The Emergence of Modularity in Biological Systems

Dirk M. Lorenz[a], **Alice Jeng**[b], and **Michael W. Deem**[a,*]

[a]Department of Physics and Astronomy, Rice University

[b]Department of Chemical, Biological, and Materials Engineering, University of Oklahoma

Abstract

In this review, we discuss modularity and hierarchy in biological systems. We review examples from protein structure, genetics, and biological networks of modular partitioning of the geometry of biological space. We review theories to explain modular organization of biology, with a focus on explaining how biology may spontaneously organize to a structured form. That is, we seek to explain how biology nucleated from among the many possibilities in chemistry. The emergence of modular organization of biological structure will be described as a symmetry-breaking phase transition, with modularity as the order parameter. Experimental support for this description will be reviewed. Examples will be presented from pathogen structure, metabolic networks, gene networks, and protein-protein interaction networks. Additional examples will be presented from ecological food networks, developmental pathways, physiology, and social networks.

> There once were two watchmakers, named Hora and Tempus, who manufactured very fine watches. Both of them were highly regarded, and the phones in their workshops rang frequently — new customers were constantly calling them. However, Hora prospered, while Tempus became poorer and poorer and finally lost his shop. What was the reason?
>
> The watches the men made consisted of about 1,000 parts each. Tempus had so constructed his that if he had one partly assembled and had to put it down — to answer the phone say— it immediately fell to pieces and had to be reassembled from the elements. The better the customers liked his watches, the more they phoned him, the more difficult it became for him to find enough uninterrupted time to finish a watch.
>
> The watches that Hora made were no less complex than those of Tempus. But he had designed them so that he could put together subassemblies of about ten elements each. Ten of these subassemblies, again, could be put together into a larger subassembly; and a system of ten of the latter sub-assemblies constituted the whole watch. Hence, when Hora had to put down a partly assembled watch in order to answer the phone, he lost only a small part of his work, and he assembled his watches in only a fraction of the man-hours it took Tempus."
>
> H. A. Simon, The Architecture of Complexity, 1962 [1].

*Corresponding author : mwdeem@rice.edu (Michael W. Deem).

1. Introduction

As Simon's classic parable of Hora and Tempus illustrates, there are advantages to assembling a complex system from modular pieces in a hierarchical fashion. This is particularly true in evolutionary adaptation, where a modular structure provides many benefits. The space of all genotypes is exponentially large, making an exhaustive search for fitness maxima impossible even on evolutionary time scales. But a system that can be decomposed into modules can be evolved one module at a time. Thus, modularity can reduce the task of searching the entire space of possibilities into a polynomial problem of searching in the subspace of modular solutions. A physicist might recognize the similarity to a separable Hamiltonian while a geneticist might describe this decomposition as a reduction of pleiotropic effects. Separating a complex system into independent components allows for the separate evolution of each component. Modules may change with limited perturbation to other modules. In addition, once these modules exist, new functions can be generated by combinatorial recombination of these modules rather than invention of new functionality from scratch.

Modularity is an important property in biology because it helps a system 'save its work' while allowing further evolution. In the natural world, one often finds modular, hierarchical structures. In this article, we will discuss the conditions under which formation of such structure may be thought of as a symmetry-breaking phase transition. We will discuss how modularity gives biological systems a greater ability to respond to change. We will review how modular structure might form spontaneously.

Modularity provides biology with a basis set to explore the space of biological possibility. From a computer science point of view, an evolving system may approximate the NP-hard problem of searching all of configuration space with a polynomial-hard problem by becoming modular and hierarchical. The analogy in physics would be achieving separability of a glassy Hamiltonian, or block diagonalization in quantum chemistry. The modular subproblems are much easier to solve, and the partial solutions are efficiently recombined to find solutions to the original problem.

What is the drawback? The drawback is that the system has placed a constraint upon the space of states to be considered. The modular and hierarchical subspace is exponentially smaller than the original space. This is the reason for the NP $\rightarrow$ P transition. Thus, there is a trade off between increased speed of convergence toward greater fitness versus a reduced density of states of solutions in a modular framework.

The advantage of modularity is commonly employed in engineering, leading some to encourage the use of modularity in evolutionary design [2]. Biological systems, however, are not designed; they are shaped by evolution. Explaining the evolutionary emergence of modularity has been a challenge, and so far no consensus has been reached [3]. Often in biology modularity is presumed to exist *a priori*, *e.g.* the genome can be decomposed into genes, even though it has been recognized that modular solutions make up only a tiny fraction of the solution space and often optimal solutions are not modular [4]. In other

words, most functions that biological systems perform could be performed better by less modular approaches.

In our review of the empirical evidence, we will show that natural and man-made systems employ modularity to a non-zero extent. That is, we will show that the polynomial approximation achieved by modularity and hierarchy has evolved in real networks. Modularity has been observed in all parts of biology on scales from proteins and genes [5] to cells [6, 7] to organs [8] to ecosystems [9, 10]. Proteins are often made up of almost independent modules, which may be exchanged through evolution. Pieces of DNA that encode these distinct protein modules have become organized and concatenated in the course of evolution [11]. Topological analysis of networks of genes or proteins has revealed modularity as well. Motifs [12] and modules [13] have been found in transcriptional regulation networks, and modules have been found across all scales in metabolic networks [14]. Animal body plans can also be decomposed into clear structural or functional units [15, 16]. Food webs also show compartmentalization [9]. Thus, a hierarchy of modules can be observed that spans many scales of biology.

