Chapter 2

A DETAILED GUIDE TO NETWORK ANALYSIS

James J. Kay
Department of Environment
and Resource Studies,
University of Waterloo
Waterloo, Ontario
N2L 361 Canada

Lee A. Graham
Department of Entomology,
Texas A&M University,
College Station,
Texas 77843
USA

Robert E. Ulanowicz
Chesapeake Biological Laboratory
University of Maryland
Solomons, Maryland 20688-0038
USA

INTRODUCTION

This chapter will introduce a number of the more frequently used network analysis techniques, many of which trace their origins from Leontief's (1936) economic input-output analysis and Shannon's information theory (1948). We begin by presenting a standard format for recording all of the flow data related to the ecosystem in a network representation. Having constructed the representation, the direct and indirect relationships that exist among the different components can then be examined in depth using a suite of network analyses that have been developed through the years by a number of researchers. These analyses consist of calculating a number of measures which synthesise some part of the information about flows of energy or materials through an ecosystem. These measures and analytical techniques cover from the microscopic level, the level of a component, right up to the macroscopic level, the level of the whole ecosystem. The measures are presented in a gradation from micro to macroscopic. The set of measures described herein provides a very rich description of an ecosystem; a description which looks at the system from many different perspectives. Clearly anyone applying these measures will have need for only those which reflect the perspectives of interest to the researcher .

All of the analytical techniques and measures discussed in this chapter have been written up as a set of computer programs (NETWRK, STRUCTURE, ENVIRON and others). This software is available

from SCOR (see elsewhere in this book, Wulff *et al.* 1989) or from the authors. Details of how to use the programs are included with the software. Also, a list of the subheadings in this chapter and the programs which do the corresponding analysis is also included. Finally, reference to the programs has been made in the text where relevant.

FLOW NETWORK REPRESENTATION OF ECOSYSTEMS

This section explains how a network representation can be used to describe the flows of mass and energy in an ecosystem. The process of building such a representation serves two purposes. It brings together all of the available data related to mass and energy flows in the ecosystem, and it identifies what information is missing. A standard data format is presented allowing for easy calculation and comparison of different ecosystems is presented.

The term "network" describes a collection of elements called *nodes*, pairs of which are joined to one another by a collection of elements called *edges*. Each node represents a compartment (biotic or abiotic) of the ecological system and is referred to by a numerically-referenced name, such as x_1 or x_6, which can also be used to reference the compartment's standing stock. The edges that connect the nodes are usually directed, the direction indicating that matter and energy flow from the initial to the terminal node. Directed edges are called *arcs*. Arcs are named using the numerical identifiers of the nodes they connect. For example, the arc connecting components x_1 and x_6 is referenced as f_{16}.

Each arc in an ecological flow network can have an associated value. This value represents the magnitude of flow that occurs from the initial to the terminal node of the arc in a given unit of time. A network of this sort is said to be weighted. Living systems also exchange matter and energy with their environment. Thus, ecological networks have at least some arcs that originate or terminate outside of the system. These are called open networks. All of the networks discussed in this book are open, weighted networks.

The flow diagram in Figure 2.1 illustrates a generalized scheme for diagramming ecological flow networks. Values are associated with either standing stocks or one of four classes of flow: (1) inputs from outside the system, (2) flows between components, (3) exports to other systems, and (4) respiration losses. Note that all flow values are presented in terms of units of mass or energy per unit area or volume per unit time, while standing stocks are measured as units of mass or energy per unit area or volume.
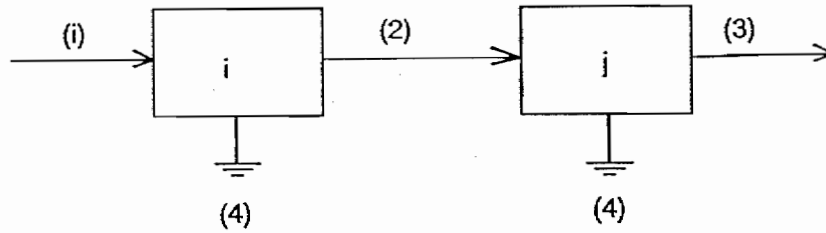
Fig 2.1. The four classes of ecological flows: (1) exogenous inputs, (2) intercompartmental exchanges, (3) exports of useable medium, and (4) dissipation of unuseable medium.

Figure 2.2a depicts a simple weighted network representing the five components of the Cone Spring ecosystem described by Tilly (1968) and quantified by Williams and Crouthamel (unpub. ms.). Flows of energy (kcal/m^2/yr) are indicated by the values that appear on the arcs, while standing stocks (kcal/m^2) are indicated inside the boxes that represent the nodes. There are two inputs that originate outside of the system. The arcs that terminate outside the system represent exports of energy in a form that can be used by other systems. The special ground symbols represent energy that is dissipated through respiration. This energy is lost from the system and is unusable by any system at the same scale (Ulanowicz 1986a).
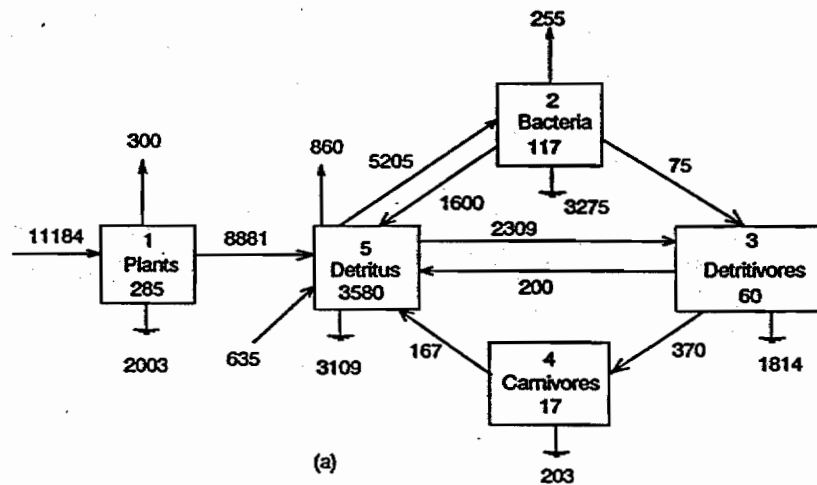

Fig 2.2a. Energy flows (kcal/m^2/yr) and densities (kcal/m^2) among the 5 components of the Cone Springs network.

Diagrammatical representations of networks are common, but can be confusing and cumbersome when used to describe large systems. A mathematical network representation which describes an ecosystem's structure and function is necessary. In this book a standardized method is used for representing ecological flow networks using a single matrix (flows between compartments) and four vectors (standing stocks, inputs, exports, and respirations). (See Figure 2.2b for an example, the Cone Spring model). This convension must be adopted by the reader in constructing networks for analysis by the software associated with this book.

$$
\begin{bmatrix}
0 & 0 & 0 & 0 & 8881 \\
0 & 0 & 75 & 0 & 1600 \\
0 & 0 & 0 & 370 & 200 \\
0 & 0 & 0 & 0 & 167 \\
0 & 5205 & 2309 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
11184 \\
0 \\
0 \\
0 \\
635
\end{bmatrix}
\begin{bmatrix}
300 \\
255 \\
0 \\
0 \\
860
\end{bmatrix}
\begin{bmatrix}
2003 \\
3275 \\
1814 \\
203 \\
3109
\end{bmatrix}
$$

$$
\quad\quad\quad F \quad\quad\quad\quad\quad\quad Z \quad\quad E \quad\quad R
$$

Fig 2.2b. The standardized method for representing ecological flow networks using a single matrix of flows between compartments (F) and three vectors representing inputs Z, exports E, and respiration, R.

The number of compartments n in the Cone Spring network is 5, 4 of which are biotic and one of which is abiotic. The vector x contains elements $x_1 \ldots x_n$ that indicate the magnitudes of the standing stocks of each compartment. In order to reap the full benefit of the analysis programs and software associated with this book, we recommend that the flow network be conceptualized and quantified so that, when ordering the nodes (assigning positions $x_1 \ldots x_n$), the biotic components are numbered before the abiotic ones. In the Cone Spring example, component 5 (abiotic detritus) is assigned the last of the 5 possible numbers. The software program, REORDER, can be used in those cases where improperly ordered network data already exists (see software documentation for details).

The vector z is used to store flows that originate from outside the system (inputs). Flows between system components are stored in the $n \times n$ exchange matrix F. The orientation of the matrix is such that flow travels from row to column components of the matrix. We adopt the From Row - To Column orientation throughout this book, but note that many of the analyses we will describe (those developed by Patten, Finn, Barber, and Bosserman) often use the To Row - From Column format. One should always be careful to note matrix orientation when examining the literature relating to specific analyses.

Losses from the compartments are denoted by the elements of the export vector e and the respiration vector r. The export and respiration vectors are sometimes summed to produce a total loss vector y.

Most network analysis techniques were originally formulated for steady state systems. This means that, for each compartment, the sum of all its inputs must equal the sum of all its outputs. Most analyses can handle non-steady state systems, but care must be taken in interpreting the results. The discontinuities and nonlinearities that exist in the interactions occurring in nature cannot always be represented by using a simple matrix description of the ecosystem.

THEORETICAL BACKGROUND

Leontief (1936, 1963) developed economic input-output analysis as a means of quantifying the amount of raw materials and industrial services required to produce a quantity of consumer goods. His work introduced the backward case analysis, concerned with tracing output or demand back to inputs. Augustinovics (1969) later developed the forward case economic analysis that traces the fate of system inputs through the system to outputs.

Hannon (1973) was the first to apply Leontief's analysis to ecological systems. Hannon was interested in determining the direct and indirect contributions of the system to a unit of output, and therefore modified Leontief's analysis to solve the equations differently. His success spawned a flurry of efforts among systems ecologists. Finn (1976) later developed a forward case analysis for analyzing ecological networks. In addition he introduced several measures of ecosystem function, including cycling, that aid in the ecological interpretation of the results of input-output analysis. Ulanowicz (1986a) has documented these analytical techniques in textbook fashion.

It is important to note that the network perspective provided by flow analysis is the foundation on which all of the analyses in this book are built

Input-Output Flow Analysis

The basics of flow analysis are quite simple (see also Costanza & Hannon 1989, this volume). Given the system information provided in the F matrix and z, e, and r vectors, one can construct a production matrix, which serves as the basic computational unit for flow analysis. The basic structure of the production matrix is given in Figure 2.3a, with the Cone Spring network presented as an example in Figure 2.3b. The important thing to realize about the production matrix is that the F22 partition is the F matrix, the F12 partition holds the system inputs (z), while the latter partition contains the e and r vectors (perhaps summed as the total loss vector y).

$$\begin{bmatrix} F11 & F12 \\ \hline F21 & F22 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 11180 & 0 & 0 & 0 & 635 \\ 2300 & 0 & 0 & 0 & 0 & 8881 \\ 3525 & 0 & 0 & 75 & 0 & 1600 \\ 1810 & 0 & 0 & 0 & 370 & 200 \\ 203 & 0 & 0 & 0 & 0 & 167 \\ 3970 & 0 & 5205 & 2309 & 0 & 0 \end{bmatrix}$$

Fig 2.3. The basic structure of the production matrix (top) and the Cone Spring network as an example (below).

We can use the production matrix to compute the throughflow, T_i, for each compartment i. This value is defined as the rate of energy or material flow through compartment i. It is calculated as the sum of the inflows to compartment i (obtained by summing the ith column of the production matrix) or the sum of the outflows (calculating the row sums) from i. In a steady state system, the two methods for calculating T_i produce the same result. For a dynamic system, throughflow is modified to include changes in storage as inflow or outflow. A positive state derivative (xd_+) is considered a loss from the system pool of mobile energy, while a negative state derivative (xd_-) is viewed as a gain to the pool (Patten et $al.$ 1976, Finn 1977). The equations for throughflow then become either, for outflows:

$$T_i = \sum_{j=1}^{n} f_{ij} + e_i + r_i + xd_{+i} \tag{1}$$

or, for inflows:

$$T_i = \sum_{j=1}^{n} f_{ij} + z_i + x_{d\text{-}i} \qquad\qquad (2)$$

Total system throughflow T (also called total system throughput) is defined as sum of all compartmental throughflows. It measures the total amount of material or energy flowing in the system, multiplied by the number of compartments that the flow is passing through. T is one of the total system properties discussed later in this chapter.

