CHAPTER

System-wide measures in ecological network analysis

3

Caner Kazanci[a,*], **Qianqian Ma**[b]

[a]*Department of Mathematics, Faculty of Engineering, University of Georgia, Athens, Georgia, USA*

[b]*Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, Provincial Key Laboratory of Applied Botany, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, PR China*

**Corresponding author: e-mail address: caner@uga.edu*

3.1 INTRODUCTION

Ecological network analysis (ENA) (Patten, 1978; Fath and Patten, 1999; Ulanowicz, 2004) is a system-oriented methodology to simulate and analyze ecosystem interactions. ENA relies on compartmental models that are constructed to represent the transactions of energy or matter within ecosystems. To facilitate the evaluation of an ecosystem, various system-wide measures have been proposed to capture its holistic properties (Jørgensen et al., 2013). Over the years, ENA has been enriched by new ecological measures, many of which are inspired by network measures or concepts utilized in other fields. Finn's cycling index (FCI) (Finn, 1977) is based on the Leontief structure matrix, initially developed for economic input–output analysis (Leontief, 1966). Ascendancy and several related measures (Ulanowicz, 1986) are based on information theory (MacArthur, 1955; Rutledge et al., 1976) and thermodynamics. First centrality measures were originally developed for and applied to social networks (Katz, 1953; Sabidussi, 1966). System-wide measures in ecology play an increasingly important role in studying and assessing various ecosystems (Patrício et al., 2004), including forest ecosystems (Schaubroeck et al., 2012), marine ecosystems (Tomczak et al., 2013), and lake ecosystems (Chrystal and Scharler, 2014).

The understanding of relationships between ecosystem properties has been in the forefront of system research in ecology for years (Jordán and Jørgensen, 2012). Unlike empirical ecological indicators (e.g., concentration of a toxic substance, rate of primary production), ENA measures are based on compartmental models of the ecosystem, and often have complicated formulations. Application of ENA measures for ecosystem studies and environmental management is not any easier than their development. Except for several basic measures (e.g., #links, #compartments, and total system throughflow (TST)), most of these measures (e.g., FCI, ascendency, and

Developments in Environmental Modelling, Volume 27, ISSN 0167-8892, http://dx.doi.org/10.1016/B978-0-444-63536-5.00003-X

development capacity) require a more profound understanding of which health aspects they are able to cover and how they can be used in environmental management (Jorgensen et al., 2005). Mastering the development, computation, and interpretation of over 40 ENA measures poses a serious challenge for the scientist intending to utilize them. Actually, once a compartmental model is built, computation of these measures is often straightforward, and multiple freely available software packages exist (Fath and Borrett, 2006; Kazanci, 2007; Schramski et al., 2011; Borrett and Lau, 2014) to handle the task. However, interpreting each value computed for each measure is a much greater challenge.

This chapter provides a detailed description of most commonly used ENA measures, and investigates their relationships. While mastering over 40 measures is an extremely difficult task, investigating the relationships between these measures may provide insightful information as to which different aspect of an ecosystem each measure represents. For example, cycling index (Finn, 1978; Ma and Kazanci, 2014) and indirect effects index (Higashi and Patten, 1986; Ma and Kazanci, 2012) are two different measures. Assuming that these two measures turn out to be very similar for two different ecosystem models, is this interesting information that requires further attention, or is it already expected?

Earlier studies of pairwise relationships of ecosystem measures have been conducted through theoretical and empirical investigations. Cohen and Briand (1984) reports that the link density (the average number of links per compartment) does not change with network size. In contrast, Havens (1992) shows link density increases with network size. Yodzis (1980) finds a somewhat slow decrease of connectance with increasing species richness, while Martinez (1992) reports that the connectance tends to remain constant across networks of different size. Higashi and Patten (1986, 1989) and Patten (1991) show that indirect effects increase with network size, connectance, FCI, and TST. Fath (2004) constructs artificial ecological networks to investigate the relationships between amplification, homogenization, synergism, and network size. Buzhdygan et al. (2012) compares 10 system-wide measures for 7 geographically close pastoral ecosystems, built based on 3 years of field research. Vermaat et al. (2009) assesses 20 food web-structure properties and finds substantial covariance exists among these properties.

In this chapter we provide a more comprehensive comparison among a much larger set of ENA measures than earlier works. This study is based on published network models of 52 ecosystems, with a variety of network sizes, flow currencies, and flow and storage magnitudes. A thorough literature search informs us that there exist around 40 commonly used system-wide measures in ecology in general. Nine of these measures are based on the topology (adjacency matrix) of the network, 26 are defined using flow rates, and 5 are based on storage information. Due to their complex formulations, it is not feasible to derive the mathematical relations between the measures in an algebraic fashion. Instead, cluster analysis is used as a statistical tool to classify the measures based on their similarities. We report our findings and compare them with observations from earlier published works.

3.2 DESCRIPTION OF SYSTEM-WIDE MEASURES

System-wide measures for ENA are based on compartmental models of ecosystems. Compartments represent various entities in the ecosystem, such as plants, animals, and nutrient sources. The flows among compartments are the transport of energy or matter within the system. Boundary input and output represent the transfer of energy or matter between the system and the environment. Each model contains the following data: environmental inputs (z), environmental outputs (y), flow matrix (F), and storage values (x). Assuming there are n compartments in the system, this data is denoted as follows:

z_i : Rate of environmental input to compartment i

y_i : Rate of environmental output from compartment i

x_i : Storage value at compartment i

A_{ij} : Indicator of flow from compartmentent j (columns of A) to compartment i (rows of A)

F_{ij} : Rate of direct flow from compartment j (columns of F) to compartment i (rows of F)

where $i, j = 1, 2, \ldots, n$. We also define a generalized flow matrix R that combines flows among compartments (F), environmental inputs (z), and outputs (y):

$$R = \begin{bmatrix} F & z \\ y & 0 \end{bmatrix}$$

R_{ij} is similar to the flow matrix F, with an additional compartment $(n+1)$ representing the environment. $F_{ij} = R_{ij}$ for $i, j = 1, 2, \cdots, n$. Throughflow T_i is the rate of material (or energy) moving through compartment i. Input throughflow is defined as the sum of flow rates into compartment i from other compartments and the environment. Similarly, output throughflow is the sum of flow rates from compartment i to other compartments and the environment. For a system at steady state, input and output throughflows are equal:

$$T_i = \sum_{j=1}^{n} F_{ij} + z_i = \sum_{j=1}^{n} F_{ji} + y_i$$

Flow intensity matrix G is obtained by normalizing the flow matrix F by the throughflows values:

$$G_{ij} = \frac{F_{ij}}{T_j}$$

G is actually a one-step probability transition matrix, where G_{ij} represents the probability of material (or energy) transferring from compartment j to compartment i per unit time step. Almost all system-wide measures are defined based on this presented information.

Depending on the information utilized, these measures can be classified into three major groups:

1. *Structure-based measures* only require the topology (adjacency matrix) of the network, as shown in Figure 3.1a.
2. *Flow-based measures* are computed based on flow rates and do not require storage information, as shown in Figure 3.1b.
3. *Storage-based measures* require storage values, in addition to flow rates, as shown in Figure 3.1c.

Our work includes 9 structure-based, 26 flow-based, and five storage-based measures. Detailed description and mathematical definitions of all 40 measures are provided here.

