Environment and Planning, 1971, volume 3, pages 395-410

Signed digraphs and the growing demand for energy†

F S Roberts
The Rand Corporation, Santa Monica, California, USA
Received 17 May 1971

Abstract. Many problems of society, including those relating to energy use, air pollution and solid-waste disposal, seem amenable to formulation using the techniques of the rapidly growing field of graph theory, in particular, the notion of a signed digraph (directed graph). This paper outlines a methodology which exploits the signed digraph for handling problems of forecasting energy demand and the effect of new technologies and institutions on that demand, and for generating and analyzing policy alternatives for meeting environmental constraints on energy use. The forecasting and policy problems are translated into signed digraph problems, and in particular to problems of so-called pulse processes on signed digraphs. Research problems related to the development of the methodology are described.

1 Introduction

Many of the problems of our complex modern society are amenable to mathematical formulation providing that we do not restrict ourselves to the traditional mathematics of the physical sciences.

In particular, a whole host of problems of society, including those relating to energy use, air pollution and solid-waste disposal, seem amenable to formulation using the techniques of the rapidly growing field of graph theory. The result of such a formulation, while not necessarily a complete solution to the problem, is often a better understanding of what the possible solutions are, or a feel for the qualitative interrelationships that underlie the problem, or an identification of significant or vulnerable points of attack.

In the following, we shall discuss the application of one particular concept of graph theory, the signed digraph, to societal problems, particularly to problems of the growing demand for energy. The methodology described seems to be applicable to a wide variety of such problems, and differs somewhat from the way in which signed digraphs have been applied to various social, political or behavioral questions in the past.

The following are among the basic problems related to the growing demand for energy. First, noting the present trends of rapidly increasing energy use, one would like to develop a methodology for forecasting whether these trends will continue and, more generally, what the future demand for energy will be. Second, one would like to understand the basic variables or forces which underlie the growing demand for energy. Third, one would like to understand the effect on future energy consumption of the introduction of new technology and the development of new institutions in society. Fourth, noting that rapidly increasing energy use is in apparent conflict with the quality of the environment, one would like to find a systematic procedure for generating alternative policies or strategies to modify energy use, so as to meet certain environmental (and other) constraints or standards. Finally, one would like to develop methods which distinguish optimal strategies among the alternatives.

† This paper was prepared under National Science Foundation Grant Number G1-44 to the Rand Corporation.

In the following, the use of the signed digraph in developing solutions to these problems is outlined. It is important to remark that in the development of a new methodology for handling difficult problems such as those described above, it is often best to be quite general. Thus, rather than concentrate on specific signed digraphs (except by way of illustration), one should develop the concept of the signed digraph in general, and formulate the various problems just described as signed digraph problems.

In this paper we commence this type of investigation. Section 2 introduces the notions of deviation-amplifying and deviation-counteracting processes. We then show in Section 3 how these ideas are captured by signed digraph concepts. Informal examples are also given in which signed digraphs are applied to solid-waste disposal problems, energy use problems, and air pollution problems. Section 4 illustrates how the signed digraph analysis can handle, in a qualitative way, some of the basic problems already posed. The two sections following delve more deeply into several of these problems by formalizing the notion of a pulse process and proving certain theorems about pulse processes. Finally, Section 7 is used to outline research problems relating to the kinds of applications of signed digraphs discussed earlier.

2 The notions of deviation-amplifying and deviation-counteracting mutual causal processes

Many processes in the world of nature or society involve several variables mutually affecting one another, either simultaneously or alternately. Sometimes these mutual causations result in negative feedback, as in the case of physiological regulation of body temperature, automatic steering devices, the law of supply and demand in economics, etc. Negative feedback from the environmental effects of power generation is resulting in controversies over power plant siting. "As the problems of air and water pollution, denudation of forested recreation areas and other environmental changes become increasingly apparent to society, the public begins to recognize or perceive the effect of these changes on human values such as health and aesthetics. This results in a strong negative feedback which militates against new facilities." (Chapman and O'Neill, 1970, p.7). Processes with negative feedback are usually stable or in equilibrium, and are characterized by the observation that *deviations are counteracted.*

Much less studied, but remarkably widespread, are those mutual causal processes which involve positive feedback, processes where *deviations are amplified.* An *initial kick* or *initial pulse*, which may very well be purely accidental, is amplified well out of proportion. This situation is unstable.

Such deviation-amplifying mutual causal processes, to use the terminology of Maruyama (1960, 1963), appear in many disciplines. The reader will get a feel for the widespread appearance of such processes if some examples are given,

Example 1. Geology: weathering of rock

"A small crack in a rock collects some water. The water freezes and makes the crack larger. A larger crack collects more water, which makes the crack still larger." (Maruyama, 1963, p.166).

Example 2. Ecology: coloration of moths

"A species of moth has predators. Because of the predators, the mutants of the moth species which have a more suitable cryptic coloration (camouflage) and cryptic behavior than the average survive better. On the other hand, those mutants of the predators which have a greater ability than the average in discovering the moth will survive better. Hence, the cryptic coloration and the cryptic behavior of the moth species improves generation after generation, and the ability of the predators to discover the moth also increases generation after generation." (Maruyama, 1963, p.168).

Example 3. Urbanology: the growth of the city
Several small farms exist in a relatively homogeneous plain. One farmer opens a tool shop. This becomes a meeting place for farmers. A food stand is established next to the tool shop. A village grows. The village facilitates the marketing of agricultural products and more farms flourish around the village. Increasing agricultural activities necessitate development of still more shops, and perhaps growth of industry. The village grows into a city. The exact location of the city is determined by the relatively accidental initial kick, which is then magnified by numerous deviation-amplifying processes (Maruyama, 1963, p.166).

Similar analysis is applied on a broad scale in the recent book *Urban Dynamics* by Forrester (1969).

Example 4. Education: quality of an academic institution
"High-quality staff in an academic institution attract high-quality students, who in turn make it easier to attract high-quality staff." (Coombs *et al.*, 1970, p.83).

Example 5. Behavioral sciences: mental illness
"An individual who is 'mentally ill' may act in ways that cause other people to reject him, and this rejection leads to greater feelings of isolation and inferiority, i.e. exacerbates the 'sickness'." (Coombs *et al.*, 1970, p.83).

Example 6. Environment: increase of energy capacity
Many environmentalists are claiming that an increase in energy capacity, say by the construction of one new power plant in a given area, is a deviation which is amplified by drawing in new industry, which in turn uses more energy and therefore creates pressures for increased energy capacity.