Fractional Coefficient and Structure Matrices

We now consider separately the *backward* (Leontief/Hannon) and *forward* case (Augustinovics/Finn) flow analyses, beginning with the backward case input or inflow analysis (also called creaon analysis in Patten *et al.* 1976). Inflow analysis traces system output back through the network to system input. As a first step, the fractional inflow matrix G' is generated for a network by dividing each f_{ij} element of the exchange matrix by T_j, the throughflow entering compartment j. The elements of this matrix indicate the fraction of compartmental throughflow that originates from a another node in the network. Ulanowicz (1986a) calls the elements of G' partial feeding coefficients, since each value represents the proportion of a component's "diet" that is derived from a particular "feeding" interaction.

The next step involves calculating the Leontief inverse (N') matrix. We will follow Hannon's practice and call N' the input structure matrix. N' and its output structure analog N" are often referred to as a fundamental matrix (Kemeny and Snell 1960) or a transitive closure matrix (Patten *et al.* 1976). The elements of the structure matrices show the direct and indirect contributions of each interaction to system output. Remember that once matter and energy enter a system, they may visit many compartments along a large number of possible pathways before exiting. The input structure matrix is obtained by subtracting G' from the identify matrix I (a matrix that is all zeros except for the diagonal elements, which have values of 1), and then inverting the result using matrix algebra:

$$N' = (I - G')^{-1} \qquad\qquad (3)$$

The elements of N' can be interpreted as (1) n'_{ij} is the amount of throughflow from i required to produce a unit of output in j, or (2) n'_{ij} is the expected number of times energy or matter that ended up in compartment j has passed through compartment i. The variances (stored in the U' matrix) associated with these expectations can also be calculated using the equations presented in Patten and Matis (1982). These variances are useful indicators of the complexity of network structure and flow distribution.

The forward-looking output or outflow analysis (also called genon analysis) is concerned with tracing the fate of inputs through the system to outputs. Because ecologists want to know what drives living

Table 2.1. Matrices produced by Input/output analysis of the Cone Spring network.
Flows travels from rows (i) to column (j) compartments in this format.

The fractional inflow matrix; G'

	1	2	3	4	5
1	.000	.000	.000	.000	.773
2	.000	.000	.032	.000	.139
3	.000	.000	.000	1.000	.017
4	.000	.000	.000	.000	.014
5	.000	1.000	.969	.000	.000

The fractions of the total inputs to j that flow directly from i.

The fractional outflow matrix; G"

	1	2	3	4	5
1	.000	.000	.000	.000	.794
2	.000	.000	.014	.000	.307
3	.000	.000	.000	.155	.084
4	.000	.000	.000	.000	.451
5	.000	.453	.201	.000	.000

The fractions of the total outputs from j that flow directly from i.

$(I-G')^{-1}$ inverse matrix (also known as N')

	1	2	3	4	5
1	1.000	.933	.933	.933	.933
2	.000	1.169	.201	.201	.169
3	.000	.039	1.039	1.039	.039
4	.000	.018	.018	1.018	.018
5	.000	1.207	1.207	1.207	1.207

Represent the fractions of donor throughflow (direct and indirect)
needed to produce one unit of recipient outflow.

$(I-G")^{-1}$ inverse matrix (also known as N")

	1	2	3	4	5
1	1.000	.434	.199	.031	.958
2	.000	1.169	.092	.143	.374
3	.000	.084	1.039	.161	.186
4	.000	.247	.113	1.018	.545
5	.000	.547	.251	.039	1.207

Represent the fractions of recipient throughflow generated by one unit of inflow
to the donor.

Table 2.1. Continued

	U' variance matrix (of N')				
	1	2	3	4	5
1	.000	.062	.0623	.062	.062
2	.000	.198	.229	.229	.198
3	.000	.040	.040	.040	.040
4	.000	.018	.179	.018	.018
5	.000	.249	.249	.249	.249

	U" variance matrix (of N")				
	1	2	3	4	5
1	.000	.393	.175	.031	.436
2	.000	.198	.091	.015	.389
3	.000	.106	.040	.141	.228
4	.000	.269	.109	.018	.473
5	.000	.433	.207	.039	.249

systems, and since outputs in ecological systems are generally dominated by respiration losses, outflow analysis has been applied more frequently in ecology. It has also been refined and redefined to produce new insights into the intricacies of ecological organization. The analysis begins with the calculation of the fractional outflow matrix $G"$ (together, $G"$ and G' are referred to as fractional or technical coefficient matrices). Rather than normalizing the f_{ij} by the throughflow T_j entering j, we normalize by the throughflow T_i leaving the donor compartment i. Now, the elements of the fractional coefficient matrix indicate the fraction of throughflow that is generated as a result of the demand from another node. The elements of the fractional outflow matrix are also called partial host coefficients by Ulanowicz (1986a).

The Augustinovics inverse matrix $N"$ is obtained in a manner similar to the Leontief inverse:

$$N" = (I - G")^{-1} \qquad (4)$$

(Due to our desire to keep matrix orientation in the from Row-to Column format, the Augustinovics inverse is seen here in its transposed form.). $N"$ is the output structure matrix for the network. Each element can be interpreted as (1) the output y_j in compartment j generated by a unit throughflow T_i in compartment i, or (2) the expected number of times a unit of matter or energy originating in i will pass through compartment j before leaving the system. $U"$ is the matrix of variances associated with the elements of $G"$. The fractional coefficient matrices (G', $G"$) and the structure (N', $N"$,) and variance (U', $U"$) matrices for the Cone Spring ecosystem are presented in Table 2.1.

We will later see how the structure matrices can be used to determine the origins and fates of any given flow in a network, including specific techniques for elucidating the underlying trophic structure. These matrices also serve as the foundation for the development of a number of information measures, many of which can be used to measure growth, development, specialization, and diversity in ecological systems.

Attributes of Individual Compartments

Network analyses produce results that permit the description of an ecological system in terms of: 1) attributes of the components, 2) relationships between the components, and 3) total system properties. In what follows, this section focuses on attributes of components, the next three sections focus on between compartment relationships, and the following three sections focus on total system properties. One of the goals of this work is to identify what ecological insights are gained from thoroughly analyzing a network and how well these insights carry over in comparisons of similar systems and very different ecosystems.

In characterizing individual compartments, the estimation of compartmental throughflows (T_i) is useful for revealing the relative rates at which the flow medium enters and leaves each component. Where standing stocks are also available, the turnover rate for each compartment can be calculated by dividing compartmental throughflow by the component's standing stock. Turnover rates are measured in units of inverse time and indicate how fast matter and energy are exiting from the compartment (because turnover rates are based on losses from compartments, they are presented as negative numbers). Fast compartments have high turnover rates compared to slow ones. The inverse of the turnover rate is turnover time, which is the time required for the equivalent amount the original standing stock to pass through the compartment. In the Cone Spring network, detritus (x_5) turns over every 113 days -a much slower rate than the living plant biomass (17 days) that supplies the detritus pool and the bacteria (8 days), detritus feeders (9 days), and carnivores (17 days) that depend on it for nourishment.

Finn (1976) and Patten *et al.* (1976) introduced a number of flow analysis measures, some of which characterize properties of individual system components. By calculating the rows sums of the N^* matrix, we obtain the outflow path length vector ply. If ply_i equals 1, the outflow from i is directly connected to compartment i. A value greater than 1 is obtained when indirect links (longer paths and/or cycles) help produce the outflow from i.

The column sums of N' produce the inflow path length (plz) vector. Each plz_j measures the penetration of exogenous and endogenous inputs to compartment j into the network.

The outflow path lengths for the Cone Spring model are 1.0 for plants, 4.37 for bacteria, 3.4 for detritus feeders, 4.4 for carnivores, and 3.37 for detritus. The inflow path lengths are 2.6 (plants), 1.6 (bacteria), 1.5 (detritus feeders), 1.9 (carnivores), and 2.06 (detritus). The inflow path length values for the animal components (including bacteria) of the ecosystem are derived entirely from endogenous inputs. For the compartments receiving exogenous inputs, the solar inputs the plants receive travel deeper into the network structure than do inputs to the detritus pool.

Finn is also responsible for introducing the first measures for estimating the amount of cycling in ecological flow networks. The diagonal elements of the structure matrices N'' and N' represent the number of visits that a unit of flow starting or ending in a given compartment makes to that same compartment. The number of straight through visits is always 1. Therefore, the cycled portion of throughflow in the ith compartment is calculated as:

$$c_{ii} = [(n_{ii}-1)/n_{ii}] / T_i .$$ (5)

The recycling efficiency is therefore expressed as:

$$r_{ci} = (n_{ii}-1) / n_{ii} .$$ (6)

The compartmental cycling efficiencies for the Cone Spring compartments are 0.79, 0.32, 0.24, 0.45, and 0.65 respectively.

BILATERAL RELATIONSHIPS

Whereas the suite of measures discussed in the last section focus on the status of single compartments as they are imbedded in the whole network, much useful information exists concerning the bilateral relationships between any two nodes of interest where the entire rest of the network serves as the conduit between them.

Input and Output Flow Environs

An input environ is the set of fractional flows generated in the system by one unit of inflow into one of the compartments receiving input from outside the system. There are as many input environs as there are compartments receiving imports. Similarly, an output environ is the set of fractional flows which occur among compartments when one unit of flows leaves one of the exporting compartments. There are as many output environs as there are compartments which export useful flows. In matrix terminology, the fractional coefficient and structure matrices for the ecosystem can be used to trace the origins or fate of any input, output, or internal flow represented in the network. Let us examine the backward case first. Because the elements of G' indicate the fraction of the recipient component's throughput that is generated from each donor compartment's throughflow, it is possible to estimate the intrasystem flow required to sustain the output from the recipient compartment. One must first

diagonalize the column of the G' matrix that corresponds to the output of interest (for example, to look at the output from the bacteria in the Cone Spring system, we would use column 2 to form a matrix whose diagonal elements are the elements of column 2 and whose other elements are 0), then multiply it by the input structure matrix N'. The resultant matrix T'_i is the unit input environ matrix for outflow y_i. The elements of the matrix are generally used to construct a diagram that indicates the location and magnitude of flow required to generate a unit of output within or from compartment i. Since all compartments in ecological systems lose medium to their environment, there will be one diagram for each compartment. The unit input environs for the Cone Spring ecosystem are show in Figure 2.4.

The unit outflow environs T'_i are constructed in a similar manner as the unit inflow environs. This time, the elements of the resultant matrix indicate the fate of a unit input into compartment i. One can
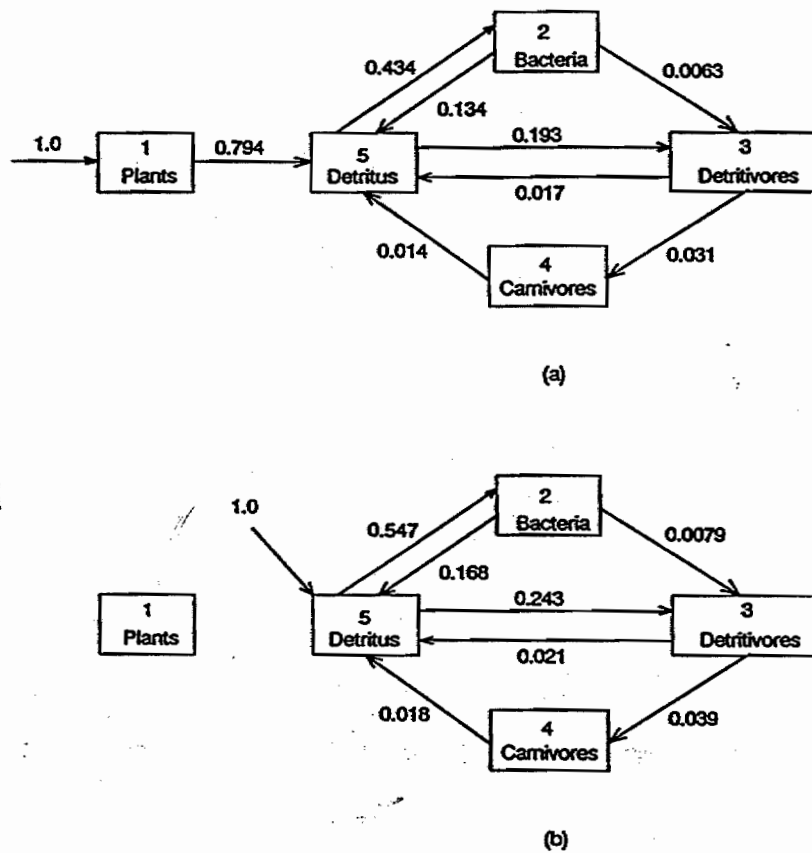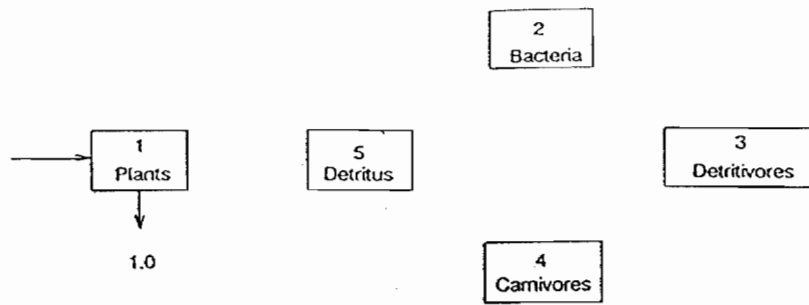

