Integrative and Comparative Biology

Integrative and Comparative Biology, volume 61, number 6, pp. 1991–2010
doi:10.1093/icb/icab069

Society for Integrative and Comparative Biology

JUMPSTART NSF INVITED

Biological Networks across Scales—The Theoretical and Empirical Foundations for Time-Varying Complex Networks that Connect Structure and Function across Levels of Biological Organization

Paul Bogdan,* Gustavo Caetano-Anollés,[1,†] Anna Jolles,[‡] Hyunju Kim,[§] James Morris,[¶] Cheryl A. Murphy,[‖] Catherine Royer,[‖‖] Edward H. Snell,[**] Adam Steinbrenner[††] and Nicholas Strausfeld[‡‡,#]

*Ming-Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California, Los Angeles, Los Angeles, CA 90007, USA; [†]Department of Crop Sciences, University of Illinois, Urbana–Champaign, Champaign, IL 61801, USA; [‡]Carlson College of Veterinary Medicine and Department of Integrative Biology, Oregon State University, Corvallis, OR 97331, USA; [§]The Beyond Center for Fundamental Concepts in Science, Arizona State University, Tempe, AZ 85287-0506, USA; [¶]Baruch Institute for Marine and Coastal Sciences, University of South Carolina, Columbia, SC 29201, USA; [‖]Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI 48824, USA; [‖‖]Department of Biological Sciences, Rensselaer Polytechnic Institute, 110 8th Street, Troy, NY 12180, USA; **Hauptman-Woodward Medical Research Institute and SUNY at Buffalo, Buffalo, NY 14203, USA; [††]Department of Biology, University of Washington, 24 Kincaid Hall, Seattle, WA 98105, USA; [‡‡]Department of Neuroscience, University of Arizona, Tucson, AZ 85721, USA

[#]Authors are listed in alphabetical order.

[1]E-mail: gca@illinois.edu

Synopsis 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 and dynamics 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. Also, subnetworks may emerge and disappear as a result of the need for the biological system to cope with for example invaders or new information flows. 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, that is, 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.

Advance Access publication May 22, 2021

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 and scales. 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 (Barabási and Oltvai 2004). ***Figure 1***A 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 (Aziz and Caetano-Anollés 2021). 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). A network can be represented with an adjacency matrix, a square matrix used to describe a finite graph, a property that is useful for spectral graph theoretical applications (Fig. 1B). The matrix becomes asymmetric when links are directed. 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 link influences the connectivity or information flow of the network (Fig. 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, and persistence).

As expected from complex systems, network abstractions in biology are often difficult to understand: (i) *Complexity:* Networks can be structurally complex when their wiring diagrams become tangles (e.g., multiple rules govern network responses to environmental perturbations); (ii) *Connectivity:* Links between nodes can have different weights, directions, and signs and can describe different kinds of interactions (e.g., link communities describing different classes of biological functions); (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

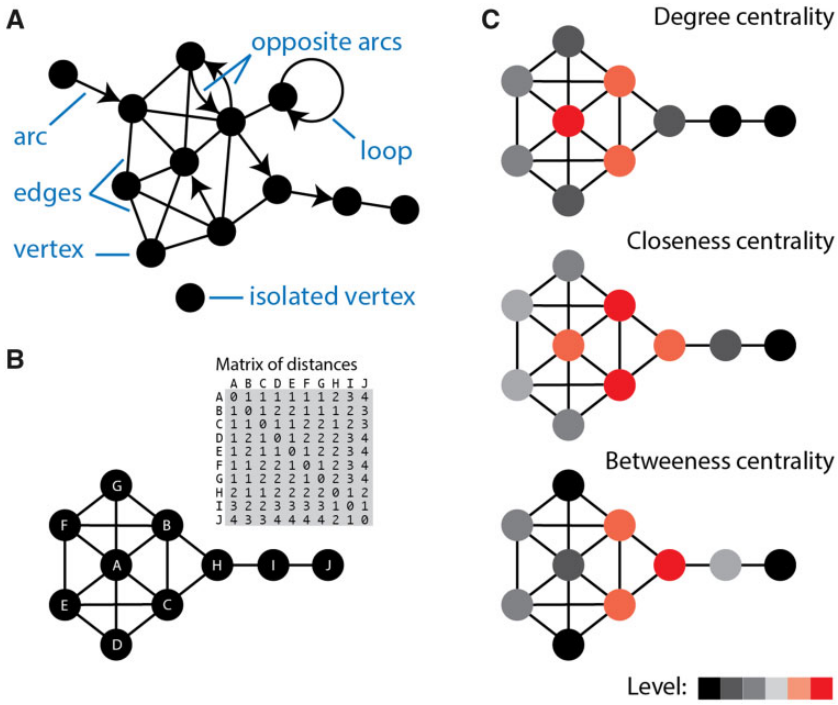

Fig. 1. A network view of biological systems. (**A**) An anatomical analysis shows that a network N is a combination of four sets, a set V of vertices (nodes), a set L of lines (links), and sets of vertex and line value functions that are mapped onto the V and L sets, respectively. Each line is associated with a pair of vertices (lines are two-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. (2021).

time (e.g., effects of canalization on network dynamics); and (v) *Dynamics:* Nodes and links can themselves portray non-linear and long-range memory/multifractal dynamic behaviors. The state of each node or link can vary in time in complicated ways in order to ensure a common collective goal unfolds in a decentralized way.

While complex, diverse, and evolving networks can effectively describe how parts are connected to each other in natural systems, the correct definition of a biological part becomes central to the network modeling exercise. For example, structural domains are considered "units" of protein structure that are useful for the taxonomical classification of the world of proteins (Caetano-Anollés et al. 2009). Domains represent 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. However, defining domains in proteins is not a trivial endeavor. Advanced machine learning methodologies of structural recognition, such as hidden Markov models (HMMs) (Eddy 2004), have been effectively used to catalog domains with automatic and manual curation approaches. However, not all domains fold into discrete structural entities within the space of possible folds (Harrison et al. 2002). Some popular domains overlap within a continuum. This "gregariousness" makes it difficult to classify the folds of certain domain structures, demanding instead the use of super-secondary structural motifs (e.g., $_{\beta}$-hairpins) as lower-level classification tools. These kinds of difficulties make constructing networks difficult when "units" cannot be consistently

defined or when they "skip" levels of structural organization. Luckily, artificial intelligence (AI) algorithms are becoming more powerful and are facilitating the classification task. AI systems learn from data and can enhance themselves by learning new heuristics or re-write supporting algorithms. These emerging strategies include ensemble learning methods such as Bayesian network approaches (e.g., model averaging and optimal classifiers), bagging classifiers (e.g., random decision forests), and stacked generalization methods that build predictive models by iterative integration (Rokash 2011). The challenge however is to bring an evolutionary rationale to computational advances, especially because units must be evolutionary for them to make sense in biology. In addition, there is real "fuzziness" in natural systems, which goes beyond the experimenter definition of nodes and links. This difficulty needs to be appropriately addressed and represents a significant barrier to integrating structure and function at different scales. Finally, fuzziness in node definitions may be inherent to the biological scale of observation and perhaps can be perturbed and measured. This could bring a measure of rationality to the "biological parts as units" problem of constructing networks.