As we shall see below, many of the processes involved in various environmental problems, including those related to the growing demand for energy as in Example 6, can be understood by pinpointing the many deviation-amplifying subprocesses that are occurring. These particular subprocesses contribute to overall instability and disequilibrium, which presumably can be counteracted by the introduction of new deviation-counteracting subprocesses, or by reversing the initial pulse of the deviation.

As Forrester (1969, pp.9, 108) points out, a complex system consists of many interacting subprocesses or feedback loops, some positive and some negative. The characteristics of the whole process can only be understood by considering the interactions among these deviation-amplifying and deviation-counteracting subprocesses. In the next section, a mathematical model is described for studying these interactions.

3 A mathematical model: the signed digraph

The notions of deviation-amplifying and deviation-counteracting processes, and the interactions of such processes, can be made precise by using a mathematical concept called a signed digraph.

A *digraph* D (short for 'directed graph') consists of a set N called the *nodes* or *points*, and an irreflexive binary relation R on N called the *relation of adjacency*. Irreflexivity of R means that no node is adjacent to itself[(1)]. The nodes are variables or objects of the process or system being studied. These might be objects (or institutions) such as people, nations, industries or minority groups; or they might be variables such as population, number of jobs available, or tons of nitrous oxide emitted in a given area. The relation R then represents some relationship among the objects or variables. For instance, if the nodes represent people, then $x\,\mathrm{R}\,y$, that is

(1) The irreflexivity assumption is mostly a matter of convenience, and can be eliminated if it turns out to be useful to have a node adjacent to itself.

(x, y) is in the relation R, might mean 'x likes y'. If the nodes are variables, then the relation $x \mathrm{R} y$ might mean 'an increase in x causes an increase in y'.

The digraph is usually represented as a simple figure in which the nodes appear as points and there is a directed line, called an *edge* of the digraph, from x to y if $x \mathrm{R} y$ holds. Figure 1 shows a digraph. Note that it is quite possible for directed lines to go in both directions, from x to y and from y to x. (Note also that, although the edges x_5, x_2 and x_6, x_3 cross, there is no node of the digraph at the crossing point.)

Digraphs have been studied extensively in the literature of combinatorial mathematics, and have been applied in a wide variety of disciplines including economics, operations research, electrical engineering, chemistry, genetics, linguistics, and the social and behavioral sciences. (See Busacker and Saaty, 1965 for a survey of applications, and Harary *et al.*, 1965 for emphasis on the applications to social and behavioral sciences.)

In accordance with Harary *et al.*, (1965), a *sequence* in a digraph D is a sequence of nodes $x_1, x_2, ..., x_n$ such that $x_1 \mathrm{R} x_2, x_2 \mathrm{R} x_3, ..., x_{n-1} \mathrm{R} x_n$. A *path* is a sequence such that all the nodes are distinct. A *cycle* is a path $x_1, x_2, ..., x_n$ such that $x_n \mathrm{R} x_1$. Finally, we shall call a sequence where $x_n \mathrm{R} x_1$ a *loop*, in conformity with some graph-theoretical usage. To give some examples, in the digraph of Figure 1, x_1, x_2, x_3, x_4 gives a path, while x_5, x_2, x_3, x_5, x_6 gives a sequence which is not a path. Also, $x_1, x_2, x_3, x_4, x_5, x_6$ gives a cycle, while $x_5, x_2, x_3, x_5, x_6, x_3$ gives a loop which is not a cycle.

Frequently, relationships can be described as being either positive or negative. If each directed edge in a digraph is given a sign + or −, the resulting object is called a *signed digraph*. (See Figure 2 for an example of a signed digraph obtained from the digraph of Figure 1.)

Signed digraphs have been studied in connection with a variety of problems dealing with balance in small groups (Cartwright and Harary, 1956; Harary, 1954, 1955, 1959; Harary *et al.*, 1965; Norman and Roberts, 1972a, 1972b; and Taylor, 1970, to give just a few references), with international relations (Harary, 1961), and so on. The nodes have usually been objects (or concepts) rather than variables. The positive relations have usually been interpreted as 'likes', 'associates with', 'tells the truth to', and so on, while the negative relations have been given the opposite interpretations.

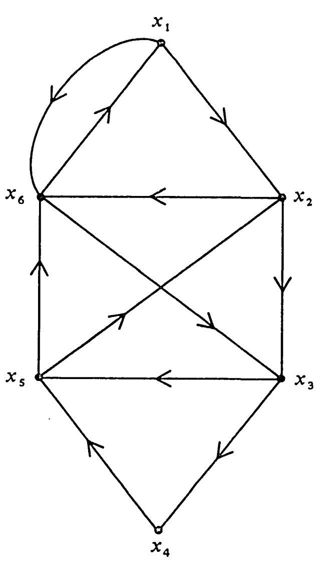

Figure 1. Simple digraph.

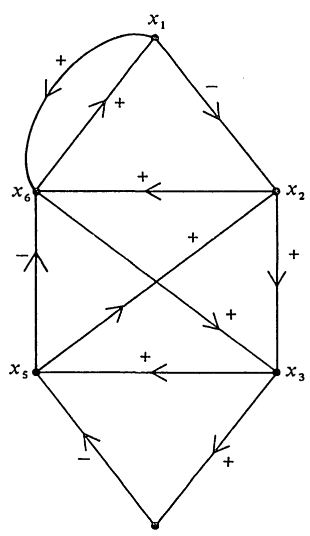

Figure 2. Simple signed digraph.

If the signs are given somewhat different interpretations, the signed digraphs can be applied to societal problems with the nodes treated as relevant variables (compare the 'level' variables of Forrester, 1969). A useful interpretation is to assume that a + sign on the directed edge x to y means that, all other things being equal, variable x 'augments' variable y, or put in other words, an increase (decrease) in variable x leads to an increase (decrease) in variable y. A − sign then means that, all other things being equal, variable x 'inhibits' variable y, or an increase (decrease) in x leads to a decrease (increase) in variable y. (In this sense, the sign might conveniently be thought of as a sort of first derivative of a function relating y to x.)