(a)

(b)
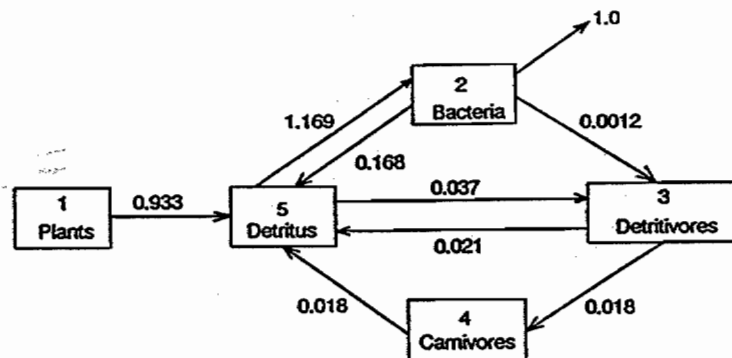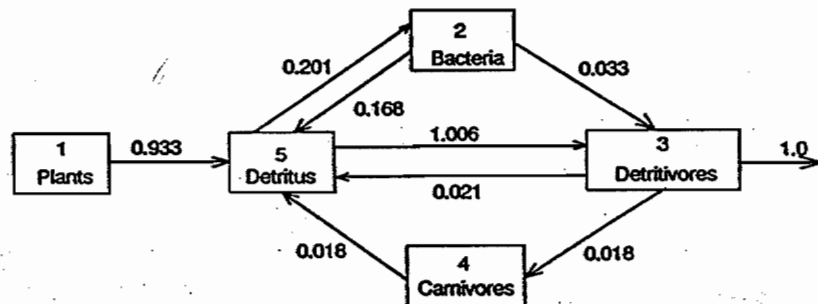
Fig 2.4. The unit input environs for the Cone Spring network.The relative amounts of internal exchanges generated by a unit input to a) Plants, b) Detritus.

(a)

(b)

(c)

Fig 2.5a-e. The unit outflow environs for the plant and detrital components of the Cone Spring network. Each figure depicts the amounts of internal flows necessary to sustain the single unit output shown. Flow units are relative to outputs.
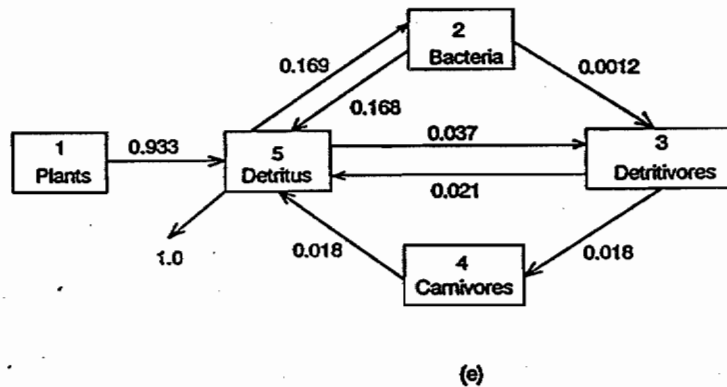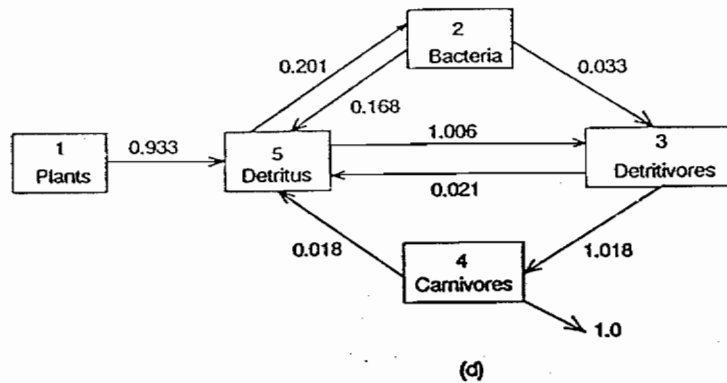
(d)

(e)

Fig 2.5 a-e. (continued).

construct unit input environs for all compartments in the system, but only the ones for compartments receiving external inputs are generally diagrammed. The unit outflow environs for the plant and detrital components of the Cone Spring model are shown in Figure 2.5.

Total Flow Matrices

Each value in the Leontief inverse matrix represents the amount of flow from i that it takes to generate one unit of output from j. Such a measure is very useful to know in the field of economics, where emphasis is upon exogenous inputs and (especially) outputs. But in ecology the exogenous outputs are dominated by respiration, and the question of how much each compartment contributes, directly and indirectly, to the sustenance of another, is of more interest. Conversely, one might wish to know the degree to which a particular species depends upon another for its material existence.

Szyrmer and Ulanowicz (1987) show how these total intermediate flows can be calculated from the components of the Leontief and Augustinovics inverses. Without going into the mathematical details here, the total flow from compartment i to j (excluding that which has been recycled back through i) is:

$$z_{ij} = [(n'_{ij} - d_{ij}) / n'_{jj}] T_j \tag{7}$$

where z''_{ij} is the total flow from i which reaches j (without recycling through i) over all pathways of all lengths, n'_{ij} are the elements of the Leontief inverse, $d_{ij} = 0$ when $i \neq j$ and $d_{ij} = 1$ when $i = j$, and T_j is the amount flowing through compartment j.

Looking backwards, it is possible to say how much of the medium arriving at j passed through i at some point in its journey through the system:

$$z'_{ij} = [(n^*_{ij} - d_{ij}) / n^*_{ii}] T_i \tag{8}$$

where z'_{ij} is the flow reaching j that at some earlier time passed through i and n^*_{ij} is the ijth element of the Augustinovics inverse.

In general, because of cycling $z'_{ij} \neq z''_{ij}$. In the Cone Spring example, of the 1675 kcal/m^2/yr leaving the bacteria, $z''_{24} = 73.02$ of those units eventually reach the carnivores; whereas of the 370 units consumed by the carnivores, $z'_{24} = 63.54$ of them passed through the bacteria along the way.

Perhaps the total flow information is most useful in its normalized form. Thus, the quotient z''_{ij} / T_i represents the fraction of the total output of i which reaches j and z'_{ij} / T_j, the fraction of j's total consumption which passed through i. In the output from the program NETWRK, the former quotients are called "total contribution coefficients", while the latter are termed "total dependency coefficients." The dependency coefficients are especially useful diagnostic indices because they represent the extended diet of the recipient compartment j. If one reads down the jth column of the normalized flow matrix, G', one sees the fractions of the actual diet of the jth species that come from the various other members of the community. Often the entries in the column are sparse, as most species derive food from no more than about 4 prey items. However, the corresponding column of the dependency coefficient matrix is usually filled with many non-zero values that portray the history of the same food. Occasionally, the extended diets reveal important differences not apparent in the rations of direct prey (Baird & Ulanowicz 1989). A Chesapeake Bay example would be the striped bass (*Monrone saxatillis*) and the bluefish (*Pomatomus saltatrix*). Both appear as pelagic carnivores. Their extended diets,

however, reveal significant niche separation. The striped bass is a recipient of energy and material that was passed up the pelagic grazing chain, while the medium reaching the bluefish had arrived mostly via the benthic community. This disparity in the extended diets provided a partial explanation of why the pesticide Kepone (a contaminant of the James River sediments) was appearing in high concentrations in the flesh of the bluefish, but was not prominent in the striped bass.

Investigators are troubled at first by the fact that the extended diet columns usually sum to a number greater than unity (i.e., they do not represent well-normalized probabilities). Their anxiety eases, however, once they realize that a given particle or quantum of energy usually passes through several compartments before reaching the predator in question. It is only natural, therefore, that the dependency coefficients represent a systematic multiple counting scheme. In fact, the sum of any column of this matrix is closely related to the average trophic status of the corresponding species (see section on trophic analysis below).

Markovian Transition Probabilities

Earlier, we introduced the fractional coefficient matrices G' and G", which partition a compartment's input and output throughflow relative to its origins or destination within the ecological network. Another interpretation treats the fractional coefficients as discrete-time finite state Markov transition probabilities (Kemeny & Snell 1960, Barber 1977a,b). A Markov process is the simplest type of stochastic process described in statistics. The condition of the ecosystem is represented as a finite set of states representing the biomasses in the compartments at one time and the flows between them. The condition of the system at any time is assumed to depend only upon the condition immediately beforehand and is independent of all time prior to that. Thus the condition of the system at one instant is assumed to contain enought information to predict the next condition or transition (Legrendre and Legendre 1983). It is called a first order process, since it is only going back one step in time to predict the next condition. Higher powers of G' and G" (for example, G'^1, G'^2,...G'^3) represent the kth order distribution of a unit of medium in each compartment to all compartments following k transitions.

One of the obvious network characteristics that has largely been ignored in previous discussions is the importance of standing stocks in determining system structure and function. In fact, only in our discussion of turnover rate and turnover time have standing stocks come into play. Barber (1977a, b) was the first to cast forward and reverse case flow analysis in terms of a Markovian model that incorporates storage as well as flow probabilities. Environ analysis (Matis & Patten 1981, Patten & Matis 1982) was later introduced as a formal methodology for computing storage partitions in addition to the usual flow partitions produced by traditional inflow and outflow analyses.

Environ analysis is founded on the development of linear donor and recipient controlled models of system dynamics. Any network (such as the Cone Spring network we use in our examples) can be described by a set of difference or differential equations. For a static linear system, this is achieved by making each flow a function of its donor compartment, $f_{ij} = a_{ij} x_i$. The linear coefficients a_{ij} can be stored in a matrix A^* so that

$$x = A^*x + z. \qquad (9)$$

where x is the vector of standing stocks and z the vector of imports into the system.

The diagonal elements of A^* are the turnover rates (expressed as negative numbers) for each compartment. Alternatively, one can normalize each flow by the standing stock of its recipient compartment. In this case, the change in standing stock is expressed using the following equation:

$$x = A'x + y. \qquad (10)$$

where y is the vector of total exports. A^* and A' (Table 2.2) are the forward and reverse linear system matrices for the ecological network.

Table 2.2. The forward (A'') and reverse (A') linear system matrices for the Cone Spring network.

	A″ matrix				
	1	2	3	4	5
1	-39.240	.000	.000	.000	31.160
2	.000	-44.640	.643	.000	13.720
3	.000	.000	-39.730	6.167	3.333
4	.000	.000	.000	-21.760	9.824
5	.000	1.454	.645	.000	-3.208

	A′ matrix				
	1	2	3	4	5
1	-39.240	.000	.000	.000	.000
2	.000	-44.640	.000	.000	44.640
3	.000	1.250	-39.730	.000	38.480
4	.000	.000	21.760	-21.760	.000
5	2.481	.447	.056	.047	-3.208

Knowing A^* and A', we can compute the discrete time Markov transition matrices P^* and P' as follows:

$$P^* = (I + hA^*) \qquad (11)$$

and

$$P' = (I + h(-A')) \qquad (12)$$

where I is the identity matrix and h is a scalar selected to guarantee that each element along the diagonal of the transition matrices is positive.

In terms of time, h represents the number of discrete points within one unit of time t. This makes sense when one remembers that the inverse of the turnover rate is turnover time, meaning that h must be small enough that it captures the dynamics of the fastest component (where t is given in whatever real time units used in formulating the flow rates for the model). The value of h must therefore be less than or equal to the value of fastest compartment's turnover time. An h of 1 means that 1 time unit (days, weeks, etc.) is sufficient for recording the turnover of standing stock of each compartment. An h of 0.25 days indicates that one must sample at least 4 discrete points within a time unit to catch the dynamics of the faster components. Each Markovian transition would correspond to 0.25 days.

An element p''_{ij} of P″ indicates the probability that an initial unit of medium will flow from i to j during t/h units of time, while an element p'_{ij} of P′ indicates the probability that the unit of matter or energy now in j traveled from i during one transition interval. The probabilities on the diagonal (p''_{ii} and p'_{ii}) then indicate the probability that the medium originated or remained in i during one transition. Table 2.3 shows the Markov transition matrices for the Cone Spring ecosystem.

Table 2.3. Forward (P″) and reverse (P′) Markov transition matrices for the Cone Spring network.

