Title: `enaR`: An R package for Ecosystem Network Analysis

Running Title: R ecosystem network analysis package

Word Count: 3146

Authors: Stuart R. Borrett, Matthew K. Lau

Addresses:

- SRB: Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC, 28403
- MKL: Department of Biological Sciences and the Merriam-Powell Center for Environmental Research, Northern Arizona University, 617 S. Beaver St., Flagstaff, AZ, 86011

Contact Details:

- Email: borretts@uncw.edu
- Phone: 910.962.2411
- Fax: 910.962.4066

enaR: An R package for Ecosystem Network Analysis

Stuart R. Borrett[a,*] and Matthew K. Lau[b]

[a] Department of Biology and Marine Biology, University of North Carolina Wilmington, Wilmington, NC

[b] Department of Biological Sciences and the Merriam-Powell Center for Environmental Research, Northern Arizona University, 617 S

[*] Corresponding author, borretts@uncw.edu

April 10, 2014

Abstract

- Network analysis is a useful approach for complex, relational datasets in many biological fields, including ecology and molecular and evolutionary biology.
- Here, we introduce `enaR`, an R package for conducting Ecosystem Network Analysis (ENA), an analytical tool set rooted in ecosystem ecology with over 30 years of development, which examines the structure and dynamics of matter and energy movement between discrete ecological compartments, such as food webs.
- In addition to describing the primary functionality of the package, we also highlight several value added features, including a library of 100 empirical ecosystem models, the ability to analyze and compare multiple models simultaneously, and connections to useful ecological network analysis tools in R.

KEYWORDS: network analysis, ecosystem, social network analysis, software, network environ analysis, ascendency, input–output analysis, food web, Ecopath, WAND

1 Introduction

Network Ecology – the study of ecological systems using network models and analyses to characterize their structure, function, and evolution – is a large and rapidly growing area of ecology (Proulx *et al.*, 2005). For example, Ings *et al.* (2009) discovered that a notable fraction of 2008 publications in 11 select journals were related to food webs (≈2.4%), mutualistic networks (≈0.9%), and host-parasitoid networks (≈0.06%). Likewise, Borrett *et al.* (2014) found that the percent of ecology and evolutionary biology papers indexed by Web of Science that could be classified as Network Ecology increased from 0.4% in 1990 to more than 5% in 2012. This rise of Network Ecology contributes to, mirrors, and builds on the more general growth of network sciences (Barabási, 2012; Borgatti & Foster, 2003; Freeman, 2004; Newman, 2003; Wasserman & Faust, 1994). However, it seems likely that the field is growing in part because the network approach is useful.

Ecosystem Network Analysis (ENA) is a branch of Network Ecology that has been used to address a range of key ecosystem questions (Borrett *et al.*, 2012; Fath & Patten, 1999; Ulanowicz, 1986). For example, in the food web of Big Cypress National Preserve (Florida, USA) Bondavalli & Ulanowicz (1999) found evidence of an indirect mutualism between the American alligator and some of its prey items, including several invertebrates, frogs, and small mammals. Applications of ENA have also lead to new insights into the classic trophic questions of "What limits food-chain length?" (Ulanowicz *et al.*, 2014) and "Are food webs modular?" (Allesina *et al.*, 2005; Borrett *et al.*, 2007; Krause, 2004). Hines *et al.* (2012) used ENA to quantify the relative importance of coupling between biogeochemical processes (e.g., nitrification + anammox) in the Cape Fear River estuary sediementary nitrogen cycle. Further, scientists have used ENA to investigate differences in urban sustainability (Bodini & Bondavalli, 2002; Bodini *et al.*, 2012; Chen & Chen, 2012; Zhang *et al.*, 2010). Collectively, this work consistently shows the power of a transactional network to generate unexpected ecological relationships that then influence the system function and evolution

(Jørgensen *et al.*, 2007; Patten, 1991; Ulanowicz, 1997).

`enaR` is an open-source software to facillitate ENA. Extant ENA software (Allesina & Bondavalli, 2004; Christensen & Walters, 2004; Fath & Borrett, 2006; Kazanci, 2007; Ulanowicz & Kay, 1991) each have crtical limitations, which led us to three primary design objectives for `enaR`. The first objective was to collect the major ENA functions into a single software package. While multiple investigators have contributed to algorithmic development (e.g., Allesina & Bondavalli, 2003; Fath & Patten, 1999; Finn, 1976; Ulanowicz, 1986; Ulanowicz & Kay, 1991), the broad set of tools is not available in a single existing software. The second objective was to increase the availability and extensibility of the software. We chose to use R in part because of its increasing popularity as an analytical tool in the biological sciences (e.g., Dixon, 2003; Metcalf *et al.*, 2012; Revell, 2012). Further, users can freely download a stable version of the package from the CRAN website (`http://cran.r-project.org/web/packages/enaR/`), and the code for every funciton in R is available from within R (e.g., `edit`(x)). In addition, `enaR` development is being managed via GitHub (`https://github.com/TheSeeLab/enaR`) to encourage collaborative development. The third design objective was to enable `enaR` users access to other network analysis tools from other disciplines. To enable this, `enaR` was designed to work directly with two existing R network analysis packages: `network` (Butts, 2008a) and `sna` (Butts, 2008b). In summary, the aim of the `enaR` package is to make ENA tools more available and easier to use, adapt, and extend.

