Biological networks across scales

The theoretical and empirical foundations for modeling, analysis and control of time-varying complex networks that connect structure and function across levels of biological organization

Paul Bogdan[1], Gustavo Caetano-Anollés[2], Anna Jolles[3], Hyunju Kim[4], James Morris[5], Cheryl Murphy[6], Catherine Royer[7], Edward H. Snell[8], Adam Steinbrenner[9], Nicholas Strausfeld[10]

[1]Ming-Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles (pbogdan@usc.edu); [2]Department of Crop Sciences, University of Illinois, Urbana-Champaign (gca@illinois.edu); [3]Department of Integrative Biology, Oregon State University, Corvallis (aejolles@gmail.com); [4]The Beyond Center, Arizona State University, Tempe (hyunju.kim@asu.edu); [5]Baruch Institute for Marine and Coastal Sciences; University of South Carolina, Columbia (morris@inlet.geol.sc.edu, or morris1520@gmail.com); [6]Department of Fisheries and Wildlife, Michigan State University, East Lansing (camurphy@msu.edu); [7]Biological Sciences, Rensselaer Polytechnic Institute, Troy (royerc@rpi.edu); [8]Hauptman-Woodward Medical Research Institute and SUNY Buffalo (esnell@hwi.buffalo.edu); [9]Department of Biology, University of Washington, Seattle (astein10@uw.edu); [10]Neuroscience, University of Arizona, Tucson (flybrain@email.arizona.edu)

Summary: Many biological systems across scales of size and complexity exhibit a time-varying complex network structure that emerges and self-organizes as a result of interactions with the environment. Network interactions optimize some intrinsic cost functions that are unknown and involve for example energy efficiency, robustness, resilience, and frailty. A wide range of networks exist in biology, from gene regulatory networks important for organismal development, protein interaction networks that govern physiology and metabolism, and neural networks that store and convey information to networks of microbes that form microbiomes within hosts, animal contact networks that underlie social systems, and networks of populations on the landscape connected by migration. Increasing availability of extensive (big) data is amplifying our ability to quantify biological networks. Similarly, theoretical methods that describe network structure are being developed. Beyond static networks representing snapshots of biological systems, collections of longitudinal data series can help either at defining and characterizing network dynamics over time or analyzing the dynamics constrained to networked architectures. Moreover, due to interactions with the environment and other biological systems, a biological network may not be fully observable. For example, due to the influence of a virus, some parts of a biological network may be hidden or may appear as an intrinsic component of the biological network while in the absence of this viral intervention the involved subnetworks may not appear at all. Also, subnetworks may emerge and disappear as a result of the need for the biological system to cope with invaders or new information flows (e.g., antibacterial resistance phenomenon). The confluence of these developments renders tractable the question of how the structure of biological networks predicts and controls network dynamics. In particular, there may be structural features that result in homeostatic networks with specific higher-order statistics (e.g., multifractal spectrum), which

maintain stability over time through robustness and/or resilience to perturbation. Alternative, plastic networks may respond to perturbation by (adaptive to catastrophic) shifts in structure. Here, we explore the opportunity for discovering universal laws connecting the structure of biological networks with their function, positioning them on the spectrum of time-evolving network structure, i.e. dynamics of networks, from highly stable to exquisitely sensitive to perturbation. If such general laws exist, they could transform our ability to predict the response of biological systems to perturbations – an increasingly urgent priority in the face of anthropogenic changes to the environment that affect life across the gamut of organizational scales.

Introduction

Nature presents us with an overwhelming plenitude of structures, the functions of which are so diverse as to suggest descriptive rules pertaining to structural-functional relationships are highly specialized. Exclusive to one or another specific domain of biological science, structure manifests in genes and development, neural circuits and integration, metabolic pathways and trophic interactions, to mention just a few. Here we attempt to address an overarching question: whether multifarious descriptions of interactions within defined biological domains find precision and unification using a language that identifies commonality of organization across all biological domains. In terms of its overall structure and dynamics might each domain present an underlying organization that suggests a universal principle of interactive connectivity across its components such that, for example, structural and dynamic interactions of elements within a defined ecology can be described using the same mathematical rules as those that describe structural and dynamic interactions of, for example, a defined part of the brain, or the genomic organization of tissue differentiation.

Biological systems can be decomposed into parts – components that combine with other components to make up a whole (Simon 1962). When parts interact with other parts of the system their interactions are constrained by space, time, information flows (including processing, transfer, and storage), and/or function, all of which are influenced by the external environment. Interactions are usually modeled with graphs, mathematical constructs that connect points known as vertices with lines (Barabasi and Oltvai 2004). Figure 1A describes the anatomy of a network. Vertices represent parts of a system and lines represent pairwise interactions between them. For example, a graph describing the combination of structural domains in multidomain proteins will connect vertices describing structural domains with lines describing the presence of domains in proteins. When connections of vertices are undirected, lines fail to point in any direction; each connection involves an unordered pair of (end) vertices. These lines are called *edges*. When connections are directed, lines point in one direction; each connection involves an ordered pair of vertices (an initial vertex and a terminal vertex). These lines are called *arcs*. Graphs become networks whenever value functions (properties or weights) are mapped onto the vertices and lines of the graphs. For consistency, we will call the vertices of the network *nodes* and the lines that connect the vertices the *links* of the networks.

Some network properties help visualize and study network structure and makeup (Wasserman and Faust 1994; Newman 2003). For example, networks can be studied with measures of network

centrality, by detecting community structure, or by dissecting their makeup. Measures of ***network centrality*** estimate how a node or edge influences the connectivity or information flow of the network (Figure 1C). Detecting ***community structure*** allows to establish groups of nodes that are more connected with themselves than with the rest. We will refer to these communities as 'modules'. A number of hierarchical clustering algorithms can efficiently detect these network modules, including the popular Girvan-Newman algorithm (Girvan and Newman 2002). Other useful algorithms include those that maximize modularity functions, extract information through random walks (e.g. infomap algorithm), use recursive percolation methods, or analyze fractal geometric (Xue and Bogdan 2017) and differential geometric (Sia et al. 2019) characteristics of complex networks. Finally, ***compositional patterns*** such as network motifs or network cliques can highlight elemental units of network makeup, which can become useful when studying the evolution of function in network structure. However, given the intrinsic stochasticity, nonergodicity and continuous interaction with the environment, the network motifs can vary over space and time scales, yet they can explain how biological systems self-program and self-optimize to achieve the collective goal (e.g., adaptation for maximizing survival, energy efficiency).

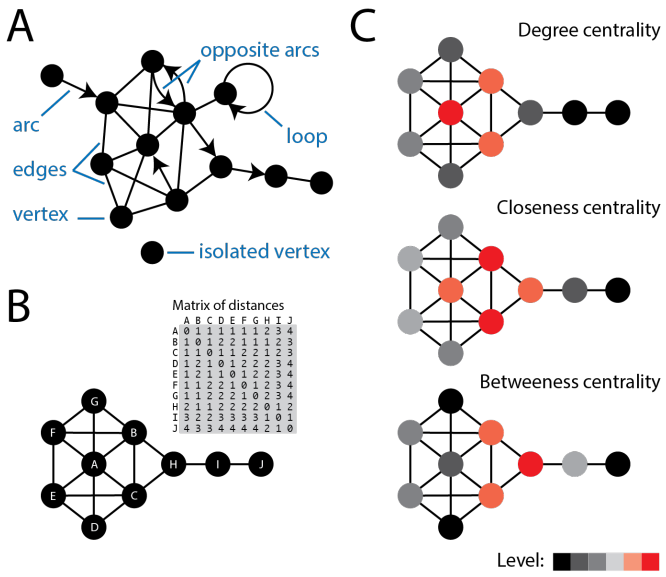


Figure 1. A network view of biological systems. **A.** An anatomical analysis shows that a network N is a combination of four sets, a set V of vertices (nodes), a set L of lines (links), and sets of vertex and line value functions that are mapped onto the V and L sets, respectively. Each line is associated with a pair of vertices (lines are 2-element subsets of V) representing edges or arcs if lines are undirected or directed, respectively. Loops are lines with identical endpoints. The illustrated network is a 'mixed network' because it contains both arcs and edges. **B.** A network can be represented with an adjacency matrix. The example network is undirected (it does not contain arcs). Consequently, its adjacency matrix is symmetric. **C.** Network centralities offer different views of the influence of nodes in a network. *Degree centrality* estimates how well a node is connected to other nodes. The degree of a node (its connections) provides a local view of network connectivity. *Closeness centrality* estimates how easy is for a node to reach other nodes. Finally, *betweenness centrality* estimates how important is a node in terms of its capacity to connect to other nodes. It offers a global view of connectivity. Other centralities (not shown) offer views of *prestige,* how important is a node in terms of the importance of its neighbors. Diagram modified from Caetano-Anollés et al. (ms. in press).

As expected from complex systems, networks abstractions in biology are often difficult to understand: (i) *Complexity:* Networks can be structurally complex when their wiring diagrams become tangles (relying on multiple rules of growing or evolution when required to cope with environmental perturbations); (ii) *Connectivity:* Links between nodes can have different weights, directions and signs and can describe different kinds of interactions (link communities); (iii) *Diversity:* Nodes and links can be diverse (e.g. biochemical networks that control cell division consist of a variety of substrates and enzymes); (iv) *Evolution:* The structure and dynamics of networks change when they grow and their wiring diagrams unfold in time. (v) *Dynamics:* Nodes and links can themselves portray non-linear and long-range memory/multifractal dynamic behaviors and the state of each node or link can vary in time in complicated ways in order to ensure in a decentralized way a common collective goal.

Network dynamics is made explicit when matter, energy, information and time flow through the network structure. These flows can be expressed in different ways, including cost, Shannon entropy, time directionality, and higher-order network statistics (Xue and Bogdan 2017). These 'flow networks' pose important conceptual and computational challenges. For example, directed networks, which induce directed connections (arcs), also induce input and output connectivity and the formation of internally connected subnetworks (cycles) that bias hierarchical structure. Moreover, the directed flows in these networks are not only time varying, but also possess multifractal characteristics. For example, the dynamics between sets of genes and linked transcription factors in gene regulatory networks exhibit fractal and long-range cross-correlated characteristics (Ghorbani et al. 2018). This implies that when a biological network is analyzed at two different time scales, its corresponding directed flow network can dramatically differ because the system is trying to concurrently process information and achieve multiple (rich) functionalities with potentially reduced/compressed set or rules of life. These cross-correlation exponents characterizing the interaction between a gene (or more genes) and a transcription factor (or more transcription factors) in gene regulatory networks are not unique and could explain the functionality achieved by a network motif or subnetwork. Also, the distribution of the cross-correlation exponents of gene regulatory networks for several types of cells can be interpreted as a measure of the complexity of their functional behavior. Consequently, one can wonder how the information processing, transfer and storage triggers the emergence of rules that govern the evolution of a time varying network (by addition, rewiring, and deletion of nodes and edges). Within this network dynamics paradigm, when aiming to understand and explain biological systems, one also requires mathematical tools to reconstruct the network structure while overcoming partial observability and 'perturbation' influences from other biological systems and environments. Since the interplay of network structure and levels of organization in biology is a crucial endeavor, studying these flow networks can uncover important regularities and principles for designing self-programming and self-optimizing synthetic biological systems.