Network dynamics is also difficult to explore. 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 a potentially reduced/compressed set or rules. These cross-correlation exponents characterizing for example 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 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 links. 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, 2021). 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 (Koorehdavoudi and Bogdan 2016; Mahmoodi et al. 2017; 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. For example, phylogenomic analyses can help uncover how evolution tailors the structure and

function of biological networks during billions of years of natural history (Aziz et al. 2016; Caetano-Anollés et al. 2019; Mughal and Caetano-Anollés 2019; Aziz and Caetano-Anollés 2021).

Objectives

The following objectives illustrate the broad scope of inquiry 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 (WWW) and metabolism show a bow–tie structure, in which inputs into a highly connected component result in a number of outputs (***Fig. 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?

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, that is, 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 Navlakha 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 or protein–protein interactions networks? Since biological systems are not isolated, are we to expect that the effects of integration space be

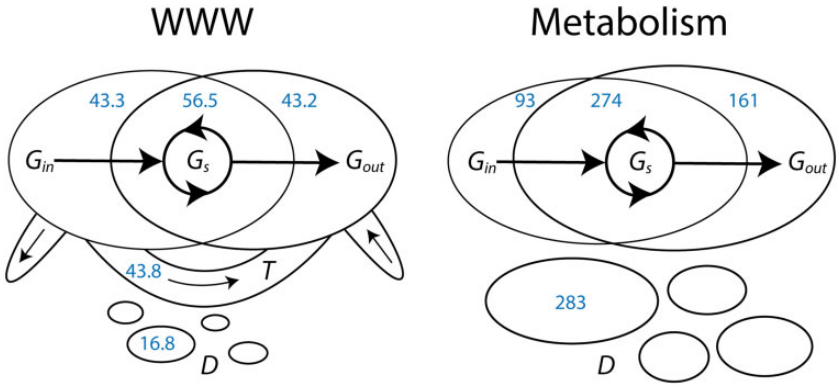

Fig. 2. The bow–tie hierarchical structure of directed networks. These networks have a giant strongly connected component (G_s), giant "in" component (G_{in}), giant "out" component (G_{out}), tendrils and tubes (T), and disconnected components (D). The number of nodes that are present in these subgraphs are listed (in blue) as millions of web resources for the WWW (Broder et al. 2000) and as connected enzymes in the metabolic networks of *Escherichia coli* (Ma and Zeng 2003). Note that metabolism lacks tendrils and tubes.

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, connectivity patterns, 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 can be studied along different timeframes and biological scales in a number of fundamental steps. The first step concerns the definition of entities (nodes) and connectivities (links), as well as rigorous computational and 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 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, and 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 using and mining phylogenetic, phylogenomic, or stratigraphic information in order to reconstruct the history of biological networks that describe evolving molecular machinery (e.g., proteome, metabolome, functionomes, signaling networks, protein–protein interactions, and 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 AI computations. A crucial step toward understanding the intelligence and the nature of optimization taking place in biology requires the investigation of the structure of evolving networks, elucidating the sources, means, and goals of specific network properties (including scale-freeness, randomicity, modularity, hierarchy, centralities, generalized fractal dimension, multifractal connectivities, and 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 toward characterizing the phase transitions of biological systems and predicting 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. For example, an evolving metabolic network that unfolds enzymatic activities on a timescale of billions of years was studied using a database that traces evolutionary information onto metabolic network structures (https://manet.illinois.edu) and bipartite network approaches that connect different levels of molecular organization (Mughal and Caetano-Anollés 2019). To illustrate, the enzymes of metabolic pathways can be grouped into "subnetworks" and "mesonetworks" following levels of the KEGG database classification (Kanehisa et al. 2004). Subnetworks encompass functionally related enzymatic pathways, while mesonetworks pool subnetworks with similar functional capabilities. For example, enzymatic pathways of nucleotide interconversion, biosynthesis, catabolism, and salvage of the subnetworks of "purine metabolism" and "pyrimidine metabolism" are grouped into the "purine metabolism" mesonetwork. ***Figure 3*** shows a time series of networks describing how enzymes are shared by "mesonetworks." These evolving networks can be used to study the recruitment of enzymatic activities in metabolic pathways. Similarly, an evolving network that links protein domains to functional