				P″ matrix		
	1	2	3	4	5	6
1	.387	.000	.000	.000	.487	.126
2	.000	.303	.010	.000	.214	.473
3	.000	.000	.379	.096	.052	.472
4	.000	.000	.000	.659	.154	.187
5	.000	.023	.010	.000	.949	.017
6	.946	.000	.000	.000	.054	.000

				P′ matrix		
	1	2	3	4	5	6
1	.387	.000	.000	.000	.000	.613
2	.000	.303	.000	.000	.698	.000
3	.000	.019	.379	.000	.601	.000
4	.000	.000	.340	.659	.000	.000
5	.039	.007	.001	.001	.949	.003
6	.195	.299	.154	.017	.335	.000

. This approach essentially relates the flow and standing stock properties of the system. However, as Barber (1978b) points out, a high resolution perspective of an ecosystem might view ecological resources as particulate in nature. Since all of the particles that comprise a component's standing stock presumably possess the same potential behavior, it follows that a model of resource flow could follow the movement of one such particle through the network, where the distribution of all particles is represented in terms of the random variables of a Markovian transition matrix, and the flow of particles among components is described as a stochastic process. The meaningful application of such a model to an ecological network requires that one exercise care in aggregating components. For example, it would be incorrect to lump a fast component with a slow one, since the probabilities for retention of medium would vary significantly between the two.

Markovian Flow Partitions

The higher powers of the P'' and P' matrices can be used to trace the fates or origins of a unit medium of flow through the network over k transitions in the same manner as that described earlier for the G'' and G' matrices. P'' and P' are time-referenced by h, so the probabilities now represent probabilities of transfer over t/h units of real time. The Leontief and Augustinovics inverses (and associated variances U'' and U') can be computed for each of the Markovian probability matrices using the following equations:

$$Q'' = (I - P'')^{-1} \tag{13}$$

and

$$Q' = (I - P')^{-1} \tag{14}$$

The elements of Q'' and Q' indicate, in the first case, the number of times material in i will visit j before leaving the system, and in the second case, the expected number of times substance in i has visited j since entering the system. When the model is parameterized so that h approaches infinity (corresponding to continuous time), the P'' and P' matrices are identical to the G'' and G' matrices, so that N'' and N' are equivalent to Q'' and Q' (Patten & Matis 1982). Patten (1985) has developed an extended environ analysis (not presented in this volume) that is a microscopic analysis for tracing the position of particles in the network through time.

A pair of Markovian unit input and output environs (E'_i and E''_i) can be computed for each system input and output by multiplying the linear coefficient matrices A' and A'' by a diagonalized column of their negative inverses (see Patten & Matis 1982). These environs are storage-referenced rather than throughflow-referenced (as are the environs T''_i and T'_i). The elements of the output environ matrix E''_i are interpreted as the flow from i to j that is generated from a unit input to i. The values in the input

environ matrix E_i^l give the quantities of flow needed to generate a unit of outflow from i. Standing stock is now treated as a source of flow for this set of environs.

Origin and Destination Probabilities

When one computes the row or column sums over all environs E_i^l and $E_i^.$, the end result is an expression of the origin (W') and destination (W") probabilities for the ecosystem resources. These matrices are shown for the Cone Spring model in Table 2.4. Each entry w'_{ij} indicates the probability that medium in i had entered the system via the inflow to j. The elements of W" give the probability that a unit of medium entering the system as z_j will exit via the outflow from j. In Cone Spring, we can see most of the exogenous energy entering from both primary production and detrital inputs will exit via the detritus and bacteria compartments. The origin probabilities show that the overwhelming majority of energy in all compartments can be traced back to the sun.

Storage Partitions and Residence Times

The most significant analytical contribution of discrete time environ analysis is its ability to estimate the mean residence times and storage partitions for network medium. The mean residence time matrices M' and M" (Table 2.5) are calculated as $-(A')^{-1}$ and $-(A')^{-1}$, respectively. When corrected for the Markov time step h (as are all the matrices computed using the ENVIRON program), the values stored in these matrices indicate the mean time that a particle now in compartment i has resided in j since entering the system (M') or the average time that resources in i will spend in j before leaving the system (M"). When referenced back to system input (z) or output (y), one can estimate what portion of each compartment's standing stock was generated by each system input (stored in matrix X') and what part of each standing stock will leave the system via each outflow (X"). For example, for the Cone Spring ecosystem (Table 2.6), all of the plant biomass is generated via the input to x_1, but the other compartments owe some of their standing stock to external inputs to the detritus pool. On the other hand, energy that began as plant biomass may ultimately leave through any of the system's outflows. However, an almost equal amount of the energy bound in the detrital biomass will ultimately exit through bacterial metabolism as that directly leaving the detritus compartment.

TROPHIC ANALYSIS

Anyone who has worked in the field of ecology for even a short while is usually impressed by the complexity of the feeding relationships within an ecosystem. The feeding "web" is often quite convoluted. Given this, one might well ask of what use is it to continue to speak of herbivores and carnivores as if the feeding relationships always could be sequenced in a linear fashion? If one focuses on any species beyond the primary producers, then one is often faced with a situation where there exists

Table 2.4. Matrices of destination (W") and origin (W') probabilities for the Cone Spring network.

			W" Matrix		
	1	2	3	4	5
1	.206	.295	.151	.017	.331
2	.000	.000	.000	.000	.000
3	.000	.000	.000	.000	.000
4	.000	.000	.000	.000	.000
5	.000	.371	.191	.021	.417

			W' Matrix		
	1	2	3	4	5
1	1.000	.000	.000	.000	.000
2	.933	.000	.000	.000	.067
3	.933	.000	.000	.000	.067
4	.933	.000	.000	.000	.067
5	.933	.000	.000	.000	.067

Table 2.5. Mean residence time matrices for Cone Spring ecystem resources

			M' Matrix		
	1	2	3	4	5
1	.025	.000	.000	.000	.000
2	.024	.026	.001	.001	.376
3	.024	.004	.026	.001	.376
4	.024	.004	.026	.047	.376
5	.024	.301	.001	.001	.376

The mean time that a particle now in compartment i has resided in j since entering the system.

			M" Matrix		
	1	2	3	4	5
1	.025	.009	.005	.001	.299
2	.000	.026	.002	.001	.117
3	.000	.002	.026	.007	.058
4	.000	.006	.285	.047	.169
5	.000	.012	.006	.002	.376

The average number of times that a particle in i will cycle to j before leaving the system.

Table 2.6. Storage partitions using environ analysis for the Cone Spring network.
Variances in parentheses.

Inflow Z_1 partitioned among compartments X_i

X 1	285.000	(81225.01)
X 2	108.819	(51919.48)
X 3	55.996	(29603.41)
X 4	15.866	(16339.81)
X 5	3340.548	(.169E+08)

Inflow Z_5 partitioned among compartments X_i

X 1	.000	(.000)
X 2	7.781	(198.309)
X 3	4.004	(116.878)
X 4	1.134	(66.068)
X 5	238.852	(57050.450)

Outflow Y_1 partitioned among compartments X_i

X 1	58.686	(3444.160)
X 2	.000	(.000)
X 3	.000	(.000)
X 4	.000	(.000)
X 5	.000	(.000)

Outflow Y_2 partitioned among compartments X_i

X 1	83.951	(8055.764)
X 2	92.469	(8550.570)
X 3	3.426	(620.549)
X 4	2.846	(931.387)
X 5	1327.794	(1763035.000)

Outflow Y_3 partitioned among compartments X_i

X 1	43.141	(2127.315)
X 2	8.160	(708.929)
X 3	47.415	(2248.191)
X 4	1.463	(245.954)
X 5	682.328	(465571.111)

Outflow Y_4 partitioned among compartments X_i

X 1	4.827	(26.641)
X 2	.913	(8.878)
X 3	5.306	(28.155)
X 4	9.491	(90.073)
X 5	76.357	(5830.469)

Outflow Y_5 partitioned among compartments X_i

X 1	94.392	(10184.020)
X 2	15.057	(2904.263)
X 3	3.852	(784.492)
X 4	3.200	(1177.450)
X 5	1492.921	(2228813.000)

a multitude of feeding pathways of different lengths that can be traced back from the animal of interest to the primary producers. As Cousins (1985) observes, "A hawk feeds at five trophic levels."

One strength of network analysis techniques is that they offer the investigator a systematic and mechanical way of keeping track of the myriad of feeding pathways upon which any given population within the network depends. Thus, it might have happened that the hawk in Cousin's network of feeding relationships received its sustenance along some 75 or more separate pathways leading back to the plants. If necessary, those pathways could be enumerated using the backtracking methods presented in the next section. However, if one also has a weighted network description of the ecosystem, then the simpler input/output methods portrayed earlier will suffice to determine how much of the flow reaching the hawk arrives after traversing pathways of different integer lengths. Suppose that an analysis reveals that 5% of the hawk's diet is plants, 35% arrives after two feeding transfers, 40% after 3, 15% after 4 and 5% after 5 exchanges. Then the hawk functions on the average trophic level 3.8 (= $.05 \times 2 + .35 \times 3 + .40 \times 4 + .15 \times 5 + .05 \times 6$). A unique number has been assigned to the trophic status of the hawk. This value can be compared with the corresponding trophic positions of other members of the community for purposes of ranking. More interestingly, any change in this value as the external conditions or community composition change could be used to help assess how well the hawk is adapting to the new conditions. (A lowered trophic status has been hypothesized to be indicative of stress on the population in question.).

Given the partitioning of flows over pathways of various lengths, one could estimate the average trophic status for each compartment in the network. Also this information is useful in constructing a picture of the trophic status of the entire community. Lindeman (1942), for example, envisioned a trophic pyramid, or concatenation of trophic relations wherein the amounts transferred to each higher level would become progressively smaller. The same matrix methods we used to portray the predator's diet, makes it possible to apportion the activity of, say, the hawk to a series of abstract trophic levels 2 thru 6 in such a way that the matter and energy of the system is conserved. The resulting hypothetical trophic chain would yield a profile of the efficiencies of trophic transfer that could serve as a diagnostic of how well the community is functioning. For example, well-developed communities could be expected to have longer trophic chains with more material and energy reaching the higher levels. Stressed communities, on the other hand, are hypothesized to possess short, but intense chains of transfers.

The pivotal question remains how to calculate the trophic apportionments. A basic method for doing this was described in Ulanowicz & Kemp (1979) and is sketched below:

Recall that the elements of the input structure matrix, g'_{ij} represent the fraction of the total input to j that flows directly from i. The utility of the matrix G' is that its successive powers quantify how much flows from i to j along all pathways of length equal to the power to which G' is raised. For example, the ijth element of $[G']^3$ (the matrix G' multiplied twice by itself) is precisely the fraction of total input to j which flowed from i along all pathways of length 3 connecting i to j. Hence, if one begins with a knowledge of how much each population acts as a primary producer, one can use the powers of G' to see how that primary production reaches the organisms at higher trophic levels after each succeeding transfer.

If z_i represents the external input to compartment i and T_i the total flow through i, then the quotient z_i/T_i will be taken to quantify the degree to which i acts as a primary producer (or supplier) of medium to the network. Consider a row vector with elements $L_i = z_i/T_i$. It follows that the product LG' will also be a row vector whose elements describe how much arrives at each node after exactly one internal transfer, that is, how much each population feeds at the second trophic level. In general, $L[G']^{m-1}$ will be a row vector giving the fraction of its total consumption which each population receives at the mth trophic level.

If the feeding network contains no directed cycles (a simple cycle is observed when the initial and terminal nodes of a sequence of arcs are identical, but none of the intermediate nodes are repeated), the sequence of row vectors just described will truncate (yield a row vector of all zeros) after at most n-1 steps, where n is the number of nodes in the feeding web. It then becomes possible to construct an $n \times n$ matrix L whose ith row is identical to $L[G']^{i-1}$. This matrix is called the Lindeman trophic transformation matrix. The ith row of L describes how much of the feeding activity of each population takes place at the ith trophic level. It usually happens that only the first m rows ($m < n$) are non-zero, hence the aggregation of n species into m levels. The jth column of L shows the fractions of the total feeding activity of population j that occur at each trophic level. In a network without cycles these columns will always sum to unity, indicating that all activity has been accounted for.

The concept of trophic level is best confined to living, feeding organisms, so that the creation of L will involve only the living species. It is for this reason that in the standard data format used in this book the living members of the community should appear first in the compartment list. This allows the trophic aggregation routine in NETWRK to operate initially only on the living food web. Pimm (1982) relates how directed cycles are rare among the livings members of an ecosystem and are usually of very small magnitude if they do occur. Hence, in the event that an ecosystem network does possess a few small cycles in the food web, these usually can be removed (see the section on cycle identification below) without significantly impacting the subsequent trophic analysis.

welfare is how much this capital actually circulates. Hannon (1973) suggests that ecosystems bear more resemblance in this regard to economic communities than to physical systems. The activity of a system is often the matter of most immediate concern, as evidenced by the intense interest in such activity as the gross national product.