Using this interpretation of sign, let us look at the signed digraph in Figure 3. This signed digraph, adapted from Maruyama (1963, p.176), describes the important relationships among a number of variables related to the solid-waste disposal problem in a city. The edge P, G is positive because an increase in the population of the city, all other things being equal, leads to an increase in the amount of garbage, while similarly a decrease in the population leads to a decrease in the amount of garbage. The edge D, P is negative because an increase in the number of diseases leads to a decrease in population, while a decrease in the number of diseases leads to an increase in population. Other signs are determined in a similar manner. (There is no edge at all from B to M, for example. This is because the augmenting or inhibiting effect of B on M is considered negligible.)

Note that the four variables P, G, B, and D form a cycle which is deviation-counteracting. An increase in any node on this cycle leads ultimately, through the other nodes on the cycle, to a decrease in this node, and vice versa. (The more people in a city, the greater the amount of garbage; the greater the amount of garbage, the more bacteria; the more bacteria, the greater the number of diseases; the greater the number of diseases, the fewer people, and so on.) On the other hand, the cycle P, M, C is deviation-amplifying. An increase (decrease) in any variable on this cycle leads ultimately to a further increase (decrease) in this variable. ("The more people in a city, the greater the pressure toward modernization; the greater the modernization the more appealing a city is to immigrate to; and the more immigration, the more people in the city." Coombs *et al.*, 1970, p.82.)

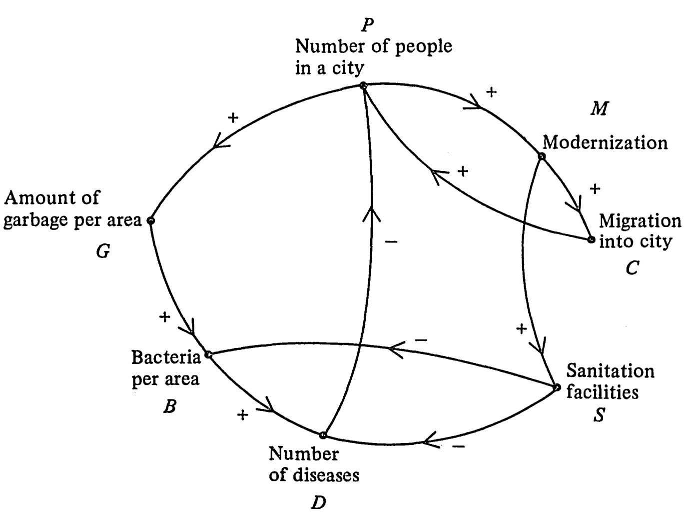

Figure 3. Signed digraph for solid-waste disposal (Maruyama, 1963).

Deviation-amplifying cycles are easy to identify:

Observation (Maruyama, 1963): A cycle is deviation-amplifying if and only if it has an even number of −signs, and it is deviation-counteracting otherwise.

It is interesting to note the connection between this notion of deviation-amplification and that of balance derived earlier in the study of small groups. Cartwright and Harary (1956) call a cycle *balanced* precisely if it has an even number of −signs. Thus, a cycle is deviation-amplifying if, and only if, it is balanced. It will be convenient to extend the Cartwright and Harary definition of balance to an arbitrary sequence, and call it balanced precisely if it has an even number of −signs.

A similar signed digraph can be described for the energy demand situation in a given area [2]. For the sake of discussion, we shall use just a small number of the variables relevant to energy demand as nodes: energy capacity, energy use, energy price, population, number of jobs, number of factories, and quality of the environment (measured by some appropriate indicator). A careful analysis of the energy demand situation would require either a much larger set of variables or would use some statistical technique, such as a cluster or factor analysis, to choose the basic variables for study out of this large class. We return to this point in Section 7.

Our signed digraph for energy demand is shown in Figure 4. For the sake of discussion, we have included only some of the most direct relationships. The sign of the edge *U, Q* is negative because, all other things being equal, an increase in energy use leads to a decrease in the quality of the environment, and a decrease in energy use leads to an increase in the quality of the environment. The edge *J, P* is positive because an increase (decrease) in the number of jobs brings in people (causes people to leave). The edge *C, F* is positive because an increase (decrease) in energy capacity tends to draw new factories into an area (stimulate factories to leave an area).

It should be noted that the cycle *U, Q, P* is deviation-counteracting. The cycles *C, R, U* and *C, F, J, P, U,* are deviation-amplifying, and indeed all cycles containing the node *C* are deviation-amplifying. We shall note the significance of this observation below.

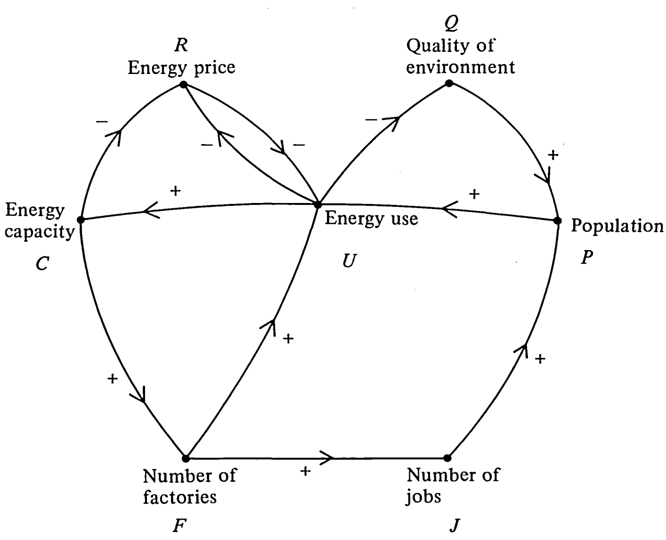

Figure 4. Signed digraph for energy demand.

[2] Figure 1 of Chapman and O'Neill (1970, p.8) could easily be made into a signed digraph which illustrates the negative feedback loop resulting in power plant siting controversies, to which we referred earlier.

Two additional signed digraphs for the study of air pollution are included for reference as Figures 5 and 6. The signed digraph in Figure 6 is an expanded version of that in Figure 5. These signed digraphs are more detailed than those of Figures 3 and 4 and show how alternative strategies for air pollution abatement might be studied in a signed digraph context. In Section 5, we describe in more detail how to assess alternative strategies for decreasing energy demand by means of signed digraphs.