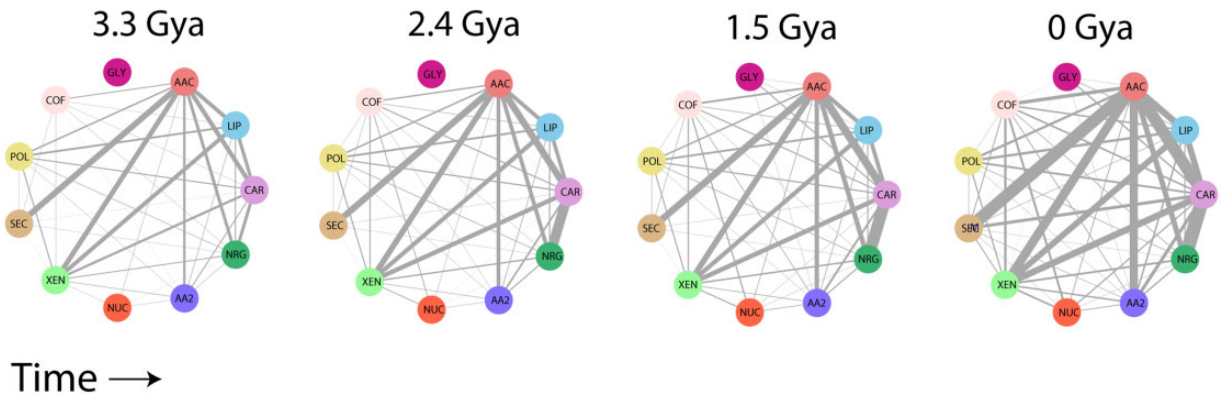

Fig. 3. The sharing of enzymes among mesonetworks at different stages of metabolic evolution. Nodes represent mesonetworks: AAC, amino acid metabolism; SEC, biosynthesis of other secondary metabolites; CAR, carbohydrate metabolism; NRG, energy metabolism; GLY, glycan biosynthesis and metabolism; LIP, lipid metabolism; COF, metabolism of cofactors and vitamins; POL, metabolism of terpenoids and polyketides; NUC, nucleotide metabolism; AA2, metabolism of other amino acids; and XEN, xenobiotics biodegradation and metabolism. Links represent sharing of enzymes, with weights proportional to their numbers. Time of networks is given in billions of years ago (Gya) and was inferred from a molecular clock of protein folds (Wang et al. 2011). Note how all mesonetworks (except GLY) are already sharing enzymes 3.3 Gya, especially AAC. Redrawn from Mughal and Caetano-Anollés (2019).

loops and defines an "elementary functionome" of protein structure was unfolded on a timescale of billions of years (Aziz et al. 2016). This allowed tracking 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?

Unknown unknowns: Tracing networks in time 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 "unknown unknowns" governing the observed biological dynamics. Consequently, to decipher and characterize biological networks over time, we need new mathematical and algorithmic tools that would reconstruct networks from partial observations, from various types of biological data sources and overcoming interventions. Examples include the use of time series data analysis on average sensitivity values of the networks, spike/event time sequences of biological activity (excitatory or inhibitory), and 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, and bacteria) may exhibit long-range memory and multi-fractal dynamic characteristics in order to cope with external perturbation and enforce a cue or rule toward 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 of estimating the number of unknown unknowns and determining 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).

Developing the framework

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), while carefully considering drawbacks from the "units in biology" problem we discussed above.

(2) Use biological sampling to define the history of biological networks (e.g., development, physiology, metabolite dynamics, and 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, and 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, and 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 toward 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, and ecosystem levels of biological organization.

How can hierarchy and other forms of network 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 (***Fig. 4***) 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 WWW (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 AI 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. Major categories of drivers include: (1) Evolutionary (e.g., life history, adaptation, canalization, and recruitment); (2) matter-energy (e.g., dissipation and budget); (3) information (e.g., entropic flow and modes); (4) structural (e.g., energy potentials and binding sites); (5) spatiotemporal (e.g., molecular and structural spaces, temporal flow); (6) trade-off solutions (e.g., economy, flexibility, robustness, and plasticity); (7) perturbation (stress)–homeostasis (some networks just developed to evaluate stress only); (8) ontogeny; (9) growth and development; (10) ecology; (11) levels of biological organization; (12) behavior; and (13) ontology (e.g., the Gene Ontology directed acyclic graph).

The following are examples of systems, from lower to higher levels of organization. They illustrate major

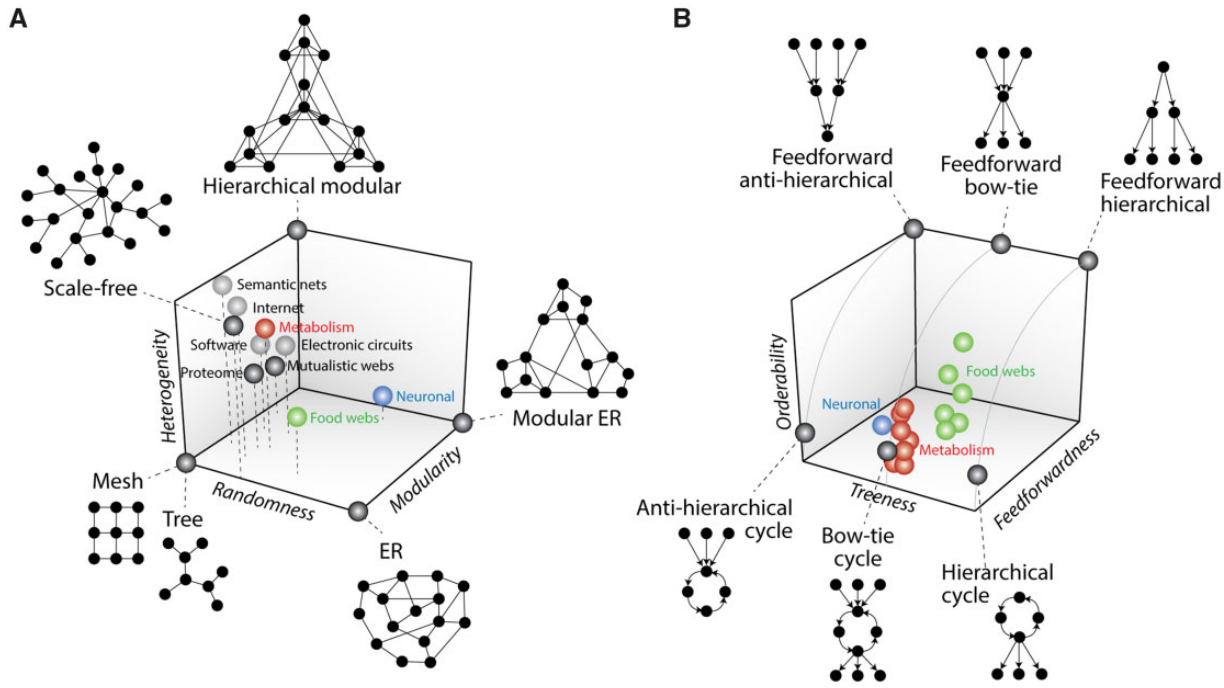

Fig. 4. Morphospaces of network structure (**A**) and hierarchy (**B**) showing the placement of toy examples of typical graphs describing archetypes of the phenotypic landscapes and real networks (metabolic, neuronal, and food web networks highlighted with colors). In one morphospace (**A**), Erdös–Rényi (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 (**B**), treeness defines the unification or diversification of hierarchical signal in the network, whereas orderability defines the centrality of cycles in network structure. Figures redrawn from Solé and Valverde (2004) and Corominas-Murtra et al. (2013).

drivers of network structure and dynamics (in parentheses). These networks are familiar to one or more of the authors and involve biological domains immediately suited for analysis using the approaches discussed above.