Grand challenge

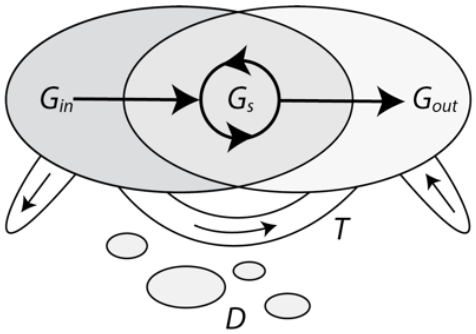
Time varying complex network abstractions provide a comprehensive graph theoretical framework with which to describe biological systems across spatiotemporal scales and levels of

organization (Caetano-Anollés et al. 2019). One important goal is to develop and rely on mathematical models and rigorous algorithmic tools to decipher time varying complex networks from heterogeneous biological measurements while overcoming challenges related to partial observability and 'perturbation' influences (Bogdan 2019; Gupta et al. 2019). Another important goal is to mine the spatiotemporal geometry and the higher-order network statistics of time varying complex networks in order to find patterns, rules, processes and models of computation (i.e., specific concurrent interplay among rules and processes) embedded in the network structure and dynamics that would help identify common organizing principles (Mahmoodi et al. 2017; Koorehdavoudi and Bogdan 2016; Balaban et al. 2018; Kim et al. 2019). Experimental and retrodictive exploration can then test theoretical frameworks and predictions. Advances in comparative and evolutionary genomics, physiology, and systems and synthetic biology can help address a number of important questions and provide potential solutions to the pluralistic and multiscale complexity of biological systems.

Objectives

The following objectives illustrate the broad scope of enquiry of our framework:

Finding commonalities in network structure across levels of organization: Simulated and real networks at different levels of organization could be compared in search for commonalities in their structural makeup and dynamics that could uncover organizing principles. As one example, directed networks such as the World Wide Web and metabolism show a bow-tie structure, in which inputs into a highly connected component result in a number of outputs (Figure 2). Depending on the networks, there will be also shunts of connectivity and disconnected components that add complexity to the makeup of these networks. Are these properties universal? Can they be studied at different levels of organization?



	G_{in}	G_s	G_{out}	T	D
WWW (x10^6)	43.3	56.5	43.2	43.8	16.8
Metabolism	93	274	161	—	283

Figure 2. The bow-tie hierarchical structure of directed networks. These networks have a giant strongly connected component (Gs), giant 'in' component (Gin), giant 'out' component (Gout), tendrils and tubes (T) and disconnected components (D). The number of nodes present in these subgraphs are listed for the WWW (Broder et al. 2000) and the metabolic networks of *Escherichia coli* (Ma and Zeng 2003). Diagram modified from Caetano-Anollés et al. (ms. in press).

Quantifying characteristics of dynamics *on* the networks to find commonalities or diversities across different types or scale of networks: To find organizing principles governing different types of networks across different scales, commonalities in structural and dynamic characteristics of the networks should be studied. One of the most distinct dynamical characteristics of biological systems is criticality. When a system is perturbed by external inputs, the perturbation may be amplified and percolated to the entire system or can have local influence, may manifest over some specific scales, or may vanish after some time. A system for the former and the latter is considered in chaotic and stable regime, respectively. Many biological systems lie between these two regimes, i.e. near critical point (Daniels et al 2018). In other words, local perturbation or signal in the biological networks is preserved in the networks. Is it possible that the dynamics of evolving networks may share commonalities or can be characterized into different classes?

Integrating the network system with external information: Systems are not isolated but depend on a superseding environment and other systems. This external integration needs to be resolved and analyzed. One way to assess integration space is to bind networks with external information such as physical or functional constraints and ask how hierarchy, modularity and other structural or dynamic properties unfold under those conditions. One interesting line of exploration that highlights integration space is the study of Rentian scaling of networks (Bassett et al. 2010; Ho and Nvlakha 2018). In the 1960s, IBM scientist E.F. Rent discovered a peculiar scaling relationship between the number of logic gates (internal components acting as network nodes) in a logical block of a computer circuit (a piece of circuit resembling a network module) and the number of circuit connections between circuit blocks (Landman and Russo 1971). This empirical relationship followed a power law with an exponent that generally ranged $0.5 < p < 0.8$, the Rent's exponent. Circuits with larger logical capacity have higher exponents. Rentian scaling relationships are robust for very large-scale integrated circuits and a number of biological networks, including neural networks. Are these scaling relationships present in networks that are spatially bound to lower degrees such as metabolism of protein-protein interactions networks? Since biological systems are not isolated, are we to expect that the effects of integration space be pervasive? This poses the additional challenge of analyzing the structure and dynamics of the integration space that wires network systems to each other.

Modes of network structure and dynamics: Morphospaces can help dissect network structure and dynamics. Morphospaces are phenotypic spaces defined by a limited number of properties that account for the most salient features of a system (Niklas et al 1994; Shoval et al. 2012). However, there is likely a multidimensional space of significant drivers of network structure and dynamics that must be uncovered. Novel deep-learning classification tools should be used to find relevant summary descriptors that are meaningful across systems. Networks do exhibit different densities, patterns of connectivities, modularity levels, hierarchical organization, and granularity, all of which could provide characteristics that may be unique to individual levels of organization in biology.

Deciphering and unfolding networks in time: Changes of network structure and dynamics with time can be studied along different timeframes and biological scales in a number of steps. First fundamental step concerns the definition of entities (nodes) and connectivities (links, arcs), as well as rigorous mathematical techniques for identifying them for each biological system while considering technological and physics-based limitations (e.g. causal influence detection, measuring signaling and Heisenberg uncertainty principle). Once nodes and links are defined, the second fundamental step consists of carefully analyzing the scarce biological sampling in order to construct a history (trajectory) of various interdependent biological networks (e.g., involving the development, physiology, metabolite dynamics, structural dynamics) that unfold over multiple time scales (i.e., including manageable timeframes from years to minutes to nanoseconds). For example, such time varying networks include those that describe gene expression patterns, signaling networks, developmental networks, the photosynthetic light harvesting complexes, food webs and neural networks. Moving at higher scales of the hierarchical organization, we need to rigorously sample the niches and populations in order to define and predict the history of ecological networks, as well as study and control their dynamics. Consequently, we need to develop new mathematical and algorithmic techniques capable to use and mine the phylogenomic or stratigraphic information in order to reconstruct the history of biological networks that describe evolving molecular machinery (e.g., proteome, metabolism, functionomes, signaling networks, protein-protein interactions, domain organization) or genes that encode this machinery. Most of these networks hold very deep evolutionary history and could provide new models of computation that biology could have discovered through evolution and inspire new trends in artificial intelligence. A crucial step towards understanding the intelligence and the nature of optimization taking place in biology requires us to investigate the structure of evolving networks, elucidating the sources, means and goals of specific network properties (e.g., scale-freeness, randomicity, modularity, hierarchy, centralities, generalized fractal dimension, multifractal connectivities, network curvature). Within this effort, the modeling of network growth and dynamics must be done according to different criteria. For example, one can use a 'morphospace' of networks where modularity, hierarchy and dynamics are made explicit (see below) to study simulated and real networks. Moreover, in order to overcome the inherent variability and stochasticity of biological systems, one can rely on characterizing the multifractal properties for establishing rigorous connections between various time varying network motifs and specific rules of life. Another important step towards characterizing the phase transitions of biological systems and predict their future interdependent dynamics requires an accurate tracing of their dynamics along evolving networks by defining (biologically relevant) events along a timeline or mapping dynamic behavior directly on the evolving networks. This is not a trivial task since in reality not all biological variables can be measured, due to emerging evolutionary behavior not all biological variables are known from the beginning (but rather discovered as the biological evolution unfolds) or the environmental perturbations grow in number, magnitude and complexity (e.g., as a function of disappearance of biological species, variations in temperature, humidity, pressure) - these are called as the 'unknown unknowns' governing the observed biological dynamics. Consequently, for deciphering and characterizing the unfolding of biological networks over time, we need new mathematical and algorithmic tools for reconstructing networks from partial observations, from various types of biological data sources and overcoming interventions (e.g., using time series data analysis on average sensitivity values of the networks,

spike/event time sequences of biological activity (excitatory or inhibitory), time sequences of partially observable subnetworks of an unknown time evolving biological network (Xue and Bogdan 2019). Moreover, specific critical nodes (e.g., neurons, cells, bacteria) may exhibit long-range memory and multi-fractal dynamic characteristics in order to cope with external perturbation and enforce a cue or rule towards a collective goal.

From a mathematical perspective, we require not only more accurate causal inference techniques to identify the multiscale interactions across biological components, but also algorithms capable to estimate the number of unknown unknowns and determine which variables exhibit either a non-Markovian dynamics (i.e., which can be modeled through a combination of fractional order derivatives) or a Markovian one (i.e., which can be encoded through integer order derivatives) (Bogdan 2019; Gupta et al. 2019). For example, an evolving metabolic network that unfolds enzymatic activities on a timescale of billions of years was studied using bipartite network approaches that connect to different levels of molecular organization (Mughal and Caetano-Anollés 2019). A database that traces evolutionary information onto metabolic network structure is available (https://manet.illinois.edu). Similarly, an evolving network that links protein domains to functional loops and defines an 'elementary functionome' of protein structure was unfolded on a timescale of billions of years (Aziz et al. 2016). This allowed to track the emergence of function in protein domain organization. At completely different timescales, physiological processes that are triggered by stress can also be dissected with networks. For example, metabolomic networks that describe the connectivity of metabolites on a timescale of hours reveal patterns of bacterial metabolic rewiring (Aziz et al. 2012). In all of these examples, hierarchical modularity, multifractal and network curvature appear as emergent properties of biological network structures. Why? Is hierarchy, multifractal characteristics and specific network curvatures a necessary consequence of the rise of modules in biology and how are those related to the functionality and rules of life? Is hierarchy associated with the rise of levels of organization?

We propose a series of activities to develop our framework:

1. Define entities (nodes) and connectivities (links, arcs) that are appropriate to each biological system (see case studies below).
2. Use biological sampling to define the history of biological networks (e.g. development, physiology, metabolite dynamics, structural dynamics) that unfold at manageable timeframes (years to minutes to nanoseconds). Example networks include networks that describe gene expression patterns, signaling networks, developmental networks, food webs and neural networks.
3. Sample niches and populations to define the history of ecological networks and study their dynamics.
4. Use phylogenomic or stratigraphic information to reconstruct the history of biological networks that describe evolving molecular machinery (e.g. proteome, metabolism, functionomes, signaling networks, protein-protein interactions, domain organization) or genes that encode this machinery. Most of these networks hold very deep evolutionary history.
5. Study the structure of evolving networks (scale-freeness, randomicity, modularity, hierarchy, centralities, generalized fractal dimension, multifractal connectivities, network curvature).

6. Model network growth and dynamics according to different criteria. For example, use a 'morphospace' of networks where modularity, hierarchy and dynamics are made explicit (see below) to study simulated and real networks.
7. Trace dynamics along evolving networks by defining events along a timeline or mapping dynamic behavior directly on the evolving networks.
8. Study the mathematical characteristics of the evolving networks (e.g., using time series data analysis on average sensitivity values of the networks, spike / event time sequences of biological activity (excitatory or inhibitory), time sequences of partially observable subnetworks of an unknown time evolving biological network (Xue and Bogdan 2019). For instance, specific critical nodes may exhibit long-range memory and multi-fractal dynamic characteristics to cope with external perturbation and enforce a cue or rule towards a collective goal.
9. Explore how networks integrate across levels of biological integration. Determine what information is lost or gained as networks incorporate information from molecular, cellular, organ, organism, population, community, ecosystem levels of biological organization.