Hannon suggested that the aggregate of all the transfers, otherwise known as the total system throughflow or throughput, should be the primary measure of system size. The equations used to calculate T depict it as the sum of all internal and exogenous inputs to system components, or the sum of all outputs (endogenous flows, exports, and respirations) of all the compartments. There remains some disagreement on exactly how to calculate T (Hannon 1973, Finn 1976). The most inclusive sum (of all external and internal flows) is now gaining favor, i.e., the sum of both of the input and output representations of T described earlier. This all-inclusive form of the total system throughflow is the same whether viewed from the input or the output perspective and remains well-defined, even if the system is not in balance. We therefore recommend its use as calculated in the programs NETWRK, ECOSYS and ENVIRON.

The non-inclusive value for T for the Cone Spring network is 30,626 kcal/m^2/yr. The all-inclusive value for total system throughput is 42,445 kcal/m^2/yr.

It should be noted that T increases with the number of subdivisions into which the system is divided and is therefore dependent upon the degree and manner of aggregation decided upon by the investigator. A special case occurs when compartments are aggregated by the program AGG which pools compartments specified by the reseacher. When two or more compartments are pooled, the flows between the previously separated compartments are modelled as loops leaving and re-entering the new pooled compartment. Under these circumstances (with "self loops") T remains the same when compartments are aggregated (see Warwick and Radford 1989, this volume, for examples).

Finn Cycling Index

Of the activity represented by T, some represents medium flowing through the system, whereas the rest of the activity consists of medium being cycled within the community. As discussed in the previous section, the exact proportions of these two components is ecologically significant. Odum (1969), for example, cites an increased amount of cycling as an important attribute of more mature ecosystems. (There are indications [Ulanowicz 1985, Schneider 1988], that the immediate response of a system to a eutrophying stress is to increase its proportion of internal cycling. This is interpreted as a sign of a healthy system coping with stress.) Most agree that the proportion of recycle in a network is a significant component of its overall status.

The reader will recall that the diagonal elements of the total flow describe exactly the amount of flow which leaves compartment i and eventually makes its way back to i. That is, it represents how much of the throughput of i is engaged in cycling. It follows that the sum of the diagonal elements quantifies the total amount of system activity devoted to cycling as:

$$T_{cyc} = \sum_{i=1}^{n} z'_{ii} = \sum_{i=1}^{n} z''_{ii} \tag{15}$$

where T_{cyc} is that part of T which represents cycled flow. Accordingly, the fraction which this amount comprises of the total system throughput is called the Finn cycling index, CI, where

$$CI = T_{cyc}/T. \tag{16}$$

In Cone Spring $T_{cyc} = 2816$ kcal/m^2/yr, giving a Finn cycling index of 0.0663.

Average Path Length

The cycling index gives the investigator an idea of how retentive a system is. There is another measure of system retention called the average path length, which combines the attribute of recycle with that of trophic length. The average path length (APL), as its names implies, is the average number of transfers an arbitrary quantum of medium will experience on its trip through (and around) the network. If Z is the sum of all the exogenous inputs from other systems, then $T - Z$ is the amount of flow occurring after the medium has entered the system. It follows that $APL = (T - Z) / Z$ will yield the average number of times this input is transferred by the system. In Cone Spring this quantity comes to 2.59. One would expect APL to rise under normal succession and to decrease in the face of stress.

TOTAL SYSTEM PROPERTIES: MEASURES OF STRUCTURE

The measures examined in this section characterize the structural self-organization of an ecosystem. By structural organization is meant the way in which the different compartments are interconnected. These interconnections are determined by who is eating what. To be more precise, a compartment can be treated as a source or as a consumer of resources. The structural organization of an ecosystem is determined by the way in which the consumers make use of the other compartments as resources. The set of sources used by a consumer is its resource niche. Characterizing the resource niche of the compartments describes the structural self-organization of an ecosystem.

A description of the resource niche must characterize each of the individual sources and the total se of sources used by the consumers. The individual sources are characterized by the quantity and quality

The entire ecosystem is generally larger than just the living components, and most networks of ecosystems flows include one or more abiotic compartments such as detritus or inorganic nutrient pools. It is through these non-living elements that the preponderance of material or energy cycling in the ecosystem occurs. Although the method of trophic partitioning does not involve the abiotic elements, the trophic aggregation would be incomplete if it did not in some way incorporate the recycled quantities. Following an assumption commonly made during the course of the IBP program, Ulanowicz (1989) assigns all the non-living components a trophic status of one. This has the effect of equating herbivory and detritivory in the trophic sense (all of which appeals to the intuition, because it is sometimes difficult to separate living from dead plant tissue). To complete the mass balance, the Lindeman transformation matrix can be expanded by d (the number of non-living members in the community) rows and columns that are all zeros except for the terminal d x d submatrix,which would contain a single one somewhere in each column (usually in the last entry).

The expanded Lindeman transformation matrix is then used to complete the trophic analysis. In the Cone Spring example, there are 4 living populations, as shown in Figure 2.2a. The corresponding Lindeman transformation matrix looks like the following:

Effective Trophic level	Plants	Bacteria	Detritivores	Carnivores
1	1	0	0	0
2	0	1	0.969	0
3	0	0	0.031	0.969
4	0	0	0	0.031

(Here the number of trophic levels just happens to equal the number of living species.) To find the effective trophic level of each of the species, one simply multiplies all the members in each row of the Lindeman matrix by the number of that row and then adds down the column. As a result, one finds that the plants are at level 1.0, the bacteria at 2.0, the detritivores at 2.03 and the carnivores at 3.03. The detritus has been assigned to level 1.0.

The trophic "aggregation" of the Cone Spring network (keeping the autotrophs and the detritus separated for the time being) is depicted in Figure 2.6. The trophic levels are designated by Roman numerals to avoid confusion with the species designations. The exports and respirations of the trophic levels are calculated by premultiplying the actual losses by the Lindeman matrix, i.e., Le and Lr, where e and r are the column vectors of exports and respirations, respectively. (These results are called the canonical exports and canonical respirations in the output from NETWRK.) In the Cone Spring example there are no exports from levels III and IV.

To calculate the interlevel transfers and recycles one employs the transformed flow matrix, LFLT in ways that need not be detailed here (Ulanowicz, 1989). The focal structure in Figure 2.6 is the "grazing chain," or the transfers from each level to the next higher up. (The Cone Spring network is peculiar in that it evinces no herbivorous grazing. The usual circumstance is for a network to possess an uninterrupted grazing chain.) In general each level returns a certain amount to the detrital pool, much of which is fed back into the grazing chain as detritivory (the 7514 kcal/m^2/yr shown flowing from D to level II.) The exogenous inputs to, exports and respirations from, and circulation within the detrital pool (none) all appear separately on Figure 2.6.
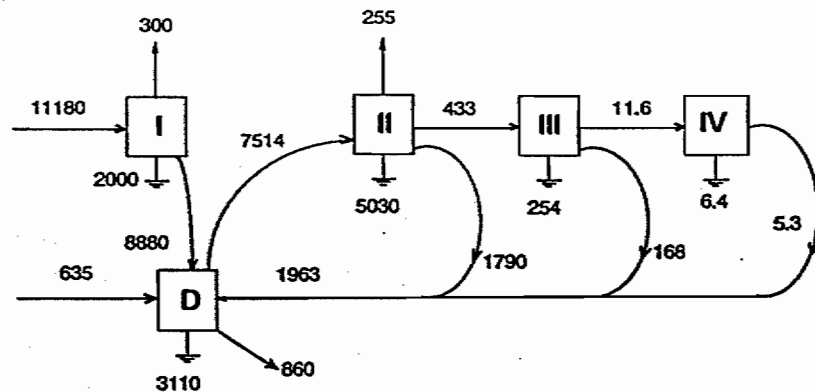

Fig 2.6. The trophic aggregation of the Cone Spring network with the autotrophs and detritus separated. The trophic levels are designed by Roman numerals, D represents the detritus pool.

If the detrital pool is merged with the autotrophs, one sees the classical trophic pyramid as portrayed in Figure 2.7. Here the interlevel transfers form a "Lindeman spine" of ever decreasing flows falling from 13,778 units into trophic level I to a scant 11.6 into trophic level IV. The uniform decrease allows one to calculate the trophic efficiency of each level as the amount it transfers to the next higher level divided by the amount it received from the one below it. In Cone Spring these efficiencies are uniformly decreasing. Experience with a variety of ecological networks shows that there is a tendency for trophic efficiencies to decrease at higher levels. The trend is not without exceptions, however, and it appears that each exception can be traced back through the Lindeman transformation matrix to an interesting peculiarity of the network being analyzed (Baird & Ulanowicz 1989).

Despite the obvious complexity of most observed foodwebs, it is apparent that network analysis can be employed to reveal an underlying canonical trophic structure that is useful for interpreting how a particular ecosystem is functioning.
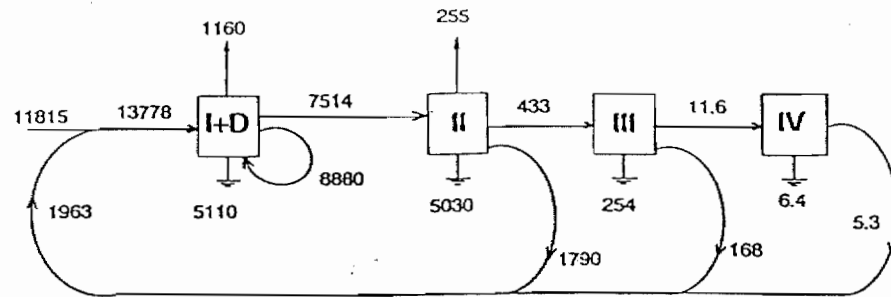
Fig 2.7. The trophic aggregation of the Cone Springs network with the autotrophs merged with the detrital pool (I + D).

CYCLE ANALYSIS

Positive feedback is a critical phenomenon in determining overall system structure. Positive feedback occurs in ecosystems when flows cycle among compartments. Compartments engaged in positive cycling will have an increase in output relative to those compartments which do not. This could result in a strong competitive advantage over non cycling compartments during growth phases. As well, positive feedback may also be a stabilizing factor in systems with a time lag. Cycles in ecosystems are an important factor contributing to their autonomous behaviour (Ulanowicz 1983, 1986a). Furthermore, it has been argued (Kay 1984) that the organization of flow in ecosystems reflects a tendency to degrade the exergy content of the flowing energy while conserving materials flowing within the system. Exergy degradation is related to straight-through flows, while material conservation is related to the cycles in the ecosystem. These are reasons for studying the cycling in ecosystem flows and in particular for separating the flow network into two components: the cycles and the straight-through flows.

The process of separating cycles from straight-through flows has two distinct aspects. The first involves identifying which components are connected together to form a cycle. This is a straightforward topological problem whose solution can be found in most elementary graph theory texts. The second aspect involves partitioning the flows among compartments into those associated with cycles and those associated with straight-through flows. This is a problem which has not been dealt with extensively and which still requires further investigation.

What follows is a brief description of the process of identifying cycles. More detailed descriptions can be found in Ulanowicz (1983, 1986a). The description of how NETWRK accomplishes this task can be found in Ulanowicz & Kay (1986). The flows between various pairs of compartments are defined by non-zero entries in the exchange matrix.

Topological Cycles and their Enumeration

The identification of the structure of the cycle, that is, what is topologically connected to what to make up the cycle, begins with a depth-first search (See Ch. 6 of Horowitz & Sahni 1984). The purpose of the search is to identify the cycle arcs. Cycle arcs are connections between compartments whose direction is from a higher compartment number to a lower compartmental number. In the Cone Spring example the arc from detritus to bacteria (from 5 to 3) is a cycle arc. If one searches for cycles starting only from those compartments which have cycle arcs into them, one is assured of finding all the cycles. This significantly cuts down on the time for searching out and enumerating all the cycles.

The next step is a backtracking search. The idea of backtracking search is to start with one compartment and see if it is connected to the next higher numbered compartment (i.e., is 3 connected to 4?) If it is connected, then we see if this next compartment (4) is connected to the lowest numbered compartment not already visited whose number is as least as great as the starting compartment number (i.e., is 4 connected to 5?). The process continues until the starting compartment is reached or until we find a compartment which is not connected to any compartment whose number is higher than the starting compartment number and which has not been visited. In the former case a list of the visited compartments constitutes a simple cycle. In the latter case we back up one compartment in our list and see if it is connected to the second next lowest numbered compartment not already visited (i.e., is 3 connected to 5, given that 3 is connected to 4, and 4 is not connected to any compartment whose number is greater than 3, nor to 3) and so on.