In this paper, we present an overview of `enaR` and highlight some of its functionality. A full description of the ENA algorithms and their use and interpretaiton is beyond the scope of this short paper, but we refer interested readers to a selection of reviews as an entry point to ENA (Fath & Borrett, 2006; Fath & Patten, 1999; Jørgensen *et al.*, 2007; Schramski *et al.*, 2011; Ulanowicz, 1997). For a more comprehensive description on how to use the `enaR` package, please refer to the package vignette: `http://cran.r-project.org/web/packages/enaR/vignettes/enaR.pdf`.

2 Overview of `enaR`

ENA is an agglomeration of algorithms developed to analyze network models of energy or matter movement in ecosystems (e.g., Fath & Patten, 1999; Hannon, 1973; Ulanowicz, 1986), but it can generally be appled to any Input-Output system that follows a thermodynamicaly conserved unit among the compartments. Thus, it is a family of related algorithms to analyze the ecosystem from several perspectives including its structure, flow, storage, and utility. Together, these analyses functions as a "macroscope" to investigate (1) whole system organization, (2) the direct and indirect effects among system components, and (3) the processes that create and sustain ecological systems. In this section we provide an overview of these algorithms and tools include in the `enaR` software. After describing the required model information, we highlight the primary ENA algorithms included in `enaR` . We then walk through an example application of the `enaR` Flow analysis.

2.1 Data Requirements and Input

ENA is a data-intensive methodology. The system is modeled as a set of compartments or network nodes that represent species, species-complexes (i.e., trophic guilds or functional groups), or non-living components of the system in which energy or matter is stored. These nodes are connected by a set of direct energy or matter transactions among the nodes, termed directed edges or links. These models also have energy–matter inputs into the system and output losses from the system. In summary, the full set of data required includes: (1) internal flows, (2) boundary inputs, (3) boundary exports, (4) boundary respiration, (5) boundary outputs, which may be the sum of exports and respiration, (6) biomass or storage values, and (7) designation of living status of each node. While all seven elements are required for a full analysis, the specific data requirements varies among the ENA algorithms.

The primary ENA algorithms in `enaR` assume the model data is presented as an R network data object defined in the `network` package. Given the data elements, users can use the `pack` function to combine the data elements into the R network data object. While a standard data format for an ENA model does not yet exist, there are two commonly used formats. First, there is the Scientific Committee for Ocean Research (SCOR) format that is the required input to NETWRK (Ulanowicz & Kay, 1991), and the second format is the Excel sheet formatted data that is the input to WAND (Allesina & Bondavalli, 2004). The `enaR` package includes a `read.scor` and a `read.wand` function to read in these common data formats (Table 1).

2.2 Visualization

Visualization of network models can be an essential analytical tool (Lima, 2011; Moody *et al.*, 2005). Because `enaR` is built on top of the `network` package and data type, it is possible to quickly create network plots of the model internal structure. Fig. 1a shows an example visualization of Dame & Patten's (1981) Oyster Reef ecosystem model. The `network` package includes three network layout algorithms: circle, Fruchterman-Reingold, and Kamada-Kawai. The Fruchterman-Reingold algorithm used here is the default. The R script to generate this visulization is included in the online supplimentary information.

2.3 Algorithm Overview

`enaR` includes many of the most commonly used ENA algorithms (Table 2), along with a number of work flow and specialty analyses (Tables 1 and 3). Note that the XX primary ENA functions begin with the prefix 'ena' followed by the specific analysis name (see Table 2). There are a total of YY funcitons in the `enaR` package.

Scharler & Fath (2009) identify two schools of ENA. The first school is based on the work of Robert Ulanowicz and colleagues at the University of Maryland (Ulanowicz, 1986, 1997, 2009).

Primarily focused on trophic ecology, this approach uses information theory and the ascendency concept that characterizes ecosystem growth and development Ulanowicz (1986, 1997). This work is often refered to as "Ecological Network Analysis" as it predates many other types of network ecology. The second school is based on the work of Bernard Patten at the University of Georgia (Fath & Patten, 1999; Matis & Patten, 1981; Patten, 1982; Patten *et al.*, 1976). Steeped in dynamic equations, simulations, and systems analysis, this approach developed around the environ concept that formalizes the concept of environment (Patten, 1978), and has often been referred to as "Network Environ Analysis." `enaR` currently captures all of the Patten School algorithms previously implemented in NEA.m (Fath & Borrett, 2006), along with some recent developments. Ulanowicz School algorithms are more limited, including the ascendency calculations (Ulanowicz, 1997) and mixed trophic impacts analyses (Ulanowicz & Puccia, 1990). We expect the package capabilities to continue to grow, especially with the assistance of new users.

2.4 Example Application

Given a network model, applying ENA algorithms with `enaR` is straight forward. We demonstrate how to use the package with an example Flow analysis on Dame & Patten's (1981) model of energy flow in an Oyster Reef ecosystem. Table 4 shows the example script. The analysis invovles: (1) loading the model data, (2) checking and balancing the model if necessary, and (3) inputing the balanced model into the analysis function. The final step is interpreting the analytical output. This is a typical workflow for ENA.

After loading the `enaR` package, the next step is to enter the model data. Here, we have extracted the model information from the paper and created a vector of node names, the flow matrix (F), inputs (z), outputs (y), and the logical vector indicating whether or not the nodes are living (Table 4). We then use the `pack` function to create the requried network data object.