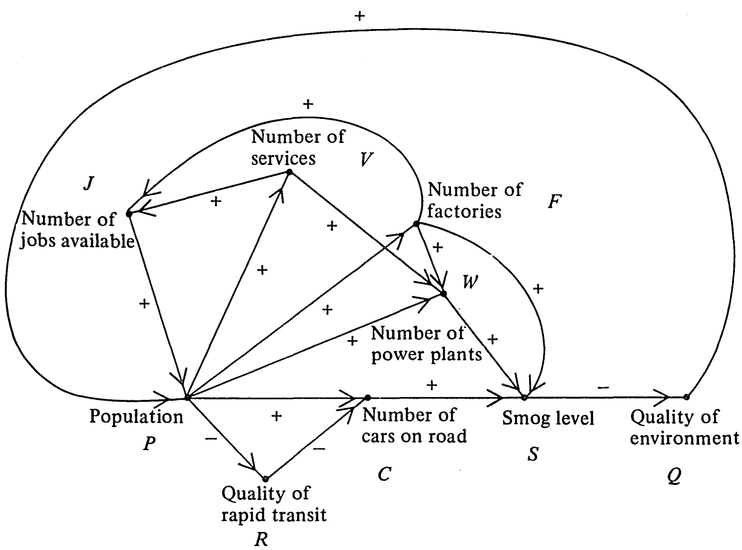

Figure 5. Simplified signed digraph for air pollution in a given region.

4 Analysis of signed digraphs

There are certain distinctions which are not taken into account in a signed digraph. For example, it seems likely that different directed edges have different strength effects. Thus, in the signed digraph for energy demand (Figure 4) it seems likely that energy price increases have much less effect on energy use than do increases in population. Hence, the directed edge *R*, *U* should be weighted differently from the directed edge *P*, *U*. In general, we can speak of a *weighted signed digraph*, where each ordered pair *x*, *y* of nodes is assigned a real number $w(x, y)$ (positive, negative or 0 if there is no edge from *x* to *y*), called the weight of the edge, and interpreted as the (signed) strength of the effect[(3)].

More generally, the (signed) strength of the effect will be a function of the level of the variables in question. Thus, with given *x* and *y*, the strength of the effect of *x* on *y* should be given by a function f_{xy} (x, y) of the values x and y of the variables *x* and *y*. For simplicity, we might assume that f_{xy} is a function of its first argument. This function might be increasing for small values of *x* and decreasing for larger values, or have other complex properties. To give an example, in the solid-waste disposal situation, the function f_{PM} has this character, with increases of population at moderate levels bringing about pressures for modernization, and increases of population at higher levels putting so much strain on a city's budget that modernization programs sit idle and considerable decay occurs.

(3) In this weighted signed digraph, we can include edges such as *Q*, *U*, which are negligible compared with those included in Figure 4. (This edge might have negative weight: for example, a decrease in air quality leads to increased use of air conditioning and hence to more use of electrical power.)

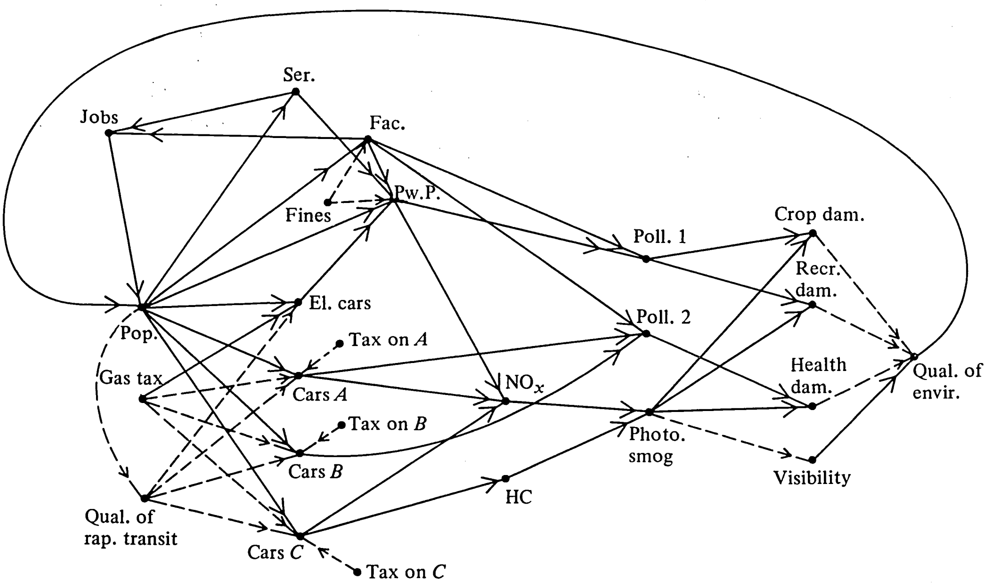

Figure 6. Signed digraph for air pollution abatement strategies, showing some of the important signed relationships. (For readability, + lines are solid, — lines are broken).

Cars *A*	number of cars of type *A* on the road each working day
Cars *B*	number of cars of type *B* on the road each working day
Cars *C*	number of cars of type *C* on the road each working day
Crop dam.	crop damage (measured in dollars perhaps)
El. cars	number of electrical cars on the road each working day
Fac.	number of factories
Fines	fines on industrial stack emissions
Gas tax	tax on gasoline
Health dam.	damage to health (measured in one of the ways suggested by Ridker, 1967, or Wolozin, 1966)
HC	total hydrocarbon emissions
Jobs	number of jobs available in the area
NO_x	total nitrogen oxide emissions
Photo. smog	photochemical smog level (daily average, say)
Poll. 1	pollutant of type 1 (average level)
Poll. 2	pollutant of type 2 (average level)
Pop.	population
Pw. P.	number of power plants
Qual. of envir.	quality of the environment (measured by some quality indicator)
Qual. of rap. transit	quality of rapid transit (measured by some measure such as those proposed in Dalkey *et al.*, 1970)
Rec. dam.	damage in lost or diminished recreational hours and damage to recreational land
Ser.	number of service-oriented businesses
Tax on *A*	tax on cars of type *A*
Tax on *B*	tax on cars of type *B*
Tax on *C*	tax on cars of type *C*
Visibility	average visibility in miles

If we are able to express in a signed digraph the relationships among the variables, it follows that the function f_{xy} is monotonic (increasing or decreasing) in its first argument. This is often a reasonable approximation, at least when the ranges of variables are somewhat restricted.

A digraph, together with a function $f_{xy}(x, y)$ for each directed pair x, y, will be called a *functional signed digraph* [4]. A weighted signed digraph is a special case where each f_{xy} is a constant function. We shall return to the weighted and functional signed digraphs presently.