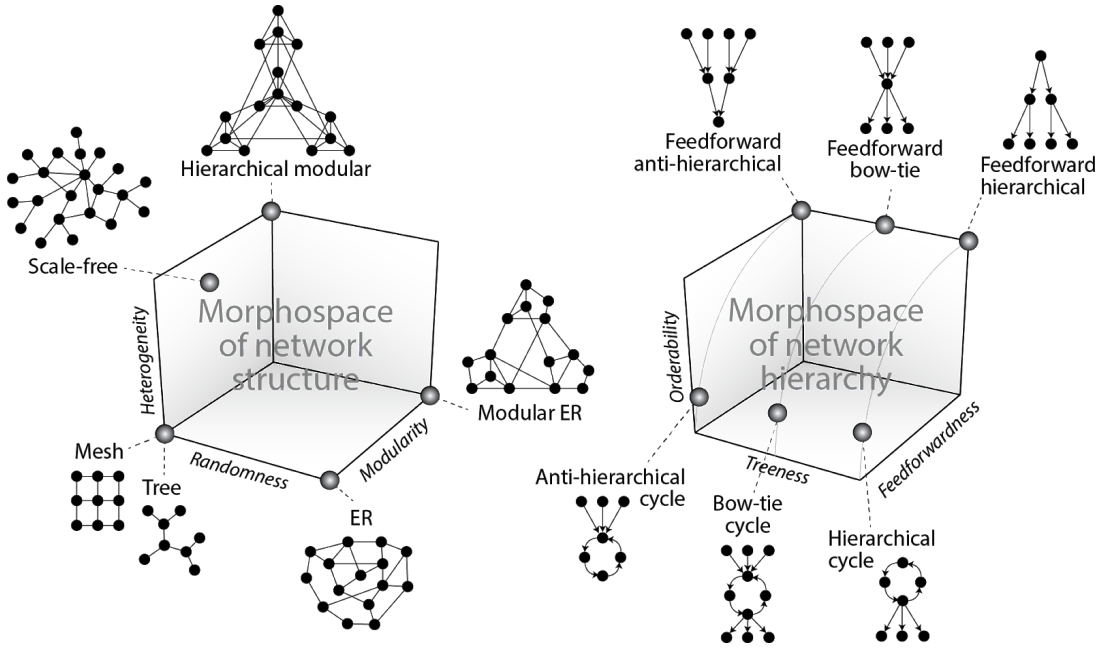


Figure 3. Morphospaces of network structure and hierarchy showing toy examples of typical graphs describing archetypes of the phenotypic landscapes. In one morphospace (left), Erdös-Rényi (ER) random graphs transform into regular graphs by decreasing randomness or into modular ER graphs by increasing modularity. Hierarchical modular structure requires both increasing modularity and heterogeneity and decreasing randomness. In another morphospace (right), treeness defines the unification or diversification of hierarchical signal in the network, whereas orderability defines the centrality of cycles in network structure. Figure modified from Caetano-Anolles et al. (2019).

How can hierarchy be explored and how its networking complexity be linked to functionality and the rules of life? A useful approach is to define a morphospace of network structure and a morphospace of network hierarchy (Figure 3) and compare how model networks generated by

simulation (satisfying specific properties in terms of multifractality and curvature / hyperbolicity) and real networks distribute in structural space. Corominas-Murtra et al. (2013) for example have shown that networks across scales exhibit a bow-tie structure that is typical of that found when studying the World Wide Web (Broder et al. 2000) or metabolic networks (Ma and Zeng 2003; Kim et al. 2019). Is this indeed a generic structure that manifests across scales? To determine when a hierarchical network was accurately identified and characterized, we require mathematical and algorithmic techniques to investigate the nonconvex free energy landscape associated with the morphospace of network hierarchy and determine the model networks that minimize the network free-energy candidates. Furthermore, being able to estimate or investigate the scale-dependent free-energy landscape from biological data could also help us determine how generic structures and the rules by which are generated manifest across spatiotemporal scales. From this perspective, the deciphering and understanding of biological systems contributes to the birth of a new branch of mathematics at the intersection of multifractal network geometry, statistical physics and optimization and potentially lead to new data science, machine learning and artificial intelligence algorithms.

Drivers of network structure and dynamics at different levels of organization: A multidimensional landscape of drivers or causal relationships are likely responsible for the structure and dynamics of biological networks. These drivers can be of different types and most likely themselves form a wire diagram of causality. Example drivers include: (a) Evolutionary (e.g. life history, adaptation, canalization, recruitment); (b) Energy (e.g. dissipation, budget); (c) Information (e.g. entropic flow, modes); (d) Mass; (e) Spatiotemporal landscape (e.g. molecular and structural spaces); (f) Trade-off solutions (e.g. economy, robustness, plasticity); (g) Perturbation (stress) - homeostasis (some networks just developed to evaluate stress only); (h) Ontogeny; (i) Growth and Development; (j) Levels of biological organization; and (k) Ecological.

The following are examples of systems, from lower to higher levels of organization. These networks are familiar to one or more of the authors and illustrate biological domains immediately suited for analysis using the approaches discussed above.

(i) Protein-protein interaction networks. Protein-protein interaction networks (PPINs), with individual proteins as nodes and physical interaction as links, are classic subjects of systems biology. PPINs have been identified for protein families, whole proteomes, and even inter-species relationships. Historically, this has been enabled by high-throughput technologies for data collection for both nodes (transcriptomics and proteomics to rapidly define all protein nodes) and edges (affinity pulldown - mass spectrometry, yeast two-hybrid, and other heterologous screens for measuring interaction strength). Modularity emerging from PPINs often correspond with specific functions: transcription, nucleosome assembly, hormone signal transduction, etc (Arabidopsis Interactome Mapping Consortium 2011). Within functional modules, certain nodes form hubs with high degrees of connectivity. In addition, articulation points that connect across modules were apparent. For example, in a recently measured cell surface Interactome for plant leucine-rich repeat ectodomains, high degree and articulation nodes are apparent and correspond with known co-receptors shared in many different immune receptor complexes (Smakowska-Luzan et al. 2018). Functional validation of these

nodes using genetic knockouts has demonstrated that hubs and articulation points have widespread immune phenotypes that affect multiple pathways (Figure 4). in contrast to peripheral nodes only required for specific recognition functions. Inter-species PPINs with factors required for pathogen virulence feature edges that predominantly connect to host hubs (Muhktar et al. 2011).

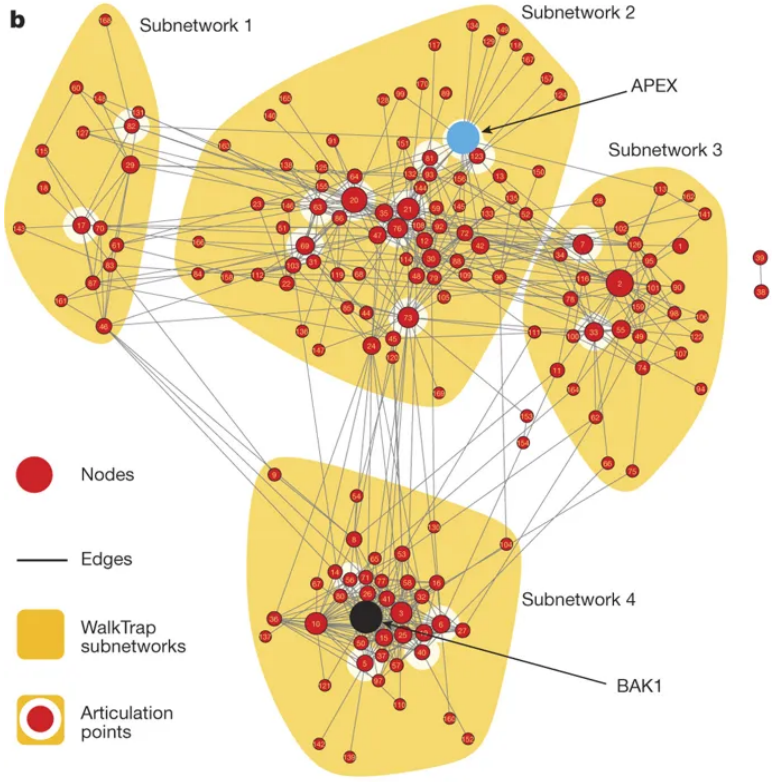


Figure 4. A leucine-rich-based cell surface interaction network (Smakowska-Luzan et al. 2018). Network analysis makes clear the existence of four distinct subnetworks and two critical nodes. The diameter of the nodes is proportional to their PageRank score.

(ii) <u>Cell cycle network (transition/developmental network)</u>. The yeast cell cycle represents a well-studied and important biological system. The network of protein factors that allow the cell to progress from one phase to the next is particularly important. The data used to make the network are the physical properties of the protein factors. Parameters of localization, concentration, dynamics, and interactions are a function of cell size. Nodes are cell cycle phases (G1, S, G2, M, cytokinesis) and the edges are the events that allow transitions from one phase to the next. Each node encompasses a sub-network; Figure 5 describes the subnetwork composing the G1-phase node. The changes in this subnetwork with time allow for progression from G1 to S phase. Note that: (a) The edges are the transitions from one phase to the next. Their thickness changes from 0 to 100% probability over time as the interactions within the module change. Once the transition occurs they revert back to zero. Reverse transitions are not allowed. (b) The stochastic interactions within each module and the changes in protein factor copy number with time determine the dynamics of the network. There is biological noise due to the stochasticity of the interactions. (c) The outputs are the

cell size at which each of the transitions occurs. (d) Changes in environment or mutations perturb the network. Extension to mammalian cells and cancer. Need to develop tools for making required measurements in less genetically modifiable systems than yeast.

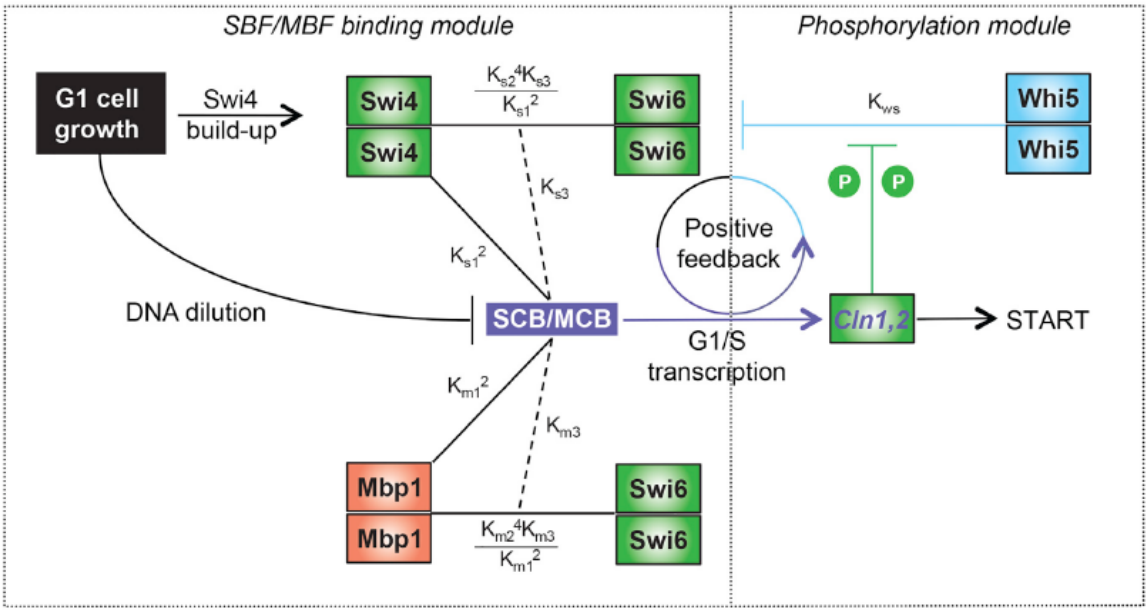


Figure 5. The subnetwork describing the G1-phase node. The transcription factors, SBF and MBF, which control the G1/S cell cycle transition in yeast, increase in copy number throughout G1, eventually saturating the G1/S target promoters. A feedback phosphorylation loop inactivates Whi5, a repressor if SBF via a cyclin dependent kinase ensures a sharp transition.