The next step is to apply the `ssCheck` function ensure that the model is at steady-state, which is one of the assumptions of the flow analysis (Fath & Borrett, 2006; Finn, 1976). If the model had not been at steady-state, we could have then appled one of four automated balancing algorithms Allesina & Bondavalli (AVG, Input-Output, Output-Input, AVG2, 2003) to force the model into a steady-state. We then apply the `enaFlow` function to the model to perform the desired ENA flow analysis. As shown with the `attributes` function, this analysis returns 4 matrices (**G**, **GP**, **N**, **NP**) and two vectors (throughflow, T, and a vector of 20 whole-network statistics, ns).

Interpreting the ENA results is the final challenge. Here, we provide a few illustrative interpretations of the Flow analysis. Starting with the whole-network flow statistics, we see that the total system throughflow (TST) of the oyster reef model is 83.6 Kcal m^{-2} d^{-1}. TST is a measure of the total activity of the system, which is often referred to as the size or power of the system. The Finn Cycling Index (FCI) indicates that 11% of this activity was generated by recycling. Further, the average path length (APL = 2.02) shows that an average input passes over two paths before exiting the system, and the ratio of indirect to direct flows (ID.F = 1.58) indicates that the indirect flow exceeds the direct flow in this system. Together, these whole network indicators show the importance of indirect interactions in the system. A next analytical step might be to apply the Utility or Mixed Trophic Impacts analyses to determine the net relationships among the ecosysetm components when we consider the direct and indirect interactions. More detailed guidance for how to interpret ENA results can be found in previously published literature (Fath & Borrett, 2006; Jørgensen *et al.*, 2007; Schramski *et al.*, 2011).

3 Value Added Features

There are several features of the `enaR` packagw beyond the core analyses that add substantive value for users. We highlight three of these featrues here.

3.1 Model Library

To facilitate new systems ecology and network science, we included a library of 100 previously published ecosystem network models with the `enaR` package. These models each trace a thermodynamically conserved unit (e.g., C, N, P) through a particular ecosystem. The models in this set are empirically-based in that the authors attempted to model a specific system and parameterized the model to some degree with empirical estimates. The library includes models used previously to test several systems ecology hypotheses (Borrett, 2013; Borrett & Salas, 2010; Borrett *et al.*, 2010; Salas & Borrett, 2011). This set has a 47% overlap with the set of models previously collected by Dr. Ulanowicz (`http://www.cbl.umces.edu/~ulan/ntwk/network.html`).

We have tentatively split these models into two classes. The most abundant class is the trophic network models. These models tend to have a food web at their core, but also include non-trophic fluxes generated by processes like death and excretion. The annual carbon flux model for the mesohaline region of the Chesapeake Bay is a typical example (Baird & Ulanowicz, 199). The second class of models focuses on biogeochemical cycling. In contrast to the trophic networks, the biogeochemical cycling models tend to have more highly aggregated nodes (more species grouped into a compartment), include more abiotic nodes that could represent chemical species (e.g., ammonia in a nitrogen cycle), have a lower dissipation rate, and therefore they tend to have more recycling (Borrett *et al.*, 2010; Christian *et al.*, 1996). Christian & Thomas's (2003) models of nitrogen cycling in the Neuse River Estuary are good examples of the class. The package vignette has a full listing of the models included along with references to their original publications (Lau *et al.*, 2013).

3.2 Batch Analysis

Advances in ecosystem ecology have been made by comparing network metrics across multiple ecosystem models. For example, Christensen (1995) applied ENA to identify and compare the maturity of 41 ecosystem models, and van Oevelen *et al.* (2011) compared the the organic matter processing of food webs in three sections of the Nazaré submarine canyon. The `enaR` tool simplifies the work flow for these types of comparison. Given a list of models like the model library, it is possible to quickly analyze multiple models using R's `lapply` function (see `help`("lapply")). This facilitates the kind of comparative network analysis often of interest to ecologists (Christian *et al.*, 2005; Monaco & Ulanowicz, 1997; Whipple *et al.*, 2007).

Batch analysis can be used in several additional ways. One application is for meta-analyses, such as tests of the generality of hypothesized ecosystem properties like network non-locality (Salas & Borrett, 2011), or to investigate how physical features might influence ENA results (Niquil *et al.*, 2012). Fig. 1b illustrates the rank-ordered network homogenization statistic for the 56 trophic-based ecosystem models in the library. Notice that the homogenization statistic is greater than one in all of these models indicating that the network of indirect interactions tend to more uniformly distribute the resources than is obvious from the direct interactions, which extends previous results of Borrett & Salas (2010) to include several new models. A second kind of application is the exploration of new ENA inter-relationships. With the collection of algorithms and the library of models, we can now investigate possible relationships among ENA indicators from different schools (Fig. 1c). A third application of batch analysis is to investigate the previously unknown empirical ranges of ENA whole-network statistics, which may be useful for interpreting results from specific applications. Fig. 2 shows the observed distribution of values for selected network statistics from the 100 models in the library easily analyzed using `lapply` and the associated `enaR` functions.

