J. theor. Biol. (1985) **116**, 321–341

Information Theoretical Analysis of the Aggregation and Hierarchical Structure of Ecological Networks†

HIRONORI HIRATA‡ AND ROBERT E. ULANOWICZ

University of Maryland, Center for Environmental and Estuarine Studies, Chesapeake Biological Laboratory, Solomons, Maryland 20688-0038, *U.S.A.*

(*Received* 19 *December* 1983, *and in revised form* 15 *May* 1985)

This paper considers the aggregation and hierarchical structure of ecological networks from the viewpoint of information theory. It is pointed out that the aggregation which maximizes the amount of information preserved during the process of aggregation bears heavily upon the concept of trophic compartments. Hierarchical structure is defined as a series of aggregation processes. How the organization at one level is related to that at the next level is elucidated by two propositions and attendant corollaries.

1. Introduction

"A model which must be capable of accounting for all the input–output behavior of a real system and be valid in all allowable experimental frames can never be fully known" (Zeigler, 1976). The starting point for considering any "system", be it flows in an ecosystem, an economic network, or a transportation grid, is perforce arbitrary to some degree; and any analysis which lumps smaller elements into a few system compartments must reckon with the errors introduced by the aggregation process.

Most previous analyses of aggregation error center around the effects of consolidation upon the *assumed* dynamics of the system (Cale & Odell, 1979, 1980; O'Neill & Gardner, 1979; O'Neill & Rust, 1979; Cale, O'Neill & Gardner, 1983). Initial compartmentalization, however, is done on the basis of observation and intuition. The question arises, therefore, as to whether it is possible to describe a systematic and rational scheme for combining compartments based solely on the *observed* transitions occurring in the system.

† Contribution No. 1635 of the Center for Environmental and Estuarine Studies, University of Maryland.

‡ Permanent address: Department of Electronics, Faculty of Engineering, Chiba University, 1-33, Yayoi-cho, Chiba-shi 260, Japan. To where correspondence should be addressed.

0022–5193/85/190321+21 $03.00/0

Observation of a system is intended to produce information about the ensemble. The information value of quantitative observations (measurements) is properly the domain of information theory. For this reason ecosystems researchers have lately been paying attention to the applications of information theory to ecosystems networks (Rutledge, Basore & Mulholland, 1976; Ulanowicz, 1979). In particular, it has been suggested (Ulanowicz, 1980, 1981) that the average mutual information inherent in the structure of an ecological flow network is a pertinent indicator of the degree of organization (development) exhibited by the given community, and Hirata & Ulanowicz (1984) formally derived such an index from the body of information theory.

Common sense dictates that one wishes to avoid, whenever possible, any combination of compartments which would serve to obscure important facets of organization in a system. This translates directly into a preference for those aggregation schemes which minimize the decrease in mutual information of the associated network. In particular, it will be shown by way of example how grouping compartments so as to minimize the loss of mutual information creates collections of species that coincide with trophic compartments defined *a priori* on intuition.

As a prelude to the search for such desirable combinations of compartments, the process of aggregation is first defined, and how the information in the network varies upon aggregation is then discussed. Thereafter, attention is turned to information in hierarchical systems. The results from the study of aggregation are used to define the notion of total information in a hierarchically-nested network.

2. Aggregation

Definition 1

N and $\bar{N}$ are the sets of elements

$$N = \{\alpha_k\}_{k=1,\dots,n} \tag{1}$$

$$\bar{N} = \{\beta_i\}_{i=1,\dots,m}, \qquad (m \leq n). \tag{2}$$

If a homomorphic mapping ϕ is made from N to $\bar{N}$ as

$$\phi: N \to \bar{N}; \tag{3}$$

then N is called the original network; $\bar{N}$, the aggregated network of N; and ϕ, the aggregation mapping.

A matrix is a convenient representation of an aggregation mapping as follows:

Definition 2

Define an aggregation matrix **S** as

$$\mathbf{S} = [s_{ik}]_{i=1,\ldots,m,k=1,\ldots,n} = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & & & \\ \vdots & & & \vdots \\ s_{m1} & & \cdots & s_m \end{bmatrix} \tag{4}$$

where

$$0 \leq s_{ik} \leq 1 \quad \text{and} \quad \sum_{i=1}^{m} s_{ik} = 1 \tag{5}$$

If all s_{ik} are either 0 or 1, the mapping is referred to below as a "discrete aggregation" (Hirata, 1978). Otherwise, it is referred as a "weighted aggregation". One may consider the discrete aggregation as a special case of the weighted aggregation (e.g., Ulanowicz & Kemp, 1979).

Those positions of s_{ik} which are not zero signify which elements should aggregated into the same group, *i*. Because the aggregated network ($\bar{N}$) of the original network (N) depends on an aggregation matrix (**S**), it may be expressed as the function of **S**, $\bar{N}(\mathbf{S})$.

Example 1

Consider an aggregation between N and $\bar{N}$ as follows:

$$\begin{aligned} N &= \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6\} \\ \bar{N} &= \{\beta_1, \beta_2, \beta_3\}. \end{aligned} \tag{6}$$

A 6×3 matrix **S** describes the aggregation of N into $\bar{N}$.