(iii) <u>Organ-level network (liver)</u>. A perturbation network (stressor - beyond homeostasis) describes pathways that converge to lead to steatosis-lipogenesis, and fatty acid uptake, efflux and oxidation (Angrish et al 2016; Knapen et al 2018; Villeneuve et al 2018). The hepatic steatosis adverse outcome pathway (AOP) network represents a network that spans scales, and includes molecular, cellular, organ-level and organismal level responses (Figure 6). The output of the network is to predict hepatic steatosis. The network is structured to represent the receptors within the liver and how activation of these receptors intersect and direct processes that when - are off balance, could induce fatty liver disease. The modularity of the network is represented by what can be measured in terms of physiological parameters (binding to receptors, and measurements of lipids, etc). The nodes in the network are called key events are largely physiologically derived. The edges are downstream effects after activation, or relationships between key events (metabolome). The strength of association of each node is estimated through Bayesian network analyses and this is a feed forward network. If sufficient perturbation of this network occurs within a specified amount of time, hepatic steatosis will occur. The network exhibits plasticity to a point of departure (at each key event), and then proceeds to the next outcome. There will be individual variability (each person is different), that could be explained by population identifiers. The network is intended to accurately represent and predict how a system will respond to perturbation, even if that

involves some degree of abstraction, simplification, or implicit embedding of more detailed underlying systems understanding (Villeneuve et al 2018).

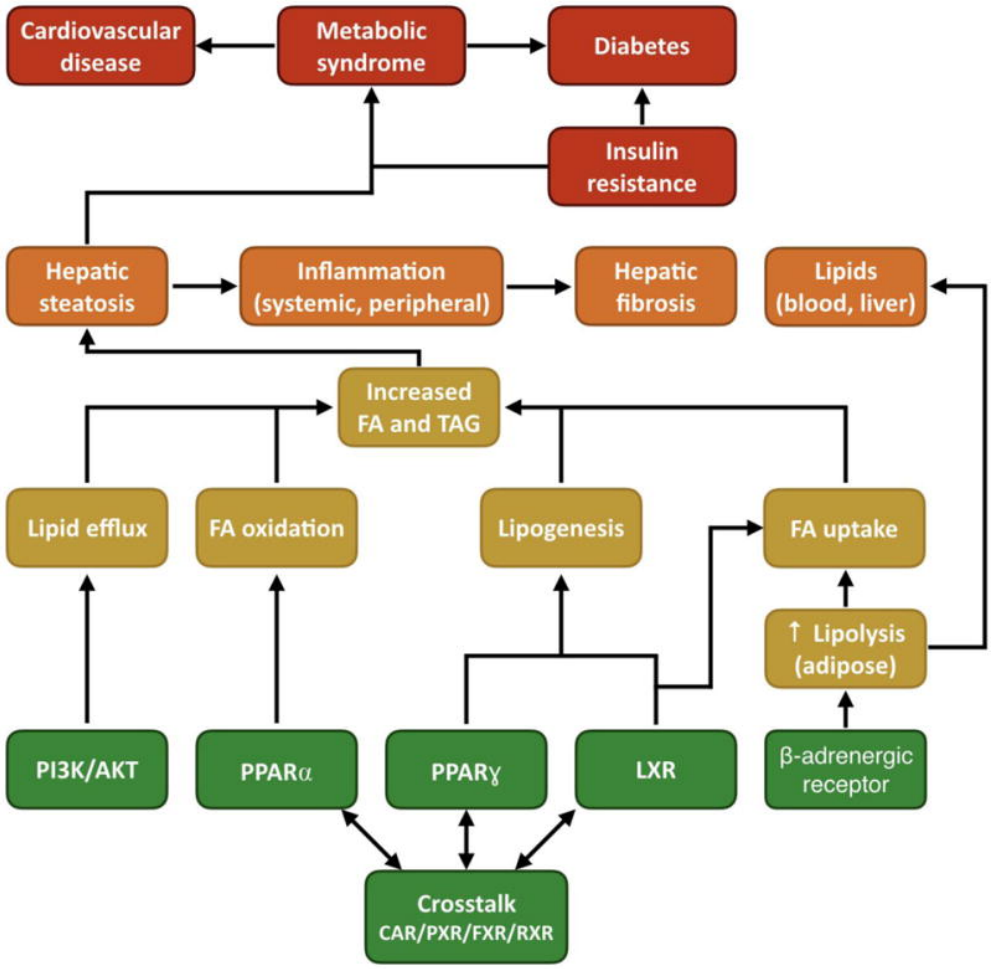
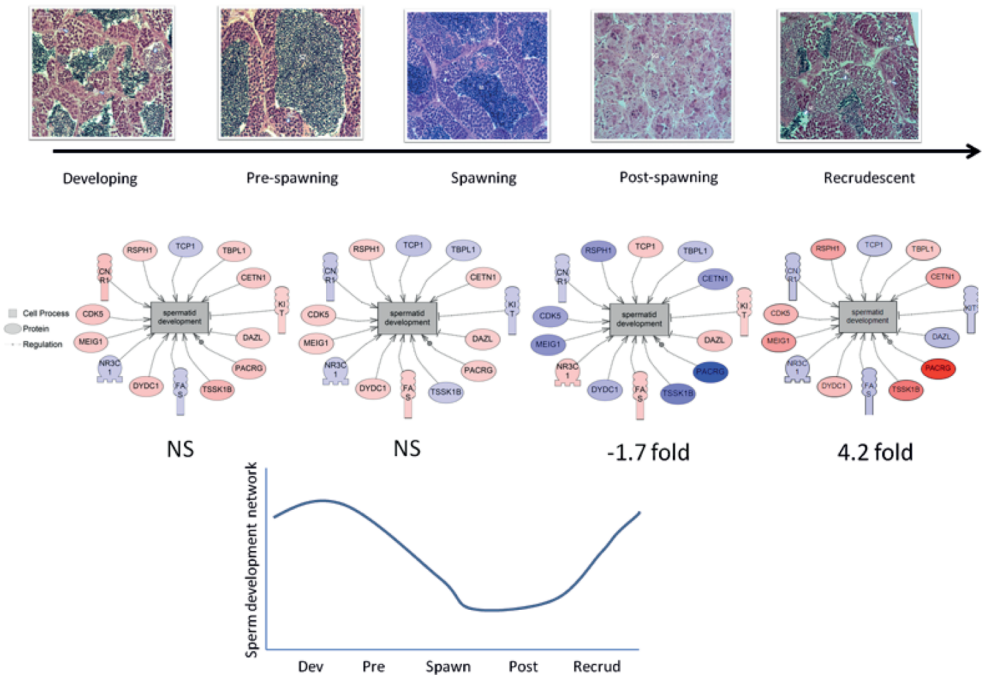


Figure 6. Network representation of metabolic disorders mediated by hepatic steatosis. The network was built to predict events that lead to hepatic steatosis from high throughput assays. The network topology converged into 4 key events (i.e., lipogenesis, and fatty acid uptake, efflux, and oxidation) that were viewed as critical paths leading to steatosis. Assays measuring these points of convergence integrate the complex interplay of upstream events and translate them into measures that are more directly related to the adverse outcome. FA = fatty acid; TAG = triacylglycerol; PI3K = phosphatidylinositol-3-kinase; AKT = protein kinase B; PPAR = peroxisome proliferator-activated receptor; LXR = liver X receptor; CAR = constitutive androstane receptor; PXR = pregnane X receptor; FXR = farnesoid X receptor; RXR = retinoid X receptor. From Knapen et al. (2018).

(iv) <u>Developmental networks</u>. Gonadal growth of male rainbow darter during periods designated as developing, pre-spawning, spawning, post-spawning and recrudescence and the transcriptional network that corresponds with each stage changes and is dependent on structure and function (Figure 7). These data suggest that there are distinct transcriptomics fingerprints for testis stages, and this study provides novel mechanistic insight into molecular signaling cascades underlying sperm maturation in fish (Bahamonde et al 2016).

A



B

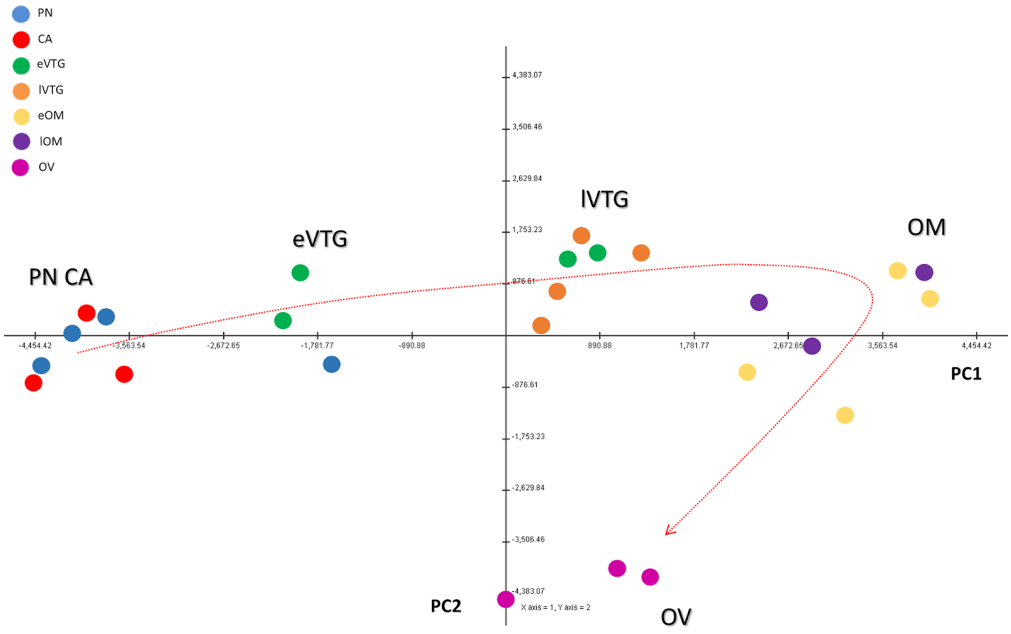


Figure 7. Gene transcriptional networks change as (A) rainbow darter testis undergoes development to maturation (Bahamonde et al 2016), and (B) as largemouth bass undergoes ovarian development - Principal component analysis (PCA) shows a clear progression of ovary development from the first to the terminal stage of ovarian development. Stages are defined as PN (perinuclear), CA (cortical alveoli), eVTG (early vitellogenesis), lVTG (late vitellogenesis), eOM (early ovarian maturation), lOM (late ovarian maturation), and OV (ovulation) (Basili et al 2018).

A gene expression network based on microarray data describing how the gonad develops demonstrates how the network changes as structure and function changes. This particular network is based on one level of organization (the transcriptome) but is classified according to the organ level changes.The genes cluster differently at each stage of gonadal development. Since this is microarray data, and not RNA-seq, some aspects of the network could be missed (Bahamonde et al 2016; Basili et al 2018).