For the moment, we shall argue that some important observations can be made simply from the signed digraph [assuming, of course, that the functions f_{xy} (x, y) are all monotonic in the first argument, so that the signed digraph accurately reflects the directed relationships]. First, from the signed digraph we can draw certain qualitative conclusions which are true regardless of strength of effects. In particular, in the energy demand example, Figure 4, we noted that every cycle containing the energy capacity node C is deviation-amplifying. Most of the shorter loops containing C are deviation-amplifying as well. This suggests that initial increases in energy capacity are amplified into a continuing increase in energy capacity, and makes precise an observation made imprecisely by many environmentalists and referred to in Example 6 of Section 2.

It also suggests that one way to counteract the tremendous growth in energy capacity is to introduce negative 'initial kicks' into some of the deviation-amplifying cycles containing C. (One way to do this is by introducing a positive initial kick into the energy price, R.) Alternatively, one could try to make some of these cycles deviation-counteracting by changing the sign of some edge. (How this is accomplished on a public policy level is, of course, a difficult question.) Other possibilities are analyzed below, and the question of strategies or policies is put on a more quantitative level.

A second advantage of the signed digraph is that it allows us to carry out at least a qualitative analysis without actually knowing how to measure certain variables. For example, it is very difficult to decide how to measure the 'quality of the environment'. Until we can settle on adequate environmental quality indicators, we can at least carry out an analysis of qualitative effects.

Thirdly, use of the signed digraph helps us to organize our data and to exhibit, in a simple revealing way, the interrelationships in question. Doing so often helps us to understand the effects of certain changes we are contemplating, or at least to appreciate the issues which must be resolved before the effects of these changes can be properly evaluated. To see this, let us look at the detailed signed digraph for air pollution shown in Figure 6. We note that increasing fines on industrial stack emissions seems like a good idea from the quality of environment point of view, but we ask what other effects this might have. We then discover, for example, that the jobs node and the factories node are on a deviation-amplifying cycle. The fines throw a negative kick into this cycle, and so the number of jobs will continue to go down in a deviation-amplifying way, as will the number of factories. This may well have a more depressing effect on the economy than we originally imagined.

Fourthly, the signed digraph allows us to make precise definitions of such problems as the forecast of energy demand, the forecast of the effect of a new technology on this demand, and the identification and choice of possible strategy alternatives for meeting environmental constraints. We shall see this in the next section.

Fifthly, through the use of the existing theory of digraphs, we can identify critical, vulnerable or significant nodes and edges (for example, cut-points, point- or line-vulnerable points and points in point bases; see Harary *et al.*, 1965 for definitions); and we can measure influence of a node or edge, balance of a system, and so on.

All the above applications, except, in part, the first, deal essentially with the static situation of the fixed signed digraph with which we start. Even more interesting is to

(4) The term functional digraph, less redundant, is used already in graph theory. See Harary *et al.*, 1965.

change the initial signed digraph in certain ways and compare the new signed digraph with the old one. Thus, for example, we can ask what might be the effect (on overall stability, or on the ultimate value of the quality of environment node) of adding a new node (perhaps a new institution) and certain relationships to it. Or we can ask abstractly what relationships should be induced on a new node to achieve increased stability, and then try to identify what real-world institution would have these relationships. We can also study the effect of changing the sign on an existing relationship. For example, changing the sign of the edge U, R in the signed digraph of Figure 4 basically corresponds to a policy which is being considered now, namely that of inverting the rate structure for electrical power. Other similar sign changes correspond to additional policy alternatives. We can also try to add new cycles (should they be deviation-amplifying or deviation-counteracting?), new edges, and so on, and analyze the effects of these changes. It is this kind of analysis of possible modifications in the existing system of relationships which is perhaps most interesting and useful.

5 Pulse processes

In order to make a somewhat deeper analysis, let us make some very specific assumptions about the passage of the 'initial kick' or 'initial pulse' which is introduced at a particular node. Let us assume for the following discussion that each node x attains a value $x(n)$ at time n. The value $x(n+1)$ is determined from $x(n)$, from an outside pulse $x^{\circ}(n+1)$ introduced at node x at time $n+1$, and from information about whether other nodes y adjacent to x went up or down at the last time period. If there is a directed edge from y to x with a positive (negative) sign, a change in y at time n is reflected (reversed) in x at time $n+1$. For example, if y, x is negative, and $y'(n)$ is a number representing the change in y at time n, then the effect of y on x at time $n+1$ is an increase in x of $-y'(n)$. (The notation reflects the obvious relation to the first derivative.) The change $y'(n)$ is given by the difference $y(n)-y(n-1)$ if $n>0$ and by $y^{\circ}(0)$ if $n=0$. To summarize, suppose we denote

$$\operatorname{sgn}(y,x) = \begin{cases} +1 & \text{if } y,x \text{ is } + \\ -1 & \text{if } y,x \text{ is } - \\ 0 & \text{if there is no edge } y,x \,. \end{cases}$$

Then we define

$$x(n+1) = x(n)+x^{\circ}(n+1)+\sum_{y} \operatorname{sgn}(y,x)y'(n) \,. \tag{1}$$

A *pulse process* on a signed digraph D is defined by Rule (1), an initial vector of values

$$X(0) = \langle x_1(0), x_2(0), ..., x_p(0)\rangle \,,$$

where $x_1, x_2, ..., x_p$ are the nodes of D, and by vectors giving the outside pulse introduced at each node at each time period. We shall denote these vectors by

$$X^{\circ}(n) = \langle x_1^{\circ}(n), x_2^{\circ}(n), ..., x_p^{\circ}(n)\rangle \,.$$

In dealing with this pulse process, we can make both the forecasting problem and the problem of forecasting effects of new technology or institutions precise and further we can define strategies for choosing alternatives to obtain desired environmental constraints and also give criteria for a strategy to be optimal. Before doing this, however, let us note certain assumptions inherent in this definition of the pulse process.