Figure 1(a) is the case of weighted aggregation shown by an aggregation matrix as

$$\mathbf{S} = \begin{bmatrix} 1 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ 0 & 0 & \frac{2}{3} & \frac{1}{3} & 1 & 1 \end{bmatrix}. \tag{7}$$

In this case

$$\begin{aligned} \beta_1 &= \alpha_1 + \tfrac{1}{2}\alpha_2 \\ \beta_2 &= \tfrac{1}{2}\alpha_2 + \tfrac{1}{3}\alpha_3 + \tfrac{2}{3}\alpha_4 \\ \beta_3 &= \tfrac{2}{3}\alpha_3 + \tfrac{1}{3}\alpha_4 + \alpha_5 + \alpha_6. \end{aligned} \tag{8}$$

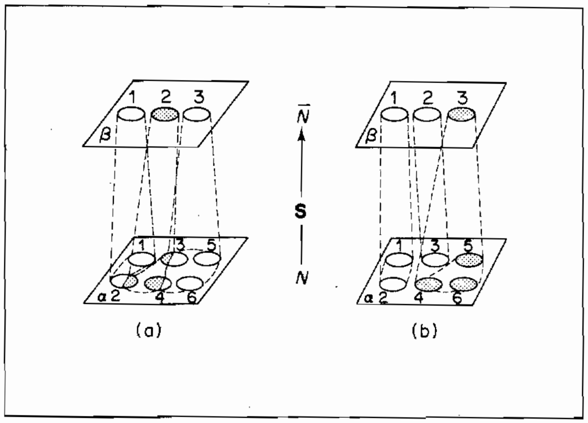

FIG. 1. Schematized aggregation. (a) weighted aggregation, (b) discrete aggregation.

Figure 1(b) is the case of discrete aggregation shown by an aggregation matrix as

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \end{bmatrix}. \tag{9}$$

In this case $\beta_1 = \alpha_1 + \alpha_2$, $\beta_2 = \alpha_3$

$$\beta_3 = \alpha_4 + \alpha_5 + \alpha_6. \tag{10}$$

3. Networks

Consider networks consisting of n compartments each of which is characterized by its throughput of a single given medium such as carbon, nitrogen, mass, energy, etc. The kth compartment is characterized as shown in Fig. 2. The symbols for the kth compartment are as follows:

- T_{kj}: the flow leaving the kth compartment and directly contributing to the jth compartment; $T_{kj} \geq 0$.
- D_k: the dissipated flow leaving the kth compartment; $D_k \geq 0$.
- E_k: the useful export leaving the kth compartment; $E_k \geq 0$.
- I_k: the input flow to the kth compartment from the outside world; $I_k \geq 0$.

The throughput of the kth compartment, T_k, is defined as

$$T_k = \sum_{j=1}^{n} T_{kj} + D_k + E_k. \tag{11}$$

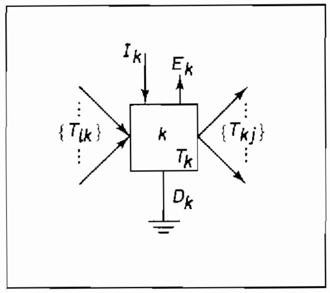

FIG. 2. The typical kth compartment of an ecological network.

For convenience one uses the vector and matrix forms of these variables as follows:

$$\begin{aligned}\mathbf{T} &= (T_k)'_{k=1,\ldots,n} = (T_1, \ldots, T_n)' \\ \mathbf{D} &= (D_k)'_{k=1,\ldots,n} \\ \mathbf{E} &= (E_k)'_{k=1,\ldots,n} \\ \mathbf{I} &= (I_k)'_{k=1,\ldots,n} \\ T^* &= [T_{kj}]_{k,j=1,\ldots,n}\end{aligned} \tag{12}$$

where the prime (′) represents the vector transpose. Equation (11) may then be expressed as

$$\mathbf{T} = T^*\mathbf{1} + \mathbf{D} + \mathbf{E} \tag{13}$$

where $\mathbf{1} = (1, \ldots, 1)'$. In a system at steady state T_k also equals $\sum_{l=1}^{n} T_{lk} + I_k$, i.e.

$$\mathbf{T} = T^*\mathbf{1} + \mathbf{D} + \mathbf{E} = T^*\mathbf{1} + \mathbf{I}. \tag{14}$$

Consider the problem of making an aggregated network, $\bar{N}(S)$, from the network, N. The bar over the symbols, e.g. $\bar{T}_i$, $\bar{T}_{il}$, $\bar{D}_i$, $\bar{E}_i$ and $\bar{I}_i$, means that they pertain to the aggregated network. One may then set

$$\begin{aligned}\bar{\mathbf{T}} &= (\bar{T}_i)'_{i=1,\ldots,m} \\ \bar{\mathbf{D}} &= (\bar{D}_i)'_{i=1,\ldots,m} \\ \bar{\mathbf{E}} &= (\bar{E}_i)'_{i=1,\ldots,m} \\ \bar{\mathbf{I}} &= (\bar{I}_i)'_{i=1,\ldots,m} \\ \bar{T} &= [T_{il}]_{i,l=1,\ldots,m}.\end{aligned} \tag{15}$$

The relation

$$\bar{\mathbf{T}} = \bar{T}^{*}\mathbf{1} + \bar{\mathbf{D}} + \bar{\mathbf{E}} \tag{16}$$

is preserved in analogy with (13).

Relations between the original network, N, and the aggregated network, $\bar{N}(S)$, are derived by using aggregation matrix $\mathbf{S}$ as shown in Proposition 1.

Proposition 1

The relations between the variables of the original network, N, and those of the aggregated network, $\bar{N}(S)$, are as follows:

$$\begin{aligned} \bar{\mathbf{T}} &= S\mathbf{T} \\ \bar{\mathbf{D}} &= S\mathbf{D} \\ \bar{\mathbf{E}} &= S\mathbf{E} \\ \bar{\mathbf{I}} &= S\mathbf{I} \\ \bar{T}^{*} &= ST^{*}S'. \end{aligned} \tag{17}$$

The model for choice within open systems is studied as shown in Fig. 3, where exogenous inputs come from the 0th compartment; exports enter the

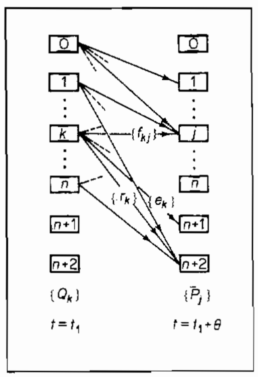

FIG. 3. A generic model for choice in open systems.