3.3 New Connections

A fourth added benifit of the `enaR` package design is that it enables network ecologists easier access to other network tools and analyses that might be useful. The `enaR` package uses the R network data structure defined in the `network` package (Butts, 2008a). This means that network ecologists using `enaR` can also use the network manipulation functions and visualization features of the `network` package. Further, the R Social Network Analysis (SNA) package, `sna`, (Butts, 2008b) also uses this network data object. This means that network ecologists can apply many of the SNA algorithms directly to their ecological network models. Fig. 1d illustrates applying the betweenness centrality function to the Chesapeake Bay trophic model (Baird & Ulanowicz, 199) and visualizing the results using a target centrality plot (Brandes *et al.*, 2003). This analysis highlights the central role of Sedimentary Particulate Carbon and bacteria in the Sediment Particulate Organic Carbon (POC) in the carbon flux of the estuary.

In addition, `enaR` can be a starting point for ecosystem network ecologists to use other R network tools. For example, the `iGraph` package provides functions to apply classic graph theory (Csardi & Nepusz, 2006). The `limSolve` package provides capabilities to infer network model fluxes from empirical data by linear inverse modeling (Soetaert *et al.*, 2009), which can also be used for uncertainty analyses of ENA (Kones *et al.*, 2009). There are a wealth of additional R package that network ecologists may find useful including `bipartite` (Dormann *et al.*, 2008), `vegan` (Dixon, 2003), `bioconductor` (Gentleman *et al.*, 2004), `Cheddar` (Hudson *et al.*, 2013), and packages in the `statnet` family (Handcock *et al.*, 2008).

4 Conclusion and Future Development

The `enaR` package is designed to address limitations of existing ENA software and facillitate wide development and use. It does this by (1) providing greater accessibility to the code (e.g., free and open source software avialable on multiple OS), (2) collecting a broad set of available ENA algorithms and workflow management functions, and (3) creating the potential for collaborative development (via GitHub and CRAN). Further, the software is extensible for individual needs and it lets users integrate ENA into a broader workflow in R in a way that is not possible in web based tools like EcoNet (Kazanci, 2007; Schramski *et al.*, 2011). Finally, it lets users have access to other network and statistical analysis tools (e.g., social network analysis) that are already part of R. These benefits come at the cost of having a steeper learning curve (e.g., users must know R), which may make `enaR` more suited to advanced practicioners.

In the near future, we anticipate two initial lines of continued development for the `enaR` package. The first is to increase the connections between the `enaR` package and other modeling and analytical tools. For example, we are currently working with colleagues to enable users of Ecopath with Ecosim (Christensen & Walters, 2004) to apply the `enaR` tools in a seamless way. We are also developing functions to connect between `enaR` and the R limSolve package (Soetaert *et al.*, 2009) for creating models using Linear Inverse Modeling and to enable uncertainty analysis (Kones *et al.*, 2009). The second line of development is to extend the package's capabilities. While it currently contains most of the many commonly used ENA algorithms used by ecologists, it is far from complete. For example, Ulanowicz's (1983) decomposition of cycles is not yet included nor is his construction for the Lindeman trophic spine (Ulanowicz & Kemp, 1979). The package could also include network model construction tools, such as least-inference methods for building models from empirical data (Ulanowicz & Scharler, 2008) and Fath's (2004) algorithm for constructing plausible ecosystems models.

In conclusion, `enaR` is an R pacakge to facilitate the use and the collaborative development of Ecosystem Network Analysis. This analysis is a branch of the broader domain of Network Ecology, which is rapidly growing in part becuase the tools and techniques let ecologists address a wide range of relational questions that lie at the core of ecology. We look forward to seeing the new ecology discovered through the use of `enaR`.

5 Acknowledgments

We would like to acknowledge and thank David Hines for contributing to the initial code. We also thank several individuals who used the earlier versions of the software and provided helpful feedback for further development including Ursula Scharler, Shaoqing Chen, Emily Oxe, and John Mejaski. In addition, we thank the many ecosystem model authors who created, shared, and published their work. This work was funded in part by the US National Science Foundation (DEB1020944, DEB0425908), an NSF Integrative Graduate Education and Research Traineeship (MKL; DGE0549505) and a UNCW Cahill award (SRB).

References

Allesina, S., Bodini, A. & Bondavalli, C. (2005) Ecological subsystems via graph theory: the role of strongly connected components. *Oikos*, **110**, 164–176.

Allesina, S. & Bondavalli, C. (2003) Steady state of ecosystem flow networks: A comparison between balancing procedures. *Ecol Model*, **165**, 221–229.

Allesina, S. & Bondavalli, C. (2004) Wand: An ecological network analysis user-friendly tool. *Environ Model Softw*, **19**, 337–340.

Baird, D. & Ulanowicz, R.E. (199) The seasonal dynamics of the Chesapeake Bay ecosystem. *Ecol Monogr*, **59**, 329–364.

Barabási, A.L. (2012) The network takeover. *Nature Physics*, **8**, 14–16.

Bodini, A. & Bondavalli, C. (2002) Towards a sustainable use of water resources: a whole-ecosystem approach using network analysis. *Int J Environmental Pollution*, **18**, 463–485.

Bodini, A., Bondavalli, C. & Allesina, S. (2012) Cities as ecosystems: Growth, development and implications for sustainability. *Ecol Model*, **245**, 185–198.

Bondavalli, C. & Ulanowicz, R.E. (1999) Unexpected effects of predators upon their prey: The case of the American alligator. *Ecosystems*, **2**, 49–63.

Borgatti, S.P. & Foster, P.C. (2003) The network paradigm in organizational research: A review and typology. *J Manage*, **29**, 991–1013.