Structure-based measures

1. *#compartments* (n): Total number of compartments in the system.
2. *#links* (m): Total number of connections among all compartments.
3. *#SCC*: Number of strongly connected components (Newman, 2009). SCC is a subset of the compartments such that (i) every compartment in the subset has a path to every other compartment and (ii) the subset is not part of some larger set with the property that every compartment can reach to every other compartment.
4. *#big SCC*: Number of strongly connected components that contain more than one compartment.
5. *Percent nodes in big SCC*: Number of compartments participating in big SCCs.
6. *Link density, or complexity*: Average number of intercompartmental links (m) per compartment.

$$\text{Link density} = \frac{m}{n}$$

7. *Connectance over direct paths*: Ratio of the number of direct links (m) to the number of possible intercompartmental links.

$$\text{Connectance over direct paths} = \frac{m}{n^2}$$

8. *Connectance over all paths*: Ratio of the number of direct and indirect links to the number of possible intercompartmental links. The difference between this measure and the previous one is due to the fact that two compartments that are not directly connected may be indirectly connected via a third (or more) compartment in between.

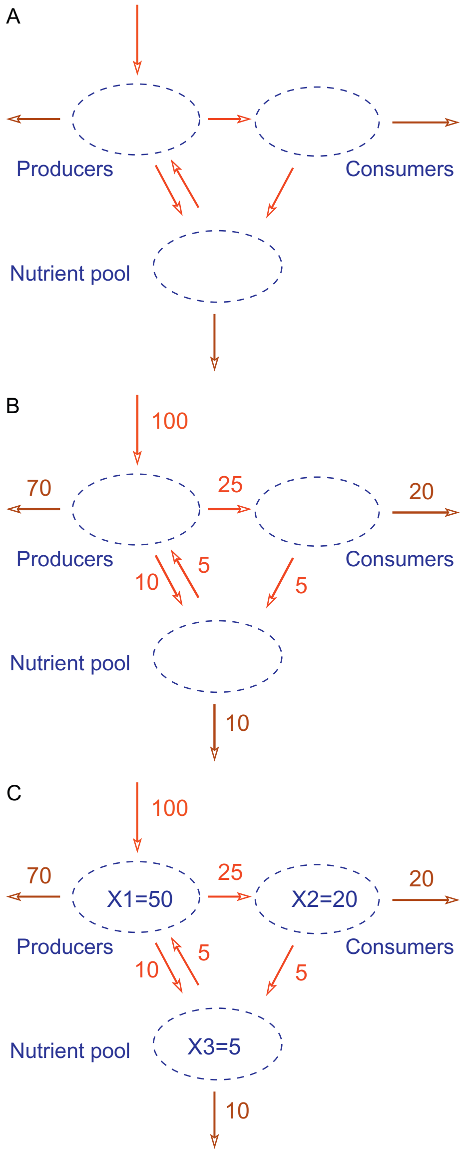

FIGURE 3.1

Ecological network types with different levels of data integration. (a) Only structural data is provided. (b) Flow rates are included in addition to network topology. (c) Storage values are included as well.

9. *Degree diversity*: Degree of a compartment (D) is the number of links that connect to it. The measure is defined by applying Shannon's information measure to the set of degrees of all compartments.

$$\text{Degree diversity} = -\sum_i \frac{D_i}{D_.} \log \frac{D_i}{D_.}$$

$D_.$ represents the sum of the degrees of all compartments. This measure is affected by the number of compartments and the evenness of degrees. Higher number of species and a more even distribution of connections result in an increase in Shannon's diversity.

Flow-based measures

1. *Total boundary input*: Sum of flows entering the system.

$$\text{Total boundary input} = \sum_i z_i$$

2. *Total internal flow*: Sum of flow rates of intercompartmental flows.

$$\text{Total internal flow} = \sum_{i,j} F_{ij}$$

3. *TST*: Sum of throughflows of all compartments in the system.

$$\text{TST} = \sum_i T_i$$

4. *Mean throughflow*: Average throughflow of all compartments.

$$\text{Mean throughflow} = \frac{\text{TST}}{n}$$

5. *Total system throughput*: Sum of all flow rates, including environmental inputs (z), environmental outputs (y), and intercompartmental flows (F). This measure is an analogue of TST.

$$\text{Total system throughput} = \sum_{i,j} F_{ij} + \sum_i z_i + \sum_i y_i = \sum_{i,j} R_{ij}$$

6. *Average path length*: Average number of compartments a unit flow material passes through before exiting the system.

$$\text{Aggradation} = \frac{\text{TST}}{\text{Total boundary input}}$$

This measure is also named network aggradation by Ulanowicz et al. (2006).

7. *I/D ratio*: Ratio of direct to indirect flows:

$$\text{I/D ratio} = \frac{\sum_i \left[\left(G^2 + G^3 + \cdots\right)T\right]}{\sum_i (GT)} = \frac{\sum_i [(N - I - G)T]}{\sum_i (GT)} \tag{3.1}$$

Here, $N = I + G + G^2 + \cdots = (I - G)^{-1}$, where I represents the identity matrix. This measure is the revised version (Ma and Kazanci, 2012) of the original definition (Patten, 1985).

8. *Indirect effects index* (*IEI*): Rescaled version of the previous measure, so that it takes values between 0 and 1:

$$\text{IEI} = \frac{(\text{I/D ratio})}{1 + (\text{I/D ratio})} \tag{3.2}$$

9. *Finn's cycling index* (*FCI*): Fraction of the TST due to cycling (TST_c) (Finn, 1978):

$$\text{FCI} = \frac{\text{TST}_\text{c}}{\text{TST}} = \frac{1}{\text{TST}} \sum_i T_i \frac{N_{ii} - 1}{N_{ii}} \tag{3.3}$$

10. *Amplification*: Number of entries of the matrix $N = (I - G)^{-1}$ that are larger than one, excluding the diagonal elements $\left(N_{ij} > 1,\ \ i \neq j\right)$.

11. *Amplification percentage*: Rescaled version of the previous measure, so that it takes values between 0 and 1:

$$\text{Amplification percentage} = \frac{\text{Amplification}}{n(n-1)}$$

12. *Synergism*: Ratio of the sum of positive entries over the sum of negative entries of the utility analysis matrix U:

$$\text{Synergism} = \frac{\sum U_{ij} \ \ \text{where } U_{ij} > 0}{-\sum U_{ij} \ \ \text{where } U_{ij} < 0}$$

Utility analysis matrix U is defined as $U = (I - D)^{-1}$, where I represents the identity matrix and the matrix D is defined as $D_{ij} = \left(F_{ij} - F_{ji}\right)/T_i$ (Patten, 1991).