(i) Protein–protein interaction networks (PPINs) (structural drivers). 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 links (affinity pulldown—mass spectrometry, yeast two-hybrid, and other heterologous screens for measuring interaction strength). Modularity emerging from PPINs often correspond with specific functions, including transcription, nucleosome assembly, and hormone signal transduction (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 (***Fig. 5***A), in contrast to peripheral nodes only required for specific recognition functions. For example, well-studied somatic embryogenesis receptor kinase (SERK) co-receptors have been shown to form the highest connectivity in the PPIN of extracellular leucine-rich repeat receptors. Inter-species PPINs with factors required for pathogen virulence feature links that predominantly connect to host hubs (Mukhtar et al. 2011).

(ii) Cell cycle network (transition-development drivers). 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 (Dorsey et al. 2018). The

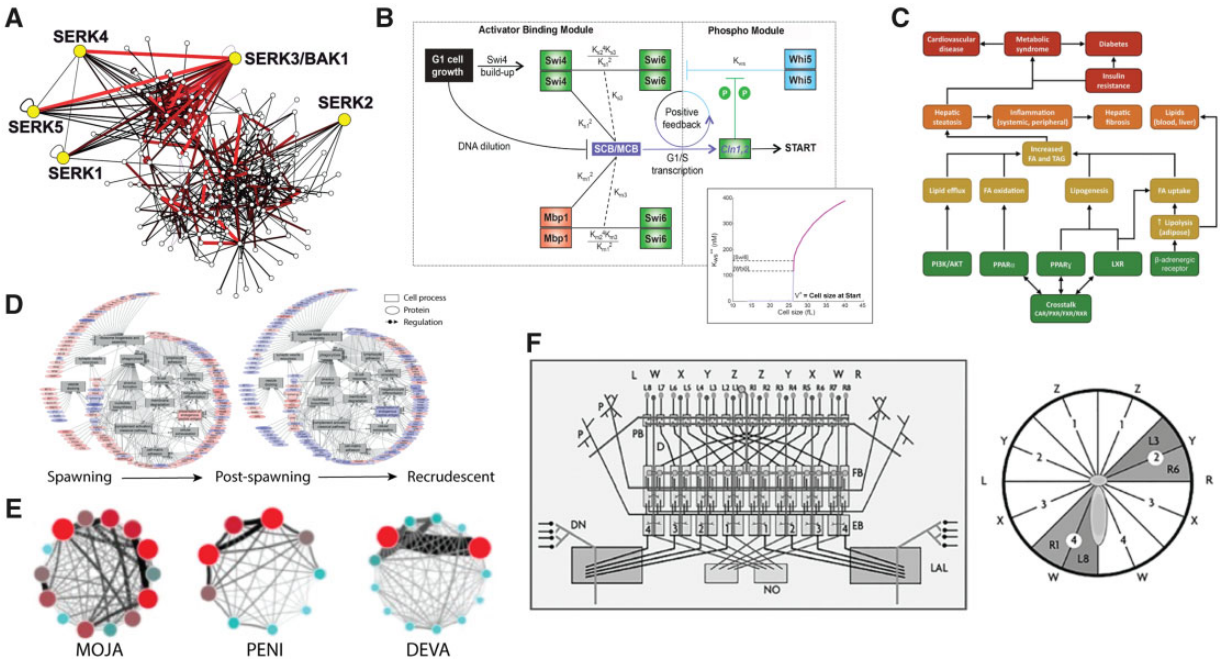

Fig. 5. Example systems visualized with network representations. (**A**) A highly connected PPIN showing significant interactions between plant leucine-rich repeat receptor ectodomains (Smakowska-Luzan et al. 2018). Subnetworks and nodes with strong and varied connectivity are apparent from network analysis. Edges indicate significant interaction between two ectodomains. Edges are thick and red colored in proportion to reported interaction strength. Extracted, yellow-colored nodes highlight highly connected SERK proteins known to be genetically required for many plant environmental responses. (**B**) A 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 (plot in the right). From Dorsey et al. (2018). (**C**) 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 four 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). (**D**) Gene transcriptional networks change as rainbow darter testis undergoes development to maturation (Bahamonde et al. 2016). (**E**) 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). (**F**) The entire sensory surround of the organism is represented in the brain's "central complex" diagrammed here. Projections of columnar neurons originating from the (upper modules W, X, Y, and Z provide sub-modules to the left [L, L8–L1] and right [R, R1–R8] of the midline that provides connections to successive computational layers EB). Computations within the PB, FB, and EB are relayed to decussating axons extending into the lateral centers (LAL), where they gate the activity of premotor neurons (DN). The proposition here is that one module represents 1/16th of the sensory envelope.

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, and cytokinesis) and the links are the events that allow transitions from one phase to the next. Each node encompasses a sub-network. Figure 5B 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: (1) The links 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. (2) 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. (3) The outputs are the cell size at which each of the transitions occurs. (4) Changes in environment or mutations perturb the network. Extension to mammalian cells and cancer demand developing tools for making required

measurements in less genetically modifiable systems than yeast.

(iii) Organ-level network (perturbation drivers). A perturbation network (stressor—beyond homeostasis) describes pathways that converge 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 (Fig. 5C). 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 intersects and direct processes that when off balance could induce fatty liver disease. The modularity of the network is represented by what can be measured in terms of physiological parameters (e.g., binding to receptors, and measurements of lipids). The nodes in the network are called key events and are largely physiologically derived. The links 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).

(iv) Developmental networks (growth and developmental drivers). 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 (Fig. 5D). These data suggest that there are distinct transcriptomic 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 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 data, some aspects of the network could be missed (Bahamonde et al. 2016; Basili et al. 2018).

(v) Microbiome networks (perturbation drivers). 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) (Berg et al. 2020). Gut microbiomes for example are well studied in humans and some animal species, usually focusing on bacterial taxa. Next generation sequencing (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 operational taxonomic units (OTUs). These microbial communities influence host health and behavior profoundly. This influence takes advantage of a range of different mechanisms, which are only beginning to be understood, 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 impact 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 OTUs are not explicitly measured. Nonetheless, exploring associations between microbiome structure and for example robustness versus 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 (ecology drivers): 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 (Buchalski et al. 2016). 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 (Bleich et al. 1990). Population extinction and recolonization have been observed, and extinction varies with elevation, precipitation, and access to water (Epps et al. 2004).

