Towards a sounder interpretation of entropy-based indicators in ecological network analysis

Alessandro Ludovisi [a,*], Ursula Michaela Scharler [b]

[a] Dipartimento di Chimica, Biologia e Biotecnologie Università degli Studi di Perugia, Perugia, Italy
[b] School of Life Sciences, University of KwaZulu-Natal, 4001 Durban, South Africa

A B S T R A C T

Various indicators rooted in the concepts of information and entropy have been proposed to be used for ecological network analysis. They are theoretically well grounded and widely used in the literature, but have always been difficult to interpret due to an apparent lack of strict relations with node and link weight. We generated several sets of 10,000 networks in order to explore such relations and work towards a sounder interpretation. The indices we explored are based on network composition (i.e., type and importance of network compartments), or network flows (i.e., type and importance of flows among compartments), including Structural Information (SI), Total System Throughput (TST), Average Mutual Information (AMI), Flow Diversity (H), and Ascendency (ASC). A correlation analysis revealed a lack of strict relationships among the responses of the investigated indicators within the simulated space of variability of the networks. However, fairly coherent patterns of response were revealed when networks were sorted by following a "bottom-up" criterion, i.e. by increasing the dominance of the large-sized top predator in the network. This ranking is reminiscent of ecosystem succession, along which the prominence of higher trophic level organisms progressively increases. In particular, the results show that a simple increase in organisms having large size and low consumption rates is potentially able to simultaneously lead to an increase of different types of information (as SI, H and AMI), thus also emphasizing the importance of bionomic traits related to body size in affecting information-related properties in a trophically connected community. The observed trends suffer from a certain dispersion of data, which was diminished by imposing specific and ecologically meaningful constraints, such as mass balancing and restriction to certain range of the ratio A/C, an index related to the viability of ecological networks. These results suggest that the identification of a set of effective constraints may help to identify improved conditions for applicability of the investigated flow-based indicators, and also provide indication on how to normalise them with respect to meaningful network properties or reference states. Thus, in order to increase confidence in the derived network metrics describing a particular ecosystem state, and thus increase their applicability, it is advisable to construct replicate networks by taking the variability of input data into account, and by applying uncertainty and sensitivity analyses.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

Ecological indicators based on the structure of ecological networks have been developed during the past several decades in order to describe ecosystem state and functioning (Jørgensen et al., 2007). They are potentially of interest for ecologists, biologists and socio-economists since changes on relatively small hierarchical scales that are mostly well documented (e.g. eutrophication,

overfishing, freshwater allocations, changes in trade volumes, etc.) can be gauged against an effect on the functioning of the entire system which can then be placed on a successional trajectory (Goerner et al., 2009; Holling, 2001). Recently, the application of network analysis indicators has been expanded to socio-economic networks to assess their efficiency and resilience (e.g. Kharrazi et al., 2013; Fang et al., 2014). From a methodological point of view, network indicators can be classified into two main classes, depending on the focus they put on the basic facets of network structure, namely composition (i.e., type and importance of network compartments), and weighted linkages (i.e., type and importance of flows among compartments). Historically, these two classes of indicators have developed quite separately, mostly because the

* Corresponding author.
 E-mail addresses: alessandro.ludovisi@unipg.it (A. Ludovisi),
scharler@ukzn.ac.za (U.M. Scharler).

methodological frameworks are different, despite their common roots in the fundamental laws of thermodynamics. From a thermodynamic standpoint, ecosystems can be viewed as self-organising systems able to maintain themselves in a far-from-equilibrium condition by exploiting the entropy exchanges with the surrounding environment (Nicolis and Prigogine, 1977; Schrödinger, 1944; Ulanowicz and Hannon, 1987). A flow of low-entropy energy is needed in order to sustain the self-organisation of organisms, trophic chains and matter cycling (Morowitz, 1968). The whole complexity of ecological networks is sustained by a basic mechanisms of self-organisation, and thus a coherent approach to the analysis of ecological networks can certainly be pursued using the tools provided by thermodynamics and related sciences. In particular, information theory (Shannon and Weaver, 1949), which is connected to thermodynamics through the statistical concept of entropy, is particularly fitting, as it can provide a wide set of tools useful for examining network complexity. Within the above theoretical framework, several entropy-based functions have been proposed as ecological orientors, i.e. indicators able to describe the stage and orientation of ecosystem development (Fath et al., 2004; Jørgensen et al., 2007; Müller and Leupelt, 1998). Among them, we can mention Structural Information (Ludovisi, 2009), Average Mutual Information and related indicators (Ulanowicz, 1986), on which the present work is mostly focused. The former (hereafter referred to as SI) has been derived within the framework of the exergy concept (Mejer and Jørgensen, 1979), and is intended to measure the information arising from the change in the apportionment of matter among ecosystem components, from the thermodynamic equilibrium to actual state. In particular, since the biotic components are extremely unlikely to be found at thermodynamic equilibrium, with a probability decreasing with body size of organisms, SI depends mostly on the composition of the biotic component, especially in terms of size spectrum and diversity. Since large-sized organisms typically dominate in late stages of ecological succession, SI is expected to increase along with ecosystem development. The soundness of SI as ecological orientor has been shown in theoretical studies (Jørgensen et al., 2010; Ludovisi, 2009; Ludovisi and Jørgensen, 2009) and its effectiveness as ecological indicator has been tested by application to simulated (Ludovisi, 2009) and real case studies, also in comparison with other thermodynamic orientors (Ludovisi, 2014; Ludovisi et al., 2012).