13. *Mutualism*: Ratio of the number of positive entries over the number of negative entries of the utility analysis matrix U (Patten, 1991):

$$\text{Mutualism} = \frac{\text{Number of tuples } (i,j) \ \ \text{where } U_{ij} > 0}{\text{Number of tuples } (i,j) \ \ \text{where } U_{ij} < 0}$$

14. *Homogenization*: Ratio of the coefficient of variation (CV) of G over N (Fath and Patten, 1999; Fath, 2004). CV is defined as the ratio of the standard deviation to the mean:

$$\text{Homogenization} = \frac{\text{CV}(G)}{\text{CV}(N)} = \frac{\text{sd}(G)/\text{mean}(G)}{\text{sd}(N)/\text{mean}(N)}$$

15. *Throughflow diversity*:

$$\text{Throughflow diversity} = -\sum_i \frac{T_i}{\text{TST}} \log \frac{T_i}{\text{TST}}$$

16. *Ascendency* (*Asc*): Average mutual information (AMI) measure (Equation 3.6) where total system throughput is used for the constant k:

$$\text{Ascencency} = \sum_{i,j} R_{ij} \log \frac{R_{ij} R_{..}}{R_{i.} R_{.j}} \quad \text{where } R_{i.} = \sum_j R_{ij},\ R_{.j} = \sum_i R_{ij} \ \text{ and } \ R_{..} = \sum_{i,j} R_{ij} \tag{3.4}$$

MacArthur (1955) applied Shannon's information measure to flow rates to define flow diversity measure H:

$$H = -k \sum_{i,j} \frac{R_{ij}}{R_{..}} \log \frac{R_{ij}}{R_{..}} \tag{3.5}$$

Rutledge et al. (1976) decomposed H into two parts: $H = \text{AMI} + H_c$. AMI quantifies the overall constraint in the system:

$$\text{AMI} = k \sum_{i,j} \frac{R_{ij}}{R_{..}} \log \frac{R_{ij} R_{..}}{R_{i.} R_{.j}} \tag{3.6}$$

Higher AMI values indicate a tighter network organization, channeling flows along more specific pathways (Ulanowicz, 1986), whereas residual diversity (H_c) gauges how unconstrained the flows remain, or how flexible the system remains to reconfigure itself:

$$H_c = -k \sum_{i,j} \frac{R_{ij}}{R_{..}} \log \frac{R_{ij}^2}{R_{i.} R_{.j}} \tag{3.7}$$

Heymans et al. (2002) proposed this measure as an indicator of system resilience.

17. *Overhead* (Φ): Residual diversity (H_c) (Equation 3.7) where total system throughput is used for the constant k:

$$\Phi = -\sum_{i,j} R_{ij} \log \frac{R_{ij}^2}{R_{i.} R_{.j}}$$

18. *Development capacity* (C): Flow diversity (Equation 3.5) where total system throughput is used for the constant k:

$$C = -\sum_{i,j} R_{ij} \log \frac{R_{ij}}{R_{..}}$$

Ascendency, overhead, and development capacity are scaled versions of AMI, residual diversity, and flow diversity, respectively. The equation $H = \text{AMI} + H_c$ implies $C = \text{Asc} + \Phi$.

19. *Ratio of ascendency to development capacity*: This measure equals the ratio of AMI/H. Higher values represent an efficient flow network that is rather vulnerable to perturbations.
20. *Ratio of overhead to development capacity*: This measure equals the ratio of H_c/H, and represents the degree of flexibility of the flow network.
21. *Internal ascendency* (Asc_I): Ascendency measure based solely on intercompartmental flows:

$$A_\text{I} = \sum_{i,j} F_{ij} \log \frac{F_{ij} F_{..}}{F_{i.} F_{.j}}$$

22. *Internal overhead* (Φ_I): Overhead measure based solely on intercompartmental flows:

$$\Phi_\text{I} = -\sum_{i,j} F_{ij} \log \frac{F_{ij}^2}{F_{i.} F_{.j}}$$

23. *Internal capacity* (C_I): Development capacity measure based solely on intercompartmental flows:

$$C_\text{I} = -\sum_{i,j} F_{ij} \log \frac{F_{ij}}{F_{..}}$$

24. *Robustness*: Characterizes the encounter between the opposing trends toward efficient operation (AMI/H) and increasing opportunity for reconfiguration ($-\log(\text{AMI}/H)$):

$$\text{Robustness} = -e \frac{\text{AMI}}{H} \log \frac{\text{AMI}}{H}$$

where e is the base of natural logarithm function.

25. *Ratio of internal ascendency to internal capacity*:

$$\frac{\text{Asc}_\text{I}}{C_\text{I}}$$

26. *Ratio of internal overhead to internal development capacity*:

$$\frac{\Phi_{\mathrm{I}}}{C_{\mathrm{I}}}$$

Storage-based measures

1. *Total system storage* (TSS): Sum of storage values of all compartments:

$$\mathrm{TSS} = \sum_i x_i$$

2. *Mean storage*: Average storage value of all compartments:

$$\text{Mean storage} = \frac{\mathrm{TSS}}{n}$$

3. *System residence time*: Average time of flow material retention in the system:

$$\text{System RT} = \frac{\mathrm{TSS}}{\text{Total boundary input}}$$

4. *Storage-based cycling index* (SCI): Fraction of TSS due to cycling ($\mathrm{TSS_c}$) (Ma and Kazanci, 2014):

$$\mathrm{SCI} = \frac{\mathrm{TSS_c}}{\mathrm{TSS}} = \frac{1}{\mathrm{TSS}} \sum_i x_i \frac{N_{ii} - 1}{N_{ii}} \tag{3.8}$$

5. *Biomass diversity*: Also called information-theoretic biodiversity, this measure is derived by MacArthur (1955) using Shannon's information measure on storage values of compartments:

$$\text{Biomass diversity} = -\sum_i \frac{x_i}{\mathrm{TSS}} \log \frac{x_i}{\mathrm{TSS}}$$

3.3 ECOSYSTEM MODELS USED FOR COMPARISON

We selected 52 ecological network models from the literature. Table 3.1 provides the reference, flow currency, network size, and mean storage for each model. Selected models have a variety of flow currencies, including carbon, nitrogen, energy, mineral, and biomass. The collection includes models with as little as 4 and as high as 124 compartments. Table 3.2 provides summary statistics (minimum, maximum,

Table 3.1 Fifty-Two Ecological Networks

ID	Models	Flow Currency	Flow Unit	#Compartment	Mean Storage
1	Aggregated baltic ecosystem (Wulff and Ulanowicz, 1989)	Carbon	mg/m^2/day	15	59.306
2	Chesapeake mesohaline ecosystem (Baird and Ulanowicz, 1989)	Carbon	mg/m^2/day	15	23.808
3	Crystal creek (Ulanowicz, 1986)	Carbon	mg/m^2/day	21	52.53
4	Pine forest	Nitrogen	kg/ha/year	6	1
5	North Sea pelagic marine ecosystem (Steele, 1974)	Energy	kcal/m^2/year	10	1
6	Generic euphotic oceanic ecosystem (Webster et al., 1975)	Mineral	kg/ha/year	6	1
7	Open ocean mixed layer	Carbon	g/m^2/year	6	1
8	Puerto Rican rain forest (Jordan et al., 1972)	Calcium	kg/ha/year	4	1
9	Generic salt marsh ecosystem (Webster et al., 1975)	Mineral	kg/ha/year	6	1
10	Silver Springs (Odum, 1957)	Energy	kcal/m^2/year	5	1
11	Freshwater stream ecosystem (Webster et al., 1975)	Mineral	kg/ha/year	6	1
12	Temperate forest (Webster et al., 1975)	Mineral	kg/ha/year	6	1
13	Tropical forest (Webster et al., 1975)	Mineral	kg/ha/year	6	1
14	Tropical rain forest (Edmisten, 1970)	Nitrogen	g/m^2/day	5	1
15	Generic tundra ecosystem (Webster et al., 1975)	Mineral	kg/ha/year	6	1
16	Upper Chesapeake Bay mesohaline ecosystem	Carbon	million ton/ year	12	1
17	Temperate estuary (Baird and Milne, 1981)	Carbon	g/m^2/year	13	1
18	Cypress dry season (Ulanowicz et al., 1997)	Carbon	g/m^2/year	68	192.49