DBH networks defined by observed levels of gene flow (Nm) vary in topology, and populations within networks vary in centrality (Fig. 5E). The Mojave (MOJA) and Death Valley (DEVA) networks are similar in size, but populations in the Mojave are more connected than in Death Valley. Centrality in the DEVA 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 (PENI) network is smaller, and has an intermediate number of strongly connected populations compared with the MOJA and DEVA networks, with slightly weaker connectivity overall compared with the other two networks. Which networks are more resilient to environmental perturbations of different types—from climatic variation to invasion of infectious agents?

(vii) Saltmarsh (ecology and perturbation drivers). Ecosystems are complex networks of interacting species with various environmental inputs of varying importance and with stabilizing feedbacks. For example, salt marsh ecosystems have existed for millennia more or less in equilibrium with sea level, and this has been possible because of negative feedback between the higher plants and flooding (Morris et al. 2002). However, the feedback can be positive and destabilizing if the rate of sea-level rise is too rapid. Focusing on the negative feedback, we know that the plants respond positively with greater net primary production (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. The result of these feedbacks is a stable (within bounds) system that has been remarkably resilient in the face of rising sea level.

(viii) Networks of the brain (behavior drivers). Simple hierarchical systems of neurons provide various levels of network complexity. It is no accident that artificial computational networks are referred to as "neural nets." They resemble connections of nerve cells. However, few 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), which was derived from a simplistic but relevant view (in 1949) of hippocampal organization. Hebb postulated that a neuron's propensity to relay information (efficacy) depends on its persistent stimulation by a presynaptic drive: when two neurons converge on the neuron and provide coincident inputs these can be sufficient to permanently change the efficacy of the postsynaptic cell's synapse. In other words, synaptic strength results from presynaptic association. Hebb's work immediately attracted researchers working on the cortex and hippocampus, both mediating in short and long-term memory (e.g., Frolov and Murav'ev 1993).

We know from descriptions of chordate and invertebrate brains that every functional domain is defined by its characteristic network arrangement-patterned synaptic connections among its constituent neurons, and its connections from and to other domains. Some functional domains show close genetic, structural, pathological, and functional similarities, which taken together imply genealogical correspondence: hence phenotypic and genotypic homology implying an origin in deep time before the divergence of lineages leading to vertebrates and invertebrates. Currently, the most interesting "real" neural networks are in the most anterior region of the brain: the vertebrate basal ganglia and hippocampus; in panarthropods the "central complex" and mushroom bodies (Wolff and Strausfeld 2016). Basal ganglia and central complexes in common (Strausfeld and Hirth 2013) coordinate motor actions by

editing outputs by orchestrating systems of inhibitory connections that selectively gate outputs relevant to a required behavior permitting information to reach circuits controlling motor neurons to muscle. Genetic deletions, or interventions of dopaminergic modulators in the network lead to Parkinson's-like pathologies in both mouse and fruit fly. Insect mushroom bodies and vertebrate hippocampus form long term associations relating to the memory of place, experience, and sentience.

The "central complex" comprises discrete computational modules supplied by high-level sensory inputs (Fig. 5F). Modules assess the bilateral weighting of sensory percepts to provide appropriate signals to controllers—the inhibitor neurons that gate motor actions. Precision of connections across the modules reflects dexterity: invariant precision of a praying mantis, but noisy connectivity in a species with moderate dexterity, such as a cockroach. In *Drosophila*, optogenetics and electrophysiology documenting the central body's role in working memory and motor control (Seelig and Jayaraman 2013; Wolff and Rubin 2018) demonstrate that this center is 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.

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, network diversity, structure, complexity, metacomplexity, causality, completeness, and universality complicate knowledge integration.

Diversity: An important barrier is the actual diversity of the nodes and links of networks. This diversity must be defined when studying, comparing, and/or integrating systems. For example, the PPINs of Fig. 5A have protein nodes connected by links describing the existence of interactions between cell surface proteins. The network of protein factors of the cell cycle of Fig. 5B describes the interaction of transcription factors and a cycle dependent kinase with promoters of crucial genes of the G1 binding and phosphorylation modules. The networks of DBH sheep populations of Fig. 5E describe how population nodes are connected in different landscapes. Connecting interactions of cell surface proteins, cell cycle regulation, and spread of genes in sheep populations showcases the complexity of trying to integrate three distinct biological systems. These interactions could be visualized with a tripartite graph, which is a special case of k-partite graphs. This general class of graphs has nodes that can be divided (partitioned or colored) into k disjoint sets (partitions or colors) and connections (links) that always connect nodes belonging to different sets. Closed k-partite graphs do not impose restrictions of the k-partite structure of connected nodes (all sets can connect to each other). Open k-partite graphs do not allow a tightly connected structure (circular in the case of tripartite graphs). The use of k-partite structures in network biology has been limited. For example, Koç et al. (2018) devised a tripartite network of gene-metabolite-pathway connectivity that linked transcriptomes to metabolism using a metabolite-centric reporter pathway analysis. However, one benefit of k-partite structures is that they can be decomposed into simple graphs; open tripartite graphs can be decomposed into one-mode and two-mode (bipartite) graph projections to improve visualization.

Structure: 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. 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, 2021), what are the evolutionary drivers of hierarchical modularity in network structure?

One example at the molecular structure level is the structural domain module of a multi-domain

protein. The organization of domain modules in proteins, which massively unfolded in a "big bang" of domain combination during the rise of multicellularity and the eukaryotic superkingdom, has been modeled with a time series of evolving networks (Aziz and Caetano-Anollés 2021). These networks unfold both hierarchy and modularity in evolution. They show significant network structure.

Structural modules also exist in cellular organization. Together with the "central complex" of the brain (Fig. 5F), the "paired mushroom" bodies are examples of networks comprising discrete modules and interactive nodes. Homologs across phyla represent divergences from a "ground pattern" network, originating about 600 million years ago according to "trace" fossils that recorded behaviors of the earliest bilateral animals. Mushroom bodies, like the hippocampus, comprise orthogonal arrangements of intersecting neurons that comprise a Hebbian-like network. Work on learning and memory in the fruitfly *Drosophila* (Heisenberg 2003) provides the most accessible system for investigating whether Hebbian-type associations apply to real-world biological learning networks. Structural studies show the mushroom body's neurons consisting of orthogonal arrangements of local interneurons intersected by converging inputs encoding various types of unimodal sensory data organized as would be a massive Hebbian network. Output neurons that encode multisensory associations allow the experimenter to "read" functional properties of the biological network.