$(n+1)$st compartment; and dissipation is collected in the $(n+2)$nd compartment. Although one could consolidate input, export and dissipation into one global compartment, they are treated separately to facilitate extension of the theory.

Now time is introduced into the analysis. The time interval for flow from one compartment to another is taken to be θ. New variables are defined as follows:

Q_k: the percentage of the total flow through the ecological network at time t_1 which passes through the kth compartment; $Q_k \geq 0$ $(k = 0, \ldots, n)$, $Q_{n+1} = Q_{n+2} = 0$.

P_j: the percentage of the total flow through the ecological network at time $t_1 + \theta$ which passes through the jth compartment; $P_0 = 0$, $P_j \geq 0$ $(j = 1, \ldots, n+2)$.

f_{kj}: the percentage of the total flow through the kth compartment at time t_1 that passes into the jth compartment between time t_1 and $t_1 + \theta$; $f_{kj} \geq 0$.

r_k: the percentage of the flow through the kth compartment which is dissipated; $r_k \geq 0$.

e_k: the percentage of the flow through the kth compartment which is exported as useful flow; $e_k \geq 0$.

The relations between these variable are provided by the equations

$$
\begin{aligned}
P_j &= \sum_{k=0}^{n} f_{kj} Q_k \qquad (j = 1, \ldots, n) \\
P_{n+1} &= \sum_{k=1}^{n} e_k Q_k \\
P_{n+2} &= \sum_{k=1}^{n} r_k Q_k
\end{aligned}
\tag{18}
$$

where

$$
\begin{aligned}
\sum_{j=1}^{n} f_{kj} + r_k + e_k &= 1 \qquad (k = 1, \ldots, n) \\
\sum_{j=1}^{n} f_{0j} &= 1
\end{aligned}
\tag{19}
$$

At steady state the following additional relations hold

$$
\begin{aligned}
Q_k &= P_k \qquad (k = 1, \ldots, n) \\
Q_0 &= P_{n+1} + P_{n+2}.
\end{aligned}
\tag{20}
$$

In network N one may identify variables Q_k, f_{kj}, r_k and e_k as follows

$$
\begin{aligned}
Q_k &= T_k/(T+I) \qquad (k = 1, \ldots, n) \\
Q_0 &= I/(T+I)
\end{aligned}
\tag{21}
$$

$$
\begin{aligned}
f_{kj} &= T_{kj}/T_k \qquad (k, j = 1, \ldots, n) \\
f_{0j} &= I_j/I \qquad (j = 1, \ldots, n) \\
f_{k0} &= 0 \qquad (k = 0, \ldots, n)
\end{aligned}
\tag{22}
$$

$$r_k = D_k/T_k \tag{23}$$

$$e_k = E_k/T_k \tag{24}$$

where

$$T = \sum_{k=1}^{n} T_k \tag{25}$$

and

$$I = \sum_{k=1}^{n} I_k. \tag{26}$$

It is a simple matter to reinterpret the model for choice shown in Fig. 3 as the classical mode of a channel (Hirata & Ulanowicz, 1984). One may consider that $A = \{a_k\}_{k=0,1,\ldots,n,n+1,n+2}$ is the set of input events with probabilities $P(a_k) = Q_k$. $B = \{b_j\}_{j=0,1,\ldots,n,n+1,n+2}$ is the set of output events with probabilities $P(b_j) = P_j$. $F = [f_{kj}]$ is the communication matrix of this channel, i.e. $P(b_j/a_k) = f_{kj}$.

The orthodox way of calculating the mutual information contained in the network is derived in an earlier paper (Hirata & Ulanowicz, 1984). For the convenience of the reader the results are repeated in Proposition 2 below. As has been mentioned by Rutledge *et al.* (1976), mutual information represents the average amount of uncertainty resolved by the knowledge of the network structure.

Proposition 2

The information, $M(N)$, contained in the structure of network N may be considered as the sum of four terms

$$M(N) = \sum_{\chi=\sigma,s,e,r} M_\chi(N) \tag{27}$$

where

$$M_\sigma(N) = \sum_{j=1}^{n} f_{oj} Q_o \log\left[f_{oj} \Big/ \left(\sum_{l=0}^{n} f_{lj} Q_l\right)\right] \geq 0 \tag{28}$$

$$M_s(N) = \sum_{k=1}^{n} \sum_{j=1}^{n} f_{kj} Q_k \log\left[f_{kj} \Big/ \left(\sum_{l=0}^{n} f_{lj} Q_l\right)\right] \geq 0 \tag{29}$$

$$M_e(N) = \sum_{k=1}^{n} e_k Q_k \log\left[e_k \Big/ \left(\sum_{l=1}^{n} e_l Q_l\right)\right] \geq 0 \tag{30}$$

$$M_r(N) = \sum_{k=1}^{n} r_k Q_k \log\left[r_k \Big/ \left(\sum_{l=1}^{n} r_l Q_l\right)\right] \geq 0 \tag{31}$$

Proof of Proposition 2. From the definition of the mutual information of a channel

$$M(N) = M(A; B) = H(A) - H(A/B) = H(B) - H(B/A) \qquad (32)$$

$$= \sum_{A,B} P(a_k, b_j) \log [P(b_j/a_k)/P(b_j)] \qquad (33)$$

Equations (27)–(31) derive from (32) because the probabilities, $P(a_k)$, $P(b_j)$, $P(b_j/a_k)$, etc. are obtained from (21)–(26). □

M_σ captures the amount of information associated with the pattern of inputs; M_s, the amount assigned to the internal network flow structure; M_e, the amount contained in the pattern of useful exports; and M_r, the amount included in the pattern of dissipative flows.

