INT. J. SYSTEMS SCI., 1990, VOL. 21, NO. 2, 429–437

# Symmetrical overhead in flow networks

ROBERT E. ULANOWICZ† and JEFFREY S. NORDEN‡

Recent work by one of the authors has identified the average mutual information and the conditional entropy as two measures from information theory that are useful in quantifying the system organization and incoherence, respectively. While the scaled average mutual information, or network ascendency, is inherently symmetrical with respect to inputs and outputs, the scaled conditional entropy, or overhead, remains asymmetrical. Employing the joint entropy, instead of the conditional entropy, to characterize the overhead, results in a symmetrical overhead and also permits the decomposition of the system capacity, or complexity, into components useful in following the response of the whole system to perturbations.

## 1. Introduction

Flow networks are very convenient respresentations of ensembles of transactions, such as might occur in economic communities, ecosystems, neural systems, and a host of other kinetic structures often designated as 'self-organizing'. Whatever self-organizing behaviour such systems may exhibit is likely to derive from the formal structure of material or energetic interactions within the community (Odum 1971, Ulanowicz 1988). Thus, a deeper understanding of developing systems should follow from any phenomenological observations on how the network structures of living systems evolve. If such phenomenology is to be scientific, it perforce must be quantitative. Hence, one seeks methods for quantifying various system-level properties of flow networks (Ulanowicz 1986).

Ecology has been a particularly fertile domain for the discussion of network properties. Almost four decades ago Odum (1953), following suggestions by his mentor, G. E. Hutchinson, pointed to the multiplicity of network pathways between two different species in an ecosystem as a structural feature that promotes the homeostasis of any bilateral interchange. In simple terms, if communication along one of the multiple paths is disrupted, it is still possible that the remaining connections can compensate to maintain the overall exchange at a viable level.

MacArthur (1955) was quick to realize that the nascent theory of information could be used to quantify the multiplicity of pathways in a system. For example, if $T_{ij}$ represents the flux from compartment $i$ to species $j$, then the fraction, $f_{ij}$, of the total flow constituted by $T_{ij}$ becomes

$$f_{ij} = \frac{T_{ij}}{\sum_k T_{ik}}$$

MacArthur used these fractions in the Shannon–Wiener index of uncertainty to define

Received 3 November 1988.

† University of Maryland, Chesapeake Biological Laboratory, Solomons, MD 20688, U.S.A.

‡ Department of Mathematical Sciences, State University of New York, Binghamton, NY 13901, U.S.A.

0020-7721/90 $3.00 

the diversity of flows, $D_m$, as

$$D_m = -\sum_{i,j} f_{ij} \log f_{ij} \tag{1}$$

where the summation is taken over all combinations of $i$ and $j$.

Unfortunately, attention in ecology soon shifted from the diversity of flows to the diversity of biomass distributions, and well over a decade of ecosystems research was dominated by futile attempts to confirm or reject the existence of a positive correlation between 'diversity and stability' (Woodwell and Smith 1969). May (1973) cooled the fervour for the topic by demonstrating that a *necessary* connection between the two properties does not exist. Because the results from so great a collective effort turned out to be so equivocal, many ecologists continue to nurture a disdain for the mention of information theory. What almost everyone failed to notice during the 60s, however, was that new concepts relevant to ecosystems had been derived by information theorists. It remained for Rutledge *et al.* (1976) to reinterpret MacArthur's ideas in terms of *conditional* probabilities. They showed that the conditional entropy was a more appropriate measure of functional redundancy in food webs.

Rutledge defined the statistical entropy, $H$, of the flows in a closed network differently from MacArthur:

$$H = -\sum_{i} Q_i \log Q_i \tag{2}$$

where

$$Q_i = \frac{\sum_k T_{ik}}{\sum_{l,m} T_{lm}}$$

is the probability that a quantum of medium is flowing through $i$. Given that a quantum is flowing through $i$, the conditional probability that it will flow next into $j$ is estimated by MacArthur's $f_{ij}$. Knowing the location of a quantum reduces one's uncertainty about where it will flow next by an amount known as the average mutual information,

$$I = \sum_{i,j} f_{ij} Q_i \log\left(\frac{f_{ij}}{\sum_k f_{kj} Q_k}\right) \tag{3}$$

This amount when deducted from the statistical entropy in (2) yields the *conditional entropy*, $D_R$, or

$$D_R = H - I \tag{4}$$

$D_R$ is the network property most related to system homeostasis. However, Rutledge's attempts to show that systems matured in the direction of higher values of $D_R$ yielded equivocal results.