First, we are assuming as above that all effects (edges) have equal (signed) strength and that the (signed) strength of the effect is independent of the value of the corresponding variables and always has the same sign. This is not a seriously limiting assumption, because what we do below can be extended to weighted signed digraphs and, at least in principle, to functional signed digraphs. We simply define the new pulse process according to the rule

$$x(n+1) = x(n) + x^{\circ}(n+1) + \sum_{y} \mathrm{f}_{yx}[y(n), x(n)]y'(n) . \qquad (2)$$

Second, and more critical, we are assuming that effects represented by one edge are reflected in one time period. This is a tremendous oversimplification. To give an example, in the signed digraph of Figure 4, the effect of a decrease in Q on P will certainly take much longer than the effect of an increase of P on U[5]. On the other hand, once again by clever use of functional signed digraphs, this difficulty can be avoided. A second function $\mathrm{g}_{xy}(\mathrm{x}, \mathrm{y})$ can be used to give the length of time for the effect to take place. (In contrast to f, the function g takes only non-negative values.) A more complicated, but not essentially different, pulse process can be defined in an appropriate way on the *double-functional signed digraph* using the functions f and g. If both f and g are constant, then we shall speak of a *double-weighted signed digraph.* Most of what is said below can probably be generalized to double-functional signed digraphs. In practice, the hard part would be the estimation of the parameters f_{xy} and g_{xy}.

Returning to the simple pulse process defined above, we can now state precisely what we mean by the forecasting problem.

The forecasting problem. Given a pulse process defined by the vectors $X(0)$, $\{X^{\circ}(n): n = 0, 1, 2, ...\}$, what is the value $z(n)$ at the particular node z at time n? [What is the value $z(n)$ at node z at time n given information on initial conditions, that is, initial values at all nodes and values of all outside pulses?]

There are many variations of the forecasting problem. One is to bring in the effect of new technology or institutions by changing the signed digraph at a particular time t, the time when this technology is introduced. The result is merely a change in Rule (1) for $n \geqslant t$. Another is to make probabilistic forecasts of exp $z(n)$ based on the expected value of t.

Environmental (or other) constraints or standards can be introduced by placing limits on certain nodes above or below which the value cannot go. This is done by introducing two vectors $m = \langle m_1, m_2, ..., m_p \rangle$ and $M = \langle M_1, M_2, ..., M_p \rangle$, representing the lower and upper constraints m_i and M_i at each point x_i, with these numbers allowed to be $-\infty$ and $+\infty$ respectively. Typical constraint vectors in the pulse process defined from the signed digraph of Figure 4 might have a specified m_Q and m_J with all other $m_i = -\infty$ and all $M_i = +\infty$.

Given these constraints, we try to find alternative policies which will meet them. Possible *policies* or *strategies* include some or all of the following:

1 introduction of pulse vectors $X^{\circ}(n)$: add pulses to specific nodes at specific times;
2 at a given time addition of a new node (institution) and new directed edges to and from it (relations of interaction of the institution with existing ones);
3 change of sign of a given edge at a given time;
4 addition of a new edge between existing nodes;
5 addition of a new cycle (deviation-amplifying or deviation-counteracting).

Once a fixed set of strategy alternatives has been defined, the policy problem can be stated precisely.

(5) Even more complicated situations might be envisioned, where an increase takes a longer time to be reflected than a decrease, or vice versa. We have disregarded such complications.

The policy problem. Version 1: Find the optimal (shortest, cheapest) strategy for attaining the (environmental) standards.

Version 2: Find the policy which maximizes the values on certain nodes (for example, nodes measuring quality of environment), subject to constraints on the values at certain other nodes.

Besides finding an optimal strategy for attaining the goals (meeting the standards), it is important to see if these standards can be maintained once reached, that is, to see if the situation in which the standards are met is a stable one, or if there will be wild oscillations. So we might wish to define 'attainable' in the policy problem as meaning we reach the goal and stay there.

It is not hard to solve the forecasting and policy problems in general. Before turning to these solutions, let us note that the pulse process on signed digraphs can be compared with similar processes studied in systems theory, control theory, simultaneous equations, input–output analysis, econometrics, matrix analysis, and so on. The visual presentation of the signed digraph, however, makes many of the relationships very clear and understandable, in a way these other methods do not. Also, although in principle it is possible, given for example a system of simultaneous equations, to find all the (indirect) paths of causation leading from one variable (for example, energy capacity) to another (for example, quality of environment), this kind of information can frequently be read off at a glance from the signed digraph. Finally, the theory of signed digraphs has stressed certain notions and concepts relating to the total system, such as balance and point basis, which are different from the basic concepts of some of these other areas, and seem potentially relevant to large social problems.

6 Solution to the forecasting and policy problems in a pulse process

Suppose $r \leqslant n$ and $x = x_1, x_2, ..., x_r, x_{r+1} = y$ is a sequence in the signed digraph $D = (N, R)$, that is, $x_1 R x_2, x_2 R x_3, ..., x_r R x_{r+1}$. The *length* of this sequence is defined to be r, the number of edges in the sequence (one less than the number of nodes). If a pulse of $+1$ is introduced at the node x at time 0, then, corresponding to this sequence, a value of $+1$ or -1 is added to the node y at time r, with the sign $+$ or $-$ given by

$$\pi = \prod_{i=1}^{r} \operatorname{sgn}(x_i, x_{i+1}) .$$

We see that π is $+$ if, and only if, the sequence has an even number of minus signs, that is, the sequence is balanced. Conversely, the only additions to (subtractions from) y obtained in $\leqslant n$ steps from a pulse at x at time 0 are obtained from sequences from x to y in D of length $\leqslant n$. Defining $S_n^+(x, y)$ and $S_n^-(x, y)$ respectively as the number of balanced and the number of unbalanced sequences of length $\leqslant n$ from x to y, we reach the conclusions given in Theorem 1.

Theorem 1 [6]. Given the pulse process with

$$X^{\circ}(n) = 0 \quad \text{for } n \neq 0 ,$$

$$x^{\circ}(0) = 1 \text{ and } z^{\circ}(0) = 0 \quad \text{for } z \neq x ,$$

we have

$$y(n) = y(0) + [S_n^+(x, y) - S_n^-(x, y)] .$$

This result can be generalized to the most general pulse process. The number $S_n^+(x, y) - S_n^-(x, y)$ is easily calculated from the initial signed digraph if we define the

[6] The author is grateful to S. Johnson for his help in formulating this theorem.

(*signed*) *adjacency matrix* $A(\mathrm{D}) = (a_{xy})$ in the standard way by

$$a_{xy} = \begin{cases} +1 & \text{if the edge } x, y \text{ is } + \\ -1 & \text{if the edge } x, y \text{ is } - \\ 0 & \text{otherwise} . \end{cases}$$

It is then easy to prove by induction that the x, y entry of A^n gives the difference between the number of balanced sequences of length n from x to y and the number of unbalanced sequences of length n from x to y. Thus, if $B_n = A+A^2+\ldots+A^n$, we have:

Theorem 2.

$S_n^+(x, y) - S_n^-(x, y)$ is given by the x, y entry of the matrix B_n.

