
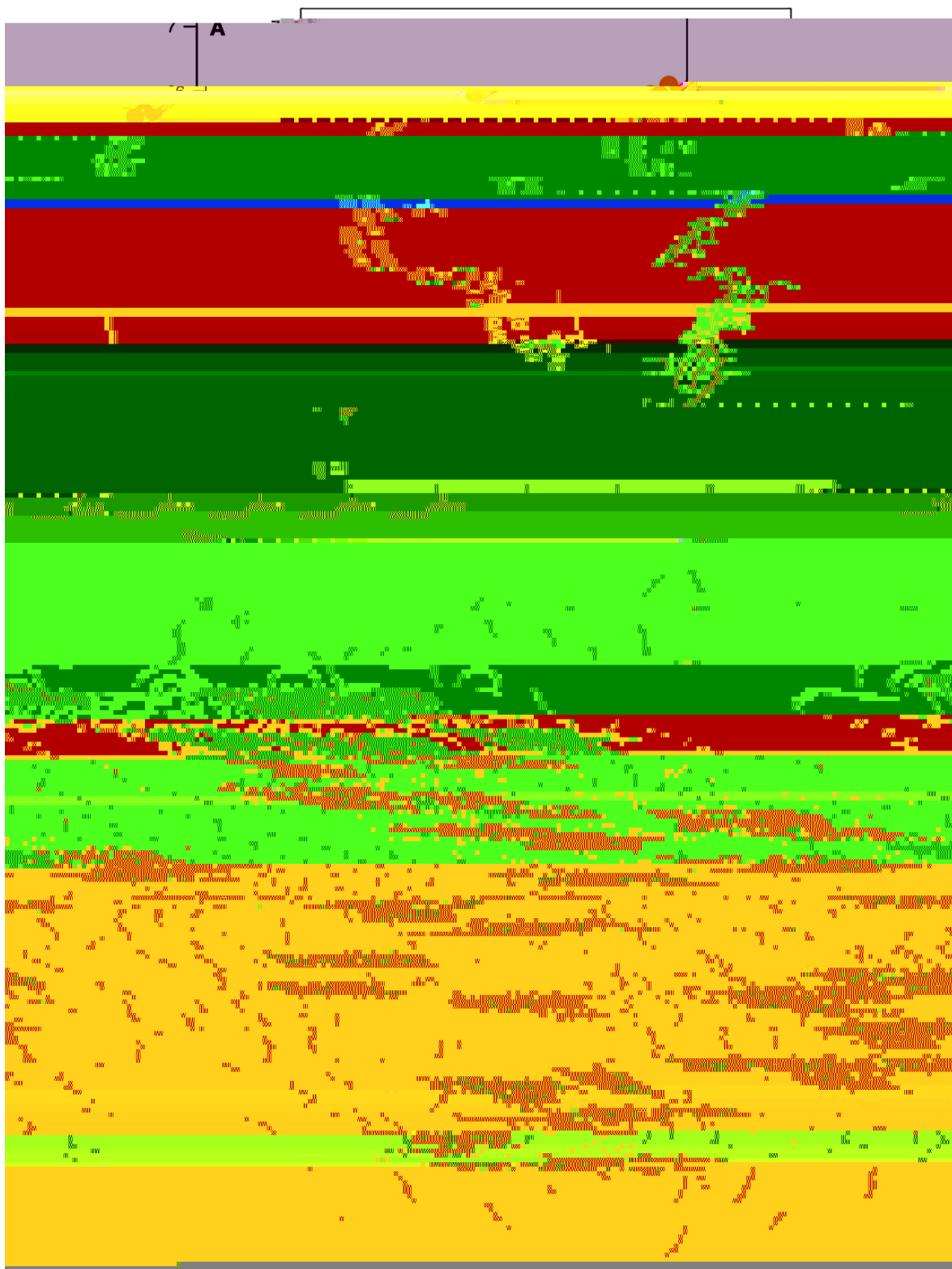


Ecological network metrics: opportunities for synthesis

MATTHEW K. LAU,^{1,†} STUART R. BORRETT,^{2,3} BENJAMIN BAISER,⁴ NICHOLAS J. GOTELLI,⁵
AND AARON M. ELLISON ¹

bank (Golley 1993, Bascompte and Jordano 2014). The application of network theory has provided a formal, mathematical framework to approach systems (Proulx et al. 2005, Bascompte and Jordano 2014) and led to the development of network ecology (Patten and Witkamp 1967, Borrett et al. 2014, Poisot et al. 2016b).