Continued

Table 3.1 Fifty-Two Ecological Networks—cont'd

ID	Models	Flow Currency	Flow Unit	#Compartment	Mean Storage
19	Cypress wet season (Ulanowicz et al., 1997)	Carbon	$g/m^2/year$	68	196.93
20	Florida Bay trophic exchange Matrix dry season (Ulanowicz et al., 1998)	Carbon	$mg/m^2/year$	125	6.0103
21	Florida Bay trophic exchange matrix wet Season (Ulanowicz et al., 1998)	Carbon	$mg/m^2/year$	125	6.3141
22	Everglades Graminoids dry season (Ulanowicz et al., 2000)	Carbon	$g/m^2/year$	66	63.716
23	Everglades Graminoids wet season (Ulanowicz et al., 2000)	Carbon	$g/m^2/year$	66	65.42
24	Mangrove estuary dry season (Ulanowicz et al., 1999)	Carbon	$g/m^2/year$	94	81.234
25	Mangrove estuary wet season (Ulanowicz et al., 1999)	Carbon	$g/m^2/year$	94	81.167
26	Bothnian Bay (Sandberg et al., 2000)	Carbon	$g/m^2/year$	12	223.54
27	Bothnian Sea (Sandberg et al., 2000)	Carbon	$g/m^2/year$	12	108.3
28	Charca Lagoon (Almunia et al., 1999)	Carbon	$mg/m^2/year$	21	1
29	Chesapeake Bay mesohaline network (Baird and Ulanowicz, 1989)	Carbon	$mg/m^2/year$	36	2.7685
30	Bothnian Sea (Sandberg et al., 2000)	Carbon	$g/m^2/year$	5	1
31	Crystal River (control) (Ulanowicz, 1986)	Carbon	$mg/m^2/day$	21	55121
32	Crystal River (thermal) (Ulanowicz, 1986)	Carbon	$mg/m^2/day$	21	35972
33	Ems estuary (Baird et al., 1991)	Carbon	$mg/m^2/day$	15	3.87E +05
34	English Channel (Brylinsky, 1972)	Energy	$kcal/m^2/day$	6	1

Table 3.1 Fifty-Two Ecological Networks—cont'd

ID	Models	Flow Currency	Flow Unit	#Compartment	Mean Storage
35	Narragansett Bay (Monaco and Ulanowicz, 1997)	Carbon	mg/m^2/year	32	8476.3
36	Georges Bank (Link et al., 2008)	Wet weight	g/m^2/year	31	10.06
37	Gulf of Maine (Link et al., 2008)	Wet weight	g/m^2/year	31	10.374
38	Lake Findley (Richey et al., 1978)	Carbon	g/m^2/year	76	0.59175
39	Lake Oneida (post -ZM) (Miehls et al., 2009a)	Carbon	g/m^2/year	74	0.2866
40	Lake Oneida (pre-ZM) (Miehls et al., 2009a)	Carbon	g/m^2/year	80	1.645
41	Lake Quinte (post-ZM) (Miehls et al., 2009b)	Carbon	g/m^2/year	74	0.30345
42	Lake Quinte (pre-ZM) (Miehls et al., 2009b)	Carbon	g/m^2/year	4	1
43	Lake Wingra (Richey et al., 1978)	Carbon	g/m^2/year	5	1
44	Middle Atlantic Bight (Link et al., 2008)	Wet weight	g/m^2/year	32	10.188
45	Marion Lake (Richey et al., 1978)	Carbon	g/m^2/year	5	1
46	Mirror Lake (Richey et al., 1978)	Carbon	g/m^2/year	5	1
47	Northern Benguela upwelling (Heymans and Baird, 2000)	Carbon	mg/m^2/day	24	9998.4
48	Oyster reef (Dame and Patten, 1981)	Energy	kcal/m^2/day	6	518.66
49	Southern New England (Link et al., 2008)	Wet weight	g/m^2/year	33	8.4498
50	Somme estuary (Rybarczyk and Nowakowski, 2003)	Carbon	mg/m^2/day	9	24.368
51	Swartkops estuary (Baird et al., 1991)	Carbon	mg/m^2/day	15	1.34E +05
52	Neuse grand average (Christian and Thomas, 2000)	Nitrogen	mmol/m^2/ season	7	825.03

Table 3.2 Summary Statistics of Basic Measures of 52 Ecosystem Models

Measures	Min	Max	Mean	Median	sd	CV (sd/Mean)
#Compartments	4	125	29.73	15	32.84	1.10
#Links	5	1969	312.25	37	530.61	1.65
TST	0.32	6.01×10^6	2.61×10^5	2.64×10^3	1.07×10^6	4.10
TSS	21.2	5.81×10^6	3.32×10^5	1.30×10^3	1.10×10^6	3.31

mean, median, standard deviation, and CV) for four basic measures: (i) number of compartments, (ii) number of links, (iii) TST, and (iv) TSS. Twenty-one out of 52 models do not contain storage data. These models are indicated with a mean storage value of 1 in Table 3.1, and are excluded from our analysis for storage-based measures.

3.4 METHODS

All 40 system-wide measures are computed for the 52 ecosystem models shown in Table 3.1. To study the relationships among the 40 measures, we conduct pairwise comparisons using Pearson product–moment correlation coefficient, which measures the linear correlation between two variables. It is computed as the covariance of the two variables divided by the product of their standard deviations. This correlation ranges from -1 to 1. As it approaches zero, there is less of a linear relationship (closer to uncorrelated). The closer the coefficient is to 1 (or -1), the stronger the positive (or negative) linear correlation between two variables.

For the 40 measures, there are totally $C(40,2) = 780$ possible pairwise relations. Figure 3.2 shows the histogram of Pearson correlation coefficients for these 780 pairwise relations. About 7.44% (58) and 0.26% (2) of the pairwise relations have Pearson correlations larger than 0.9 and less than -0.9, respectively, indicating a significant presence of strong linear correlations among the 40 system-wide measures. However, due to the large number of pairwise relations, it is not feasible to cover each such pair individually. Instead, we use cluster analysis to represent and visualize the relationships among these measures.