Ulanowicz (1980) refocused Rutledge's treatment to emphasize $I$ as the cardinal attribute of a developing network. In fact, when $I$ is given physical dimensions by multiplying it by the total amount of flow in the system, $T\,(=\Sigma_{i,j} T_{ij})$, then the resulting 'ascendency',

$$A = TI \tag{5}$$

correlates well with most of Odum's (1969) 24 properties of 'mature' ecosystems. It appears that, in the absence of major perturbations, systems mature in the direction of increasing ascendency.

In order to assess the limits to increasing ascendency, Ulanowicz multiplied Rutledge's statistical entropy, $H$, by $T$ and observed that a basic result from information theory requires that

$$C \geqslant A \geqslant 0 \tag{6}$$

where $C = TH$. That is, the factor $C$ serves as an upper bound on $A$—a circumstance that led Ulanowicz to coin the name 'development capacity' for $C$. Thus, the ascendency appears to grow towards its theoretical limit, $C$, but is destined to fall short of that limit by a non-negative amount, $C - A$, which thereby takes on the appearance of the system *overhead*.

## 2. Completeness and symmetry

If one limits the discussion, as Rutledge did, to materially closed systems, then the overhead stems entirely from pathway redundancy. However, closed systems are theoretical abstractions, so that Ulanowicz (1980) sought to extend the Rutledge formalism to include open systems as well. In the process he showed that the newly-defined system overhead is augmented by losses from the system, such as exports and respirations.

Unfortunately, Ulanowicz's original formulation of the average mutual information in open systems employed incomplete probabilities. Among other difficulties, this means that the ascendency would often take on significantly different values when calculated in terms of the inputs to each node from those computed on the basis of their outputs. Likewise, the overhead would be highly asymmetric with respect to the input and output perspectives. This assymetry in the ascendency was obviated by Hirata and Ulanowicz (1984) by redefining the normalization factor (the total system throughput) used to estimate the probabilities. They defined separate compartments to serve as virtual origins and sinks for exogenous transfers. In an $n$-compartment system they attached the index zero to the source of exogenous inputs and $n + 1$ to the virtual repository for medium leaving the system. (To be more accurate, the labels $n + 1$ and $n + 2$ were used to designate the sinks for usable exports and dissipations, respectively; there are excellent hierarchical reasons for making this distinction—Ulanowicz (1986)—which will be invoked later in this article; however, for now it is more convenient to consider that all losses from the system flow into compartment $n + 1$.) Thus, $T_{0j}$ refers to the exogenous inputs into compartment $j$, and $T_{i,n+1}$ designates all the losses from component $i$. The modified total system throughput becomes the aggregate of all $T_{ij}$, where $i$ and $j$ both range from 0 to $n + 1$. Two new terms, $Q_0$ and $Q_{n+1}$ appear in the equation for statistical entropy (2), and $f_{0j}$ and $f_{i,n+1}$ also contribute to new terms in the revised average mutual information (3).

A significant advantage of the new ascendency is that it is entirely symmetrical with respect to inputs and outputs. For example, if one defines $g_{ij}$ as the fraction of total input to $i$ that flows from $j$, then it can be shown that

$$\begin{aligned} I &= \sum_{i,j=0}^{n+1} f_{ij} Q_i \log \left( \frac{f_{ij}}{\sum_k f_{kj} Q_k} \right) \\ &= \sum_{i,j=0}^{n+1} g_{ji} Q'_i \log \left( \frac{g_{ji}}{\sum_k g_{jk} Q'_k} \right) \end{aligned} \tag{7}$$

where

$$Q'_i = \sum_k T_{ki}/T$$

That is, the ascendency ($A = T \times I$) will have the same value regardless of whether it is 'looked at' from the input or the output perspective (i.e. one could reverse the directions of the arrows and not affect the numerical results). Furthermore, the mutual information in (7) remains well defined, even if the inputs and outputs around each compartment do not balance. The ascendency may now be calculated for networks not in the steady state. It can be shown that the mutual information in a balanced network always exceeds those of any nearby unbalanced configurations (Ulanowicz 1986).