Network ecology can be defined as the use of network models and analyses to investigate the structure, function, and evolution of ecological systems at many scales and levels of organization (Borrett et al. 2012, Eklöf et al. 2012). The influx of network thinking throughout ecology, and ecology's contribution to the development of network science highlight the assertion that "networks are everywhere" (Lima 2011). And, as one would expect, the field has grown rapidly, from 1% of the primary ecological literature in 1991 to over 6% in 2017 (Fig. 1A). Some examples include the following: applying network theory to population dynamics and spread of infectious diseases (May 2006); description and analysis of networks of proteins in adult organisms (Stumpf et al. 2007) or during development (Hollenberg 2007); expanding classical food-web to include parasites and non-trophic interactions (Ings et al. 2009, Kéfi et al. 2012); investigating animal movement patterns (Lédée et al. 2016) and the spatial structure of metapopulations (Holstein et al.



(Fig. 1. Continued)

ecology has grown rapidly since the turn of the last century but has been developing in isolated sub-fields. (A) Plot showing the increase in “network ecology” keywords in the literature from 1991 to current (updated using search developed by Borrett et al. [2014]). (B) Contour plot of common topics in network ecology with peaks indicating clusters of related topics. The regions are labeled with the most common terms found in the clusters. From Borrett et al. (2014), reproduced with permission.

methods based on multivariate, correlative approaches (Legendre et al. 2012). While some approaches to studying subsets of species incorporate the underlying pattern of direct and indirect links (e.g., modules [sensu Holt 1997, Holt and Hoopes 2005]), the majority do not. Such limitations repeatedly have led to calls for the application of “network thinking” to ecological questions (e.g., Patten and Witkamp 1967, Urban and Keitt 2001, Proulx et al. 2005, Ings et al. 2009, Golubski et al. 2016, Jacoby and Freeman 2016, QUINTESENCE Consortium et al. 2016). There are now many resources for learning about network ecology and network theory in general, and we point the reader in the direction of excellent reviews in this area (Proulx et al. 2005, Bascompte and Jordano 2007, Ings et al. 2009, Borrett et al. 2012, Brandes et al. 2013) and more comprehensive introductions (Brandes and Erlebach 2005, Newman 2010, Estrada 2015).

Network ecology employs network theory to quantify the structure of ecological interactions. All networks consist of sets of interacting nodes (e.g. species, non-living nutrient pools, habitat patches) whose relationships are represented by edges (e.g., nutrient or energy transfers, pollination, movement of individuals). Conceptually, a network is a set of things or objects with connections among them. Stated mathematically, a network is a generic relational model comprised of a set of objects represented by nodes or vertices (N) and a set of edges (E) that map one or more relationships among the nodes, $G = (N, E)$. A canonical ecological example of a network is a food-web diagram, in which the nodes represent species, groups of species, or non-living resources, and the edges map the relationship who-eats-whom.

The analysis of networks is inherently hierarchical, ranging from the entire network down to individual nodes and edges. Depending on the characteristics and level of detail of the information provided for a given model, there is a large

number of network analyses and metrics that can be used to characterize the system at multiple levels (similar to Hines and Borrett 2014, Wasserman and Faust 1994), including: (1) the whole-network level (i.e., the entire network), (2) the sub-network level (i.e., groups of two or more nodes and their edges), and (3) the individual node or edge level (Fig. 2).

Network-level metrics integrate information over the entire set of nodes and edges. For example, the number of nodes (e.g., the species richness of a food-web) and the density of connections or connectance are both network-level statistics used to describe the overall complexity of a network and have been investigated by ecologists for over 40 yr (May 1972, Allesina and Tang 2012).

Sub-network-level analyses focus on identifying specific subsets of nodes and edges. There are a variety of groups that have different names (e.g., module, motif, cluster, clique, environ) and different methods for measurement. Sub-networks often represent more tractable and meaningful units of study than individual nodes and edges on the one hand or entire networks on the other. For example, in landscape and population ecology, the preferential movement of individuals and genes (edges) between habitat patches (nodes) has implications for conservation of populations and the design of preserves (Calabrese and Fagan 2004, Holt and Hoopes 2005, Fletcher et al. 2013). Also, both nodes and edges can be divided into classes. An example of this is the bipartite graph, in which interactions occur primarily between, rather than within, each class or “part” of the community. A bipartite network has only two classes of nodes, such as in a pollination network in which the community is divided into plants being pollinated and insects that do the pollination (Petanidou et al. 2008). In this network, edges representing pollination visits can only map between two nodes in the different classes.

Metrics at the individual node or edge level quantify differences in relative importance. Whether we are interested in an individual or species that transmits disease, species whose removal will result in secondary extinctions, or key habitat patches that connect fragmented landscapes, identifying important nodes is a critical component of network analysis. Another type of node or edge-level metric classifies nodes or edges according to their roles within a network. This classification can use information from differing levels. Additionally, nodes and edges can have variable characteristics. Edges can be weighted

Table 1. Ecological network metric summary and classification.