(v) Microbiome networks: A microbiome is a community of microbes (which can include bacterial, protozoal and viral taxa - "virome") that inhabit a particular organ / tissue of a host (typically an animal or plant) - e.g. gut microbiomes are well studied in humans and some animal species, usually focusing on bacterial taxa. NGS technologies enable quantitative descriptions of such communities in great detail, including phylogenetic distinctions below the species level (in any case, the species concept is rather fraught for microbes), delivering relative abundances of thousands of taxonomic units. These microbial communities influence host health and behaviour profoundly, and via a range of different mechanisms, which are only beginning to be understood, as are the ontogeny of microbiomes within their hosts, and their dynamics throughout the host's lifetime. The responses of microbiome communities to perturbations, such as antimicrobial agents, infections, or changes in host diet are of particular relevance to understanding their relevance to host health, and harnessing this knowledge for therapeutic use. Microbiome communities are well represented as networks of species, characterized by co-occurrence, though typically interactions of ITUs are not explicitly measured. Nonetheless, exploring associations between microbiome structure and e.g. robustness vs plasticity over time and under different regimes of disturbance / perturbation could be a powerful approach to understand patterns of health and disease, across different host species and disease phenotypes, as driven by variation in microbiomes.

(vi) Networks of populations on the landscape: Natural populations often occur as fragmented metapopulations - networks of populations linked by dispersal and migration. Fragmented population structure may occur naturally, due to patchy distribution of suitable habitat, such as mountaintops, ponds, or in the case of humans and their animals, cities and farms. In addition, anthropogenic transformation can alter the structure of population networks, increasing or decreasing the movement of organisms among patches (connectivity). For example, human traffic can connect populations by translocating organisms, while habitat loss can isolate populations in protected areas or climatic refugia. Understanding how changes in population network topology affect the resilience / robustness of the component populations to environmental change (also: disease spread) is an increasingly urgent priority, as we continue to launch inadvertent experiments manipulating landscape connectivity.

Desert bighorn (DBH) sheep present a compelling model system: DBH inhabit mountain ranges where higher precipitation and lower temperatures provide higher forage quality, and where steep, open terrain allows them to visually locate and avoid predators. DBH are thus segregated into relatively independent populations by the naturally fragmented distribution of

mountainous terrain, creating a metapopulation-like structure in which local population sizes range from tens to a few hundred individuals and genetic drift is strong but variable. Population extinction and recolonization have been observed, and extinction varies with elevation, precipitation, and access to water.

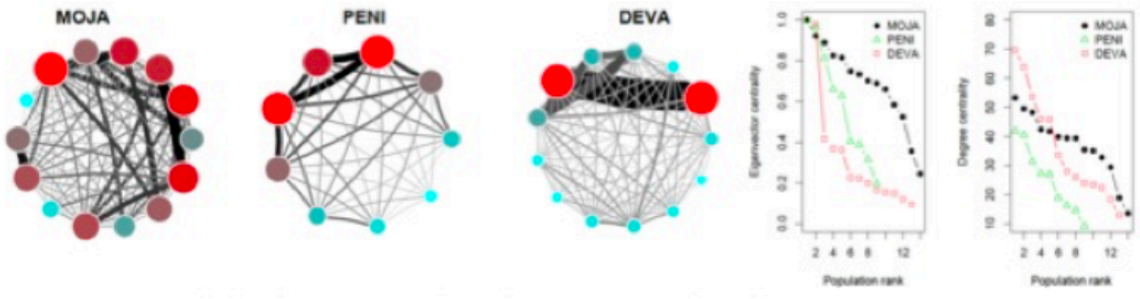


Figure 8. The Mojave (MOKA), Death Valley (DEVA) and Peninsular (PENI) networks vary in network metrics. Nodes in the network represent populations: node size and color are proportional to eigenvector centrality. Edge weight is proportional to levels of gene flow (Nm).

Desert bighorn networks defined by observed levels of gene flow (Nm) vary in topology, and populations within networks vary in centrality (Figure 8). The Mojave and Death Valley networks are similar in size, but populations in the Mojave are more connected than in Death Valley. Centrality in the DV system is far more polarized, with just two very strongly connected populations contrasting 11 fairly isolated ranges; whereas in the Mojave, the gradient in population centrality is much smoother. The Peninsular Range network is smaller, and has an intermediate number of strongly connected populations compared to the MOJA and DEVA networks, with slightly weaker connectivity overall compared to the other two networks. Key questions: which networks are more resilient to environmental perturbations of different types - from climatic variation to invasion of infectious agents?

(vii) <u>Saltmarsh</u>: A saltmarsh ecosystem is a complex network of interacting species with various environmental inputs of varying importance and with stabilizing feedbacks. It's an open system with almost infinite complexity. However, the foodweb of a saltmarsh can be described simply, provided one accepts that the microbial, algal, and detritivore communities each can be aggregated (e.g. Fig 9). For example, the microbial community is typically represented as one node representing the decomposers. In fact, the higher plant community actually is simple as it is dominated by a monoculture of the grass *Spartina alterniflora*. The primary producers are primarily limited by nitrogen and, commonly, secondarily by phosphorous. They also respond positively or negatively to flooding, depending on the relative elevation of the marsh surface within the tidal frame (its position between high tide and mean sea level.

We know of salt marshes that have existed for 4 millennia, and this has been possible because of negative feedback between the higher plants and flooding (Morris et al. 2002). The feedback can be positive and destabilizing if the rate of sea-level rise is too rapid, and this is becoming a problem today as sea-level rise accelerates. But, focusing on the negative

feedback, we know that the plants respond positively with greater NPP when sea level rises, provided the relative elevation of the marsh is high. When NPP rises, biogenic soil volume and sediment trapping increase, which raises the elevation of the marsh, maintaining equilibrium.

Feedback also exists between the plants and their microbiome (Fig.10). The most important decomposers in marsh sediments are sulfate reducing bacteria (Howarth and Teal 1979). Marsh sediments are anoxic. The sulfate reducing bacteria use sulfate as an electron acceptor. The plants release acetate from their roots, probably in proportion to their NPP. Acetate is a perfect substrate for sulfate reducing bacteria.

Sulfide is a byproduct of sulfate reduction. The sulfide reacts with insoluble Fe-P compounds to form insoluble FeS or FeS_2, which releases P precisely at the time of year with plant growth is maximal. Moreover, the sulfate reducers possess a NIF gene, and probably the fixation of nitrogen also increases when the sulfate reducers are stimulated. The details are not entirely clear, but we know for certain that the concentration of these rate limiting nutrients of N and P cycle seasonally and are synchronized with the growth rates of the primary producers.

The result of these feedbacks is a stable (within bounds) system that has been remarkably resilient in the face of rising sea level. Stabilizing feedbacks can be found in other complex systems, but apparently not in simple systems. Feedback is an emergent property of complex systems.

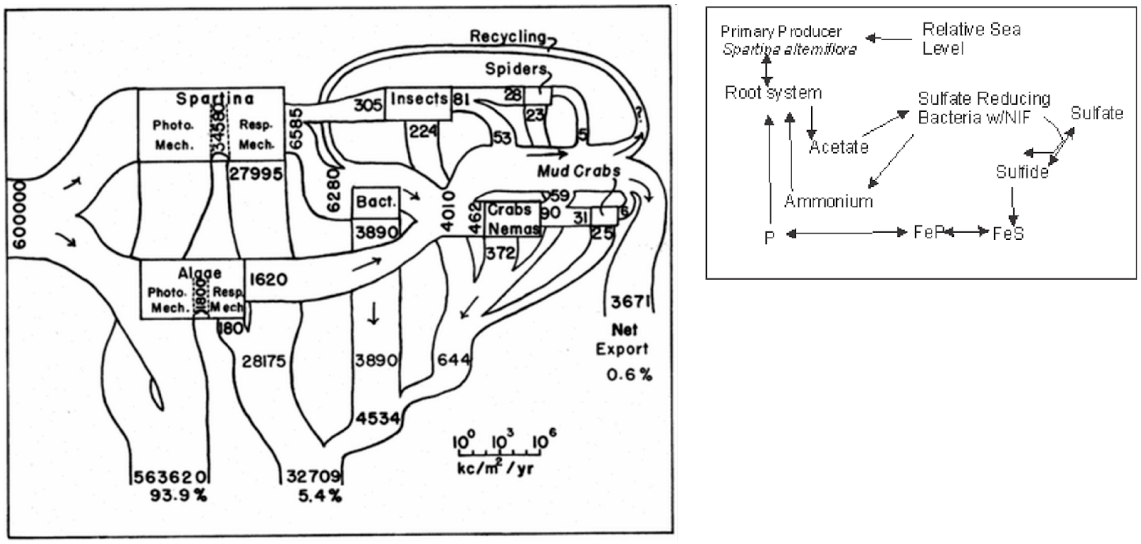


Figure 9. The simplified energy flow in a saltmarsh ecosystem (from John Teal 1962). Note the high degree of aggregation of trophic levels. The inset in the right shows a diagram of the saltmarsh network connections that result in stabilizing feedback.

(viii) <u>Simple and hierarchical systems of neurons that provide various levels of network complexity</u>: It is no accident that certain artificial computational networks are referred to as "neural nets," due to their assumed similarity to connections established by systems of nerve cells. However, few biological neuronal connectivities have been reverse-engineered to

predictive computational networks. An exception is Donald Hebb's introduction of associative learning networks based on synaptic (nodal) strengthening (Herz et al. 1988) that derive from a relatively simplistic but still relevant view (then, in 1949) of hippocampal organization, and which is also applicable to the iconic learning and memory centers of the *Drosophila* brain (Heisenberg 1993) and those of other panarthropods.

Yet, as we know from many descriptions of brains of both chordate and invertebrates every functional domain of the brain is defined by its characteristic network arrangement – the patterned synaptic connections amongst its constituent neurons, and its connections from and to other domains. Furthermore, certain of these functional domains have matching roles in vertebrates and nonvertebrates; these show genetic, structural, pathological and functional correspondence, which taken together simply genealogical correspondence: hence phenotypic and genotypic homology: thus, an origin in deep time before the divergence of Ecdysozoa and Deuterostomata.

Presently, the most interesting of such "real" neural networks resides in the most anterior region of the brain: in vertebrates it is the basal ganglia and hippocampus; in panarthropods it is the "central complex" and mushroom bodies. Basal ganglia and central complexes serve to coordinate motor actions by editing outputs from the brain by orchestrating systems of inhibitory connections that selectively gate those outputs relevant to a required behavior. These are permitted to relay information to the appropriate effectors (motor neurons to muscle). Genetic deletions, or interventions of specific dopaminergic modulators in the network leads to Parkinson's-like pathologies in both arthropod and vertebrate (Drosophila and mouse). Mushroom bodies and hippocampus form long term associations relating to the memory of place and experiences.

The "central complex" comprises 16-18 discrete computational modules supplied by high-level sensory inputs; modules assess the bilateral weighting of sensory percepts to provide appropriate signals to controllers - the inhibitor neurons that gate motor actions. At one extreme this ground pattern organization exists in the smallest and simplest hexapods (Collembola, that inhabit leaf litter on the forest floor) and in the most dexterous climbing raptorial species, praying mantis; differences between the two being the number of constituent neurons, and hence connections, that comprise each module. These neurons originate from a defined set of stem cells (ganglion mother cells) that repeatedly divide to provide those neurons. In Collembola their divisions provide 2-4 neurons/module; in praying mantis their divisions provide scores of neurons/module. Precision of connections across the modules further reflects dexterity: invariant precision of a praying mantis, but noisy connectivity on a related species with moderate dexterity, such as a cockroach.

In *Drosophila*, this network has been extensively studied both using optogenetics and electrophysiology. The associated behavioral repertoire is documented (Seelig and Jayaraman,2013; Wolff and Rubin, 2018). The central body is thus a paradigmatic neural network ready for deeper study using mathematical network analysis. Prediction of network

activity under precise parameters can be compared with experimental data. Outputs generated by the artificial network will be testable using behavior or optogenetics.