Figure 6 schematizes 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. 2004). 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. A mushroom body comprises hundreds 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 are the rules underlying how these subsets interact such that memories interact, achieve contextual valences, and form *post hoc* memory modifications: all functions expected in sentient organisms that obtain an understanding of dynamic ecologies. What is recognized from behavioral studies across species is that 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 2000 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

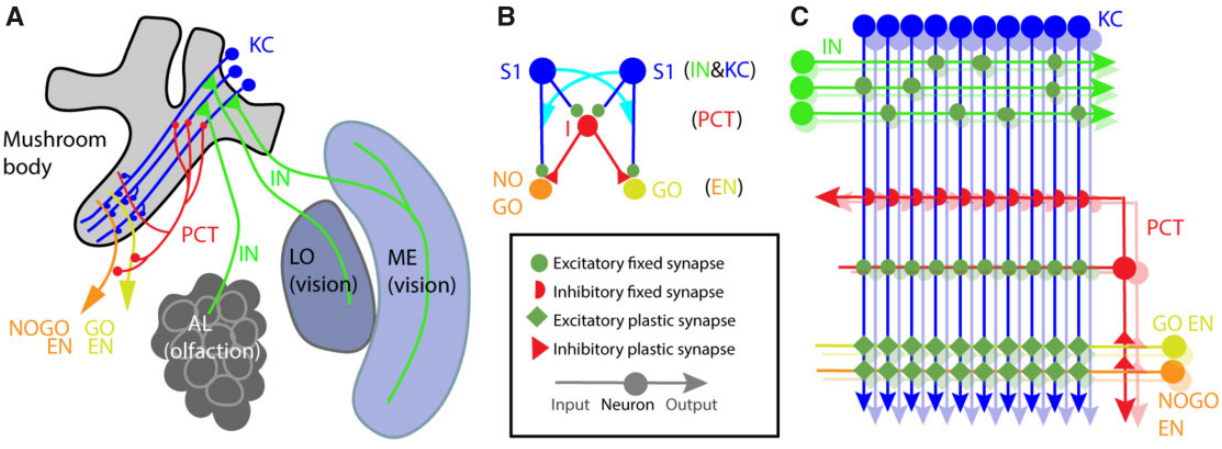

Fig. 6. Models of the mushroom bodies. (**A**) Neuroanatomy: MB Mushroom Bodies; AL Antennal Lobe glomeruli (circles); ME & LO visual neuropils. Relevant neural pathways are shown and labelled for comparison with the model. (**B**) Reduced model; neuron classes indicated at right and side of sub-figure. (**C**) Full model. For explanation see text from Cope et al. (2018).

explain and indeed imitate sensory associations and memory acquisition.

Complexity: Since systems are structured into highly integrated subsystems (Simon 1962), there will be need to integrate networks both *across* and *within* scales. For that purpose, we can take advantage of Simon's "near-decomposability" of systems (Simon 1997), which allows for "long-term behavior to be studied on an aggregative basis without concern for internal details of the parts, and allows the short-term behavior of each part to be studied independently of the behavior of the other parts." In some cases, it may be straightforward to dissect complexity scales because each part of the nearly-decomposable system will have strong internal links among its subparts (see Fig. 5B). In other cases, there could be significant difficulties because hierarchy and modularity could be loosely linked in the systems.

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 7A*** 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 Fig. 7B. 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 (xs) is positive. We can ask questions about connectivity and total system throughput (TST), stability, ascendency (Ulanowicz 1980), fractal dimension, and size.

The hope is that we can arrive at generalities about real networks by analyzing the properties of artificial networks. From a universe of >5000 random food webs composed of as many as 2200 taxa, it was demonstrated that the probability of generating a feasible network declined rapidly as the number of taxa exceeded 400. Flow diversity increased asymptotically, that is, flows became more uniform (Morris et al. 2005). Ulanowicz (2002) used an information-theoretic homolog of the May–Wigner stability criterion to hypothesize a maximal connection per taxon of about 3. From the computer-generated networks, 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 and not so different from that predicted by the May–Wigner criterion. The explanation may be the limit imposed by gross primary production on energy flow, like the limits that resource space places on the distribution of species (MacArthur 1957). These examples suggest there are fundamental relationships between network structure and function.

Meta-complexity: 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 PPINs 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 or link identifiers? It is here where epistemology and ontology must interface.

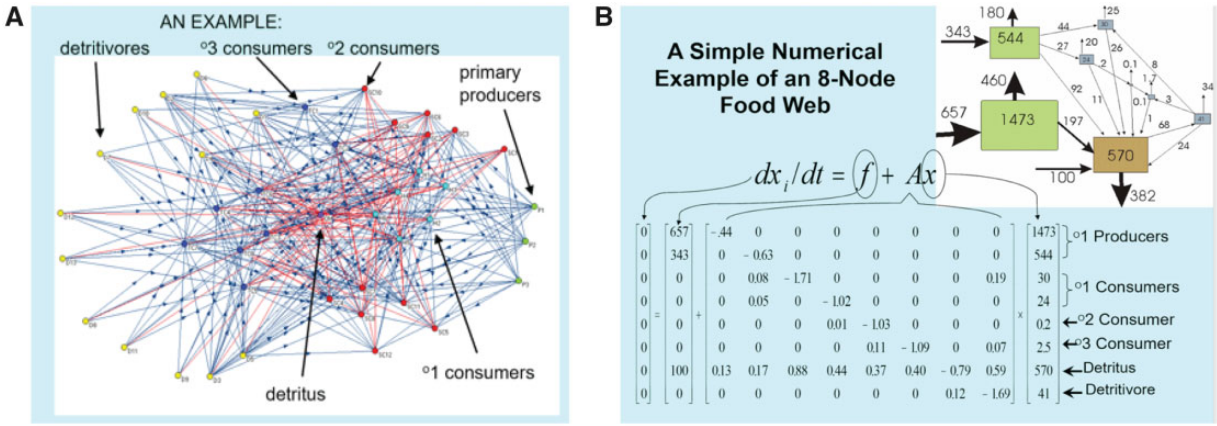

Fig. 7. Generating artificial food webs by *in silico* modeling. (**A**) Foodweb generated by populating a transfer matrix with transfer coefficients and solving for the equilibrium solution. (**B**) Methodology used to generate modeled food webs (described in the text).