Sub.discipline	Level	Metric	Concept	References
General	W	Density	The proportion of possible edges that are actually associated with nodes; called Connectance in Food-web ecology	
General	N	Centrality	Multiple ways to characterize the relative importance of nodes	Wasserman and Faust (1994)
General	N	Degree	Number of edges connected to a given node, which is a type of local centrality	
General	N	Eigenvector centrality	Global centrality metric based on number of walks that travel through a node	Bonacich (1987)
General	W	Centrality distribution	Shape of the frequency distribution of edges among nodes	Barabási and Albert (1999); Dunne et al. (2002)
General	W	Centralization	The concentration (vs. evenness) of centrality among the nodes	Freeman (1979)
General	W	Graph diameter	The longest path between any two nodes in a graph	Barabási et al. (2000); Urban and Keitt (2001)
General	W	Modularity	Degree to which edges are distributed within rather than between distinct sets of nodes	Newman (2010)
General	G	Motifs	Small sets of nodes with similar distributions of edges	Milo et al. (2002)
General	W	Link density	Average number of edges per node	Martinez (1992)
Community	N	Temperature	Measures the nestedness of a bipartite network	Ulrich and Gotelli (2007)
Community	W	Co-occurrence	Degree of overlapping spatial or temporal distributions of species relative to a null model	Gotelli (2000)
Community	N	Indicator species	The degree to which the abundance of a taxonomic group responds to an environmental gradient	
Community	W	Nestedness	The degree to which interactions can be arranged into subsets of the larger community	
Community	W	Evenness	Deviation of the distribution of observed abundances relative to an even distribution among taxonomic groups in a community	
Community	W	Diversity	Distribution of abundances among taxonomic groups in an observed community	
Community	W	Richness	The number of taxonomic groups in a community	
Community	W	Stability	The change in the abundances of taxonomic groups across a set of observations	
Food-web	N	Removal importance	The degree to which removal of a compartment or species produces subsequent removals in the ecosystem	Borrvall et al. (2000); Dunne et al. (2002); Eklöf and Ebenman (2006); Solé and Montoya (2001)
General	N	Connectance	Proportion of realized out of possible edges	Pimm (1982); Vermaat et al. (2009)
Food-web	G	Food-chain length	The number of feeding relationships among a set of compartments in a food-web	Post et al. (2000); Ulanowicz et al. (2014)
Ecosystem	W	Finn cycling index	Degree to which matter or energy passes through the same set of compartments	Finn (1980)

been called network aggradation (Jørgensen et al. 2000). In economics, average path length is known as the multiplier effect (Samuelson 1948).

Another kind of redundancy is the creation and use of multiple statistics that measure the same or very similar network aspects. A clear example of this is inherent in the proliferation of centrality measures to indicate node or edge importance. Network scientists have shown that many centrality metrics are correlated (Newman 2006, Jordán et al. 2007, Valente et al. 2008). Likewise, Borrett and Osidele (2007) found that nine commonly reported ecosystem network analysis metrics covaried in 90 plausible parameterizations of a model of phosphorus biogeochemical cycling for Lake Lanier, Georgia, but that all these metrics were associated strongly with only two underlying factors. However, even a perfect correlation does not mean that two metrics have identical properties, and they still may diverge in different models. Therefore, it is important to have mathematically based comparisons of metrics (Borgatti and Everett 2006, Borrett 2013, Kazanci and Ma 2015, Ludovisi and

Scharler 2017). It is incumbent on network ecologists to establish clearly the independence and uniqueness of the descriptive metrics used.

From the perspective of the broader field of

Box 1

Benchmarking Ecological Models

The most basic test is to feed the algorithm a set of “random” matrices to make sure that the frequency of statistically significant results is no greater than 5%. Otherwise, the algorithm is vulnerable to a Type I statistical error (incorrectly rejecting a true null hypothesis). However, specifying a matrix produced by random sampling errors is not so easy. By definition, if a null model algorithm is used to generate the random matrices, then no more than 5% of them should be statistically significant (unless there were programming errors). For binary matrices, two log-normal distributions can be used to generate realistic heterogeneity in row and column totals, while still maintaining additive effects for cell occurrence probabilities (Ulrich and Gotelli 2010). “Structured” matrices are needed to test for Type II errors (incorrectly accepting a false null hypothesis), and these require a careful consideration of exactly what sort of pattern or mecha-

et al. 2003). For example, randomized networks have been used to link motifs (Milo et al. 2002) to network assembly (Baiser et al. 2016), stability (Allesina and Pascual 2008, Borrelli et al. 2015),

and persistence in food-webs (Stouffer and Bascompte 2010).