4. Aggregation of Networks

Consider how the process of aggregation (weighted aggregation and discrete aggregation) affects the variation of information. For the aggregated network $\bar{N}(S)$, $\bar{A} = \{\bar{a}_i\}_{i=0,1,\ldots,m,m+1,m+2}$; $P(\bar{a}_i) = \bar{Q}_i$; $\bar{B} = \{\bar{b}_l\}_{l=0,1,\ldots,m,m+1,m+2}$; $P(\bar{b}_l) = \bar{P}_l$ and $\bar{F} = [\bar{f}_{il}]$ are defined in the same ways as for the original network N.

Proposition 3

During the process of network aggregation (both weighted and discrete) the following relation is maintained

$$M(N) \geq M(\bar{N}) \qquad (34)$$

Proof of Proposition 3. Relation (34) as it pertains to discrete aggregation was proved by Theil (1967). One may extend this result to the more general case of weighted aggregation by straightforward modification of his proof. The revisions necessary to extend Theil's theorem appear in Appendix 1.

Proposition 3 shows that information cannot increase (it is generally lost) during the process of aggregation, i.e. the difference $M(N) - M(\bar{N})$ is never negative. This loss of information may be considered a cost, J, of the aggregation

$$J = M(N) - M[\bar{N}(S)] \qquad (35)$$

and this cost, J, must always be compared with the economic cost of not aggregating, i.e. the cost of gathering extra data for or otherwise analyzing the expanded system. Very often these latter costs are fixed by fiscal and

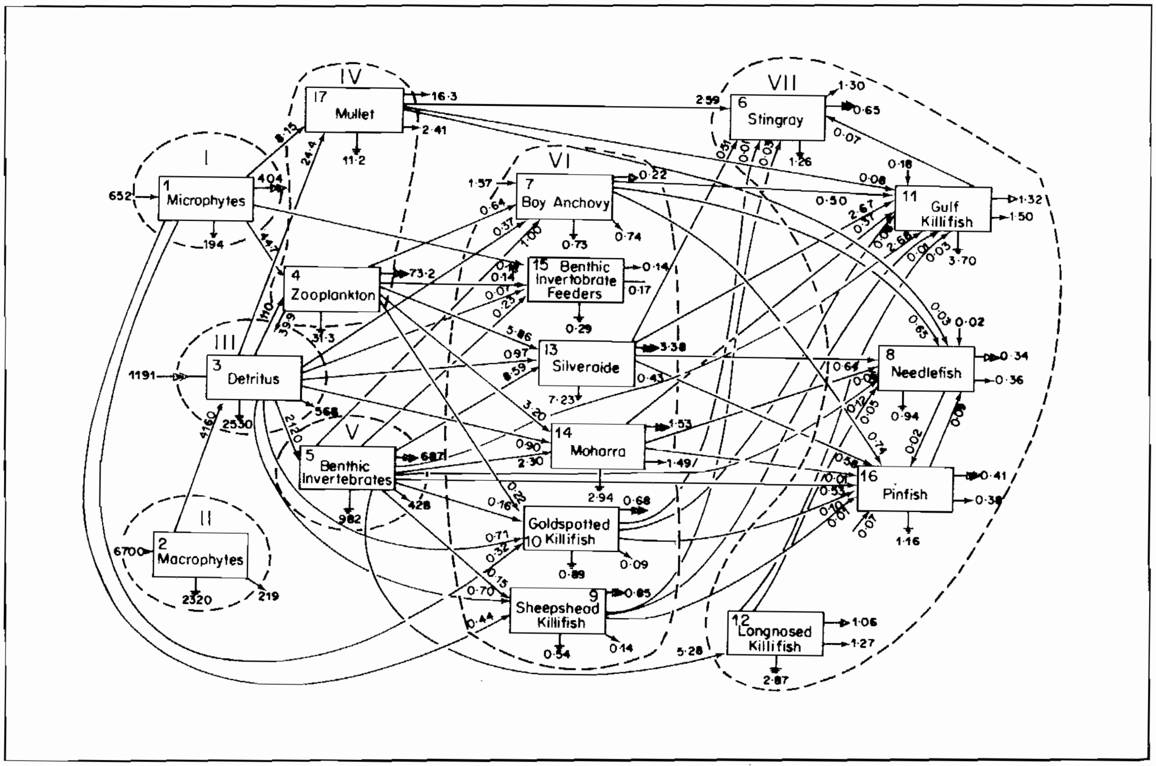

FIG. 4. A schematic of the carbon flows among 17 taxa of a marsh gut ecosystem, Crystal River, Florida (after M. Homer and W. M. Kemp, unpublished). All flow are in mg carbon/m^2 day. Linked arrows (—▷▷) represent returns to detritus, ground symbols ($\perp$) are respiratory losses, simple arrows not terminating in another box depict exports form the system, and simple arrows not originating from another box are exogenous inputs. Dotted lines indicate optimal aggregation into seven compartments. The balance is within three figure accuracy.

manpower constraints, so that the problem usually boils down to the dimension of the aggregated system, say m, being determined by the available resources (e.g. manpower, funds, etc.) and the aggregation down to m compartments being chosen so as to minimize the loss of information, i.e. J is minimized by choosing the optimal S^* such that

$$J^* = \min_{\{S\}} J = M(N) - M[\bar{N}(S^*)]. \tag{36}$$

An example of discrete aggregation now follows.

Example 2

The flows among 17 compartments of a tidal marsh stream ecosystem (in milligrams carbon/m^2-d) were measured by Homer & Kemp (unpublished ms) and are shown schematically in Fig. 4. This system of flows was subjected to an impartial stepwise aggregation of compartments to reduce the number components one at a time until the desired level (7) was reached. (There were too many combinations of 17 things taken 7 at a time to test all possible simultaneous contractions of the number of boxes. Thus, the optimal aggregation was estimated in a stepwise fashion condensing only two boxes at each step until the total was reduced to 7. Because all possible aggregation pairs were considered at each step, the search algorithm was impartial.)

