as the *reachability digraph* of D . Each point in D exists in D_r ; moreover, the line (P_i, P_j) exists in D_r if and only if P_i can reach P_j in D . If D is graph hierarchic, then D_r will have no symmetric lines in it. That is, if the line (P_i, P_j) exists in D_r then the line (P_j, P_i) does not. A violation to this condition exists every time a symmetric line exists in D_r . The degree of hierarchy, then, is defined as:

$$\text{Graph hierarchy} = 1 - \left[\frac{V}{\text{Max}V} \right]$$

where V is the number of unordered pairs of points in D_r that are symmetrically linked (that is, where P_i is linked to P_j and P_j is linked to P_i), and $\text{Max}V$ is the number of unordered pairs of points in D_r where P_i is linked to P_j or P_j is linked to P_i .

Graph hierarchy exists to the extent that the relations are strictly ordered. For example, hierarchy occurs if relations are determined by status, prestige, or formal authority. Informal relations, such as advice relations, can be ordered, but are not necessarily so. An outtree (such as the organizational chart) is perfectly hierarchical. At the other extreme, if there

²The term *hierarchy* is used differently here than by Simon. Nothing in Simon's work specifies asymmetry of relations. However, the term's use in this chapter corresponds more closely to the common use of the word. From here on, I will use the term *hierarchy* to refer to graph hierarchy rather than Simon's definition.

is no status in a system, then no graph hierarchy is likely to emerge in the informal relations.

A mechanistic organization is likely to be very status ridden (Shrader, Lincoln, & Hoffman, 1989). Members are likely to go up the organization for advice. To the extent that this is true, a mechanistic organization will be characterized by a high degree of hierarchy in advice relations. In an organic organization, on the other hand, status is more diffuse, and project leaders may not hesitate to seek advice from subordinates or someone from a different group in the organization. In such an environment, advice relations might not be hierarchically arranged at all.

3. Graph Efficiency. One of the conditions of an outtree is that the underlying graph is connected and contains exactly $N-1$ lines. Fewer lines than that and the digraph disconnects into components. More lines than that creates multiple paths and cycles between points. In a sense, these multiple paths are redundant in graph-theory terms, and they disrupt the stoic, bare-bones nature of the pure outtree structure.

The technical definition of the *graph efficiency* condition is: In the underlying graph (G_1, G_2 , etc.) of each component (D_1, D_2 , etc.) of digraph D , there are exactly N_n-1 links, where N_n is the number of nodes in the corresponding component D_n . Because fewer than N_n-1 links is not possible (because that would break the component into subcomponents), violations occur to the extent that more than this minimum number of links is present. The degree of graph efficiency is defined as:

$$\text{Graph efficiency} = 1 - \left[\frac{V}{\text{Max}V} \right]$$

where V is the number of links in excess of N_n-1 , summed over all components, and $\text{Max}V$ is the maximum number of links in excess of N_n-1 possible, summed over all components.

Links are not without costs in a social system. They take time and resources to maintain. Thus, the concept of graph efficiency characterizes how dense the network is beyond that barely needed to keep the social group even indirectly connected to one another.

Graph inefficiency should not be confused with social inefficiency or economic inefficiency. To say that a group is graph inefficient simply implies it has more than the $N-1$ links required to remain connected. In fact, in a high-tech, organic organization faced with a dynamic and unpredictable environment, graph inefficiencies may be called for to facilitate the quick cross-fertilization of innovative ideas (Shrader et al., 1989). Thus, graph efficiency reflects the cost of a dense network; it avoids answering the question about the benefits of such a network.

Nonetheless, some conjectures could be made about the relationship