On the other hand, Average Mutual Information and related indicators are based on the flow structure of a network and vary according to link weight and connectivity in the system. The basic functions of this group of indicators – Average Mutual Information (AMI) and Flow diversity (H) – are based on information-theory concepts in ecological networks describing the diversity of interactions between ecosystem components (H) and the organisational degree of these interactions in a system (AMI) (e.g. Ulanowicz, 1986). Historically, ecosystems were theorised to strive for higher efficiency (higher AMI) throughout succession (e.g. Ulanowicz, 1997); currently it is accepted that systems with excessive (high AMI) or too little efficiency (low AMI) are less likely to persist. Extremely high efficiency is connected to brittleness and higher chance of collapse, whereas too little efficiency with stagnation (Goerner et al., 2009; Ulanowicz et al., 2009). Since the flow diversity and efficiency are interrelated, the indices hold important information on the viability of a network. At its lowest flow diversity, a particular ecosystem is at its most efficient, as the number of flows is at its minimum and therefore the constraint on flows is at its highest (transfers are highly efficient). Real world ecosystems have been shown to incorporate aspects of both efficiency and inefficiency in the energy transfer in their flow structure and occupy a place that trade off efficiency against stagnation to arrive at a viable state (e.g. Ulanowicz, 1986; Goerner et al., 2009). In the ecological network analysis literature, AMI and H are often scaled

against the system size, which is expressed as the sum of all flows (Total System Throughput, or TST). The resulting indices, Ascendency (ASC = AMI × TST) and Development Capacity (DC = H × TST) are therefore highly influenced by the total magnitude of flows in a system (TST). To cancel out the influence of this scaling, a ratio of Ascendency and Development Capacity (A/C, equals AMI/H) has been widely used to describe the developmental state of a particular system, since DC is the theoretical upper limit of ASC (for a flow configuration of only one in- and outgoing flow per node).

Although the above flow-based indices have a sound theoretical basis, their interpretation for real world systems is hampered by the lack of a thorough understanding on their behaviour, as systematic investigations on this aspect are few. Investigations on theoretically constructed networks with varying size (number of taxa) and connectivity (major connections per taxon) revealed that as flow diversity (H) increases, the potential AMI value also increases (Morris et al., 2005). Overall, the study concluded that "*the true significance of these metrics may not be realised within our current means of characterising food webs*", hinting at the apparent lack of strict relations. The aggregation of the empirical Chesapeake Bay model indicated a positive relationship between ascendency and network size (number of nodes) in that smaller ascendency values were calculated for smaller sized networks (Abarca-Arenas and Ulanowicz, 2002). Similarly, Allesina et al. (2005) pointed out the effect of detritus compartments on flow-based values, however neither of the two aggregation studies offered strict relationships or a detailed description on the behaviours of the flow-based indices that is practical for further interpretation of the indices calculated for empirical networks. In fact, mass balancing a network can have a larger influence on network indices compared to aggregation (Baird et al., 2009). Aggregation scenarios of the Sylt-Rømø Bight network produced variations of network indices of ≤7%, whereas the difference in index values between balanced and unbalanced networks differed between 26 and 105%, with Flow Diversity and Ascendency showing the highest variability (Baird et al., 2009).

The above flow-based indices have previously been dealt with in detail on a conceptual basis (Fath et al., 2001, 2004; Jørgensen and Ulanowicz, 2009; Patten, 1995), or on a limited number of theoretical case studies, but it is as yet unclear whether there are generalities in responses with changing ecosystem configurations. The clearest results from these studies are a general increase of the flow-based ascendency with ecosystem succession (Patten, 1995; Ulanowicz et al., 2006), or with Total System Throughflow (Fath et al., 2001).

Although indices are prominent in the network analysis literature and part of several software packages presently in use, e.g. enaR (Borrett and Lau, 2014), Ecopath (Christensen et al., 2005), WAND (Allesina and Bondavalli, 2004), EcoNet (Kazanci, 2007), and NETWRK (Ulanowicz and Kay, 1991), a framework of generally applicable responses of the indices and their relationships is as yet not defined.

In order to work towards a framework for a sounder interpretation of the above described network indicators (Structural Information −SI, Total System Throughput − TST, Average Mutual Information − AMI, Flow diversity − H, Ascendency − ASC, A/C) their responses on sets of artificial networks of an analogous size, but different biomass allocation to nodes and flow distribution are here comparatively investigated. Small artificial 4-node networks were specifically designed and examined in order to facilitate tractability of large sets of replicates and interpretation. Realistic constraints on consumption rates of predators were imposed in order to cover a realistic spectrum of variability in the flow structure as a function of changing biomass allocation among nodes. The effect of multiplicity of nodes (biodiversity) was also investigated by adding a supplementary node to each single trophic level included in the basic networks. The relationships among the