Cluster analysis is a widely used method to partition a set of objects into two or more clusters based on their similarities (Johnson and Wichern, 2002). The set of objects in this work are the 40 system-wide measures. Measures grouped in the same cluster are more similar to each other than those in different clusters. The similarity of measures is assessed by a distance metric defined between the measures. Smaller distance between two measures indicates higher similarity. This distance metric can be defined in various ways, such as the Euclidean distance, $1-$correlation, and $1-$abs (correlation). The notation "abs(x)" represents the absolute value of x. In this work,

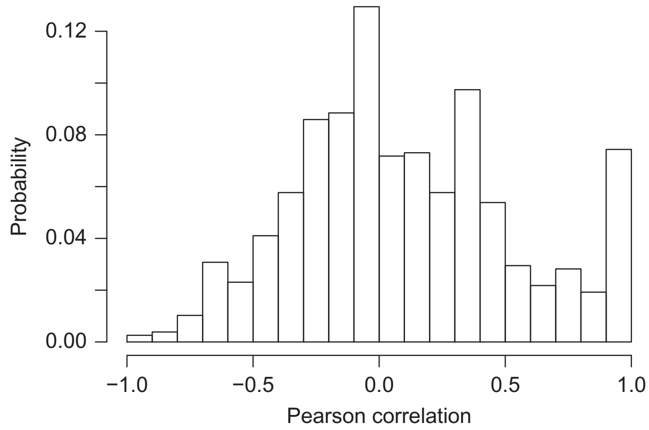

FIGURE 3.2

The histogram of Pearson product–moment correlation coefficients of all pairwise relations.

we adopt 1 − abs(Pearson correlation) as the distance between any two measures, because smaller values of 1 − abs(Pearson correlation) indicate higher (either positive or negative) correlation or similarity between two measures.

After selecting the distance metric, various methods are available to build clusters, such as single linkage, complete linkage, average linkage, Ward's method, and centroid method. There is no definitive answer as to which method is the best choice, as each method has its own advantages and disadvantages. For this work, we use the simplest and most efficient method, single linkage, also known as the nearest neighbor technique. This method is capable of finding irregular-shaped clusters, yet it suffers from the so-called chaining effects (Johnson and Wichern, 2002). The defining feature of this method is that the distance between clusters is specified as the distance between the closest pair of measures in these two clusters. The procedure works as follows:

1. Start with 40 clusters where each measure is one cluster. The distance between any two measures is 1 − abs(Pearson correlation).
2. Place the two measures with the smallest distance into a single cluster.
3. Define the distance between two clusters as the distance between the closest pair of measures from these two clusters.
4. Merge the two nearest clusters into a single cluster.
5. Repeat steps 3 and 4 until all 40 measures are within one cluster.

Figure 3.3 shows the cluster dendrogram using 1 − abs(Pearson correlation) as the distance metric. The y-axis represents the distance between clusters (or between measures if there is only one measure in each cluster). Letter r will be used to represent the Pearson correlation coefficient of two measures.

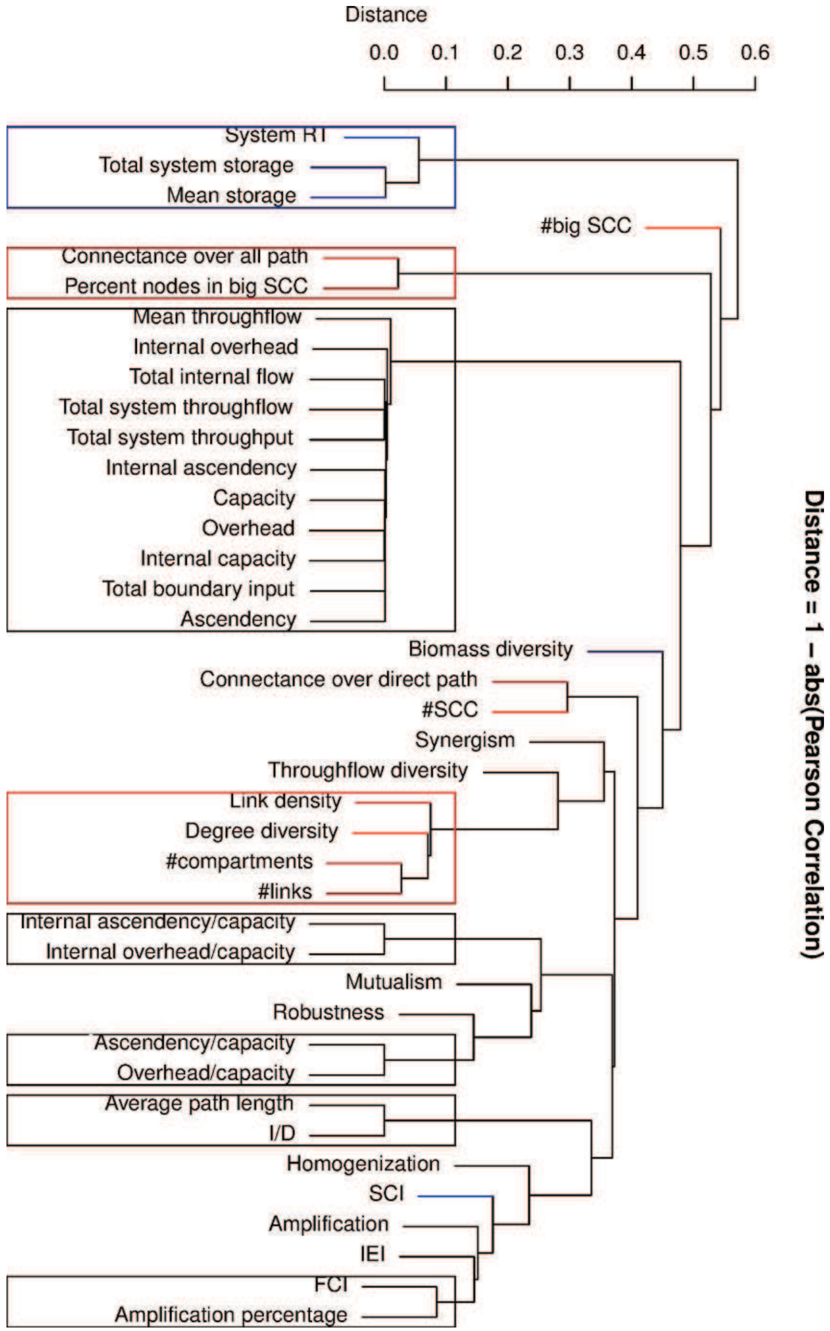

FIGURE 3.3

Cluster dendrogram of system-wide measures based on the following distance metric: 1 – abs (Pearson correlation). At a distance of 0.1, all clusters with more than one measure are bordered with rectangles.

3.5 OBSERVATIONS AND DISCUSSION

Figure 3.3 shows measures grouped in rectangles based on how similar they are (distance < 0.1). We observe that each cluster contains only a single type of measure: structure-based, flow-based, or storage-based. In other words, there is no cross correlation between different measure types. Therefore we discuss our observations of each measure type separately.

3.5.1 CLUSTERS OF STRUCTURE-BASED MEASURES

There are two clusters that contain structure-based measures. The larger one has four measures: link density, degree diversity, #compartments (network size), and #links. The high correlations among these four measures indicate that with increase of network size, link density, degree diversity, and #links increase as well. It is expected that degree diversity and #links increase with network size. However, it is not totally clear why link density increases with network size.

Link density and another related measure, connectance, have been studied extensively. There is a debate as to how they change with network size. For models in Table 3.1, Figure 3.4 shows that the link density (m/n) increases with network size, while the connectance (m/n^2) decreases slightly with network size. This indicates the total number of links (m) increases faster than the network size n, but slower than n^2.

An important observation about this four-measure cluster is that only these four structure-based measures are significantly affected by network size. Using artificial networks, Fath (2004) shows how four flow-based system-wide measures (amplification, homogenization, synergism, and indirect effects) change with network size. However, our statistical analysis based on actual ecosystem models shows that amplification ($r = -0.24$), homogenization ($r = -0.31$), synergism ($r = -0.44$), and indirect effects ($r = -0.34$), have weak relationships with network size.