The stepwise aggregations yielding the least decrease in network information are indicated by the dotted perimeters in Fig. 4, and the final condensed network appears in Fig. 5. The resulting groupings make good intuitive sense. In fact the underlying diagram for Fig. 4 was prepared for other purposes several months before the stepwise aggregation algorithm was even written. At that time the boxes in the diagram were juxtaposed so as to intuitively group boxes with similar apparent functions in proximity to one another. The fact that a blind search confirmed such *a priori* guesses should not be regarded as either suspicious or mysterious. Rather, the results serve to support the suppositions that mathematical information measures are excellent representations of what is meant by information and that average mutual information, in turn, captures the intuitive notion of organization.

The organization imposed by the minimization algorithm might best be characterized as trophic structure. If in Fig. 5 one distinguishes pathways of active feeding (heavy lines) from passive, detrital flows (fine lines) it becomes possible to decompose the network into two acyclic subgraphs as shown in Figs 6(a) and 6(b). The trophic identities of the aggregated compartments in Fig. 6(a) are apparent. The microphytes (I) and detritus (III) provide a food base for the pelagic herbivores (IV) and benthic

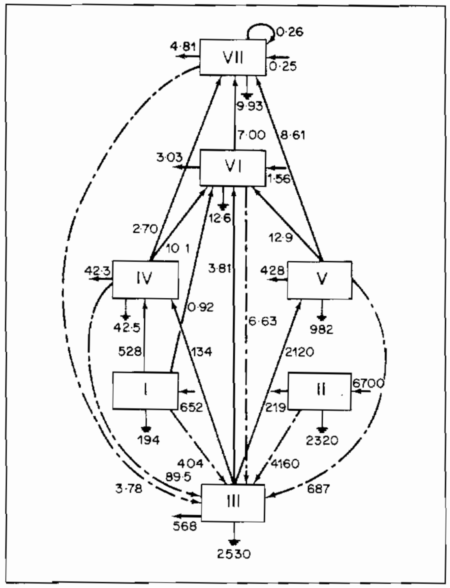

FIG. 5. The system in Fig. 4 after aggregation into seven compartments.

herbivores (V). These in turn are fed upon by the carnivores (VI), and the top carnivores-omnivores (VII).

In retrospect it is not too difficult to see why minimizing the decrease in network mutual information should lump compartments in a fashion similar to trophic levels. Implicit in the idea of a trophic level is the notion of function redundancy, i.e. several ecological components performing the

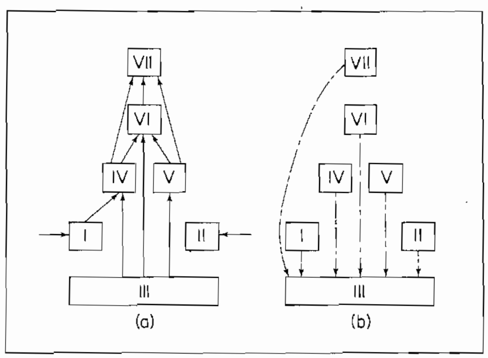

FIG. 6. A decomposition of the system in Fig. 5 into acyclic subgraphs. (a) the structure of the grazing chain, (b) the detrital returns.

same function as regards the transfer of material and energy. Functionally redundant components act in parallel with little transfer among themselves. For example, none of the species which are mapped into compartment (VI) communicate with one another. The same goes for the components of (IV) and there is negligible transfer among those species comprising (VII).

The average mutual information appears in equation (32) as the difference between two terms—$H(A)$, the entropy of the flows, and $H(A/B)$ the so-called conditional entropy of the flows. Rutledge *et al.* (1976) point out how parallel pathways are the chief contributors to the conditional entropy. If one wishes to minimize the decrease in $M(N)$, one way of achieving that objective is to force as large a decrease as possible upon the conditional entropy. Hence, parallel pathways are eliminated by aggregation into what resembles trophic levels.

5. Information Theoretical Discussion of Hierarchical Structure

The information contained in the structure of a network without reference to hierarchy was derived in a previous paper (Hirata & Ulanowicz, 1984) and has been addresssed again in section 3 of this paper. In this section, the information contained in the structure of a hierarchically nested set of networks is derived. Real systems usually have several specific levels, each characterized by different functions. For example, in ecosystems, one may consider events at the level of the cell, the organ, the organism, the species, the community, etc.

The network with p hierarchical levels is expressed by the symbol $\mathbf{N}$. The network showing the ith hierarchical level is N_i. So $\mathbf{N}$ consists of $\{N_i\}_{i=1,\ldots,p}$. In a nested hierarchical structure, $\mathbf{N}$, an upper level is considered as the aggregated network of lower levels as shown in Fig. 7, i.e. the network

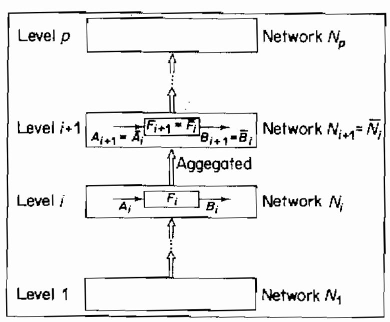

FIG. 7. Schematic representation of hierarchical structure. Network $\mathbf{N}=\{\text{Network } N_i\}_{i=1,\ldots,p}$.

which shows the ith level, N_i, is the aggregated version of the network at the $(i-1)$st level, N_{i-1}.

For simplicity, a nested hierarchy consisting of only two levels, $\mathbf{N}=\{N_i\}_{i=1,2}$, will be discussed, and the results will then be extended to that with n levels.