Borrett, S.R. (2013) Throughflow centrality is a global indicator of the functional importance of species in ecosystems. *Ecol Indic*, **32**, 182–196.

Borrett, S.R., Christian, R.R. & Ulanowicz, R.E. (2012) Network ecology. A.H. El-Shaarawi & W.W. Piegorsch, eds., *Encyclopedia of Environmetrics*, pp. 1767–1772. John Wiley & Sons, 2nd edition.

Borrett, S.R., Fath, B.D. & Patten, B.C. (2007) Functional integration of ecological networks through pathway proliferation. *J Theor Biol*, **245**, 98–111.

Borrett, S.R., Moody, J. & Edelmann, A. (2014) The rise of network ecology: maps of the topic diversity and scientific collaboration. *Ecol Model*, **in press**.

Borrett, S.R. & Salas, A.K. (2010) Evidence for resource homogenization in 50 trophic ecosystem networks. *Ecol Model*, **221**, 1710–1716.

Borrett, S.R., Whipple, S.J. & Patten, B.C. (2010) Rapid development of indirect effects in ecological networks. *Oikos*, **119**, 1136–1148.

Brandes, U., Kenis, P. & Wagner, D. (2003) Communicating centrality in policy network drawings. *IEEE Transactions on Visualization and Computer Graphics*, **9**, 241–253.

Butts, C. (2008a) network: A package for managing relational data in R. *J Stat Softw*, **24**.

Butts, C. (2008b) Social network analysis with sna. *J Stat Softw*, **24**, 1–51.

Chen, S. & Chen, B. (2012) Network environ perspective for urban metabolism and carbon emissions: A case study of Vienna, Austria. *Environ Sci Tech*, **46**, 4498–4506.

Christensen, V. (1995) Ecosystem maturity—towards quantification. *Ecol Model*, **77**, 3–32.

Christensen, V. & Walters, C.J. (2004) Ecopath with Ecosim: Methods, capabilities and limitations. *Ecol Model*, **172**, 109–139.

Christian, R.R., Baird, D., Luczkovich, J., Johnson, J.C., Scharler, U.M. & Ulanowicz, R.E. (2005) Role of network analysis in comparative ecosystem ecology of estuaries. A. Belgrano, J. Scharler U. M. Dunne & R. Ulanowicz, eds., *Aquatic Food Webs: An Ecosystem Approach*, pp. 25–40. Oxford University Press, New York, NY.

Christian, R.R., Fores, E., Comin, F., Viaroli, P., Naldi, M. & Ferrari, I. (1996) Nitrogen cycling networks of coastal ecosystems: influence of trophic status and primary producer form. *Ecol Model*, **87**, 111–129.

Christian, R.R. & Thomas, C.R. (2003) Network analysis of nitrogen inputs and cycling in the Neuse River Estuary, North Carolina, USA. *Estuaries*, **26**, 815–828.

Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal*, **Complex Systems**, 1695.

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

Dixon, P. (2003) VEGAN, a package of R functions for community ecology. *Journal of Vegetation Science*, **14**, 927–930.

Dormann, C.F., Gruber, B. & Fründ, J. (2008) Introducing the bipartite package: analysing ecological networks. *R News*, **8**, 8–11.

Fann, S.L. & Borrett, S.R. (2012) Environ centrality reveals the tendency of indirect effects to homogenize the functional importance of species in ecosystems. *J Theor Biol*, **294**, 74–86.

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

Fath, B.D. & Borrett, S.R. (2006) A Matlab© function for network environ analysis. *Environ Model Softw*, **21**, 375–405.

Fath, B.D. & Patten, B.C. (1999) Review of the foundations of network environ analysis. *Ecosystems*, **2**, 167–179.

Finn, J.T. (1976) Measures of ecosystem structure and function derived from analysis of flows. *J Theor Biol*, **56**, 363–380.

Freeman, L.C. (2004) *The development of social network analysis: A study in the sociology of science*. Empirical Press Vancouver.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J. *et al.* (2004) Bioconductor: open software development for computational biology and bioinformatics. *Genome biology*, **5**, R80.

Handcock, M., Hunter, D., Butts, C., Goodreau, S. & Morris, M. (2008) statnet: Software tools for the representation, visualization, analysis and simulation of network data. *J Stat Softw*, **24**, 1548.

Hannon, B. (1973) The structure of ecosystems. *J Theor Biol*, **41**, 535–546.

Hines, D.E., Lisa, J.A., Song, B., Tobias, C.R. & Borrett, S.R. (2012) A network model shows the importance of coupled processes in the microbial N cycle in the Cape Fear River estuary. *Estuar Coast Shelf Sci*, **106**, 45–57.

Hudson, L.N., Emerson, R., Jenkins, G.B., Layer, K., Ledger, M.E., Pichler, D.E., Thompson, M.S.A., O'Gorman, E.J., Woodward, G. & Reuman, D.C. (2013) Cheddar: analysis and visualisation of ecological communities in R. *Methods Ecol Evol*, **4**, 99–104.

Ings, T.C., Montoya, J.M., Bascompte, J., Blüthgen, N., Brown, L., Dormann, C.F., Edwards, F., Figueroa, D., Jacob, U., Jones, J.I., Lauridsen, R.B., Ledger, M.E., Lewis, H.M., Olesen, J.M., van Veen, F.J.F. & Warren, P. H. nad Woodward, G. (2009) Review: Ecological networks–beyond food webs. *J Anim Ecol*, **78**, 253–269.