This is an algorithm which is best described by pictures rather than words. The idea is to treat each compartment as a dot. The dots are numbered. Compartments i and j are connected if the ijth element of the exchange matrix is non-zero. One tries to find a path, along the arcs connecting the dots, which starts and finishes with the same dot and which contains no other dot more than once. The trick is that from any dot you may only try to follow a path either to the lowest numbered dot (whose number is greater than the original dot) which you have not already tried, or to the original dot.

For the Cone Spring example there are 5 single cycles (See Figure 2.8). For the Crystal River ecosystem of 17 compartments (see Ulanowicz 1986a) there are 119 cycles. The number of possible cycles goes

up as approximately the factorial of the number of compartments. Luckily, most ecosystems are very sparse, that is, very few of all possible connections are actually made.
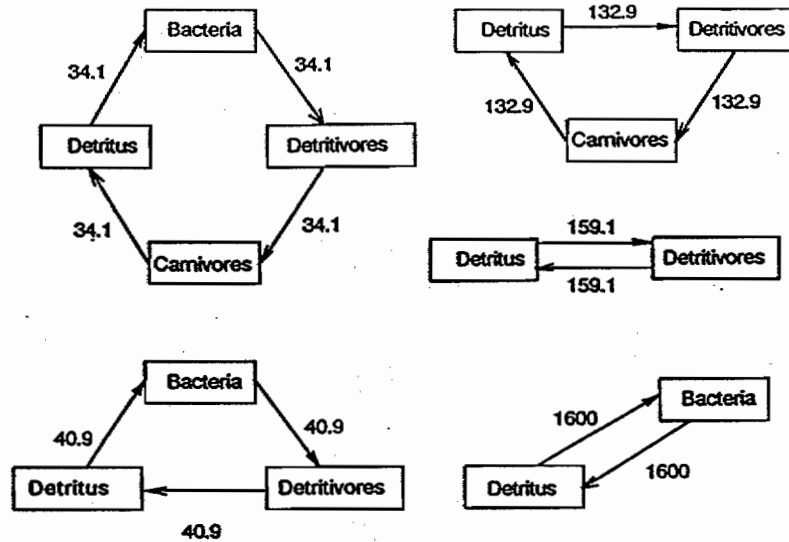

Fig 2.8. The five simple, directed cycles of energy flow in the Cone Spring network weighted according to circuit completion probabilities.

Functional Cycles and their Removal

Once the structure of the cycles has been identified, the next step is to separate the flow network into 1) arcs associated with cycles and 2) arcs associated with straight-through flow. This process begins by identifying the critical arc in each cycle. In each cycle, there is one arc which has the smallest flow through it. If this arc is removed and the value of the flow through it is subtracted from the flow in each of the other arcs in the cycle, then the cycle is effectively removed from the flow network. There is no other arc in the cycle whose flow rate could be set to zero (thus breaking the cycle) without some flows in some cycle arcs becoming negative. Thus, the critical arc is the cycle arc with the smallest flow through it.

It is most likely that an arc is part of more than one cycle. The collection of all the cycles which share the same critical arc is called a nexus. When a critical arc is removed from the network, all the cycles in the nexus are broken. The portion of the flow through the critical arc which is subtracted from each of the arcs in each of the cycles, is determined by calculating the circuit probabilities. This is done by multiplying the g'_{ij} for each cycle arc to obtain a weight for the whole cycle. The weight for each cycle is then summed for all the cycles in the nexus to give the nexus weight. The portion of the flow through the critical arc which is subtracted from each arc of a particular cycle is that cycle's weight divided by the nexus weight.

The process of removing the cycles from the network begins by finding the smallest critical arc and its nexus. The cycles in the nexus are then eliminated. This is done by subtracting the appropriate amount from each entry in the exchange matrix associated with a cycle arc from the nexus. Once a nexus has been eliminated, the updated exchange matrix is again searched for the smallest critical arc. This arc's nexus is then removed and so on until all the cycles have been eliminated. The exchange matrix which remains at the end of this process represents the network of straight-through flows. The matrix which represents the cycles is found by subtracting the straight-through flow exchange matrix from the whole network exchange matrix.

Issues for Further Research

The algorithm used in this process is not as robust as one might like. If the compartments are numbered differently, the resulting cycles might be slightly different. Where the value of the smallest critical arc is shared by more than one arc, the order in which these arcs are removed is arbitrary. (The order in which arcs are removed affects the overall partitioning of cycles and straight-through flow.). There is a problem when a cycle has more than one arc with the value of the critical arc (multiple critical arcs). It is not clear if these problems can be resolved in a satisfactory way. However, Ulanowicz and Kay have run a number of networks through the algorithm and found that these problems affect a very small percentage of the cycles.

How one decides on a critical arc and the partitioning of its flow is a subject of debate. By subtracting the total value of the smallest arc one is assuming that all the flow through it is part of a cycle and none contributes to straight-through flow. Using the circuit probabilities to assign the flow to the various cycles in a nexus is only one of several possible schemes. However, the schemes used in the partitioning algorithm are the most intuitively satisfying ones proposed so far.

Finally, partitioning the network into two components (cycles and straight-through flows) is a reductionist analysis which assumes superposition. It could be argued that the synergistic effects in the network make it irreducible. Exercises, such as this book, which test the analytical techniques suggested in this chapter on a broad spectrum of ecosystems, will ultimately determine the utility of this analysis.

TOTAL SYSTEM PROPERTIES: MEASURES OF FLOW

The order in which analyses are being discussed reflects an increase in the degree to which the associated variables characterize the total system. This progression culminates now in the consideration of how to quantify attributes of the whole system. That is, how might one encapsulate a property of all of the bilateral relationships into a single number? Such a number would resemble a state variable in thermodynamics. It would describe the overall status of the system, but not always its exact configuration. Many configurations can give rise to the same value of the system attribute. Thus, one should not expect whole system properties to specify exactly what is occurring at, say, the predator-prey level. The prediction capability of a whole-system index as to what is occurring at finer scales is minimal at best.

What is gained, however, is the potential for weaving several of these global variables into a coherent picture of ecosystem dynamics at the macroscopic level, the possible discovery of a new phenomenological principle that describes ecodynamics in the same sense that the laws of thermodynamics describe the evolution of macroscopic physical systems. Of course, it may eventually turn out that some of the variables to be described will turn out to be epiphenomenological, and do not quantify truly emergent properties of ecosystems. Such possibility notwithstanding, the verdict on this issue is still far from being written, and experimentation with these quantities is today one of the most exciting and potentially fruitful endeavors in all of science. Given the tools presented here, the reader is invited to apply these ideas to his/her ecosystem of interest and to join in the excitement of the search.

Total System Throughflow

One of the first questions one might ask about an ecosystem is how big it is. Size is generally reckoned in terms of how much material or energy an object possesses. Thus, one could begin by calculating the total system biomass, or the total energy content of all the ecosystem parts.

While these values are certainly germane to the concept of system size, they are by no means its exclusive determinants. For example, it is of distinct secondary interest to an economist how much money resides within a particular economic community. What appears to matter more to the general

of resources extracted from them. The set of sources is characterized by the relative sustenance provided by each source to each consumer. The use of the set of sources, from the perspective of the whole ecosystem, must also be characterized.

Mathematically, the flows in the ecosystem can be represented as follows: If F is the matrix of resource flow between compartments, the total output from a compartment (outj) is given by:

$$out_j = \sum_{i=1}^{n} f_{ji} \tag{17}$$

Similarly, the total input to a compartment (in_i) is given by:

$$in_i = \sum_{j=1}^{n} f_{ji} \tag{18}$$

Then the proportion of the total outputs from all compartments which flows from i is given by:

$$P_i[X_s] = out_i / \sum_{j=1}^{n} out_j \tag{19}$$

and the proportion of the total flow into compartment j which comes from compartment i is given by:

$$P_i[X_s / X_{cj}] = f_{ij} / in_j . \tag{20}$$

Similarly, the proportion of the total inputs to all compartments which flow into i is given by:

$$P_i[X_c] = in_i / \sum_{j=1}^{n} in_j \tag{21}$$

X_s is associated with output from a compartment (i.e. it acts as a source) and $P_i[X_s]$ can be taken as the probability of a quantum of resource coming from compartment i. Similarly X_c is associated with input to a compartment (i.e. acts as a consumer) and $P_i[X_c]$ can be taken as the probability of a quantum of resource flowing into compartment i. The probability distribution $P_i[X_s]$ describes the compartments in the system as sources and the probability distribution $P_i[X_c]$ describes compartments as consumers.

It should be noted that only flows into living compartments are considered. It is not sensible to talk about the resource needs of a non-living compartment. Non-living compartments act as sources and sinks in ecosystems but not as consumers. These definitions also apply to the general non-steady state situation.

Measures of resource niche and structural organization (S & D)

Using these definitions and the information theory measures described in the appendix of this chapter, two measures of resource niche and structural organization can be defined:

$$D = H[X_s] \text{ or } - \sum_{i=1}^{n} P_i[X_s] \log P_i[X_s] \tag{22}$$

and

$$S = H[X_s / X_c] \tag{23}$$

or

$$S = - \sum_{i=1}^{n} P_i[X_c] \sum_{j=1}^{n} P_j[X_s / X_{ci}] \log P_j[X_s / X_{ci}]; \tag{24}$$

where $0 < S < D < \log n$.

D measures the diversity with which compartments are used as sources by other living compartments. As the utilization of compartments as sources becomes more evenly distributed over all the compartments acting as sources, D increases. D, in a sense, indicates the degree to which the potentially available sources of resources are utilized in the particular ecosystem being examined.

S measures the diversity of sources utilized by each living compartment, averaged over all living, consuming compartments. When S is large, the living compartments are on average generalists, that is they use many sources of food evenly. When S is small, the compartments are on average specialists. They have a small number of sources, one of which is used heavily. In effect, S measures the average amount of choice in picking sources exercised by each living compartment during the time the ecosystem is observed. This is equivalent to measuring the average resource niche breadth. Note that this interpretation is particularly meaningful when the compartments correspond to individual species.

If the flow rates do reflect the proportion of available resource coming from each source, then S measures the average relative amount of sustenance provided to each living compartment by each source. That is, in an inverse sense, S indicates the average dependence of any one living compartment on the available sources. (The larger S, the less the compartments depend on any one source.) S measures the diversity of ecological interaction between the compartments which occurs through the food web.

In summary, D measures the diversity with which the compartment's outputs contribute to the total resource flows in the system. S measures the average resource niche breadth and, hence, indicates how specialized the compartments are. Together, S and D characterize the resource niche, that is how sources are being utilized by the compartments.

The structural organization of an ecosystem reflects the food consumption patters of the species which make up the compartments. The pressures which shape structural organization (e. g. positive feedback) act to modify the patterns of source utilization by the individual compartments. These patterns can be monitored using S. These pressures will also affect the overall utilization of sources in the ecosystem. This can be monitored by D. Thus, the structural self-organization of the ecosystem can be monitored using D and S.

Applications of S & D

Following are some examples to illustrate the utility of S and D. Figure 2.9 shows three hypothetical examples. The compartments labelled X are the compartments as sources and those labelled Y are the same compartments as consumers. Assume that the same amount of resource flows through each link. In system (a), each compartment is as specialized as possible ($S = 0$) and there is no diversity of source use ($D = 0$). In system (b), the compartments are still specialized, but the use of sources is as diversified as possible ($D = \log 3$). In system (c), the sources are used as evenly as possible ($D = \log 3$) and the one consuming compartment is a generalist, eating from each source equally ($S = \log 3$).

For the Cone Spring example, $S = 0.0419$ and $D = 0.2409$. These values alone do not provide us with much information. S and D are useful for relative comparisons of similar systems. They are not particularly useful as absolute measures of a system state in isolation from other states. When comparing values of S and D it should be remembered that they are logarithmic functions, small changes in value reflect large changes in the distribution of flows in the system.

Two examples explored later in this book, show the utility of S and D. One is a warm core ring system (Ducklow et al. 1989). Table 2.7 shows the values of S and D for spring, summer and fall. The diversity (D) with which the available sources are used remains constant. However, S changes quite dramatically (recalling that S and D are logarithmic). Initially, the average resource niche is somewhat diverse. In summer the resource niche breadth becomes quite broad (generalists) but by fall has become quite narrow (specialized). That D remains fairly constant means that the change in resource niche breadth does not reflect a change by consumers to a single source.

Table 2.7. Temporal variation of D and S.