3.5.2 CLUSTERS OF FLOW-BASED MEASURES

In Figure 3.3, there are totally five clusters of flow-based measures. Measures in two clusters (ascendency/capacity and overhead/capacity, internal ascendency/capacity, and internal overhead/capacity) have perfect negative linear correlations ($r = -1$), which is expected due to their formulation.

The largest cluster contains 11 measures, including mean throughflow, internal overhead, total internal flow, TST, total system throughput, internal ascendency, capacity, overhead, internal capacity, total boundary input, and ascendency. Five of these measures (mean throughflow, total internal flow, TST, total system throughput, and total boundary input) are either the sum or the mean of some flow rates in the system; therefore, their high correlation is expected.

Interestingly, capacity, ascendency, overhead, and their internal versions are also classified into this cluster. The scatterplot in Figure 3.5a clearly shows that capacity, ascendency, and overhead increase linearly with respect to total system throughput.

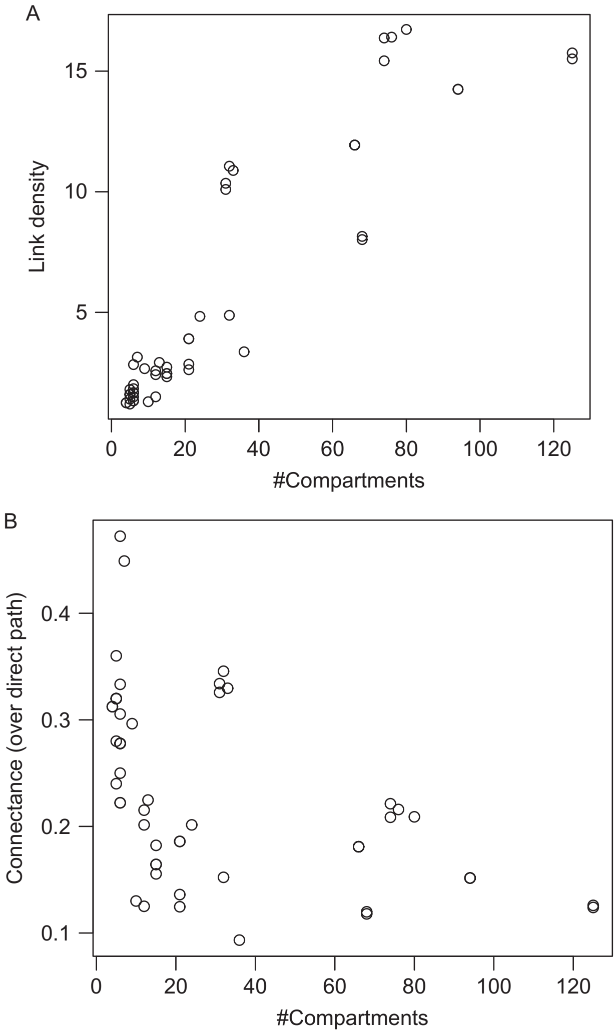

FIGURE 3.4

(a) Link density versus #compartments; (b) connectance versus #compartments.

Similarly, there exist strong positive linear relations between internal versions of capacity, ascendency, and overhead with respect to total internal flow (Figure 3.5b). Ulanowicz (2011) points out that ascendency quantifies the overall constraints of the ecosystem, or how tightly the network is organized, and overhead gauges how unconstrained the flows remain, or how flexible the system remains to reconfigure itself. Due to their high correlations with total system throughput, ascendency and overhead may not be as useful in quantifying the network organization as Ulanowicz (2011) indicates. For example, different networks with the same total

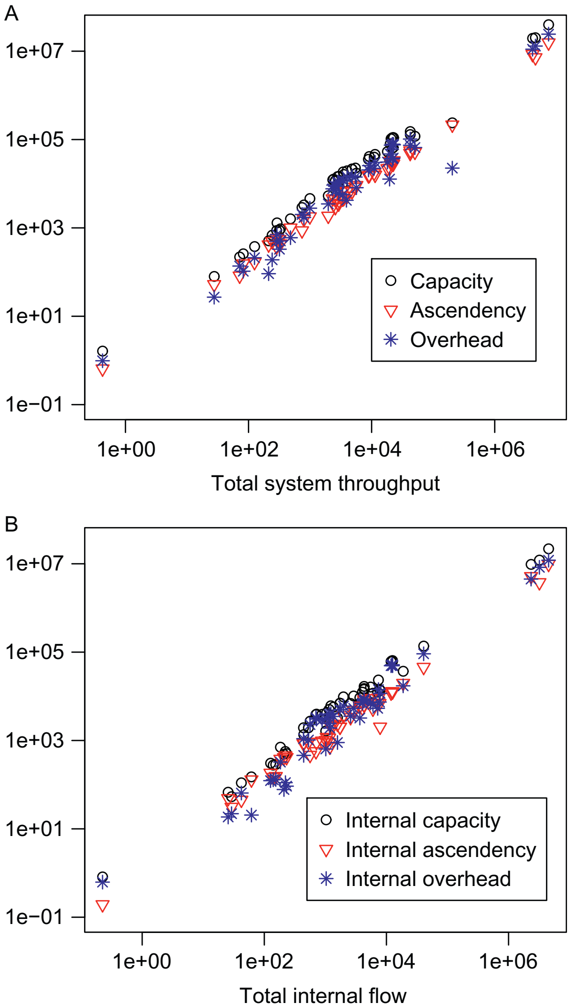

FIGURE 3.5

(a) Capacity, ascendency, and overhead versus total system throughput; (b) Internal capacity, internal ascendency, and internal overhead versus total internal flow.

system throughput may have very different network organizations, yet this difference will not be captured by ascendency or overhead.

On the other hand, according to Ulanowicz (2009, 2011), the ratio ascendency/capacity (equivalent to average mutual information/flow diversity) represents the degree of organization and overhead/capacity (equivalent to residual diversity/flow diversity) represents the degree of flexibility. Our statistical analysis shows that these two ratios truly provide new information, and are not correlated with system throughflow and any other measure.

The second cluster includes two measures: average path length and the ratio of indirect to direct effects (I/D). Higashi and Patten (1986, 1989) and Patten (1991) indicate that indirect effects increase with network size, connectance, FCI, and TST. However, according to our analysis, network size ($r = -0.20$), connectance ($r = 0.20$), and TST ($r = -0.04$) have very weak correlations with I/D ratio. Pearson correlation coefficient for FCI is $r = 0.67$. Actually, average path length ($r = 0.99$) is the measure that has the strongest correlation with I/D. We should also note that average path length ($r = -0.19$) also does not change much with network size. This is mainly because the number of trophic levels in most durable natural ecosystems is about four (Matsuno and Ono, 1996). Without considering cycling, the maximum path length for ecosystems of any size should be around four as well. The average path length should be even smaller. Thus, the increase of path length is not due to increasing network size, but due to the occurrence of cycling.