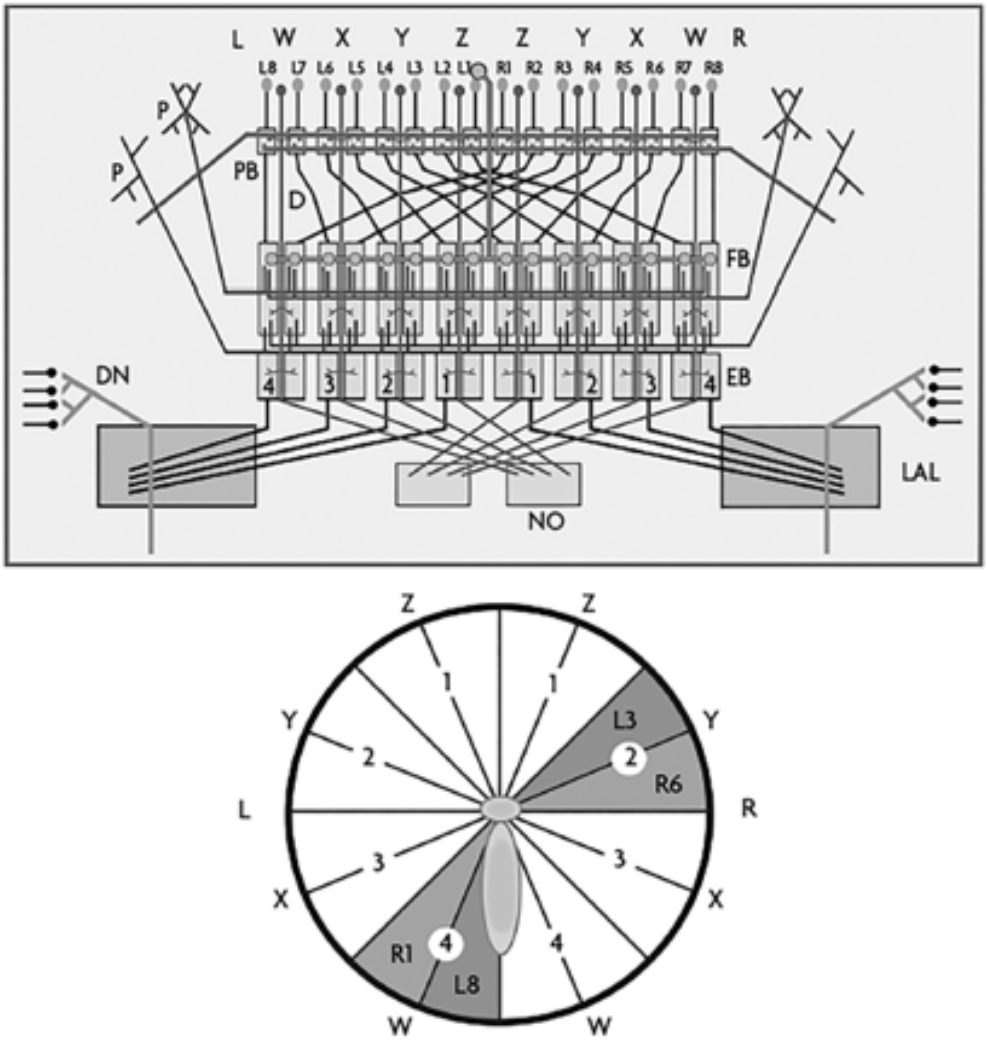


Figure 10. The entire sensory surround of the organism is represented in the brain's "central complex" diagrammed here. Principal components and projections of columnar neurons originating from the protocerebral bridge (upper module row formed by the progeny of W, X, Y and Z stem cells provide modules to the left (L, L8–L1) and right (R, R1–R8) of the midline. Each provides connections to the next modular assembly into successive structures: the fan-shaped (FB) and ellipsoid bodies (EB). Information from computations within the PB, FB and EB are relayed to systems of decussating axons that extend into the lateral accessory lobes (LAL), where they interact to gate the activity of premotor descending neurons (DN). The proposition here is that on each side of the protocerebral bridge a module represents one-eighth of half of the sensory envelope around the insect. Modules nearest the midline of the protocerebral bridge (L1 and R1) represent the front of the envelope; the most lateral modules of the protocerebral bridge represent the back of the envelope. The representation in module L4 of the central body is supplied by axons from the protocerebral bridge modules R1 and L8, representing the most posterior segment on the left of the animal and the most anterior segment on the right. The left module 2 of the central body is supplied by axons from modules L3 and R6 from the bridge, representing the segment 90° on the right and 90° on the left.

The “central complex” is just one example of a defined network comprised of discrete modules and interactive nodes. Another is the paired ‘mushroom bodies,” which because they also have homologues across Phyla represent additional ancient networks, likely originating about 547 million years ago according to the record of “trace” fossils that have recorded behaviors of bilateral animals for which there is no other fossil record.

Indeed, the modular composition of entire Drosophila brain is amenable to network modelling at various levels of elaboration: networks comprising “simple” nodes; networks comprising nodes that themselves are composed of small nodes (having fewer connections), and so on.

As experimentally approachable as the central complex network, the insect mushroom bodies have been recognized since 1850 as neural centers required for learning and memory. Largest in social insects and obligate parasitoid insects, these centers are now recognized as being crucial for allocentric memory (memory of place) and for plastic long-term memory where learned associations are modifiable post hoc by sensory inputs representing positive or negative valences. Further, the organization of neurons in the mushroom bodies correspond in detail to that hallmarking the vertebrate hippocampus, with which the mushroom bodies share immunological correspondence as well as corresponding multiple gene expression (Wolff and Strausfeld 2016).

The mushroom bodies, like the hippocampus, comprise orthogonal arrangements of intersecting neurons that comprise a Hebbian-like network. This moniker derives from the work of the Canadian Psychologist Donald Hebb who postulated that a neuron’s propensity to relay information (efficacy) depends on its persistent stimulation by a presynaptic drive; and when two neurons converge on the neuron and provide coincident inputs these can be sufficient to cause a permanent change in the efficacy of the postsynaptic cell’s synapse: thus, a gain of synaptic strength in response to presynaptic association. Hebb’s theoretical work was immediately attractive to researchers working on cortical systems, hippocampus and other centers that were already known to participate in short and long term memory (e.g. Frolov and Muravév 1993). Work on honey bee learning and memory, and especially on *Drosophila*, provides the most accessible system to investigate whether the Hebbian-type associative model actually applies to real-world biological learning networks expected to underlie, for example, cortical or hippocampal functions.

Neuroanatomical studies reveal that the intrinsic organization of the mushroom body is organized pretty much as would a massive Hebbian network, consisting as it does of orthogonal arrangements of local interneurons intersected by converging inputs encoding various types of unimodal sensory data and output neurons that relay constructed multisensory associated information.

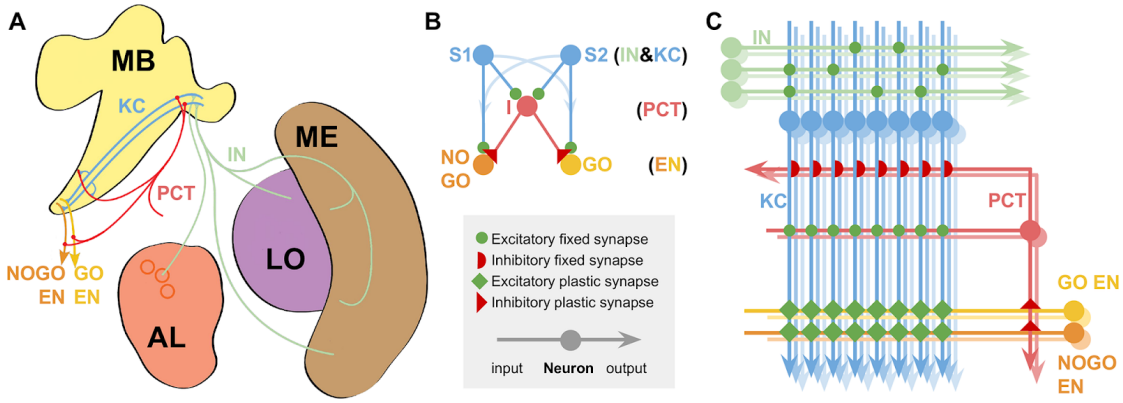


Figure 11. Models of the mushroom bodies based on known neuroanatomy. A. Neuroanatomy: MB Mushroom Bodies; AL Antennal Lobe glomeruli (circles); ME & LO Medulla and Lobula optic neuropils. The relevant neural pathways are shown and labelled for comparison with the model. **B.** Reduced model; neuron classes indicated at right and side of sub-figure. **C.** Full model, showing the model connectivity and indicating the approximate relative numbers of each neuron type. Colour coding and labels are preserved throughout all the diagrams for clarity. Excitatory and inhibitory connections indicated as in figure legend. Key of neuron types: KC, Kenyon Cells; PCT, Protocerebellar Tract neurons; IN, Input Neurons (olfactory or visual); EN, Extrinsic MB Neurons from the GO and NOGO subpopulations, where the subpopulation with the highest sum activity defines the behavioural choice in the experimental protocol. Figure and legend from Cope et al. (2018).

Figure 11 schematizes neural components contributing to such multisensory associations. Different modalities (e.g. visual from the visual centers (ME, LO) or olfactory from the antennal center (AL)) encode high level sensory data that can contribute to sensory associations mediated by Hebbian type circuits (panel B) provided by thousands of parallel fibers (panel C) that intersect these sensory inputs (Huerta et al. 2014). Short term synaptic plasticity is achieved by converging sensory inputs inducing a strengthening (positive - GO) or weakening (negative - NOGO) modification of synaptic sites that signal to output neurons. Permanent reinforcement (long term memory) may be established by repetitive convergent inputs to the networks leading to suppression or facilitation of circuits contributing to the release or suppression of downstream motor actions.

Mushroom bodies comprise thousands of such networks, many of which are clustered together in discrete domains, suggesting hub-like organizations of learning modules. While much is known about the physiology of discrete subsets of neurons in these centers, what is not known is the rules underlying how these subsets these interact with each other such that memories interact, achieve contextual valences, and form post hoc memory modifications: all functions expected in organisms that obtain temporal (clock-related) understanding of dynamic ecologies. Despite this level of ignorance what is recognized from behavioral studies is that, across species, memories are infinitely plastic, even manipulable.

Current studies on mushroom bodies are focused on connectomics: the total reconstruction of neural network using serial section reconstruction of every one of the approximately 2,000 parallel fibers and all their synaptic interactions with incoming and outgoing neurons (Eichler

et al., 2017). The many terabytes of data representing hypercomplex network organization present interesting challenges in interpretation and understanding these memory systems in terms of reconstructing functional “real world” representations that can explain and indeed imitate sensory associations and memory acquisition.