	MAY	AUG	OCT
D	1.2	1.2	1.2
S	0.75	1.0	0.5
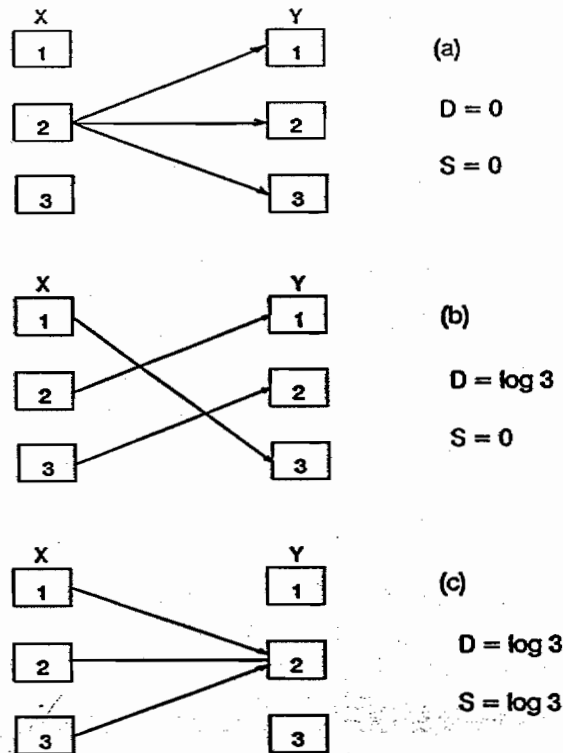
Fig 2.9. Three hypothetical examples of flow organization to illustate the utility of S and D.

An example from an upwelling (Field *et al.* 1989) is shown in Figure 2.10. In this example, S and D increase together during the initial stages of succession after the bloom peaks. This reflects the increased availability of resources. As the system moves from the beginning of the final stage of succession to its culmination, S falls off rapidly as the food web is dominated by specialists while D increases slightly. As the system dies out S continues to fall off sharply while D declines gradually. Further examples and discussion of the use of S and D can be found in Kay (1984).
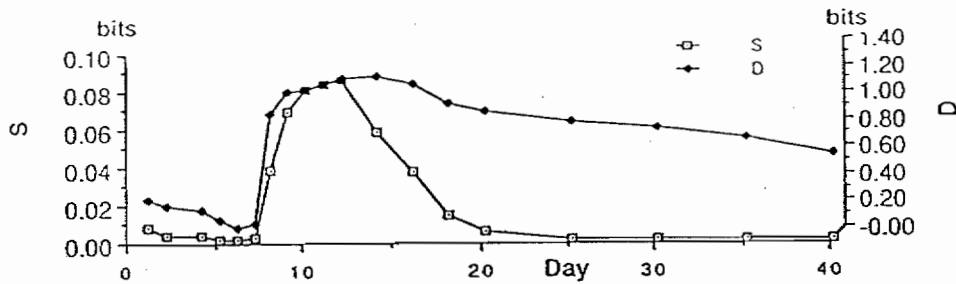
Fig 2.10. The change in S and D with time of a plankton system after an upwelling event.

As an ecosystem develops (i.e., during succession) it has been argued that species will become specialists. Specialization is linked to a more efficient use of the available resources. It would be expected that S would decrease, i.e., the resource niche would narrow with ecosystem development. It would be expected that, on the average, the sources in an ecosystem would be used more evenly, since this would normally result in a more total use of the available resources. Hence, it would be expected that D would increase with development. The upwelling system (Field *et al.* 1989) is an example of this behaviour.

The problem of aggregation in ecosystem modelling has been troublesome for a long time. There is always a trade-off in the aggregation of species into compartments. The more aggregated the system is, the less it tells you about the real system and the smaller the amount of empirical information required. In some situations, it is desirable to compare flow networks of different ecosystems. This is only possible if the ecosystems are compartmentalized in the same way. One issue is how to aggregate compartments without losing the essence of the flow network. It was suggested that S and D could be used in this context as optimizing functions. One would aggregate compartments in a way which minimized the changes in S and D. In this way the degree of connectivity within the flow network would be preserved.

Ascendency and Related Measures

The foregoing indices S and D are intended by their originator (Kay 1984) to quantify the topological attributes of ecosystem networks. They represent a continuation of the efforts of Rutledge *et al.* (1976) to use information theory to measure attributes of network structure. Ulanowicz (1980) had also

recognized the strength of the Rutledge approach for attaching a number to intensive attributes of ecosystems networks. Ulanowicz argued that ecosystem networks also have extensive properties, and that a full quantitative description of the successional process must include both extensive and intensive factors.

Accordingly, Ulanowicz took the average mutual information of a flow network and scaled this property by the popular measure of system activity - the total system throughflow T. The resultant product was called the network ascendency and defined as:

$$A = T \sum_{i=1}^{n} \sum_{j=1}^{n} (f_{ij}/T) \log (f_{ij} T/T_i T_j) . \tag{25}$$

The reader will notice that A is the product of size (T with the dimensions of flow of medium) times "organization" (the summed terms with the dimensions of information, e.g., bits).

That the average mutual information is related to organization follows from the notion that within an organized network there exists less ambiguity about the pathways over which medium flows. That is, an organized network is well articulated. An example of increasing network articulation is given in Figure 2.11. All three hypothetical configurations have identical total system throughputs of 96 units. Figure 2.11a is wholly unarticulated. At any of the four nodes there is maximal uncertainty about where a quantum will next flow. In Figure 2.11b there is only half as much uncertainty, for only a single choice between two recipients is necessary. Configuration b is better articulated than a. Finally, Figure 2.11c is maximally articulated. If a quantum is at any node, there is no uncertainty about where it will flow next.

Ascendency was tailored to encapsulate numerous phenomenological observations into a single quantitative statement. Odum (1969) had listed 24 separate attributes of more "mature" ecosystems. These criteria can be grouped into four generic categories: (a) more speciation, (b) finer specialization, (c) longer retention and (d) more cycling. All other things being equal, an improvement in any one of these generic properties of a network serves to increase its ascendency.

Ascendency can become a quantitative tool in diagnosing ecosystem change at the level of the whole system. Maturing ecosystems should, on the average, increase in ascendency. Early in system development the rise in ascendency will be due mostly to an increase in aggregate activity, T, while later on the continuing rise will mostly be caused by the finer articulation resulting from competition between parallel pathways. A decrease in ascendency is almost always indicative of a stress on the system. A small class of perturbations, however, can actually serve to increase the overall ascendency.

A sudden increase in exogenous inputs can actually increase the ascendency by disproportionately raising the total throughflow T, while at the same time decreasing the mutual information factor. Such a set of circumstances is usually called eutrophication and leaves its characteristic signature on the rise in ascendency (Ulanowicz 1986b).
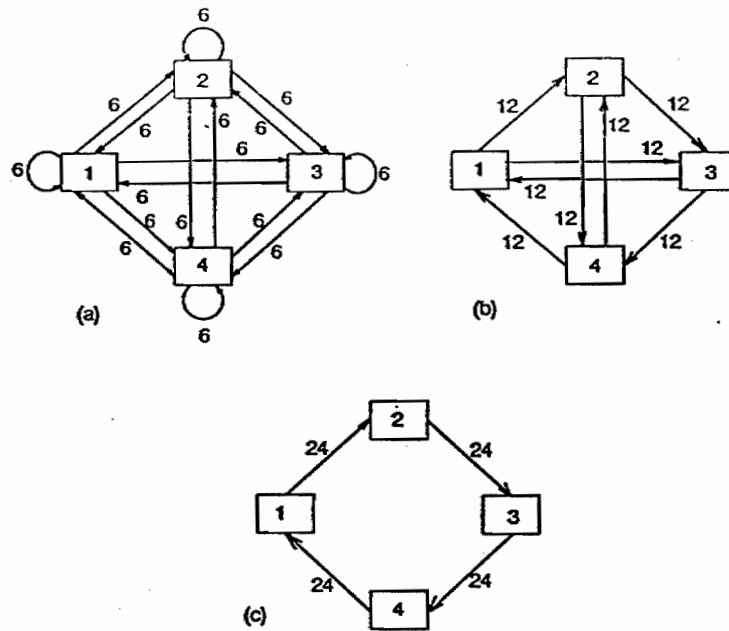

Fig 2.11. Three artificial, closed networks having the same total system throughputs but differing in their degrees of articulation. (a) The maximally connected but minimally articulated configuration. (b) An intermediate level of articulation. (c) The maximally articulated configuration.

There must be limits on the increasing ascendency, as no system can grow without bounds. To better narrate these limits in quantitative terms, one may define five other scaled information indices for any arbitrary network. One upper limit to the mutual information of the network is the entropy of flows defined earlier. One may scale this entropy by T to define a quantity known as the Development Capacity,

$$C = -T \sum_{i=1}^{n} (T_i/T) \log(T_i/T) . \tag{26}$$

In Figure 2.11c the maximally articulated network became identically equal to its upper limit, C. That was a hypothetical and unreal situation; however, and in any actual network there will always remain a positive difference $C - A$, which is referred to as the system's overhead.

Part of the overhead is generated by the exogenous transfers. It is possible to identify three separate components attributable to the inputs, exports and dissipations, respectively. A fourth component represents the residual uncertainty associated with the presence of multiple or parallel pathways among the network components, and it is termed the redundancy.

If one excludes from the capacity and the ascendency those terms generated by the exogenous transfers, one is left with what may be termed the internal capacity and the internal overhead. These measures may be of interest if one is focusing on endogenous system behavior. For algebraic reasons not important here, the internal overhead (the difference between the internal capacity and the internal ascendency) contains terms generated by the exports and respirations. These components are called the tribute and dissipation, respectively. They, along with the aforementioned redundancy, comprise the internal overhead.

The value of the development capacity of the Cone Spring network, along with its four components is:

Development Capacity	=	93,172	kcal-bits m^2/yr
Ascendency	=	56,725	"
Overhead on Inputs	=	2,652	"
Overhead on Exports	=	1,920	"
Overhead on Respiration	=	21,364	"
Redundancy	=	10,511	"

The reader will notice that the ascendency and the four overhead terms all sum to give the capacity. The corresponding decomposition of the internal capacity looks like:

Internal Capacity	=	71,372	kcal-bits m^2/yr
Internal Ascendency	=	29,332	"
Tribute	=	2,971	"
Dissipation	=	28,558	"
Redundancy	=	10,510	"

A more complete exposition of the meaning of the individual overhead terms and the algebraic details involved in the decomposition of the capacity are all given in Ulanowicz (1986b).

Ascendency, S and D: The Differences

Ascendency (A), S, and D are all measures of organization. A is meant as a measure of total ecosystem development. S and D are meant as measures of structural organization. S measures the average (over all compartments) diversity of source use by each compartment and D measures the diversity of outflow use over the whole ecosystem. A is meant as an ecosystem level measure (as is D), and not as a compartment level measure (which S is). This is the reason Kay uses two measures while Ulanowicz uses one.

A, S, and D are related mathematically as follows. A is the product of the total throughput (T), and the mutual information (I). Looking at equation 4 in the appendix , $I = D - S$. Thus,

$$A = T\ (D\text{-}S\ ). \qquad (27)$$

It should be noted that S and D are logarithmic measures. A is the product of a linear function and a log function. Thus, A will be more sensitive to system changes which affect the value of T relative to those changes which affect S and D. It is possible that significant changes in S and D will not be reflected by significant changes in A. However, changes in total throughput (T) are not measured by S and D as they are with A.

A, S, and D are meant to be used in different situations and are therefore different even though they are mathematically related. See Mann et al. (1989) for arguments as to when to use which of these measures of ecosystem development.

CONCLUDING COMMENTS

In this chapter an overview of commonly used ecosystem network measures was provided. These measures are all static and provide a snapshot of the state of the ecosystem flow network. The measures tell us about the compartments, their relationships and overall system status. All of the measures describe only the flow in an ecosystem and not the driving forces behind the flows. The picture of how an ecosystem food web functions will always be incomplete until both the flows and driving forces are known and network thermodynamic models of the ecosystems can be built.

The realities of available data limit the utility of all the measures presented. The authors of this chapter found that the limit on the utility of these measures is the lack of data of the detail (level of aggregation, temporal and spatial refinement) necessary to make the measures useful for real world applications. It is our hope that by making the measures more accessible to field ecologists, through computer software like those associated with this book, that further collection of the data needed for the application of the measures will be stimulated. Only if this happens will theoreticians be able to refine the measures so as to give environmental managers the tools they need to monitor the health of our ecosystems.

REFERENCES

Aczel, J., & Z. Daroczy 1975. *On Measures of Information and Their Characterizations*, Academic Press, New York.