In order to highlight the symmetry in the revised ascendency, one may multiply either form of $I$ in (7) by the total system throughput, $T$, to yield:

$$A = \sum_{i,j=0}^{n+1} T_{ij} \log\left[\frac{T_{ij}T}{(T_i T'_j)}\right] \tag{8}$$

where

$$T_i = \sum_k T_{ik}$$

and

$$T'_j = \sum_k T_{kj}$$

Although the revised network ascendency is well behaved, the modified overhead remains asymmetrical with respect to inputs and outputs. Recalling that the overhead is defined to be the difference between the development capacity (the scaled statistical entropy) and the mutual information, it becomes clear that the asymmetry in the overhead derives from the asymmetry inherent in the statistical entropy used to define system capacity. Neither the MacArthur statistical entropy (1) nor the Rutledge form (2) is symmetrical with respect to inputs and outputs.

## 3. Symmetrical overhead

It is rather easy to show (Abramson 1963) that the statistical *joint entropy*, $H_j$, is symmetric with respect to both inputs and outputs:

$$\begin{aligned} H_j &= -\sum_{i,j=0}^{n+1} f_{ij}Q_i \log(f_{ij}Q_i) \\ &= -\sum_{i,j=0}^{n+1} g_{ji}Q'_j \log(g_{ji}Q'_i) \\ &= -\sum_{i,j=0}^{n+1} \left(\frac{T_{ij}}{T}\right) \log\left(\frac{T_{ij}}{T}\right) \end{aligned} \tag{9}$$

The index is so named because the argument of the logarithm is an estimator of the joint probability that a quantum of medium passes through $i$ and $j$ in succession. Defining a development capacity based on this statistical entropy, $C_J = T \times H_J$, yields an overhead, $\theta$, that is the difference between two symmetric terms, and thus is itself symmetric.

There remains a secondary, but not insignificant, benefit to defining the system capacity in terms of the joint entropy. Namely, the segregation of terms generated by exogenous inputs, internal transfers, and losses can be accomplished simultaneously within the ascendency, the overhead, and the capacity. Referring to Table 1, the subscript 0 refers to that part of the summation process generated by the exogenous inputs, i.e.

$$\left.\begin{aligned} A_0 &= \sum_{j=1}^{n} T_{0j} \log\left[\frac{T_{0j}T}{(T_0 T_j')}\right] \\ \theta_0 &= -\sum_{j=1}^{n} T_{0j} \log\left[\frac{T_{0j}^2}{T_0 T_j'}\right] \\ C_0 &= -\sum_{j=1}^{n} T_{0j} \log\left(\frac{T_{0j}}{T}\right) \end{aligned}\right\} \quad (10)$$

and the components with subscript $E$ are the corresponding elements of $A$, $\theta$ and $C$ generated by the $T_{i,n+1}$. Those components with subscript $I$ are generated by the internal transfers, i.e.

$$\left.\begin{aligned} A_I &= \sum_{i,j=1}^{n} T_{ij} \log\left[\frac{T_{ij}T}{T_i T_j'}\right] \\ \theta_I &= -\sum_{i,j=1}^{n} T_{ij} \log\left[\frac{T_{ij}^2}{T_i T_j'}\right] \\ C_I &= -\sum_{i,j=1}^{n} T_{ij} \log\left(\frac{T_{ij}}{T}\right) \end{aligned}\right\} \quad (11)$$

Thus, summing across the first two components in each row of Table 1 always yields the third, and summing down the top three entries in each column yields the fourth. (In Table 1 and hereinafter the subscript $J$ has been dropped from the capacity, so that writing simply $C$ will denote the capacity based on the joint entropy.)