3.5.3 CLUSTERS OF STORAGE-BASED MEASURES

The only storage-based cluster contains three measures: system residence time (system RT), TSS, and mean storage. Strong linear correlation ($r = 0.99$) between TSS and mean storage is expected due to their formulations. High correlation between system residence time and TSS ($r = 0.93$) indicates that the flow material tends to stay longer in the system as system storage increases, which makes sense. The high correlation between system residence time and TSS agrees with recent findings (Schramski et al., 2015) that the residence time of carbon in both individual organisms and entire ecosystems increases with increasing system biomass.

These three storage-based measures (TSS, mean storage, and system RT) are analogous to three flow-based measures (TST, mean throughflow, and average path length). The key difference is that the former three take into account the residence time in each compartment, while the latter three do not. However, TSS ($r = -0.05$) is almost uncorrelated with TST. Mean storage and mean throughflow are also uncorrelated ($r = -0.06$). As we have pointed out earlier, there is no significant cross correlation between structure-based, flow-based, and storage-based measures. Other structure-based and flow-based measures are also not correlated with these three storage-based measures. This weak correlation indicates that storage introduces new information, and it is not feasible to estimate storage values using flow rates or network topology. Interestingly, few system-wide measures in existence utilize storage values. Therefore, future work focusing on developing novel storage-based measures may help capture new holistic properties of ecosystem models.

REFERENCES

Almunia, J., Basterretxea, G., Aistegui, J., Ulanowicz, R., 1999. Benthic-pelagic switching in a coastal subtropical lagoon. Estuar. Coast. Shelf Sci. 49 (3), 363–384.

Baird, D., Milne, H., 1981. Energy flow in the Ythan Estuary, Aberdeenshire, Scotland. Estuar. Coast. Shelf Sci. 13 (4), 455–472.

Baird, D., Ulanowicz, R., 1989. The seasonal dynamics of the Chesapeake Bay ecosystem. Ecol. Monogr. 59, 329–364.

Baird, D., McGlade, J., Ulanowicz, R., Baird, D., McGlade, J., Ulanowicz, R., 1991. The comparative ecology of six marine ecosystems. Philos. Trans. R. Soc. Lond. Ser. B Biol. Sci. 333 (1266), 15–29.

Borrett, S.R., Lau, M.K., 2014. enar: an r package for ecosystem network analysis. Methods Ecol. Evol. 5 (11), 1206–1213.

Brylinsky, M., 1972. Steady-state sensitivity analysis of energy flow in a marine ecosystem. Syst. Anal. Simul. Ecol. 2, 81–101.

Buzhdygan, O.Y., Patteny, B.C., Rudenko, S.S., 2012. Trophic network analysis: comparison of system-wide properties. Models Ecol. Hierarchy Mol. Ecosphere 25, 181.

Christian, R.R., Thomas, C.R., 2000. Neuse River Estuary Modeling and Monitoring Project Stage 1: Network Analysis for Evaluating the Consequences of Nitrogen Loading. Water Resources Research Institute of the University of North Carolina.

Chrystal, R.A., Scharler, U.M., 2014. Network analysis indices reflect extreme hydrodynamic conditions in a shallow estuarine lake (Lake St Lucia), South Africa. Ecol. Indic. 38, 130–140.

Cohen, J.E., Briand, F., 1984. Trophic links of community food webs. Proc. Natl. Acad. Sci. U.S.A. 81 (13), 4105–4109.

Dame, R.F., Patten, B.C., 1981. Analysis of energy flows in an intertidal oyster reef. Mar. Ecol. Prog. Ser. 5 (2), 115–124.

Edmisten, J., 1970. Preliminary studies of the nitrogen budget of a tropical rain forest. In: Odum, H., Pigeon, R. (Eds.), A Tropical Rain Forest. In: TID-24270, USAEC Technical Information Center, Oak Ridge, Tennessee, pp. 211–215.

Fath, B., 2004. Network analysis applied to large-scale cyber-ecosystems. Ecol. Model. 171 (4), 329–337.

Fath, B., Borrett, S., 2006. A matlab® function for network environ analysis. Environ. Model Softw. 21 (3), 375–405.

Fath, B.D., Patten, B.C., 1999. Review of the foundations of network environ analysis. Ecosystems 2, 167.

Finn, J., 1977. Flow Analysis: A Method for Tracing Flows through Ecosystem Models. Ph.D. thesis. University of Georgia, Athens, GA, USA.

Finn, J.T., 1978. Cycling index: a general definition for cycling in compartment models. In: Adriano, D., Brisbin, I. (Eds.), Environmental Chemistry and Cycling Processes. In: U.S. Dep. Energy Symp., vol. 45. National Technical Information Center, Springfield, VA, pp. 148–164.

Havens, K., 1992. Scale and structure in natural food webs. Science 257 (5073), 1107–1109.

Heymans, J., Baird, D., 2000. A carbon flow model and network analysis of the northern Benguela upwelling system, Namibia. Ecol. Model. 126 (1), 9–32.

Heymans, J., Ulanowicz, R., Bondavalli, C., 2002. Network analysis of the South Florida Everglades graminoid marshes and comparison with nearby cypress ecosystems. Ecol. Model. 149 (1), 5–23.

Higashi, M., Patten, B.C., 1986. Further aspects of the analysis of indirect effects in ecosystems. Ecol. Model. 31 (1), 69–77.

Higashi, M., Patten, B.C., 1989. Dominance of indirect causality in ecosystems. Am. Nat. 133 (2), 288.

Johnsonbe, R.A., Wichern, D.W., 2002. Applied Multivariate Statistical Analysis, vol. 5. Prentice Hall, Upper Saddle River, NJ.

Jordán, F., Jørgensen, S.E., 2012. Models of the Ecological Hierarchy: From Molecules to the Ecosphere, vol. 25. Elsevier, Amsterdam, Netherlands.