The hierarchical structure with two levels, $\mathbf{N}=\{N_1, N_2\}$ can be reinterpreted as the classical model of a channel illustrated in Fig. 8. According

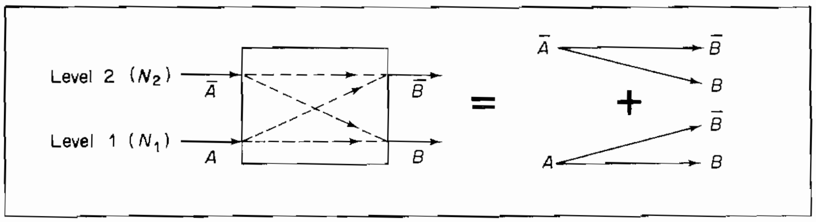

FIG. 8. Schematic diagram of a generic channel for the hierarchical structure with two levels. $\mathbf{N}=\{N_1, N_2\}$.

to the orthodox way of calculating mutual information (e.g. Abramson, 1963), one may identify the mutual information contained in the network with two hierarchical levels in the manner shown in Proposition 4. This quantity represents the average amount of uncertainty resolved by the knowledge of the structure.

Proposition 4

The information, $M(\mathbf{N})$, contained in the network structure with two hierarchical levels, $\mathbf{N}=\{N_1, N_2\}$ is:

$$M(\mathbf{N})=M(N_1)+M(N_2)+\sum_{i=1}^{2}\sum_{\substack{j=1\\(i\neq j)}}^{2} M[N_i\rightarrow N_j] \tag{37}$$

where $M(N_i)\geq 0$ $(i=1,2)$ and $M[N_i\rightarrow N_j]\geq 0$ $(i,j=1,2,i\neq j)$. $M(N_i)$ is the information associated with the network structure showing the ith level and has already been defined in Proposition 3. $M[N_i\rightarrow N_j]$ shows the information of the mutual relations between N_i and N_j. Its concrete form is revealed in the proof of this proposition.

Proof of Proposition 4. This is proved in Appendix 2.

Information $M(\mathbf{N})$ consists of four terms, $M(N_1)$, $M(N_2)$, $M[N_1\rightarrow N_2]$ and $M[N_2\rightarrow N_1]$ as shown in Proposition 4. The former two terms show the information contained in network structure of each level. The latter two terms are the information associated with the mutual relations between two levels. $M[N_i \to N_j]$ express how much information one can obtain about the structure of level N_i through knowledge about level N_j.

A hierarchical structure wherein each level is a discrete aggregation of the next lower level is a very important special case having numerous applications. Concerning this special case the following proposition can be obtained.

Corollary 1

If the upper level of a 2-level hierarchy is made by discrete aggregation of the first level, the information, $M(N)$, contained in the network structure $\mathbf{N}=\{N_1, N_2\}$, is as follows:

$$M(\mathbf{N}) = M(N_1) + M(N_2) + M[N_2 \to N_1] \tag{38}$$

Here

$$M[N_2 \to N_1] = \sum_{q=0}^{m} \sum_{j=0}^{m} \sum_{i,k}^{*} Q_i f_{ik} \log \frac{\sum_{l=0}^{n} \sum_{r,t}^{*} Q_l Q_r f_{lt} f_{rk}}{\sum_{l=0}^{n} \sum_{r,t}^{*} Q_l Q_r f_{lk} f_{rt}} \tag{39}$$

where

$$\sum_{r,t}^{*} = \sum_{a_r \varepsilon S^{-1}(\bar{a}_q)} \sum_{b_t \varepsilon S^{-1}(\bar{b}_j)}$$

and $x \in S^{-1}(\bar{x})$ means that x belongs to the group of elements mapped into $\bar{x}$.

Proof of Corollary 1. This is proved in Appendix 3.

Corollary 1 shows that the upper level does not contain any direct information about the lower level because $M[N_1 \to N_2]=0$. It is readily noticed that the information about the upper level, $M(N_2)$, implicitly includes the information on the lower level.

Proposition 4 and Corollary 1 can be easily extended to a general hierarchical structure with p levels.

Proposition 5

The information, $M(\mathbf{N})$, contained in the network structure with p hierarchical levels, $\mathbf{N}=\{N_i\}_{i=1,\ldots,p}$, is (Fig. 9)

$$M(\mathbf{N}) = \sum_{i=1}^{p} M(N_i) + \sum_{i=1}^{p} M[N_i \to \sum_{\substack{j=1 \\ (j \neq i)}}^{p} N_j 1 \tag{40}$$

where $M[N_i \to \sum_{j=1(j\neq i)}^{p} N_j]$ is the information on the mutual relations

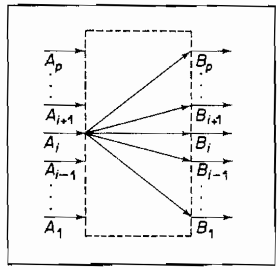

FIG. 9. Schematic diagram of a generic channel for the hierarchical structure with p levels. $\mathbf{N}=\{N_i\}_{i=1,\ldots,p}$.

between N_i and the rest of levels $(N_1, \ldots, N_{i-1}, N_{i+1}, \ldots, N_p)$ and its form is clearly defined in the proof.

Proof of Proposition 5. This is proved in Appendix 4.

In Proposition 5, $M[N_i \to \sum_{j=1,(j\neq i)}^{p} N_j]$ shows how much information about level N_i the rest of levels $(N_1, \ldots, N_{i-1}, N_{i+1}, \ldots, N_p)$ contains.

Corollary 2

If each succeeding level is made by discrete aggregation of the preceding lower level; the information $M(\mathbf{N})$, contained in the network structure with p hierarchical levels, $N=\{N_1, \ldots, N_p\}$, is as follows:

$$M(\mathbf{N}) = \sum_{i=1}^{p} M(N_i) + \sum_{i=1}^{p} M\left[N_i \to \sum_{\substack{j=1 \\ (i>j)}}^{p} N_j \right]. \tag{41}$$

One should notice that the difference between Proposition 5 and Corollary 2 is that $(i \neq j)$ has been replaced by $(i > j)$. This shows that in a strictly reductionistic system any upper level does not contain the direct information on lower levels.