The separation effected in Table 1 may seem unremarkable, until one realizes that no other choice for system capacity allows for such a clean separation of overhead. For example, using Rutledge's statistical entropy to define a system capacity will result in an internal overhead (the counterpart to $\theta_I$) that contains terms generated by exogenous transfers. (By 'generated' is meant the transfer in question appears as a multiplier of the logarithmic term. The separation in Table 1 is not meant to imply that the terms are entirely independent of each other. All the flows are actually implicit in each logarithmic argument.)

| | | |
|---|---|---|
| $A_0$ | $\theta_0$ | $C_0$ |
| $A_I$ | $\theta_I$ | $C_I$ |
| $A_E$ | $\theta_E$ | $C_E$ |
| $A$ | $\theta$ | $C$ |

Table 1. Two-way decomposition of the network development capacity into functional components. Subscripts 0, $I$ and $E$ refer to imports, internal flows, and losses, respectively. The last member in each row or column is the sum of its preceding components.

From a practical standpoint, the separation into distinct components allows one to focus on a particular section of the network. For example, one might be most interested in the internal development of the system, whereupon knowing the value of the internal capacity and how it is apportioned between the internal ascendency and the internal overhead should give clues as to the status of internal development. It should be noted that, although the whole system indices are intrinsically non-negative, some of the components of the ascendency, could hypothetically become negative. For example, if external exchanges become very much larger than the aggregated internal exchanges, it is possible that the internal ascendency could become negative. (The capacities and overheads are guaranteed to be termwise non-negative.) In the author's experience, negative internal ascendencies have been observed only for contrived, anomolous, hypothetical networks. This last observation prompts the speculation that a positive internal ascendency might be a prerequisite for the continued development of a natural system.

The reader may have noticed that the symmetries just discussed tend to obscure irreversibility and the direction of time as elements of the system description. That is, if one were to reverse hypothetically the direction of every flow in the network and recalculate all the components in Table 1, the only difference would be an exchange of row one with row three. If one were asked to choose which configuration represented the real system, one would probably choose that wherein $\theta_E > 0$. However, there is nothing in the second law of thermodynamics that would guarantee the correctness of this choice.

This ambiguity underscores the utility of the hierarchical distinction made by Ulanowicz (1980): that medium which leaves the system, but is still of use to another system at the same hierarchical level, is called 'export'. Export is thus distinguished from 'dissipation', which is the medium leaving the community that no longer can serve as an import to any comparable system. Using the convention established by Hirata and Ulanowicz (1984), exports will be assumed to flow to a sink compartment labelled $n+1$ and dissipations are imagined to pass into unit $n+2$. Exports and dissipations generate separate terms in the ascendency, overhead and capacity as follows:

$$\left.\begin{aligned} A_E &= \sum_{i=1}^{n} T_{i,n+1} \log\left[\frac{T_{i,n+1}T}{T_i T'_{n+1}}\right] \\ \theta_E &= -\sum_{i=1}^{n} T_{i,n+1} \log\left[\frac{T^2_{i,n+1}}{T_i T'_{n+1}}\right] \\ C_E &= -\sum_{i=1}^{n} T_{i,n+1} \log\left(\frac{T_{i,n+1}}{T}\right) \end{aligned}\right\} \quad (12)$$

and

$$\left.\begin{aligned} A_S &= \sum_{i=1}^{n} T_{i,n+2} \log\left[\frac{T_{i,n+2}T}{T_i T'_{n+2}}\right] \\ \theta_S &= -\sum_{i=1}^{n} T_{i,n+2} \log\left[\frac{T^2_{i,n+2}}{T_i T'_{n+2}}\right] \\ C_S &= -\sum_{i=1}^{n} T_{i,n+2} \log\left(\frac{T_{i,n+2}}{T}\right) \end{aligned}\right\} \quad (13)$$

Table 1 would thereby be expanded by adding a fourth row to accommodate $A_S$, $\theta_S$ and $C_S$. The distinction between the two types of export is hierarchical in nature (see below) and adds another element of asymmetry between inputs and outputs that helps to distinguish the direction in which irreversible events are transpiring.

As discussed by Ulanowicz (1986), the role of overhead is to provide limits on the increase in ascendency and simultaneously to reflect the system's 'strength in reserve' from which it can draw to meet unexpected perturbations. For example, $A_0$ could increase at the expense of $\theta_0$. To do so, however, the system would either have to diminish the magnitude of its imports (and thereby starve itself) or focus on the most easily tapped resource to the exclusion of all others. The latter reconfiguration would leave the entire system vulnerable to any disruption in its single avenue of sustenance. Therefore, it does not benefit the system in the long run to reduce $\theta_0$ below some point. The critical level of $\theta_0$ is determined by the spectrum of disturbances that befall the system.