Barriers and challenges

The “networks across scales” grand challenge attempts to find common network structures and/or common network dynamic behaviors that unify biological systems across levels of organization. But how can we find organizing principles that are common across biology when systems range from interactions of genes or metabolites to descriptions of entire ecosystems? Such a grand objective of finding common organizing principles that span molecular makeup to planetary macrostructure is limited by a multitude of barriers that must be overcome. For example, an important barrier is the actual diversity of network nodes and links. This diversity must be defined when studying and comparing systems. Because life requires explaining continuous change and a multitude of overlapping processes, a framework of causal explanations has the potential to uncover life’s multilayered complexity. We could call these processes ‘activities’ and the temporal ordering of dependencies between complexity layers ‘causation.’ Within this philosophical framework, nodes can represent the structure and dynamics of *immanent* entities (events) that span the spatiotemporal confine or *transcendent* entities that are abstract in nature. We can call these nodes “causal relata” and the directed links that connect them “causal relations”. Beginning with David Lewis, causal networks have been modeled by incorporating probabilistic or Bayesian network approaches and causal and counterfactual inference (Pearl 2000). These kinds of approaches are powerful. They are currently impacting the emerging field of Artificial Intelligence (AI). However, effective integration approaches must be sought, perhaps using experiments, predictive computational methods, theoretical and mathematical frameworks, and the exploration of functions and constraints with philosophical approaches.

Another barrier is the meta-complexity of the systems that must be modeled. For example, nodes can represent a variety of entities: objects, agents, relationships, scaffolding, events, dynamics, and aggregations. To illustrate, proteins in protein-protein interaction networks can be considered objects but also agents. Molecular functions in the direct acyclic graphs of Gene Ontology can be considered events. Similarly, links can become structured, revealing complexity in biological networks (Ahn et al. 2010). Link communities thus express additional meta-complexity. Can all these entities be scale invariant? Would it be possible to develop a common vernacular? If so, would there be a way to classify specific node identifiers? It is here where epistemology and ontology must interface.

Biological systems are structured. The behavior, interactions and goals of subsets of parts may differ from the rest of the system. One kind of structure that is common is the ‘module’. Modules are sets of integrated parts that cooperate to perform a task and interact more extensively with each other than with other parts or modules of the system (Hartwell et al. 1999). Modules are generally defined within structural, functional, and historic contexts. For example, a structural domain of a multi-domain protein is a higher-level structural module. Domains are arrangements

of elements of secondary structure that fold into well-packed and compact structural units of the polypeptide chain. Domains are also functional modules. They fold and function largely independently, contribute to overall protein stability by establishing a multiplicity of intramolecular interactions, and generally host specific molecular functions. More importantly, domains are also evolutionary units. They have been shown to be evolutionarily conserved and present in different molecular and functional contexts throughout the protein world. Since many networks study how modules organize into systems, the contextual definition of a module poses a problem for constructing biological networks. Modules are also at the heart of our understanding of robustness, the capacity of a biological entity to persist under the uncertainties of change. Can we generate a general theoretical framework for biological modules across spatial, functional and temporal scales? Since modularity appears linked to hierarchy in biological systems (reviewed in Caetano-Anollés et al. 2019), what are the evolutionary drivers of hierarchical modularity in network structure?

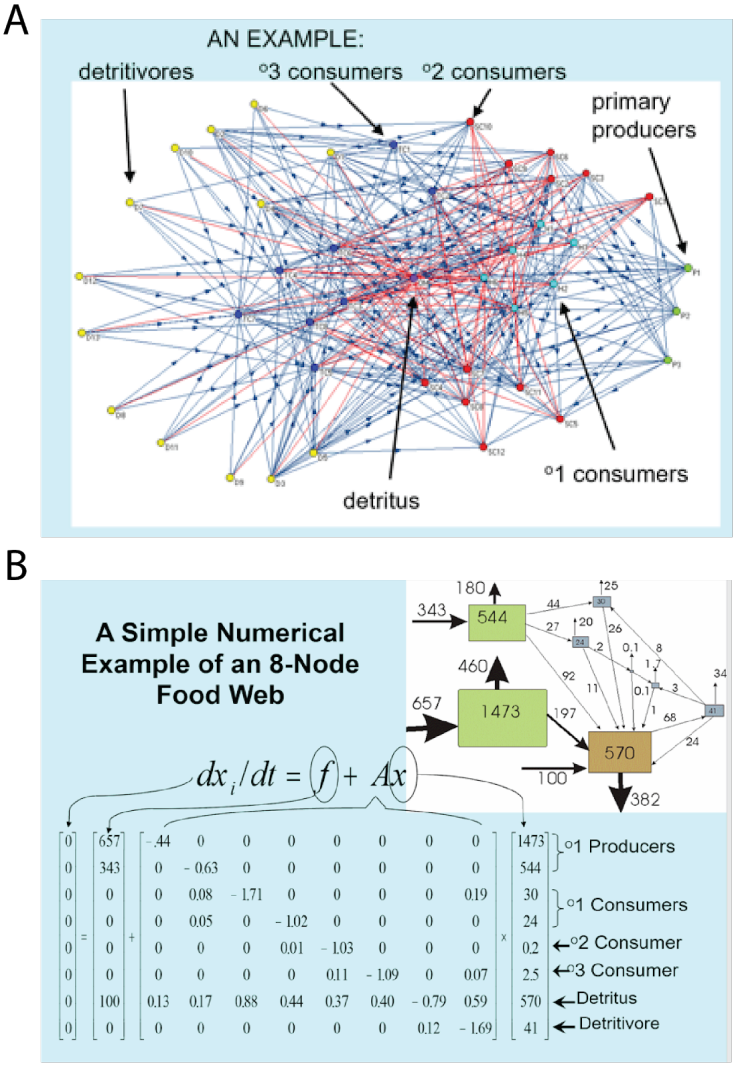


Figure 12. Generating artificial food webs by in silico modeling. A. Foodweb generated by populating a transfer matrix with transfer coefficients and solving for the equilibrium solution. **B.** Methodology used to generate modeled food webs (described in the text).

Barriers to describing very complex networks (e.g. ecosystems) can be overcome by analyzing the properties of random networks generated in silico and using what we learn to understand real networks. FIgure 12A shows an example of a feasible food web generated by populating a transfer matrix with transfer coefficients and solving for the equilibrium solution. A network is feasible if the solutions are all positive. The methodology is illustrated in Figure 12B. After the matrix dimensions are set, the random inputs (f) and transfer coefficients (A) are generated, and the solution to dx/dt=0 is determined. The foodweb is a feasible one if the solution (x's) are positive. We can ask questions about connectivity and total system throughput (TST), stability, ascendency (Ulanowicz 1980), fractal dimension, and size. We posit that we can arrive at generalities about real networks by analyzing the properties of artificial networks.

From a universe of >5,000 random food webs composed of as many as 2,200 taxa, Morris et al. (2005) found that the probability of generating a feasible network declined rapidly as the number of taxa exceeded 400 and flow diversity increased asymptotically, i.e., flows became more uniform. The average number of major flows per taxon (flows greater than 5% of the total input flows) was 2.1, similar to those of real food webs. Gross primary production, which was limited to 1,000 kcal m-2 yr-1 imposes a limit on flow in a manner similar to the resource space that MacArthur (1957) argued should limit the distribution of species. Ulanowicz (2002) used an information-theoretic homolog of the May–Wigner stability criterion to posit a maximal connection per taxon of about 3. These examples suggest there is a fundamental relationship between network structure and function and show the utility of artificially generated networks.

The development of case studies that explore and look for common threads in the structure and dynamics of networks could be promising. Commonalities that are predictive for example along economy, robustness or plasticity axes or within morphospaces could be identified and then extended to the study of a broader range of systems. However, the methodological problem of 'gappy' or incomplete data sets and the issue of 'snapshots' complicate any endeavor. Following the genomic revolution, biology has been able to define entire repertoires of biological entities (e.g. genes, metabolites, fold structures, molecular functions). While certain explorations have been comprehensive many others are lagging behind. For example, the universe of proteins can be described with a finite set of folds and fold superfamilies summarizing the overall 3-dimensional atomic design of proteins. The SCOP (Murzin et al. 1995) and CATH (Orengo et al. 1997) databases, the gold standards of protein classification, show protein folds group into 2,026 SCOP (http://scop.mrc-lmb.cam.ac.uk) and 6,631 CATH (https://www.cathdb.info) superfamilies. These numbers have reached a plateau, strongly suggesting that most structural designs have been sampled through structural genomic efforts. In sharp contrast, the world of species and our understanding of the Tree of Life is far from complete (Hug et al. 2016). Considerable 'dark matter' exists at both the level of cellular organisms and viruses. These uncertainties raise a number of important questions. Are networks biased by the experimental knowledge or focus on individual components and are there situations where key nodes are not represented because nobody has really studied them? Are there methods that can identify gaps or normalize over emphasized nodes? Another methodological problem is the issue of 'snapshots'. Numerous experimental approaches provide single measures within a continuum of change. For example, the crystallographic acquisition of 3-dimensional atomic structures has been stored in the RCSB

Protein Data Bank (PDB) repository (https://www.rcsb.org). As of December 23, 2019, there are 158,787 biological macromolecular structures available in the database, which has been growing at a significant pace (>10,000 PDB entries per year). Despite these significant accomplishments, PDB entries represent conformational 'snapshots' that give little justice to the conformational molecular landscape of proteins and nucleic acids. There is now hope that cryogenic electron microscopy (Cryo-EM) may pave the way to wide-encompassing conformational views. This example highlights the problems of acquisition of longitudinal data that can describe the dynamics of numerous biological processes at different timescales. Consequently, there will be a need for analytical tools that can manage 'big data', including longitudinal datasets, and can make use of different data flows in a unified methodological framework.

Finally, there is the problem that not all data types can be modeled with networks, and the issue of using networks across biological scales. Simplification must occur if information from multiple levels of biological integration are incorporated into a network (eg. hepatic steatosis), or if the network changes over time because of development or evolution, and a rigorous evaluation of the assumptions and rules underlying network simplification is required.

Broader Impacts

Studying biological networks across scales is by definition broad impact in terms of the immediate knowledge that it generates from a large-scale study. The practicalities of constraining this to a tractable approach develop new algorithmic techniques to link information, determine the influence of different levels of noise on the knowledge produced from that information, and the reliability of that knowledge. While leading to a set of rules, it allows those rules to be defined in their applicability and rigor. The approach uses nature as the data set to define how a system works. Where theoretical modeling does not agree with experiment, it helps find signal in noise and defines areas where new knowledge is awaiting discovery.

Nature has had a long time to conduct its own system experiments. By studying the nature of how those systems develop and interact across different scales it allows a more concrete understanding of the impact of perturbations on those systems, whether it be a large-scale shift in environment, e.g. ocean pH, average temperature shifts, advance of an invasive species, or small scale such as the extinction of a rare species, or the mutation of an amino acid. This in turn sets guidelines to prioritize the response to these changes so that resources can be devoted to mitigate influences that cause the maximum impact.

The nature of the study extends beyond biology. Nature can be seen as the ultimate laboratory setting to test network and systems performance with the experiment having the ultimate metric of success - life or extinction. The results and rules established can be extended to non-biological systems, e.g. redundancy in automation, self-organization for transport within a city, response to perturbation in a system, transient approaches that activate, etc. It is not too strong to say that this could lead to a totally new approach to network and systems science in both the physical world, but also in the computational arena.

Reintegrating biology

To effectively study a network across scales, a network of experts in each of those scales (and individual research areas) needs to be created. A common language is needed to link those experts and a backbone organization established to ensure that the effort is focused on the questions and not the administration. This mirrors the concept of collective impact where a common agenda, shared measurement systems, mutually reinforcing activities, continuous communication, and a backbone organization, maximize limited resources to produce maximal output (Kania and Kramer 2011). By design, formulation around a collective impact model reintegrates separate disciplines and expertise into a common goal.

The common agenda is to establish a framework that provides:

- Longitudinal empirical network data across a broad range of biological systems and scales, ideally including both observational and experimental approaches.
- Analytical expertise to analyze these datasets asking common questions and using common tools.
- Modeling expertise to construct parallel sets of general network dynamic models, putting into context and providing generality to the set of empirical studies.
- Space-time for empirical and theoretical project leaders to come together to synthesize findings, identifying commonalities and differences across systems.
- Measurable outcomes to test, improve, and verify the approach.