Jørgensen, S.E., Fath, B.D., Bastianoni, S., Marques, J.C., Müller, F., Nielsen, S., Patten, B.C., Tiezzi, E. & Ulanowicz, R.E. (2007) *A new ecology: Systems perspective*. Elsevier, Amsterdam.

Kazanci, C. (2007) EcoNet: A new software for ecological modeling, simulation and network analysis. *Ecol Model*, **208**, 3–8.

Kones, J.K., Soetaert, K., van Oevelen, D. & Owino, J.O. (2009) Are network indices robust indicators of food web functioning? a Monte Carlo approach. *Ecol Model*, **220**, 370–382.

Krause, A. (2004) *The role of compartments in food-web structure and changes following biological invasions in southeast Lake Michigan*. Ph.d., Michigan State University.

Lau, M.K., Borrett, S.R. & Hines, D.E. (2013) *enaR: Tools for ecological network analysis in R.* R package version 2.5.

Lima, M. (2011) *Visual complexity: mapping patterns of information.* Princeton Architectural Press.

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

Metcalf, C.J.E., McMahon, S.M., Salguero-Gómez, R. & Jongejans, E. (2012) IPMpack: an R package for integral projection models. *Methods Ecol Evol*, **4**, 195–200.

Monaco, M.E. & Ulanowicz, R.E. (1997) Comparative ecosystem trophic structure of three us mid-Atlantic estuaries. *Mar Ecol Prog Ser*, **161**, 239–254.

Moody, J., McFarland, D. & Bender-deMoll, S. (2005) Dynamic network visualization. *American Journal of Sociology*, **110**, 1206–1241.

Newman, M. (2003) The structure and function of complex networks. *SIAM review*, **45**, 167–256.

Niquil, N., Chaumillon, E., Johnson, G., Bertin, X., Grami, B., David, V., Bacher, C., Asmus, H., Baird, D. & Asmus, R. (2012) The effect of physical drivers on ecosystem indices derived from ecological network analysis: Comparison across estuarine ecosystems. *Estuar Coast Shelf Sci*, **108**, 132–143.

Patten, B.C. (1978) Systems approach to the concept of environment. *Ohio J Sci*, **78**, 206–222.

Patten, B.C. (1982) Environs: Relativistic elementary particles for ecology. *Am Nat*, **119**, 179–219.

Patten, B.C. (1991) Network ecology: Indirect determination of the life–environment relationship in ecosystems. M. Higashi & T. Burns, eds., *Theoretical Studies of Ecosystems: The Network Perspective*, pp. 288–351. Cambridge University Press, New York.

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

Proulx, S.R., Promislow, D.E.L. & Phillips, P.C. (2005) Network thinking in ecology and evolution. *Trends Ecol Evol*, **20**, 345–353.

Revell, L.J. (2012) phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*, **3**, 217–223.

Salas, A.K. & Borrett, S.R. (2011) Evidence for dominance of indirect effects in 50 trophic ecosystem networks. *Ecol Model*, **222**, 1192–1204.

Scharler, U. & Fath, B. (2009) Comparing network analysis methodologies for consumer–resource relations at species and ecosystems scales. *Ecol Model*, **220**, 3210–3218.

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

Soetaert, K., Van den Meersche, K. & van Oevelen, D. (2009) *limSolve: Solving Linear Inverse Models.* R package version 1.5.1.

Ulanowicz, R.E. (1983) Identifying the structure of cycling in ecosystems. *Math Biosci*, **65**, 219–237.

Ulanowicz, R.E. (1986) *Growth and Development: Ecosystems Phenomenology.* Springer–Verlag, New York.

Ulanowicz, R.E. (1997) *Ecology, the Ascendent Perspective.* Columbia University Press, New York.

Ulanowicz, R.E. (2009) *A third window, Natural life beyond Newton and Darwin.* Templeton Foundation Press, West Conshohocken, PA.

Ulanowicz, R.E., Holt, R.D. & Barfield, M. (2014) Limits on ecosystem trophic complexity: insights from ecological network analysis. *Ecol Lett*, **17**, 127–136.

Ulanowicz, R.E. & Kay, J. (1991) A package for the analysis of ecosystem flow networks. *Environmental Software*, **6**, 131–142.

Ulanowicz, R.E. & Kemp, W.M. (1979) Toward canonical trophic aggregations. *Am Nat*, **114**, 871–883.

Ulanowicz, R.E. & Puccia, C.J. (1990) Mixed trophic impacts in ecosystems. *Coenoses*, **5**, 7–16.

Ulanowicz, R.E. & Scharler, U.M. (2008) Least-inference methods for constructing networks of trophic flows. *Ecol Model*, **210**, 278–286.

van Oevelen, D., Soetaert, K., García, R., de Stigter, H.C., Cunha, M.R., Pusceddu, A. & Danovaro, R. (2011) Canyon conditions impact carbon flows in food webs of three sections of the nazaré canyon. *Deep-Sea Res Pt II*, **58**, 2461–2476.

Wasserman, S. & Faust, K. (1994) *Social network analysis: Methods and applications.* Cambridge University Press, Cambridge; New York.