Essentially the same result holds if instead of a signed digraph we have a weighted signed digraph and we define the pulse process using the weights $\mathrm{w}(x, y)$ by replacing Rule (1) of Section 5 by the rule

$$x(n+1) = x(n)+x^{\circ}(n+1)+\sum_y \mathrm{w}(y, x)y'(n) . \tag{3}$$

Then if a pulse of +1 is introduced at node x at time 0 and $x = x_1, x_2, \ldots, x_r, x_{r+1} = y$ is a sequence from x to y in the signed digraph, we have a value $\prod_{i=1}^{r} \mathrm{w}(x_i, x_{i+1})$ added to y at time r. If the adjacency matrix is now replaced by the matrix $W = [w(x, y)]$, we find that $y(n)$ is given by $y(0)+$ the x, y entry of V_n, where $V_n = W+W^2+\ldots+W^n$.

The forecasting problem can also be solved, *in principle*, for the pulse process on a functional signed digraph, which is defined by Rule (2) of the previous section, but here computation becomes rather involved. Even in the pulse process where a pulse of 1 is introduced at just one node, x, and only at time 0, we already get a rather complicated formula for the values at time 2. Namely, we have

$$y(2) = y(1)+\sum_z \mathrm{f}_{zy}[z(1), y(1)]z'(1) .$$

Now since $x'(0) = 1$, $u'(0) = 0$ for $u \neq x$, we have

$$z(1) = z(0)+\mathrm{f}_{xz}[x(0), z(0)]$$

and so

$$z'(1) = \mathrm{f}_{xz}[x(0), z(0)] .$$

Thus,

$$y(2) = y(0)+\mathrm{f}_{xy}[x(0), y(0)]+ \sum_z \mathrm{f}_{zy}\{z(0)+\mathrm{f}_{xz}[x(0), z(0)], y(0)+\mathrm{f}_{xy}[x(0), y(0)]\}\mathrm{f}_{xz}[x(0), z(0)] .$$

Pulse processes on functional signed digraphs may not have simple closed form solutions such as those given in Theorems 1 and 2, but they can probably be handled by existing computer techniques. These techniques would even allow inclusion of such complex strategies as: add a new node with specified relations if the value of a given node exceeds a certain number (7).

Turning to the policy problem, we again note that this problem is solvable using the matrix B_n, if we know the standards are attainable. Given a desired value (set of values) on a node x (set of nodes), suppose for simplicity that the allowable strategies consist of those where we put an initial pulse of 1 at one node, and make

(7) M. Berman (private communication).

no additional changes. To decide which strategy will enable us to reach the desired value(s) fastest, simply calculate the matrices B_1, B_2, ..., until a matrix is obtained for which the y, x entry is at least (at most) the desired value for some y. The first y which satisfies this condition gives the solution[8]. If the standards are not attainable, the procedure never ends. A similar algorithm applies to a comparison of more complicated strategies or to the choice of a strategy with least cost (if each introduction of an outside pulse or other change has a cost associated with it). The policy problem is solvable in a similar manner in the weighted signed digraph case and is, in principle, solvable in the functional signed digraph case as well.

7 Research problems

The methodology described in earlier sections is in fact, only hinted at, and indeed much of it still needs to be developed. There are two major areas for research. The first deals with the analysis of signed digraphs once these are defined. Here, there are many mathematical problems which will need to be solved as the methodology is developed. There are also questions of interpretation of results, and these can probably best be tackled by applying the methodology at an early stage to a detailed signed digraph. The second major area of research deals with the development of a methodology for building a signed digraph for analysis. The questions here relate to the gathering of data and to the development of practical techniques for choosing nodes, defining weights, and so on.

Concentrating first on the mathematical problems relating to the analysis of signed digraphs, we note that the techniques for studying pulse processes described in Section 6 need to be extended to functional and double-functional signed digraphs. Algorithms need to be developed for solving the forecasting and policy problems. On the other hand, more work also needs to be done on signed digraphs themselves, with the aim of understanding exactly what kinds of conclusions are meaningful if only the signed relationships without weights or functions are used.

The algorithms described in Section 6 for solving the policy problem apply only if we know a solution exists, that is, a strategy exists for attaining the desired goals[9]. Criteria for the existence of such a solution for a given (weighted, functional, double-functional) signed digraph need to be redeveloped. Related to this, the solutions found by the algorithms of Section 6 are solutions only in the sense of reaching desired values. There is no way of telling whether the process will stay within the constraints once the values are reached or whether the solution is stable in any of a number of senses. This notion of stability raises several interesting questions. It will be very useful to devise a precise definition for the stability of a (weighted, functional, double-functional) signed digraph at a given state in a pulse process. It will then be possible to find criteria for stability and to find algorithms which give stable solutions, not just solutions, to the policy problem.

If we turn to a more specific strategy or policy problem, one of the strategies of greatest interest would be the addition of a new node with certain relations to other nodes. An algorithm for identifying what relations to induce on this node in order to attain certain constraints most quickly (least expensively) would be extremely helpful.

In discussions of strategy or policy alternatives, it would be very useful to identify ahead of time which nodes (sets of nodes) or edges (sets of edges) are critical to the investigation. Appropriate definitions of critical node (sets) and edge (sets) should be devised; probably one of the standard graph-theoretical concepts such as 'vulnerable', 'cutting', 'deletion-minimal', and so on, would suffice.

(8) This does not necessarily give a solution which is stable. We return to this point again in the next section.

(9) Thus, it may not be fair even to call them algorithms.

Tying in even more with existing graph-theoretical literature, we find that there seems to be a close relationship between the theory of pulse processes and the theory of balance, as hinted by Theorem 1. It would probably be quite fruitful to explore this relationship in more detail; perhaps some of the concepts of balance theory or some of the measures of relative balance which have been developed will be related to notions of stability under a pulse process.

In addition to work on some of these theoretical problems, it will probably be quite important to confront practical problems of signed digraph analysis by applying the methodology quite early to some specific signed digraphs. Even starting with a simple signed digraph for energy demand, such as that in Figure 4, one could fruitfully tabulate policy alternatives generated by a change of appropriate signs, addition of specified nodes, and so on; one could try to work out which alternatives would be useful, and then try to identify what real-world policies would correspond to these alternatives. Undoubtedly, several potentially useful policy alternatives, different from any that have been suggested so far, could be generated in this way.