Jordan, C., Kline, J., Sasscer, D., 1972. Relative stability of mineral cycles in forest ecosystems. Am. Nat. 106, 237–253.
Jorgensen, S.E., Xu, F.-L., Costanza, R., 2005. Handbook of Ecological Indicators for Assessment of Ecosystem Health. CRC Press, Boca Raton, Florida.
Jørgensen, S.E., Burkhard, B., Müller, F., 2013. Twenty volumes of ecological indicators—an accounting short review. Ecol. Indic. 28, 4–9.
Katz, L., 1953. A new status index derived from sociometric analysis. Psychometrika 18 (1), 39–43. early centrality.
Kazanci, C., 2007. Econet: a new software for ecological modeling, simulation and network analysis. Ecol. Model. 208 (1), 3–8.
Leontief, W., 1966. Input–Output Economics. Oxford University Press, USA.
Link, J., Overholtz, W., O'Reilly, J., Green, J., Dow, D., Palka, D., Legault, C., Vitaliano, J., Guida, V., Fogarty, M., et al., 2008. The Northeast US continental shelf energy modeling and analysis exercise (EMAX): ecological network model development and basic ecosystem metrics. J. Mar. Syst. 74 (1), 453–474.
Ma, Q., Kazanci, C., 2012. Analysis of indirect effects within ecosystem models using pathway-based methodology. Ecol. Model. 252, 238–245.
Ma, Q., Kazanci, C., 2014. How much of the storage in the ecosystem is due to cycling? J. Theor. Biol. 357, 134–142.
MacArthur, R., 1955. Fluctuations of animal populations and a measure of community stability. Ecology 36 (3), 533–536.
Martinez, N.D., 1992. Constant connectance in community food webs. Am. Nat. 139, 1208–1218.
Matsuno, K., Ono, N., 1996. How many trophic levels are there? J. Theor. Biol. 180 (2), 105–109.
Miehls, A., Mason, D., Frank, K., Krause, A., Peacor, S., Taylor, W., 2009a. Invasive species impacts on ecosystem structure and function: a comparison of Oneida Lake, New York, USA, before and after zebra mussel invasion. Ecol. Model. 220 (22), 3194–3209.
Miehls, A., Mason, D., Frank, K., Krause, A., Peacor, S., Taylor, W., 2009b. Invasive species impacts on ecosystem structure and function: a comparison of the Bay of Quinte, Canada, and Oneida Lake, USA, before and after zebra mussel invasion. Ecol. Model. 220 (22), 3182–3193.
Monaco, M., Ulanowicz, R., 1997. Comparative ecosystem trophic structure of three US mid-Atlantic estuaries. Mar. Ecol. Prog. Ser. 161, 239–254.
Newman, M., 2009. Networks: An Introduction. Oxford University Press, Oxford.
Odum, H.T., 1957. Trophic structure and productivity of silver springs, florida. Ecol. Monogr. 27, 55–112.
Patrício, J., Ulanowicz, R., Pardal, M.A., Marques, J.C., 2004. Ascendency as an ecological indicator: a case study of estuarine pulse eutrophication. Estuar. Coast. Shelf Sci. 60 (1), 23–35.
Patten, B.C., 1978. Systems approach to the concept of environment. Ohio J. Sci. 78 (4), 206–222.
Patten, B.C., 1985. Energy cycling, length of food chains, and direct versus indirect effects in ecosystems. Ecosystem theory for biological oceanography. Can. Bull. Fish. Aquat. Sci. 213, 119–138.
Patten, B., 1991. Network ecology: indirect determination of the life-environment relationship in ecosystems. In: Burns, T.P., Higashi, M (Eds.), Theoretical Studies of Ecosystems. Cambridge University Press, Cambridge, pp. 288–351.

Richey, J., Wissmar, R., Devol, A., Likens, G., Eaton, J., Wetzel, R., Odum, W., Johnson, N., Loucks, O., Prentki, R., et al., 1978. Carbon flow in four lake ecosystems: a structural approach. Science 202 (4373), 1183–1186.

Rutledge, R.W., Basore, B.L., Mulholland, R.J., 1976. Ecological stability: an information theory viewpoint. J. Theor. Biol. 57 (2), 355–371.

Rybarczyk, A., Nowakowski, B., 2003. A neutral network–hardware implementation using fpga. Found. Comput. Decis. Sci. 28 (1), 29–40.

Sabidussi, G., 1966. The centrality index of a graph. Psychometrika 31 (4), 581–603. early centrality.

Sandberg, J., Elmgren, R., Wulff, F., 2000. Carbon flows in Baltic Sea food webs—a re-evaluation using a mass balance approach. J. Mar. Syst. 25 (3), 249–260.

Schaubroeck, T., Staelens, J., Verheyen, K., Muys, B., Dewulf, J., 2012. Improved ecological network analysis for environmental sustainability assessment; a case study on a forest ecosystem. Ecol. Model. 247, 144–156.

Schramski, J., Kazanci, C., Tollner, E., 2011. Network environ theory, simulation, and EcoNet 2.0. Environ. Model Softw. 26 (4), 419–428.

Schramski, J.R., Dell, A.I., Grady, J.M., Sibly, R.M., Brown, J.H., 2015. Metabolic theory predicts whole-ecosystem properties. Proc. Natl. Acad. Sci. U.S.A. 112 (8), 2617–2622.

Steele, J., 1974. The Structure of Marine Ecosystems. Harvard University Press, Cambridge, Massachusetts.

Tomczak, M.T., Heymans, J.J., Yletyinen, J., Niiranen, S., Otto, S.A., Blenckner, T., 2013. Ecological network indicators of ecosystem status and change in the Baltic Sea. PLoS One 8 (10), e75439.

Ulanowicz, R., 1986. Growth and Development: Ecosystems Phenomenology. Springer, New York.

Ulanowicz, R.E., 2004. Quantitative methods for ecological network analysis. Comput. Biol. Chem. 28, 321–339.

Ulanowicz, R.E., 2009. The dual nature of ecosystem dynamics. Ecol. Model. 220 (16), 1886–1892.

Ulanowicz, R., 2011. Seeking sustainability with both eyes open. http://www.cbl.umces.edu/ulan/pubs/EyesOpen.pdf.

Ulanowicz, R., Bondavalli, C., Egnotovich, M., 1997. Network analysis of trophic dynamics in South Florida ecosystem, FY 96: the cypress wetland ecosystem. Annual report to the United States Geological Service Biological Resources Division. Ref. No. [UMCES] CBL 97-075. Chesapeake Biological Laboratory, University of Maryland.

Ulanowicz, R.E., Bondavalli, C., Egnotovich, M.S., 1998. Network analysis of trophic dynamics in South Florida ecosystem, FY 97: the Florida Bay ecosystem. In: Annual report to the United States Geological Service Biological Resources Division. Ref. No. [UMCES] CBL 98–123. Chesapeake Biological Laboratory, Maryland, USA.

Ulanowicz, R.E., Bondavalli, C., Heymans, J.J., Egnotovich, M.S., 1999. Network analysis of trophic dynamics in South Florida ecosystem, FY 98: the mangrove ecosystem. In: Annual report to the United States Geological Service Biological Resources Division. Ref. No. [UMCES] CBL 99–0073. Chesapeake Biological Laboratory, Maryland, USA.

Ulanowicz, R.E., Bondavalli, C., Heymans, J.J., Egnotovich, M.S., 2000. Network analysis of trophic dynamics in South Florida ecosystem, FY 99: the graminoid ecosystem. In: Annual report to the United States Geological Service Biological Resources Division. Ref. No. [UMCES] CBL 00–017. Chesapeake Biological Laboratory, Maryland, USA.

Ulanowicz, R.E., Jørgensen, S.E., Fath, B.D., 2006. Exergy, information and aggradation: an ecosystems reconciliation. Ecol. Model. 198 (3), 520–524.

Vermaat, J.E., Dunne, J.A., Gilbert, A.J., 2009. Major dimensions in food-web structure properties. Ecology 90 (1), 278–282.

Webster, J.R., Waide, J.B., Patten, B.C., 1975. Nutrient cycling and the stability of ecosystems. In: Howell, F.G., Gentry, J.B., Smith, M.H. (Eds.), Mineral Cycling in Southeastern Ecosystems. ERDA CONF-740513. pp. 1–27. Reprinted as pages 136–162. In: Shugart, H.H., O'Neill, R.V. (Eds.), Systems Ecology. Benchmark Papers in Ecology, Dowden, Hutchinson, and Ross, Inc, Stroudsburg, PA.

Wulff, F., Ulanowicz, R., 1989. A comparative anatomy of the Baltic Sea and Chesapeake Bay, ecosystems. In: Wulff, F., Field, J.G., Mann, K.H. (Eds.), Network Analysis in Marine Ecology. Springer-Verlag, Berlin, pp. 232–256.

Yodzis, P., 1980. The connectance of real ecosystems. Nature 284, 544–545.