In addition to the random matrix approaches of null and ER models, there are other, more

complex algorithms that are used to generate structured matrices. Perhaps one of the best known in network theory is the Barabasi-Albert (BA, Barabási and Albert 1999) model, which adds nodes and edges to a growing network with a greater probability of adding edges to nodes with a higher degree. The BA algorithm is similar to ecological network algorithms that generate non-random structure, because of either direct influence or similar processes operating in systems of interest. Some of these models include processes of “preferential attachment” in which organisms tend to interact with the same, common species. Food-web modeling algorithms also have been developed that use a trait-based approach (e.g., Allesina and Pascual 2009), consumer–resource models (Yodzis and Innes 1992), niches (Williams and Martinez 2000), cyber-ecosystem algorithms (Fath 2004), and cascade models (Cohen and Ucazuk 1992, Allesina and Pascual 2009, Allesina and Tang 2012).

The statistical behavior of some models and metrics can be understood analytically. For example, the networks generated by the BA algorithm display degree distributions that approximate a power-law distribution, like many real-world “scale-free” networks (Albert et al. 2002). If the network is sparse (i.e., $[K \ll N^2]$), the degree distribution of the network should follow a Poisson distribution. However, as new models and metrics are introduced, new benchmarking should be done and compared to previous results. Newman et al. (2016) is one example of how benchmarking can be used for investigating processes operating on ecological networks. Ludovisi and Scharler (2017) advocate the same approach for the analysis of network models in general. The benchmark (Eugster and Leisch 2008) package in R (R Core Team 2017) is a general algorithm-testing software package that provides a useful starting point.

REPRODUCIBILITY: OPEN-DATA, OPEN-SOURCE, AND PROVENANCE

As analyses of network models increase in computational intensity, there is a concomitant increase in the need for new tools to track and share key computational details. This need is compounded when models incorporate data from multiple sources or analyses involve random

processes. The combination of the volume of data and computational intensity of studies of ecological networks further increases the burden on ecologists to provide information needed to adequately reproduce data sets, analyses, and results. As the sharing and reproducibility of scientific studies are both essential for advances to have lasting impact, finding easier, faster, and generally more convenient ways to record and report relevant information for ecological network studies is imperative for advancing the field.

Sharing data and open-source code have become established in ecology, and network ecologists are now producing more network models and data (e.g., Fig. 1A). These include not only ecological interaction networks, but also an influx of other relevant networks, including ecological genomic networks generated by next-generation, high-throughput sequencing technologies (Langfelder and Horvath 2008, Zinkgraf et al. 2017). There are now multiple web-accessible scientific databases (e.g., National Center for Biotechnology Information [NCBI], Data Dryad, Dataverse) and at least four databases have been constructed specifically to curate ecological network data: including “Kelpforest” (Beas-Luna et al. 2014), “The Web of Life” (Fortuna et al. 2014), “Mangal” ecological network database (Poisot et al. 2015), and the “Interaction Web Database” (<https://www.nceas.ucsb.edu/interactionweb/resources.html>).

The increase in ecological network data is linked to an increasing rate of shared analytical code and other open-source software. It is now

the ecosystem network analysis literature (Borrett and Lau 2014, Lau et al. 2015).

Although ecology has long had a culture of keeping records of important research details, such as field and laboratory notebooks, these practices put all of the burden of recording “metadata” on the researcher. Manual record-keeping methods, even when conforming to metadata standards (e.g., Ecological Markup Language [EML], see Boose et al. 2007), do not take advantage of the power of the computational environment. Data-provenance methods aim to provide a means to collect formalized information about computational processes, ideally in a way that aids the reproducibility of studies with minimal impact on the day-to-day activities of researchers (Boose et al. 2007). These techniques have been applied in other areas of research and could provide an effective means for documenting the source and9w.6(n)-401.1(ab6y)-48g-

benchmarking. Multi-trophic networks provide a

- food-web structure. *Trends in Ecology and Evolution* 27:436–442.
- Gotelli, N. J., and G. R. Graves. 1996. Null models in ecology. *Ecology* 14:368.
- Gotelli, N. J., and W. Ulrich. 2012. Statistical challenges in null model analysis. *Oikos* 121:171–180.
- Graham, M. H., and P. K. Dayton. 2002. On the evolution of ecological ideas: paradigms and scientific progress. *Ecology* 83:1481–1489.
- Gregg, E. J., and K. M. A. Chan. 2014. Leaps of faith: How implicit assumptions compromise the utility of ecosystem models for decision-making. *BioScience* 65:43–54.
- Grilli, J., T. Rogers, and S. Allesina. 2016. Modularity and stability in ecological communities. *Nature Communications* 7:12031–12041.
- Guesnet, V., G. Lassalle, A. Chaalali, K. Kearney, B. Saint-B'

Post, D. M., M. L. Pace, and N. G. Hairston. 2000.