Whipple, S.J., Borrett, S.R., Patten, B.C., Gattie, D.K., Schramski, J.R. & Bata, S.A. (2007) Indirect effects and distributed control in ecosystems: Comparative network environ analysis of a seven-compartment model of nitrogen flow in the Neuse River Estuary, USA—time series analysis. *Ecol Model*, **206**, 1–17.

Zhang, Y., Yang, Z.F., Fath, B.D. & Li, S.S. (2010) Ecological network analysis of an urban energy metabolic system: Model development, and a case study of four Chinese cities. *Ecol Model*, **221**, 1865–1879.

6 Tables

Table 1: Selected data input, management, and export functions in `enaR`.

Function	Description	Example Reference
pack	This function lets the user combine model elements into a network data object	
unpack	Extracts the invidual model elements (e.g., flows, inputs, outputs) from the network data object	
read.scor	Creates a network data object from a SCOR formatted data file	Ulanowicz & Kay (1991)
read.wand	Creates a network data object from a WAND formatted data file	Allesina & Bondavalli (2004)
ssCheck	Checks to see if the model is at steady-state	
balance	Applies one of four balancing algorithms to a model not at steady-state	Allesina & Bondavalli (2003)
force.balance	Runs balancing algorithm as many times as necessary to balance the model	
write.nea	Writes the model data to the file format used as input for NEA.m	Fath & Borrett (2006)

Table 2: Ecosystem Network Analysis functions in `enaR`.

Function	Description	Example Reference
`enaStructure`	ENA Structural analysis returns the adjacency matrix and multiple common descriptive statistics (e.g., number of nodes, connectance, pathway proliferation rate)	Borrett *et al.* (2007)
`enaFlow`	Calculates node throughflow and input and output oriented direct and integral flow intensity matrices. It also returns multiple whole network descriptive statistics including Total System Throughflow, Finn Cycling Index, and Average Path Length	Finn (1976)
`enaAscendency`	Performs ascendency analysis on the model flows and returns whole-network statistics including the average mutual information, Ascendency, Capacity, and Overhead.	Ulanowicz (1997)
`enaStorage`	ENA Storage analysis considers how the model fluxes generate the node stroage (e.g., biomass) in the system. This function returns the input and output oriented direct and integral storage matrices.	Matis & Patten (1981)
`enaUtility`	ENA Utility analysis investigates the direct relationships among the network nodes as well as the integral relationships when all of the indirect interactions are considered.	Patten (1991)
`enaMTI`	Mixed Trophic Impacts assesses the net relationships among species in a food web.	Ulanowicz & Puccia (1990)
`enaControl`	Control analysis determines the relative control one node exerts on another through the transaction network.	Dame & Patten (1981)
`enaEnviron`	Returns the n unit and n realized input and output environs of the model.	Patten (1978)

Table 3: Selected `enaR`auxillary functions and analyses.

Function	Description	Example Reference
Speciality Analyses		
enaAll	Runs all of the primary ENA algorithms	
get.ns	Returns the whole-network statistics from enaStructure, enaFlow, enaAscendency, enaStorage, and enaUtility	
eigenCentrality	Calculates the average eigenvalue centrality for any input matrix	
environCentrality	Returns the input, output, and average environ centralities for a matrix	Fann & Borrett (2012)
TET	Returns the total environ throughflows	Whipple *et al.* (2007)
TES	Returns the total enivron storages	Matis & Patten (1981)
Auxillary Functions		
get.orient	Determine the oreintation of the results (row-to-column vs. School)	
set.orient	Set the oreintation of the results (row-to-column vs. School)	
mExp	This function les users calculate matrix exponents.	

Table 4: Example code for applying `enaR` Flow analysis to Dame & Patten's (1981) oyster reef model.

```
> library(enaR)                 # load package
> m <- read.scor("oyster.dat")  # read model data from SCOR formatted file
> m <- balance(m)               # balance model using AVG2 algorithm
[1] BALANCED
> u <- unpack(m)                # unpack model data to illustrate components
> attributes(u)
$names
[1] "F"      "z"      "r"      "e"      "y"      "X"      "Living"

> F <- enaFlow(m)               # perform ENA flow analysis
> attributes(F)                  # show analysis objects created
$names
[1] "T"  "G"  "GP" "N"  "NP" "ns"

> F$ns                          # show flow analysis network statistics
     Boundary     TST     TSTp      APL       FCI       BFI       DFI       IFI
[1,]    41.47 83.5833 125.0533 2.015512 0.1101686 0.4961517 0.1950689 0.3087794
        ID.F   ID.F.I   ID.F.O    HMG.I    HMG.O AMP.I AMP.O mode0.F  mode1.F
[1,] 1.582925 1.716607 1.534181 2.051826 1.891638     3     1   41.47 32.90504
      mode2.F  mode3.F mode4.F
[1,] 9.208256 32.90504   41.47
>
```