Meta-complexity also manifests in the diversity of the functions (e.g., differential equations) that are mapped onto nodes and links. Mapping functions to links often define the non-linear dynamic behaviors of matter-energy and/or information traveling between nodes through a vector of state variables. A diversity of dynamics can therefore unfold in link communities. For example, link communities of metabolism could define reversible and irreversible metabolic reactions and transport processes. These processes can be dissected with sets of non-linear equations, which cannot be solved analytically but can be visualized in an abstract *n*-dimensional state space with a "velocity" vector field. The challenge is therefore to mine steady states of the multi-dimensional space (e.g., fixed-point attractors, chaotic aperiodic motions, and close loop attractors) to understand the landscape of dynamic behaviors of biological systems.

Causality: 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 AI field. However, effective integration approaches must be sought, perhaps using experiments, predictive computational methods, theoretical and mathematical approaches, and the exploration of functions and constraints with philosophical approaches. One example is modeling causal interdependent non-linear dynamics with multivariate discrete dynamical systems (automata networks). In particular, Boolean networks are canonical models that have been applied to a number of complex systems very successfully. To capture redundancies in system dynamics of biochemical regulatory and signaling interactions, a mathematical framework called the "effective graph" for example was capable of synthesizing both network structure and dynamics in a weighted graph representation of discrete multivariate systems (Gates et al. 2021).

Completeness: 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, flexibility, 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, and 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 three-dimensional atomic design of structural domains. The SCOP (Murzin et al. 1995) and CATH (Orengo et al. 1997) databases, the gold standards of protein classification, show that protein folds group into 2705 SCOP (http://scop.mrc-lmb.cam.ac.uk) and 5481 CATH (https://www.cathdb.info) well-curated superfamilies (as of April 29, 2021). These numbers are reaching 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 three-dimensional atomic structures has been stored in the RCSB Protein Data Bank (PDB) repository (https://www.rcsb.org). Currently, there are 177,219 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.

Universality: Finally, there is the problem that not all data types can be modeled with networks. This difficulty challenges the concept of networks across biological scales. Simplification must occur if information from multiple levels of biological integration are incorporated into a network (e.g., 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 include developing new algorithmic techniques to link information, determining the influence of different levels of noise on the knowledge produced from that information, and evaluating 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, our approach 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 and 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, for example, redundancy in automation, self-organization for transport within a city, response to perturbation in a system, and transient approaches that activate. 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 collaboratives that provide:

- Longitudinal empirical network data across a broad range of biological systems and scales, ideally including observational, experimental, computational, and theoretical 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 transparent 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.

Data availability statement

No new data were generated or analysed in support of this research.

Acknowledgments

The ideas elaborated in this "vision" paper originated in the NSF Reintegrating Biology Jumpstart meeting that was organized by the National Science Foundation and took place in San Diego, California, December 4–6, 2019. We thank Clint Epps and Ben Dalziel for providing genetic data and network figure for deset bighorn sheep.

Funding

This work was supported by the National Science Foundation [Career Award CPS/CNS-1453860, CCF-1837131, MCB-1936775, and CNS-1932620 to P.B.; MCB-0343126, MCB-074983607, OISE-1172791, and DBI-1041233 to G.C.-A.; DEB-1911994 to A.J.; PHY-1505048 to H.K.; DEB-1654853 to J.M.; PHY-1806638 to C.R.; DBI-1231306 to E.H.S.; and IOS-1754798 to N.S.], Defense Advanced Research Projects Agency [Young Faculty Award and DARPA Director Award, N66001-17-1-4044 to P.B.], National Institute of Food and Agriculture [ILLU-802-909 and ILLU-483-625 to G.C.-A.; 1014468 to C.A.M.], Northrop Grumman [grant to P.B.], and National Center for Supercomputing Applications [allocations to G.C.-A.]. The views, opinions, and/or findings contained in this article are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied by the funding agencies.

Conflict of interest

Authors declare there are no conflicts of interest.

References

Ahn YY, Bagrow JP, Lehmann S. 2010. Link communities reveal multiscale complexity in networks. Nature 466:761–5.

Angrish MM, Kaiser JP, McQueen CA, Chorley BN. 2016. Tipping the balance: hepatotoxicity and the 4 apical key events of hepatic steatosis. Toxicol Sci 150:261–8.

Arabidopsis Interactome Mapping Consortium. 2011. Evidence for network evolution in an *Arabidopsis* interactome map. Science 333:601–7.

Aziz MF, Chan P, Osorio JS, Minhas BF, Parekatt V, Caetano-Anollés G. 2012. Stress induces biphasic-rewiring and modularization patterns in metabolomics networks of *Escherichia coli*. IEEE Intl Conf Bioinf Biomed 2012:593–7.

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.

Aziz MF, Caetano-Anollés G. 2021. Evolution of networks of protein domain organization. Sci Rep published online (doi: 10.21203/rs.3.rs-119891/v1).

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: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. Nat Rev 5:101–13.

Basili D, Zhang JL, Herbert J, Hebert J, Kroll K, Denslow ND, Martyniuk CJ, Falciani F, Antczak P. 2018. In silico computational transcriptomics reveals novel endocrine disruptors in largemouth bass (*Micropterus salmoides*). Environ Sci Technol 52:7553–65.

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:e1000748.

Berg G, Rybakova D, Fischer DCernava T, Vergès M-CC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH, et al. 2020. Microbiome definition re-visited: old concepts and new challenges. Microbiome 8:103.

Bleich VC, Wehausen JD, Holl SA. 1990. Desert-dwelling mountain sheep: conservation implications of a naturally fragmented distribution. Conserv Biol 4:383–90.

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. Comput Netw 33:309–20.

Buchalski MR, Sacks BN, Gille DA, Penedo MCT, Enest HB, Morrison SA, Boyce WM. 2016. Phylogeographic and population genetic structure of bighorn sheep (*Ovis canadiensis*) in North American deserts. J Mammol 97:823–38.

Caetano-Anollés G, Wang M, Caetano-Anollés D, Mittenthal JE. 2009. The origin, evolution and structure of the protein world. Biochem J 417:621–37.

Caetano-Anollés G, Aziz MF, Mughal F, Gräter F, Koç I, Caetano-Anollés K, Caetano-Anollés D. 2019. Emergence of hierarchical modularity in evolving networks uncovered by phylogenomic analysis. Evol Bioinformatics 15:1176934319872980.

Caetano-Anollés G, Mughal F, Aziz MF, Koç I, Caetano-Anollés K, Caetano-Anollés D, Mittenthal JE. 2021. A double tale of module creation in evolving networks. In: Caetano-Anollés G, editor. Untangling molecular biodiversity. Singapore: World Scientific. p. 91–168.