The second major area of research is associated with developing a methodology for building a signed digraph. Some attempt should be made to develop a detailed signed digraph for energy demand in a systematic manner, preferably at least a double-weighted signed digraph. One approach which suggests itself is as follows. At the first stage, a 'free response' session could be held to identify potential relevant variables. These might be limited to one 'sector', such as transportation, industrial, or residential. They could include demand variables, supply variables, environmental factors, potentially relevant economic factors and indicators, aesthetic variables, social factors, and so on.

For computational purposes, some limitation on the number of variables to be considered in the signed digraph would be necessary. The existing set of variables would undoubtedly have to be replaced by a smaller set. A smaller set of the 'most important' variables could probably be chosen on an intuitive basis at first, but more appropriately, some sort of ranking procedure could be used. Alternatively, a cluster analysis could be performed to define a small set of groups or clusters of closely related variables, which could then be used to define a set of new (complex, multi-dimensional) variables for use in the signed digraph. The ranking or clustering might be performed in a manner similar to that used in the Rand studies on factors relating to quality of life carried out for the US Department of Transportation (compare Dalkey *et al.*, 1970, pp.30ff; Dalkey and Rourke, 1971).

Next, for the purpose of the analysis, some attempt could be made to define crude means of measuring the variables which are selected. At this stage, some attention should probably be paid to the theory of scaling (as developed in Coombs, 1964 or Torgerson, 1958), and to modern techniques in the theory of measurement (as in Krantz *et al.*, 1971, or Suppes and Zinnes, 1963).

The size of the set of variables could here be further limited if a factor analysis were performed to pinpoint a small set of underlying variables whose values determine the values of all the others, and this small set of variables only were used in the final signed digraph.

The fundamental variables, once chosen, would become the nodes of a double-weighted signed digraph. For simplicity, one could start with a complete signed digraph, that is, one in which all edges are present, and eliminate certain edges later by assigning them weights of 0. Estimation of the (signed) weight to place on a directed edge or on a time effect would then become the crucial problem. It would be a very difficult problem, and in some sense could form the basis for a study in and of itself. At first, these weights would probably have to be estimated quite roughly. Alternatively, such weights could be estimated by employing a group of experts and performing several iterations using the Delphi technique developed at Rand (Dalkey, 1969; Brown *et al.*, 1969). There are also some statistical techniques, based on time series data, which might be useful here. These include regression analysis, in particular lag correlation methods, *k*-stage estimation, and causal analysis (C. Morris, private communication).

Finally, once the double-weighted signed digraph were defined, various strategies could be tested, including changing of certain weights, introduction of new nodes with certain new relationships, and so on. Many of the concepts developed earlier could be applied to this specific situation.

References

Brown, B., Cochran, S., Dalkey, N. C., 1969, "The Delphi method, II: structure of experiments", RM-5957-PR, The Rand Corporation, Santa Monica, California.
Busacker, R. G., Saaty, T. L., 1965, *Finite Graphs and Networks* (McGraw-Hill, New York).
Cartwright, D., Harary, F., 1956, "Structural balance: a generalization of Heider's theory", *Psychological Review*, **63**, 277-293.
Chapman, D., O'Neill, R. V., 1970, "Ecology and resource economics: an integration and application of theory to environmental problems", General Report ORNL-4641, Oak Ridge National Laboratory.
Coombs, C. H., 1964, *A Theory of Data* (John Wiley, New York).
Coombs, C. H., Dawes, R. M., Tversky, A., 1970, *Mathematical Psychology: An Elementary Introduction* (Prentice-Hall, Englewood Cliffs, NJ).
Dalkey, N. C., 1969, "The Delphi method: an experimental study of group opinion", RM-5888-PR, The Rand Corporation, Santa Monica, California.
Dalkey, N. C., Lewis, R., Snyder, D., 1970, "Measurement and analysis of the quality of life: with exploratory illustrations of applications to career and transportation choices", RM-6228-DOT, The Rand Corporation, Santa Monica, California.
Dalkey, N. C., Rourke, D. L., 1971, "Experimental assessment of Delphi procedures with group value judgments", R-612-ARPA, The Rand Corporation, Santa Monica, California.
Forrester, J. W., 1969, *Urban Dynamics* (MIT Press, Cambridge, Mass.).
Harary, F., 1954, "On the notion of balance of a signed graph", *Michigan Mathematical Journal*, **2**, 143-146.
Harary, F., 1955, "On local balance and *N*-balance in signed graphs", *Michigan Mathematical Journal*, **3**, 37-41.
Harary, F., 1959, "On the measurement of structural balance", *Behavioral Science*, **4**, 316-323.
Harary, F., 1961, "A structural analysis of the situation in the Middle East in 1956", *Journal of Conflict Resolution*, **5**, 167-178.
Harary, F., Norman, R. Z., Cartwright, D., 1965, *Structural Models: An Introduction to the Theory of Directed Graphs* (John Wiley, New York).
Krantz, D., Luce, R., Suppes, P., Tversky, A., 1971, *Foundations of Measurement* (Academic Press, New York).
Maruyama, M., 1960, "Morphogenesis and morphostasis", *Methodos*, **12**, 251-296.
Maruyama, M., 1963, "The second cybernetics: deviation-amplifying mutual causal processes", *American Scientist*, **51**, 164-179.
Norman, R. Z., Roberts, F. S., 1972a, "A derivation of a measure of relative balance for signed digraphs and a characterization of extensive ratio systems", RM-6201/2-PR, The Rand Corporation, Santa Monica, California: *Journal of Mathematical Psychology* (forthcoming).
Norman, R. Z., Roberts, F. S., 1972b, "A measure of relative balance for social structures", in *Structural Theories in Progress*, Volume 2, Eds. J. Berger, M. Zelditch, B. Anderson (Houghton-Mifflin, New York), forthcoming.
Ridker, R. G., 1967, *Economic Costs of Air Pollution: Studies in Measurement* (Praeger, New York).
Suppes, P., Zinnes, J. L., 1963, "Basic measurement theory", in *Handbook of Mathematical Psychology*, Volume 1, Eds. R. D. Luce, R. R. Bush, E. Galanter (John Wiley, New York), pp.1-76.
Taylor, H. F., 1970, *Balance in Small Groups* (Van Nostrand-Reinhold, New York).
Torgerson, W. S., 1958, *Theory and Methods of Scaling* (John Wiley, New York).
Wolozin, H. (Ed.), 1966, *The Economics of Air Pollution* (W. A. Norton, New York).