7 Figures

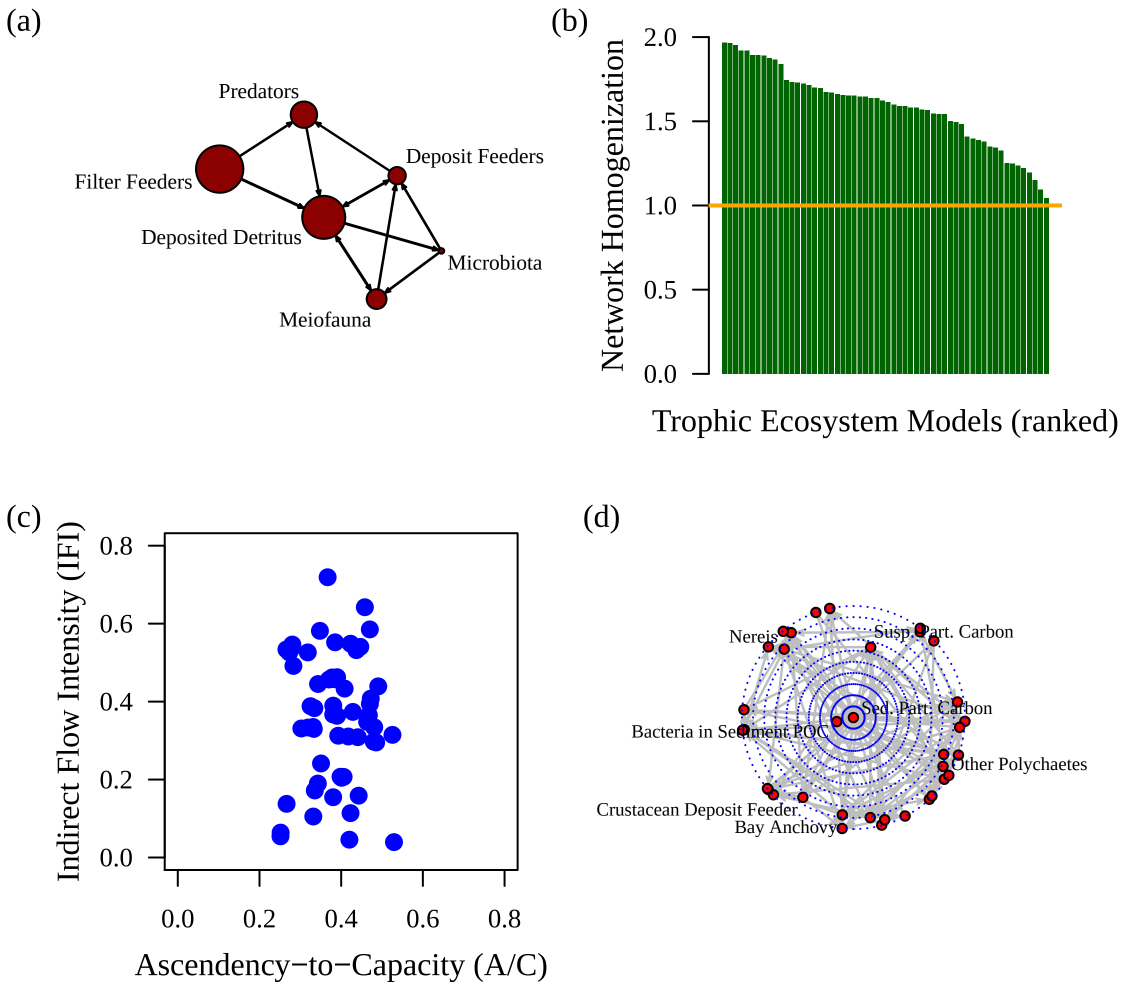


Figure 1: Example of analysis and visualizations created with `enaR` (a) network digraph of the internal flows of an oyster reef ecosystem model (Dame & Patten, 1981), (b) network homogenization statistic for 56 trophic ecosystem models (rank-ordered), (c) scatter plot showing the relationship between the ascendency-to-capacity ratio and the indirect flow index for the 56 trophic ecosystem models included in the package, and (d) target plot of the betweenness centrality from social network analysis calculated for the xx nodes of the Chesapeake Bay ecosystem model (Baird & Ulanowicz, 199).

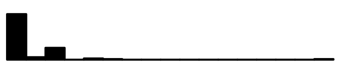
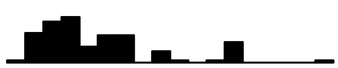
Statistic	Min	Distribution	Max	Median	Mean	CV
n	4		125	15	26.66	1.02
C	0.05		0.45	0.22	0.25	0.51
LD	1		16.91	3.14	4.58	0.89
lam1A	0		14.17	3.27	4.27	0.76
FCI	0		0.98	0.26	0.38	0.86
APL	1.37		186.25	3.67	20	1.91
IFI	0.04		0.99	0.53	0.56	0.52
HMG.O	1.04		13.07	1.78	2.3	0.83
AMP.O	0		323	6.5	19.77	1.91
AMI	1		2.25	1.57	1.58	0.21
ASC.CAP	0.25		0.75	0.39	0.42	0.28
synergism.F	2.41		60.51	3.95	5.69	1.1
mutualism.F	0.6		4	1.16	1.43	0.5

Figure 2: Distributions of selected ENA network statistics from the 100 empirically-based ecosystem models included in `enaR`. The results are summarized using a histogram showing the distribution of the values of each network statistic between the observed minimum and maximum values. The median, mean, and coefficient of variation (ratio of standard deviation and mean) values are also reported. The network statistics are the number of nodes (n), the connectance ($C = L/n^2$), link density ($LD = L/n$), pathway proliferation rate (lam1A), Finn cycling index (FCI), average path length (APL), indirect flow intensity (IFI), output oriented network homogenization ratio (HMG.O), output-oriented network amplification ratio (AMP.O), average mutual information (AMI), the ascendency-to-capacity ratio (ASC.CAP), flow-based network synergism (synergism.F) and mutualism (mutualism.F).