A shared measurement system necessarily requires a shared language across different disciplines. There are ontology approaches to this that help understanding of the results but guiding the experimental and analysis approach is more difficult. As a scientific endeavor we are more used to constructing hypotheses and testing those hypotheses - the scientific method. We must ask ourselves which aspects of information need to be retained to link biological scales. For example, if we are trying to understand the dynamics of a microbiome community, and/or its outputs that affect the host: Is it taxonomic composition that is the most informative, or is it transcript or protein products of the microbial community? This could potentially be addressed by constructing competing hypotheses (or different networks) that essentially represent the same community but using different data flows, and then asking which of the networks presents predictable dynamics or best predicts outputs.

Mutually reinforcing activities are critical. With multiple disciplines involved in a common goal those disciplines must communicate to interact. This requires physical interaction (scientific meetings), educational interaction (common training), and knowledge interaction (summaries of the knowledge produced as it is produced). The resources of the effort must be understandable by all, at least at the most basic level of being able to know what they are, how to use them, and what to look for in the output.

Continuous communication is linked to mutually reinforcing activities. For maximum efficiency in understanding a network of disparate information across scales and times, communication is

critical. That includes the free flow of information, the establishment of mutual respect and trust between different research thrusts, and output that the interested public can follow to understand progress that is being made.

Finally, the most important part is backbone support. This includes a strategic leadership that sets the goals and guides the direction, monitoring of progress in meeting goals, provision of resources that can help achieve goals, and maintaining the common direction, language, communication, and legacy involved in producing and preserving the knowledge produced.

Reintegrating biology is a necessity to study biological networks across scales.

References

Ahn YY, Bagrow JP, Lehmann S (2010) Link communities reveal multiscale complexity in networks. Nature 466:761-765.

Angrish MM, Kaiser JP, McQueen CA, Chorley BN (2016). Tipping the balance: hepatotoxicity and the 4 apical key events of hepatic steatosis. Toxicological Sciences *150*(2):261-268.

Arabidopsis Interactome Mapping Consortium (2011) Evidence for network evolution in an *Arabidopsis* interactome map. Science 333:601-607.

Aziz MF et al. (2012) Stress induces biphasic-rewiring and modularization patterns in metabolomics networks of *Escherichia coli*. IEEE Intl Conf Bioinf Biomed. 2012:593-597.

Aziz MF, Caetano-Anollés K, Caetano-Anollés G (2016) The early history and emergence of molecular functions and modular scale-free network behavior. Sci Rep. 6:25058.

Bahamonde PA, McMaster ME, Servos MR, Martyniuk CJ, Munkittrick KR (2016). Characterizing transcriptional networks in male rainbow darter (*Etheostoma caeruleum*) that regulate testis development over a complete reproductive cycle. PLoS One *11*(11):e0164722.

Balaban V, Lim S, Gupta G, Boedicker J, Bogdan P (2018) Quantifying emergence and self-organisation of Enterobacter cloacae microbial communities. Sci Rep 8: 12416.

Barabási A-L, Oltvai ZN (2004) Network biology: understanding the cell's functional organization. Nature Reviews 5:101-113.

Basili D, Zhang JL, Herbert J, et al. (2018) In Silico Computational Transcriptomics Reveals Novel Endocrine Disruptors in Largemouth Bass (*Micropterus salmoides*). Environmental Science Technology *52*(13):7553-7565.

Bassett DS, Greenfield DL, Meyer-Lindenberg A, Weinberger DR, Moore SW, Bullmore ET (2010) Efficient physical embedding of topologically complex information processing networks in brains and computer circuits. PLoS Comput Biol 6(4):e1000748.

Bogdan P (2019) Taming the unknown unknowns in Complex Systems: Challenges and opportunities for modeling, analysis and control of complex (biological) collectives. Front Physiol 10:1452.

Broder A, Kumar R, Maghoul F, Raghavan P, Rajagopalan S, Stata R, Tomkins A, Wiener J (2000) Graph structure in the Web. Comp Networks 33:309-320.

Caetano-Anolles et al. (2019) Emergence of hierarchical modularity in evolving networks uncovered by phylogenomic analysis. Evol Bioinformatics 15:1176934319872980.

Cope AJ, Vasilaki E, Minors D, Sabo C, Marshall JAR, Barron AB (2018) Abstract concept learning in a simple neural network inspired by the insect brain. PLoS Comput Biol 14(9):e1006435. https://doi.org/10.1371/journal.pcbi.100643

Corominas-Murtra et al. (2013) On the origins of hierarchy in complex networks. Proc Natl Acad Sci USA 110:13316-13321.

Daniels BC, Kim H, Moore D, Zhou S, Smith HB, Karas B, Kauffman SA, Walker SI (2018) Criticality distinguishes the ensemble of biological regulatory networks. Physical Review Letters 121(13):138102.

Eichler, K., Li, F., Litwin-Kumar, A. et al. (2017) The complete connectome of a learning and memory centre in an insect brain. Nature 548:175–182.

Frolov AA, Muravév IP (1993). Informational characteristics of neural networks capable of associative learning based on Hebbian plasticity. Network 4:495–536.

Ghorbani M, Jonckheere E, Bogdan P (2018) Gene expression is not random: Scaling, long-range cross-dependence, and fractal characteristics of gene regulatory networks. Front Physiol 9:1446.

Girvan M, Newman M EJ (2002) Community structure in social and biological networks. Proc Natl Acad Sci USA 99:7821–7826.

Gupta G, Pequito S, Bogdan P (2019) Learning latent fractional dynamics with unknown unknowns, American Control Conference.

Hartwell LH, Hopfield JJ, Leibler S, Murray AW (1999) From molecular to modular cell biology. Nature 401: c47–c52.

Hebb D O (1949) The Organization of Behaviour. Wiley, New York, NY.

Heisenberg M (2003) Mushroom body memoirs: from maps to models. Nature Neurosci Rev 4:266–275.

Herz A. et. al. (1988). The Hebb Rule: Storing Static and Dynamic Objects in an Associative Neural Network. EPL 7 663

Ho JJ, Navlakha S (2018) Evidence of Rentian scaling of functional modules in diverse biological networks. Neural Comput 30(8):2210-2244.

Howarth, RW, Teal JM (1979) Sulfate reduction in a New England salt marsh. Limnology and Oceanography 24: 999–1013.

Huerta R, Nowotny T, García-Sanchez M, Abarbanel HDI, Rabinovich MI (2004) Learning Classification in the Olfactory System of Insects. Neural Computation 16:1601–1640.

Hug LA, Baker BJ, Anantharaman K, et al. (2016) A new view of the tree of life. Nat Microbiol 1:16048.

Ito et al. (2013) Systematic analysis of neural projections reveals clonal composition of the *Drosophila* brain. Current Biology 23 (8):644-655.

Kania J, Kramer M (2011) Collective Impact. Stanford Social Innovation Review 36-41.

Kim H, Smith HB, Mathis C, Raymond J, Walker SI (2019) Universal scaling across biochemical networks on Earth. Sci Adv 5(1):eaau0149.

Koorehdavoudi H, Bogdan P (2016) A statistical physics characterization of the complex systems dynamics: Quantifying complexity from spatio-temporal interactions. Sci Rep 6:27602.

Knapen D, Angrish MM, Fortin MC, Katsiadaki I, Leonard M, Margiotta-Casaluci L, ... Zhang X (2018) Adverse outcome pathway networks I: development and applications. Environmental Toxicology Chemistry 37(6):1723-1733.

Landman BS, Russo RL (1971) On a pin versus block relationship for partitions of logic graphs. IEEE Transactions on Computers C-20(12):1469–1479.

Ma H-W, Zeng A-P (2003) The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics 19(11):1423–1430.

MacArthur RH (1957) On the relative abundance of bird species. Proc Natl Acad Sci USA 43:293-295.

Mahmoodi K, West BJ, Grigolini P (2017) Self-organizing complex networks: individual versus global rules. Front Physiol 8: 478.

Morris JT, Christian RR, Ulanowicz RE (2005) Analysis of size and complexity of randomly constructed food webs by information theoretic metrics. pp. 73-85, In: Aquatic Food Webs: an Ecosystem Approach; Belgrano A, Scharler UM, Dunne J, Ulanowicz RE (Eds), Oxford University Press.

Morris JT, Sundareshwar PV, Nietch CT, Kjerfve B, Cahoon DR (2002) Responses of coastal wetlands to rising sea level. Ecology 83:2869-2877.

Mughal F, Caetano-Anollés G (2019) MANET 3.0: Hierarchy and modularity in evolving metabolic networks. PLoS One 14(10): e0224201.

Mukhtar et al. (2011) Independently evolved virulence effectors converge onto hubs in a plant immune system network. Science 333:596-601.

Murzin A, Brenner SE, Hubbard T, Clothia C (1995) SCOP: a structural classification of proteins for the investigation of sequences and structures. J Mol Biol 247:536-540.

Newman MEJ (2003) The structure and function of complex networks. SIAM Rev 45:167-256.

Niklas KJ, Wright S, Simpson GG (1994) Morphological evolution through complex domains of fitness. Proc Natl Acad Sci USA 91:6772-6779.

Orengo C, Michie A, Jones S, Jones D, Swindells M, Thornton JM (1997) CATH – a hierarchic classification of protein domain structures. Structure 5:1093–1109.

Pearl J (2000) Causality. Cambridge University Press, Cambridge.

Seelig JD, Jayaraman V. (2013). Feature detection and orientation tuning in the *Drosophila* central complex. Nature 503:262-266.

Shoval O, Sheftel H, Shinar G, Hart Y, Ramote O, Mayo A, Dekel E, Kavanagh K, Along U (2012) Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. Science 336:1157-1160.

Sia J, Jonckheere E, Bogdan P (2019) Ollivier-Ricci curvature-based method to community detection in complex networks, Sci Rep 9:9800.

Simon HA (1962) The architecture of complexity. Proc Am Phil Soc 106: 467–482.

Smakowska-Luzan et al. (2018) An extracellular network of *Arabidopsis* leucine-rich repeat receptor kinsases. Nature 553:342-346.

Teal J (1962) Energy flow in the salt marsh ecosystem of Georgia. Ecology 43:639-649.

Ulanowicz RE (1980) An hypothesis on the development of natural communities. Journal of Theoretical Biology 85:223-245.

Ulanowicz RE (2002) The balance between adaptability and adaptation. *BioSystems* 64: 13–22.

Villeneuve DL, Angrish MM, Fortin MC, Katsiadaki I, Leonard M, Margiotta-Casaluci L, ... Zhang X (2018). Adverse outcome pathway networks II: network analytics. Environmental Toxicology Chemistry 37(6):1734-1748.

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

Wolff GH, Strausfeld NJ (2016) Genealogical correspondence of a forebrain centre implies an executive brain in the protostome–deuterostome bilaterian ancestor. Phil Trans R Soc B 371:20150055.

Wolff T, Rubin G. (2018). Neuroarchitecture of the *Drosophila* central complex: A catalog of nodulus and asymmetrical body neurons and a revision of the protocerebral bridge catalog. J Comp Neurology 526:2585-2611.

Xue Y, Bogdan P (2017) Reliable multi-fractal characterization of weighted complex networks: Algorithms and implications. Sci Rep 7: 7487.

Xue Y, Bogdan P (2019) Reconstructing missing complex networks against adversarial interventions, Nature Communications 10:1738.