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:e1006435.

Corominas-Murtra B, Goñi J, Solé RV, Rodriguez-Caso C. 2013. On the origins of hierarchy in complex networks. Proc Natl Acad Sci USA 110:13316–21.

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. Phys Rev Lett 121:138102.

Dorsey S, Tollis S, Cheng J, Black L, Notley S, Tyers M, Royer CA. 2018. G1/S transcription factor copy number is a growth determinant of cell cycle commitment in yeast. Cell Syst 6:539–54.

Eddy SR. 2004. What is a hidden Markov model? Nat Biotechnol 22:1315–6.

Eichler K, Li F, Litwin-Kumar A, Park Y, Andrade I, Schneider-Mizell CM, Saumweber T, Huser A, Eschbach C, Gerber B, et al. 2017. The complete connectome of a learning and memory centre in an insect brain. Nature 548:175–82.

Epps CW, McCullough DR, Wehausen JD, Bleich VC, Rechel J. 2004. Effects of climate change on population persistence of desert-dwelling mountain sheep in California. Conserv Biol 18:102–13.

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

Gates AJ, Correia RB, Wang X, Rocha LM. 2021. The effective graph reveals redundancy, canalization and control pathways in biochemical regulation and signaling. Proc Natl Acad Sci USA 118:e2022598118.

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 MEJ. 2002. Community structure in social and biological networks. Proc Natl Acad Sci USA 99:7821–6.

Gupta G, Pequito S, Bogdan P. 2019. Learning latent fractional dynamics with unknown unknowns. American Control Conference (doi: 10.23919/ACC.2019.8815074).

Harrison A, Pearl F, Mot R, Thornton J, Orengo C. 2002. Quantifying the similarities within fold space. J Mol Biol 323:909–26.

Hartwell LH, Hopfield JJ, Leibler S, Murray AW. 1999. From molecular to modular cell biology. Nature 402:c47–52.

Hebb DO. 1949. The organization of behaviour. New York (NY): Wiley.

Heisenberg M. 2003. Mushroom body memoirs: from maps to models. Nat Neurosci Rev 4:266–75.

Herz A, Sulzer B, Kühn R, van Hemmen JL. 1988. The Hebb rule: storing static and dynamic objects in an associative neural network. EPL 7:663–9.

Ho JJ, Navlakha S. 2018. Evidence of Rentian scaling of functional modules in diverse biological networks. Neural Comput 30:2210–44.

Huerta R, Nowotny T, García-Sanchez M, Abarbanel HDI, Rabinovich MI. 2004. Learning classification in the olfactory system of insects. Neural Comput 16:1601–40.

Hug LA, Baker BJ, Anantharaman K, Anantharam K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hernsdorf AW, Amano Y, et al. 2016. A new view of the tree of life. Nat Microbiol 1:16048.

Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. 2004. The KEGG resource for deciphering the genome. Nucleic Acids Res 32:D277–80.

Kania J, Kramer M. 2011. Collective impact. Stanford Social Innovation Review, Winter issue. Stanford University. p. 36–41.

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

Koç I, Yuksel I, Caetano-Anollees G. 2018. Metabolite-centric reporter pathway and tripartite network analysis of *Arabidopsis* under cold stress. Front Bioeng Biotechnol 6:121.

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, Munn S, O'Brien JM, Polesch N, Smith LC, et al. 2018. Adverse outcome pathway networks I: development and applications. Environ Toxicol Chem 37:1723–33.

Landman BS, Russo RL. 1971. On a pin versus block relationship for partitions of logic graphs. IEEE Trans Comput 20:1469–79.

Ma H-W, Zeng A-P. 2003. The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics 19:1423–30.

MacArthur RH. 1957. On the relative abundance of bird species. Proc Natl Acad Sci USA 43:293–5.

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. In: Belgrano A, Scharler UM, Dunne J, Ulanowicz RE, editors. Aquatic food webs: an ecosystem approach. Oxford: Oxford University Press. p. 73–85.

Morris JT, Sundareshwar PV, Nietch CT, Kjerfve B, Cahoon DR. 2002. Responses of coastal wetlands to rising sea level. Ecology 83:2869–77.

Mughal F, Caetano-Anollés G. 2019. MANET 3.0: hierarchy and modularity in evolving metabolic networks. PLoS ONE 14:e0224201.

Mukhtar MS, Carvunis A-R, Dreze M, Epple P, Steinbrenner J, Moore J, Tasan M, Galli M, Hao T, Nishimura MT, et al.; European Union Effectoromics Consortium. 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 247:536–40.

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 91:6772–9.

Orengo C, Michie A, Jones S, Jones D, Swindells M, Thornton JM. 1997. CATH—a hierarchic classification of protein domain structures. Structure 5:1093–109.

Pearl J. 2000. Causality. Cambridge: Cambridge University Press.

Rokash L. 2011. Ensemble-based classifiers. Artif Intell Rev 33:1–39.

Seelig JD, Jayaraman V. 2013. Feature detection and orientation tuning in the *Drosophila* central complex. Nature 503:262–6.

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 336:1157–60.

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 106:467–82.

Simon HA. 1997. Models of bounded rationality: empirically grounded economic reason. Vol. 3. Cambridge (MA): MIT Press.

Smakowska-Luzan E, Mott AG, Parys K, Stegmann M, Howton TC, Layeghifard M, Neuhold J, Lehner A, Kong K, Grunwald K, et al. 2018. An extracellular network of *Arabidopsis* leucine-rich repeat receptor kinases. Nature 553:342–6.

Solé RV, Valverde S. 2004. Information theory of complex networks: on evolution and architectural constraints. Lect Notes Phys 650:189–207.

Strausfeld NJ, Hirth F. 2013. Deep homology of arthropod central complex and vertebrate basal ganglia. Science 340:157–61.

Ulanowicz RE. 1980. An hypothesis on the development of natural communities. J Theor Biol 85:223–45.

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, Munn S, O'Brien JM, Pollesch NL, Smith LC, et al. 2018. Adverse outcome pathway networks II: network analytics. Environ Toxicol Chem 37:1734–48.

Wang M, Jiang Y-Y, Kim KM, Qu G, Ji H-F, Mittenthal JE, Zhang H-Y, Caetano-Anollés G. 2011. A universal molecular clock of protein folds and its power in tracing the early history of aerobic metabolism and planet oxygenation. Mol Biol Evol 28:567–82.

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

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 Neurol 526:2585–611.

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. Nat Commun 10:1738.