Similarly, a reduction in each of the other three components remains feasible only up to a point. The potential reduction in $\theta_I$ is limited by the rigors of the environment. $\theta_I$ arises from the multiplicity of pathways connecting any two system components. When few or no alternative pathways are available to compensate for flow reductions caused by stochastic disturbances along any subset of connections, the continued existence of the receiving element stands in jeopardy. With $\theta_E$ the limit is reached whenever any further reduction in a remaining export causes a proportionate decrease in one of its imports to which the given export is linked via an autocatalytic pathway in the next higher hierarchical level.

## 4. Cone spring example

A benchmark example employed by many engaged in ecosystem network analysis is the simple configuration of energy transfers among the five ecological components of Cone Spring (Williams and Crouthamel, unpublished) as shown in Fig. 1. The full suite of components of the newly-defined development capacity for Cone Spring are listed in Table 2(*a*).

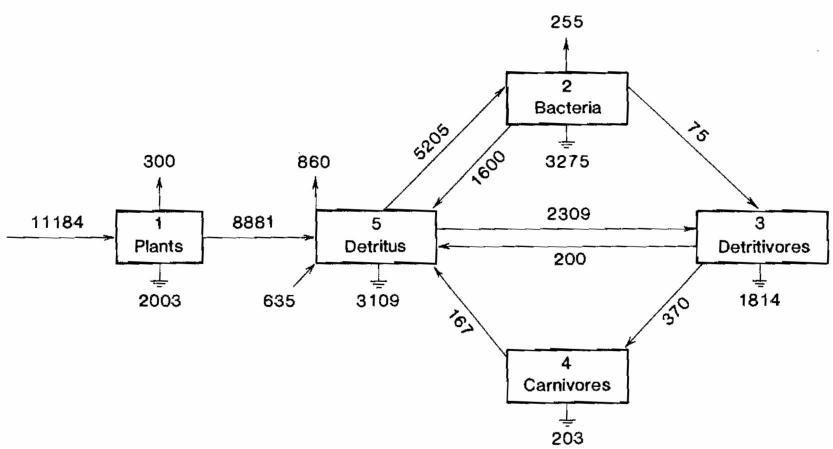


Figure 1. Observed network of energy flows in kcal $m^{-2}$ $y^{-1}$ among the five major features of the Cone Spring ecosystem (Williams and Cronthamel, unpublished).

| | | Ascendency | Overhead | Capacity |
|---|---|---|---|---|
| (a) | Inputs | 19 148 | 6222 | 25 370 |
| | Internal | 29 332 | 29 832 | 59 164 |
| | Exports | 1052 | 7811 | 8863 |
| | Dissipations | 7194 | 35 274 | 42 468 |
| | Totals | 56 726 | 79 139 | 135 865 |
| (b) | Inputs | 21 712 | 0 | 21 712 |
| | Internal | 36 507 | 24 939 | 61 446 |
| | Exports | 820 | 1062 | 1882 |
| | Dissipations | 7254 | 37 489 | 44 743 |
| | Totals | 66 293 | 63 490 | 129 783 |

Table 2. The functional components of the development capacity for (*a*) the Cone Spring network of energy flows depicted in Fig. 1; (*b*) the hypothetically streamlined version shown in Fig. 2. All values are in kcal bits $m^{-2}$ $y^{-1}$.

In some fundamental respects, this simple example is representative of many more complicated networks that have been observed. For example, the points of entry for exogenous inputs are fewer than the routes of egress from the system. Furthermore, those compartments receiving exogenous imports usually have few other inputs from internal components. The first observation means that the overhead terms generated by exports and dissipations encumber larger fractions of their capacities (88% and 83%, respectively). In the second case, most of the capacity generated by the inputs appears as ascendency (75%) and the overhead on imports is relatively small. Usually, the crucial point is how the internal ascendency and the internal overhead (redundancy) are apportioned. Here the split is very nearly 50–50%.