Much of ecological (and, in general, biological) research is conducted under the implicit assumptions that give rise to the last corollary. It should be clear from the last two propositions that the strictly-nested hierarchies is but a narrow example of a much broader set of possible perspectives. In principle, it should be possible to acquire data on biological hierarchies to test whether (41) is a sufficient description of the information in hierarchies, or whether it is necessary to invoke (40).

6. Summary

Mutual information appears to be instrumental to the most desirable way to aggregate systems compartments in the absence of a priori assumptions about network dynamics. The lexicon used to describe the lumping of compartments may be applied to the definition of the total information inherent in an hierarchical system. In general, the total system information is affected by interactions among all levels of the hierarchy and can be simplified only when the hierarchy is strictly nested, that is, consists of a cascade of discrete, homomorphic mappings.

This work was supported by a grant from the System Theory and Operations Research Program of the National Science Foundation (ECS-8110035). The Computer Science Center of the University of Maryland also contributed free computer time. The authors are grateful to Dr Alan J. Goldman and Dr Albert Cheung of the Johns Hopkins University for their helpful comments during the course of the development of this article.

REFERENCES

ABRAMSON, N. (1963). *Information Theory and Coding.* New York: McGraw-Hill.

CALE, W. G. & ODELL, P. L. (1979). In: *Theoretical Systems Ecology.* (Halfon, E.A. ed.). pp. 55–77. New York: Academic Press.

CALE, W. G. & ODELL, P. L. (1980). *Math. Biosci.* **49,** 121.

CALE, W. G., O'NEILL, R. V. & GARDNER, R. H. (1983). *J. theor. Biol.* **100,** 539.

HIRATA, H. (1978). *Int. J. Systems Sci.* **9,** 515.

HIRATA, H. & ULANOWICZ, R. E. (1984). *Int. J. Systems Sci.* **15,** 261.

O'NEILL, R. V. & GARDNER, R. H. (1979). In: *Methodology in Systems Modelling and Simulation.* (Ziegler, B.P., Elzas, M. S., Klir, G. J. & Oren, T. J. eds), pp. 447–463. Amsterdam: North-Holland.

O'NEILL, R. V. & RUST, B. W. (1979). *Ecological Modelling* **7,** 91.

RUTLEDGE, R. W., BASORE, B. L. & MULHOLLAND, R. J. (1976). *J. theor. Biol.* **57,** 355.

THEIL, H. (1967). *Economics and Information Theory.* Amsterdam: North-Holland.

ULANOWICZ, R. E. (1979). *Oecologia* **43,** 295.

ULANOWICZ, R. E. (1980). *J. theor. Biol.* **85,** 223.

ULANOWICZ, R. E. (1981). In: *A unified theory of self-organization, developments in environmental modelling, 1: Energy and Ecological Modelling* (Mitsch, W. J. *et al.*, eds.), pp. 649–652. Amsterdam: Elsevier.

ULANOWICZ, R. E. & KEMP, W. M. (1979). *Am. Nat.* **114,** 871.

ZEIGLER, B. P. (1976). In: *Systems Analysis and Simulation in Ecology,* Volume IV. (Patten, B. C. ed.), pp. 299–311. New York: Academic Press.

APPENDIX 1

Proof of Proposition 3

This proof is patterned after Theil (1967), but requires a more general set of definitions:

Define p_{kj}, $\bar{p}_{il}$ and $\hat{S}$ as follows:

$$\begin{aligned} p_{kj} &= P(a_k, b_j) = f_{kj}Q_k \\ \bar{p}_{il} &= P(\bar{a}_i, \bar{b}_l) = \bar{f}_{il}\bar{Q}_i \end{aligned} \tag{A1}$$

$$\hat{S} = [\hat{s}_{pq}]_{p=0,1,\ldots,m,m+1,m+2,q=0,1,\ldots,n,n+1,n+2}$$

$$= \left[\begin{array}{c|ccc|cc} 1 & 0 & \cdots & 0 & 0 & 0 \\ \hline 0 & & & & 0 & 0 \\ \vdots & & S & & \vdots & \vdots \\ 0 & & & & 0 & 0 \\ \hline 0 & 0 & \cdots & 0 & 1 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{array}\right] \tag{A2}$$

The following relations among these quantities hold.

$$\begin{aligned} \bar{p}_{il} &= \sum_{k=0}^{n+2} \sum_{j=0}^{n+2} \hat{s}_{ik} p_{kj} \hat{s}_{lj} \\ \bar{p}_{i.} &= \sum_{k=0}^{n+2} \hat{s}_{ik} p_{k} \\ \bar{p}_{.l} &= \sum_{j=0}^{n+2} p_{.j} \hat{s}_{lj}. \end{aligned} \tag{A3}$$

By further defining

$$\begin{aligned} r_{ij} &= \sum_{k=0}^{n+2} \hat{s}_{ik} p_{kj} \\ c_{kl} &= \sum_{j=0}^{n+2} p_{kj} \hat{s}_{lj} \end{aligned} \tag{A4}$$

and noticing that

$$\sum_{i=1}^{m} s_{ik} = 1 \tag{A5}$$

it becomes possible to rewrite the loss of information as

$$M(N) - M(\bar{N}) = \sum_{i=0}^{m+2} \sum_{l=0}^{m+2} \bar{p}_{il}(M_I + M_{II} + M_{III}) \tag{A6}$$

where

$$\begin{aligned} M_I &= \sum_{j=0}^{n+2} \frac{r_{ij}\hat{s}_{lj}}{\bar{p}_{il}} \log \frac{r_{ij}\hat{s}_{lj}/\bar{p}_{il}}{p_{.j}\hat{s}_{lj}/\bar{p}_{.l}} \\ M_{II} &= \sum_{k=0}^{n+2} \frac{\hat{s}_{ik}c_{kl}}{\bar{p}_{il}} \log \frac{\hat{s}_{ik}c_{kl}/\bar{p}_{il}}{\hat{s}_{ik}p_{k.}/\bar{p}_{i.}} \\ M_{III} &= \sum_{k=0}^{n+2} \sum_{j=0}^{n+2} \frac{\hat{s}_{ik}p_{kj}\hat{s}_{lj}}{\bar{p}_{il}} \log \frac{\hat{s}_{ik}p_{kj}\hat{s}_{lj}/\bar{p}_{il}}{(r_{ij}\hat{s}_{lj}/\bar{p}_{il})(\hat{s}_{ik}c_{kl}/\bar{p}_{il})}. \end{aligned} \tag{A7}$$