Augustinovics, M. 1981. Methods of international and intertemporal comparison of structur,. in A. P. Carter and A. Brody (Ed.) *Contributions to Input-Output Analysis*, Vol. 1, North Holland, Amsterdam.

Baird, D. & R. E. Ulanowicz. 1989. The seasonal dynamics of the Chesapeake Bay ecosystem, Ecol. Monogr (in press)

Barber, M. C. 1978a. A retrospective Markovian model of ecosystem resource flow, Ecol. Mod. 5: 125-135.

Barber, M. C. 1978b. A Markovian model for ecosystem flow analysis, Ecol. Mod. 5:193-206.

Barber, M. C. 1979. A note concerning time parameterization of Markovian models of ecosystem flow analysis, Ecol. Mod. 6: 323-328.

Bosserman, R. W. 1981. Sensitivity techniques for examination of input-output flow analyses, in W.J. Mitsch & J.M. Klopatek (eds), Energy and Ecological Modeling, Elsevier Publishers, Amsterdam.

Costanza, R. & B. Hannon 1989. Dealing with the "mixed units" problem in ecosystem network analysis. In: Wulff, F., J.G. Field & K.H. Mann (eds), *Network Analysis in Marine Ecology: Methods and Applications*, Lecture Notes on Coastal and Estuarine Studies, Springer-Verlag, New York.

Cousins, S. 1985. Ecologists build pyramids again, New Scientist 107: 50-54.

Ducklow, H. W, M. J. R. Fasham & A. F. Vézina 1989. Derivation and analysis of flow networks for open oceanic plankton systems, in: Wulff, F., J. G. Field & K. H. Mann (eds), *Network Analysis in Marine Ecology: Methods and Applications*, Lecture Notes on Coastal and Estuarine Studies, Springer-Verlag, New York.

Field, J. G., C. L. Moloney & C. G. Attwood 1989. Network analysis of simulated succession after an upwelling event, in: Wulff, F., J. G. Field & K. H. Mann (eds), *Network Analysis in Marine Ecology: Methods and Applications*, Lecture Notes on Coastal and Estuarine Studies, Springer-Verlag, New York.

Finn, J. T. 1976. Measures of ecosystem structure and function derived from analysis of flows, J. Theor. Biol. 56: 363-380.

Finn, J. T. 1977. Flow analysis: A method for tracing flows through ecosystem models., PhD Thesis, Department of Zoology, University of Georgia, Athens, GA.

Hannon, B. 1973. The structure of ecosystems, J. Theor. Biol. 41: 535-546.

Horowitz E., & S. Sahni 1984. *Fundamentals of Data Structures in Pascal*, Computer Science Press.

Kay, J. J. 1984. *Self-organization in living systems*, Ph.D.Thesis, Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada.

Kemeny, J. G. & J. L. Snell 1960. *Finite Markov Chains*, D. Von Nostrand Co, Princeton, NJ.

Legendre, L & P. Legendre 1983. *Numerical Ecology*, Elsevier, Amsterdam, 419 pp.

Leontief, W.W. 1936. Quantitative input-output relations in the economic system of the United States, Rev. Econ. Stat. 18: 105-125.

Leontief, W.W. 1963. *The structure of the American Economy, 1919-1939*, 2nd edition. Oxford University Press, New York.

Lindeman, R. L. 1942. The trophic-dynamic aspect of ecology, Ecology 23: 399-418.

Mann, K. H., J. G. Field & F. Wulff 1989. Network analysis in marine ecology: an assessment, in: Wulff, F., J. G. Field & K. H. Mann (eds), *Network Analysis in Marine Ecology: Methods and Applications*. Lecture Notes on Coastal and Estuarine Studies, Springer-Verlag, New York.

Matis, J. H. and B. C. Patten 1981. Environ analysis of linear compartmental systems: the static, time invariant case, Bull. Int. Stat. Inst. 48: 527-565.

Odum, E. P. 1969. The strategy of ecosystem development, Science 164: 262-270.

Patten, B. C., R. W. Bosserman, J. T. Finn, & W. G. Cale 1976. Propagation of cause in ecosystems, pp. 457-579, in: B. C. Patten (ed), *Systems Analysis and Simulation in Ecology, Vol 4*, Academic Press, New York.

Patten, B. C. & J. H. Matis. 1982. The water environs of the Okefenokee Swamp: an application of static linear environ analysis, Ecol. Mod. 16: 1-50.

Patten, B. C. 1985. Energy cycling, length of food chains, and direct versus indirect effects in ecosystems, pp. 119-138, in: R. E. Ulanowicz and T. Platt (eds) *Ecosystem Theory for Biological Oceanography*, Fisheries & Oceans Canada .

Pimm, S. L. 1982. *Food Webs*, Chapman and Hall, London.

Rutledge, R.W., B.L. Basorre, & R. J. Mulholland 1976. Ecological stability: An information theory viewpoint, J. Theor. Biol. 57:355-571.

Schneider, E. D. 1988. Entropy, ecosystems and evolution, To appear in: B.Weber and D. Depew (eds): *Entropy, Information and Evolution*, MIT Press, Cambridge, MA.

Shannon, C. 1948. The mathematical theory of communication, Bell System Technical Journal, Reprinted 1949: *The Mathematical Theory of Communication*, University of Illinois Press, Urbana, Illinois.

Szyrmer, J. & R. E. Ulanowicz. 1986. Total flow in ecosystems, Ecol. Mod 35:123-136.

Tilly, L. J. 1968. The structure and dynamics of Cone Springs, Ecol. Monogr. 38:169-197

Ulanowicz, R. E. 1983. Identifying the structure of cycling in ecosystems, Math.Biosci. 65: 219-237.

Ulanowicz, R. E. 1984. Community measures of marine food networks and their possible applications, pp 23-47, in: Fasham, M. J. R. (ed.): *Flows of energy and materials in marine ecosystems*, Plenum Publishing Corp, New York.

Ulanowicz, R. E. 1986b. A phenomenological perspective of ecological development, pp 73-81, in: T. M. Poston & R.Purdy (eds.) *Aquatic Toxicology and Environmental Fate, Vol 9*. ASTM 921, Am. Soc. for Testing and Materials, Philadelphia, PA.

Ulanowicz, R.E. 1986a. *Growth and Development: Ecosystems Phenomenology*, Springer-Verlag, New York.

Ulanowicz, R. E. 1989. Ecosystem trophic foundations: Lindeman exonerata, To appear in: B. C. Patten & S.E. Jorgensen (eds.), *Progress in Systems Ecology*, Prentice-Hall, New York.

Ulanowicz, R. E. & W. M. Kemp 1979. Towards canonical trophic aggregations. Amer. Nat. 114: 871-883.

Ulanowicz, R. E., & J. J. Kay 1986. NETWRK 3 : A package of computer algorithms to analyze ecological flow networks, UMCEES Ref No. 82-7C CBL, University of Maryland, Solomons, MD.

Warwick, R. M. & P. J. Radford 1989. Analysis of the flow networks in an estuarine benthic community, in: Wulff, F., J. G. Field & K. H. Mann (eds), *Network Analysis in Marine Ecology: Methods and Applications*, Lecture Notes on Coastal and Estuarine Studies, Springer-Verlag, New York.

Wulff, F., J. G. Field & K. H. Mann 1989 (eds). *Network Analysis in Marine Ecology: Methods and applications*, Lecture Notes in Coastal and Estuarine Studies, Springer-Verlag, New York.

Yaglom, A. M., & I. M. Yaglom 1960. *Probability and Information* (Russian). Fitzmatgiz, Moscow.

APPENDIX 2A

Review of Information Theory

Consider an experiment A, [1] which consists of observing which of N possible outcomes occurs. Suppose that the probability of the ith outcome being observed is $P_i[A]$. Then the average uncertainty an observer has about which event (i.e., which outcome of the experiment) will occur is given by:

$$H[A] = \sum_{i=1}^{n} -P_i[A] \log P_i[A]; \quad 0 \le H[A] \le \log N \quad \quad (1)$$

$H[A]$ also measures the information gained, on average, from observing which event occurs. It increases as the number of possible events (N) increases. It also increases as the likelihoods of the occurrences of the events become more similar, that is, as the differences between the $P_i[A]$ decrease. Maximum uncertainty occurs when each event appears equally likely ($P_i[A] = 1/N$). In this case $H[A]$ = log N. The uncertainty ($H[A]$) decreases as one or a few events become more likely (larger $P_i[A]$) relative to the other events. The function $H[A]$ is known as the **Shannon Entropy**. It has certain well known properties such as Symmetry, Normality, Expansibility, Decisivity, Strong Additivity, Recursivity, Maximality, and Sub-Additivity. (See Aczel & Daroczy, 1975.)

Now consider a second experiment B which consists of observing which of M possible outcomes occurs. The probability of the jth outcome is $P_j[B]$. The possibility of experiment A influencing experiment B, or B influencing A, can be explored using the conditional probabilities $P_i[A/B_j]$, $P_j[B/A_i]$ [2]. These indicate the probability of an outcome of one experiment occurring given that the outcome of the other experiment is known. Let us assume that the probability distributions are complete. Then the observer's average uncertainty about experiment B, given that he already knows the outcome of A is:

$$H[B/A] = \sum_{i=1}^{n} -P_i[A] \sum_{j=1}^{m} P_j[B/A_i] \log P_j[B/A_i]; 0 \le H[B/A] \le H[B] \quad \quad (2)$$

This is the **Conditional Entropy**. It also measures the average information the observer gains from observing event B given he already knows which event A occurred. The inequality states that the observer's maximum conditional uncertainty about B is the same as the uncertainty (Shannon entropy) he would have if he had no a priori information. The minimum conditional uncertainty ($H[B/A] = 0$) occurs when the outcome of A completely predicts the outcome of B.

The composite event A + B has a complete probability distribution associated with it, such that the probability of ai and bj being the outcome of the two experiments is $P_{ij}[A+B] = P_j[B/A_i]P_i[A] = P_i[A/B_j]P_j[B]$. The observer's average Uncertainty About The Composite Event is given by (1).

$$H[A,B] = \sum_{i=1}^{n} \sum_{j=1}^{m} -P_i[A] \, P_j[B/A_i] \log(P_i[A] \, P_j[B/A_i]).$$

Using (1) and (2) it can be shown that:

$$H[A,B] = H[A] + H[B/A] \; ; \; 0 \leq H[A,B] \leq H[A] + H[B] \qquad (3)$$

The inequality states that the maximum average uncertainty about the composite event occurs when the outcomes of experiment A and B are independent. Note that Bayes Theorem implies that H[A,B] is symmetric, that is $H[A,B] = H[B,A]$.

The **Mutual Information** is defined as:

$$I[A,B] = H[A] + H[B] - H[A,B]$$

This is symmetric. Using equation (3) and inequality (2) leads to:

$$I[A,B] = H[B] - H[B/A]; \; 0 \leq I[A,B] \leq H[B] \qquad (4)$$

I[A,B] measures the reduction in the observer's average uncertainty about event B due to his knowing which event A occurred (or vice versa). In other words it measures the average information gained about the outcome of experiment B from performing experiment A. It is a maximum when H[B/A] = 0, that is the outcome of A completely predicts the outcome of B. It is a minimum when H[B/A] = H[B], the outcome of A tells the observer nothing about the outcome of B. It is not an entropy measure in the strict mathematical sense. It increases as the average conditional uncertainty decreases.

The final measure to be discussed is called the redundancy. The redundancy (R) is defined as:

$$R = \frac{F((\max I[A,B] - I[A,B])}{(\max I[A,B]))} \qquad R = 0 \leq R \leq 1 \qquad (5)$$

(maxI[A,B] = the maximum value I[A,B] can take on for the given experiments A and B.)

The redundancy is a measure of how useful it is to perform experiment A if the experimenter's objective is to decrease his uncertainty about experiment B. Performing experiment A is said to be redundant

if it does not provide any additional information about experiment B. In this case the redundancy is a maximum and $I[A,B] = 0$. In the other extreme the redundancy is a minimum when performing experiment A gives maximum information, that is tells the experimenter the outcome, a priori, of experiment B. Performing experiment A is not redundant relative to determining the outcome of B and $I[A,B] = maxI[A,B] = H[B]$.

Some confusion has occurred over this measure. This is due to incorrectly interpreting R as a measure of the redundancy of performing experiment B after having performed experiment A. The reason for this interpretation, even though it is incorrect mathematically, is that it seems reasonable given the English language use of the word redundancy. However, the name "redundancy" for R makes English language sense in the context of network structures. For further reading Yaglom & Yaglom (1960) is recommended.

Endnote

[1] Experiment and event are used interchangeably.

[2] $P_i[A/B_j]$ = probability that outcome A_i occurs given that B_j has occurred.