To see how the components of the capacity might respond to changes in network structure, the total system throughput in Fig. 1 was re-routed to give the slightly more 'streamlined' or 'efficient' hypothetical configuration shown in Fig. 2. Two types of changes were made to the Cone Spring network to create this revised structure. Firstly, the five cycles found in the observed network were reduced to only two cycles without altering the aggregate throughput. This aggregation of cycles also simplified

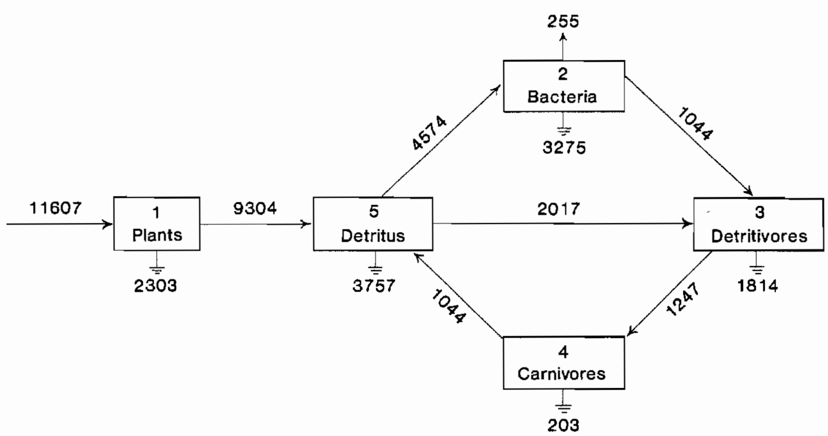


Figure 2. Hypothetical 'streamlined' version of the Cone Spring network maintaining the same system throughput, but diminishing the number of cycles and exogenous transfers.

the internal transfers by eliminating two of the actual flows. Secondly, the imports and exports were combined into one each, in such a way as to keep the total throughput constant.

As can be seen from the resulting components of the capacity listed in Table 2(*b*), the ratio of the internal ascendency to the redundancy rose to 60:40. The overhead on imports vanished, because the hypothetical system is dependent upon a single input. The export capacity fell almost five-fold with the disappearance of two of the exports, and the percentage of export capacity encumbered by overhead decreased to 56%. In contrast, the dissipation capacity and its components changed relatively little.

## 5. Summary and conclusions

Using the joint entropy to compute the development capacity of a flow network results in a systems overhead that is symmetrical with respect to inputs and outputs. Furthermore, the capacity defined by the joint entropy can be decomposed readily into 14 separate components that are useful as congeneric indices that quantify the topological changes in the network structure.

ACKNOWLEDGMENTS

The authors wish to acknowledge the Department of the Navy (Contract N00024-86-C-5188) as the source of support for this work.

Colin Atwood provided numerous helpful suggestions.

REFERENCES

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

HIRATA, H., and ULANOWICZ, R., 1984, Information theoretical analysis of ecological networks. *Int. J. Systems Sci.*, **5**, 261.

MACARTHUR, R. H., 1955, Fluctuations of animal populations and a measure of community stability. *Ecology*, **36**, 533.

MAY, R. M., 1973, *Stability and Complexity in Model Ecosystems* (Princeton, NJ: Princeton University Press).

ODUM, E. P., 1953, *Fundamentals of Ecology* (Philadelphia: Saunders); 1969, The strategy of ecosystem development. *Science, N.Y.*, **164**, 262.

ODUM, H. T., 1971, *Environment, Power, and Society* (New York: Wiley).

RUTLEDGE, R. W., BASORRE, B. L., and MULHOLLAND, R. J., 1976, Ecological stability: an information theory viewpoint. *J. theor. Biol.*, **57**, 355.

ULANOWICZ, R. E., 1980, An hypothesis on the development of natural communities. *J. theor. Biol.*, **85**, 223; 1986, *Growth and Development: Ecosystems Phenomenology*. (New York: Springer-Verlag); 1988, Formal agency in ecosystem development. *Search for Ecosystem Principles: A Network Perspective*, edited by M. Higashi and T. P. Burns (New York: Oxford University Press), to be published.

WOODWELL, G. M., and SMITH, H. H., 1969, Diversity and stability in ecological systems. *U.S. Brookhaven Symposium on Biology*, New York, Vol. 22.