Because $\sum_{j=0}^{n+2} r_{ij}\hat{s}_{lj}/\bar{p}_{il} = \sum_{j=0}^{n+2} p_{.j}\hat{s}_{lj}/\bar{p}_{.l} = 1$, one may employ Shannon's inequality to show that M_{I} is non-negative. The non-negativity of M_{II} and M_{III} follows in similar fashion. Therefore, from (A6),

$$M(N) \geq M(\bar{N}) \tag{A8}$$

with equality pertaining only when

$$\begin{aligned} r_{ij}/\bar{p}_{il} &= p_{.j}/\bar{p}_{.l} \\ c_{kl}/\bar{p}_{il} &= p_{k.}/\bar{p}_{i.} \\ p_{kj}/\bar{p}_{il} &= r_{ij}c_{kl}/(\bar{p}_{il})^2. \end{aligned} \tag{A9}$$

□

APPENDIX 2

Proof of Proposition 4

From the definition of mutual information of the channel depicted in Fig. 8,

$$M(\mathrm{N}) = M(A;\, B, \bar{B}) + M(\bar{A};\, \bar{B}, B) \tag{A10}$$

where

$$M(X;\, Y, Z) = \sum_{XYZ} P(x, y, z) \log [P(x, y, z)/P(x)P(y, z)]. \tag{A11}$$

$M(X;\, Y, Z)$ is the average amount of information about X provided by observation of both Y and Z. Equation (A11) can be rewritten as follows

$$M(X;\, Y, Z) = M(X;\, Y) + M(X;\, Z/Y) \tag{A12}$$

where

$$M(X;\, Y) = \sum_{XY} P(x, y) \log [P(x, y)/P(x)P(y)] \tag{A13}$$

$$M(X;\, Z/Y) = \sum_{XYZ} P(x, y, z) \log [P(x, y, z)/P(x/y)P(y, z)]. \tag{A14}$$

The first term on the right of (A12) is the mutual information between X and Y. The second term on the right of (A12) is the mutual information between X and Z given that Y is known.

Using (A12), equation (A10) can be rewritten as follows

$$M(\mathrm{N}) = M(A;\, B) + M(\bar{A};\, \bar{B}) + M(A;\, \bar{B}/B) + M(\bar{A};\, B/\bar{B}). \tag{A15}$$

From the definition of $M(N_i)$

$$\begin{aligned} M(N_1) &= M(A; B) \\ M(N_2) &= M(\bar{A}; \bar{B}) \end{aligned} \tag{A16}$$

and the following notations are defined.

$$\begin{aligned} M[N_1 \to N_2] &= M(A; \bar{B}/B) \\ M[N_2 \to N_1] &= M(\bar{A}; B/\bar{B}) \end{aligned} \tag{A17}$$

Therefore, equation (37) follows from (A15)–(A17) □

APPENDIX 3

Proof of Corollary 1

From the definition of a discrete aggregation, the following relations are apparent: If $P(b, \bar{b}) = 1$, then $P(a/b, \bar{b}) = P(a/b)$. If $P(b, \bar{b}) = 0$, then $P(\bar{a}, \bar{b}, b) = 0$. Therefore

$$\begin{aligned} M[N_1 \to N_2] &= M(A; \bar{B}/B) \\ &= \sum_{AB\bar{B}} P(a, b, \bar{b}) \log [P(a/b, \bar{b})/P(a/b)] \\ &= 0 \end{aligned} \tag{A18}$$

The third term of (38) is calculated as

$$\begin{aligned} M[N_2 \to N_1] &= M(\bar{A}; B/\bar{B}) \\ &= \sum_{\bar{A}\bar{B}B} P(\bar{a}, \bar{b}, b) \log [P(\bar{a}/\bar{b}, b)/P(\bar{a}/\bar{b})]. \end{aligned} \tag{A19}$$ □

APPENDIX 4

Proof of Proposition 5

From the definition of mutual information of the channel illustrated in Fig. 9

$$M(\mathbf{N}) = \sum_{i=1}^{p} M(A_i; B_1, \ldots, B_p) \tag{A20}$$

where

$$\begin{aligned} &M(A_i; B_1, \ldots, B_p) \\ &\quad = M(A_i; B_i) + \sum_{\substack{j=1 \\ (j \neq i)}}^{p} M(A_i; B_j/B_i, B_1, \ldots, B_{i-1}, B_{i+1}, \ldots, B_{j-1}). \end{aligned} \tag{A21}$$

The following notation is defined

$$M\left[N_i \rightarrow \sum_{\substack{j=1 \\ (j \neq i)}}^{p} N_j\right] = [\text{the second term of (A21)}]. \tag{A22}$$

From (A21), (A22) and the definition of $M(N_i)$, Equation (40) is derived.

The second term of (A21) may be rewritten as

$$M(A; B_1/B_i) + \sum_{\substack{j=2 \\ (j \neq i)}}^{p} H(A; B_j/B_1) \tag{A23}$$

because B_1 is the finest set in all $\{B_i\}_{i=1,\ldots,p}$. Here it should be noted that there are several ways to express the amount $M(A_i; B_1, \ldots, B_p)$. Equation (A21) is not necessarily unique. □