While most biologists agree on the existence of modularity, many different definitions are in use [17]. A systems biologist might describe modules from a graph-theoretical point of view as groups of nodes that are more strongly intraconnected than interconnected [3], a geneticist might consider a set of co-expressed or co-regulated genes a module [5, 18], and an evolutionary biologist might look for conserved sequences or structures [19].

Many theories have been proposed to explain how and under which conditions modularity emerges. Some of the theories argue that selection is not essential for modularity [20] while others have explained the emergence of modularity through direct or indirect fitness benefits such as enhanced evolvability [21], facilitated horizontal gene transfer [22, 23], or improved robustness [24]. We hypothesize that a changing environment selects for adaptable frameworks, and that competition among different evolutionary frameworks leads to selection of structures with the most efficient dynamics, which are the modular ones. Here we review theories for the emergence of modularity and provide empirical evidence for its emergence under conditions for which modularity is expected to arise spontaneously.

This article is organized as follows. In section 2 we review different theoretical descriptions for how modularity might emerge. The section culminates with a theory for the spontaneous emergence of modularity based upon three axioms: a changing environment, exchange of genetic material between individuals, and a rugged fitness landscape. In section 3 we review experimental observations of modularity. We give examples of modular systems in pathogens, metabolic and genetic networks, and protein-protein interactions. We discuss modularity in ecological networks, physiology, and social networks. In each case, we emphasize the spontaneous nature of the emergence of modularity, and how environmental pressures have lead to increasingly modular systems. We conclude in section 4.

2. Theoretical Models

2.1. Neutral Models

Neutral theory is a base case, a null model, of evolutionary theory. In neutral theories, evolution is considered to be an unbiased random walk through the state space of all possibilities. While real evolution is clearly shaped by selection, and so evolutionary trajectories are biased, neutral theory remains a bastion of theoretical effort.

2.1.1. Duplication-Differentiation—Most neutral theories for the emergence of modularity have focused on the idea of duplication. If parts of a system are duplicated, the result will be more modular than the original system. It has been shown, for example, that artificial networks created by a duplication operator can have a hierarchical modular structure that is similar to the structure observed in the yeast protein-protein interaction network [25]. A parameter of this duplication-differentiation process controls a phase transition between a highly-connected graph and a sparsely connected graph. Close to the critical value, the resulting networks are scale-free, small world network, with a modular structure [20]. Solé and Fernandez suggested that natural selection might have tuned this parameter so that networks are sparse but completely connected. Modularity would then emerge as a byproduct without any selection pressure. It was also shown that the distribution of subgraphs in the resulting networks matches that observed in the human interactome, the yeast proteome, and a subset of human transcription factors [20].

While the previous studies explored the purely topological growth of networks without regard for biological function, Soyer extended the idea of duplication-differentiation to include a constant selective pressure [26]. In his model, the requirement that a pathway be able to respond to two different signals leads to the emergence of modularity in regulatory pathways. However, this modularity is hard to maintain because of drift to non-modular systems with equal fitness. Including horizontal gene transfer into the model may affect the results.

A challenge to theories for the emergence of modularity based on gene duplication comes from an empirical study [27] of the evolution of transcriptional regulation in *E. Coli*. There, it was shown that most of the transcription factors in *E. Coli* did not evolve by gene duplication but rather by horizontal gene transfer. The same study also observed two trends in the evolution of gene regulatory networks that seem inconsistent with nearly neutral theories. Many similarities between paralogs can be shown to be the result of convergent evolution. They have not been conserved from the common ancestor. In addition, the regulation of genes that were horizontally transferred tend to be more complex than that of native genes. These results cannot be explained by neutral theories.

2.1.2. Neutral Modular Restructuring—Force *et al*. have developed a near-neutral model for the emergence of genotypic modularity based on mutation, duplication, and genetic drift [28]. In a first step, pleiotropic constraints may be reduced by neutral changes in gene architecture without altering the phenotype. This can happen if functions that were regulated together evolve to be regulated independently. The benefits of this restructuring in regulation may not be realized immediately, especially in a constant environment. However,

if the environment changes, this genotypic modularity may provide a selective advantage which may promote phenotypic modularity. Force *et al.* stress the distinction between the neutral change of the genomic architecture and its effects on the subsequent phenotypic evolution.

Neutral models may provide a theory for the initial appearance of modular structures, but they cannot explain why such a structure would persist in the presence of more optimal nonmodular structures [29].

2.2. Models Involving Natural Selection

Modularity contributes positively to fitness by several indirect means. Modular systems are more robust because the effect of perturbations can be contained within a module. A failure of one part does not affect the entire system. Modularity also enhances evolvability because it allows different parts to be optimized separately without impairing the functioning of other parts. In addition, once modules exist, they can be reused to facilitate further evolutionary adaptation. New functionality does not have to be created from scratch, but rather can result from different combination of existing modules. Furthermore, a rewiring of modules can be achieved quickly in response to environmental perturbation. Finally, modularity makes the exchange of genetic information much easier. Because of these benefits, modularity speeds up evolution and can thus be selected for directly or indirectly.

Studies on a smooth fitness landscape, however, have failed to capture these benefits of modularity. Orr studied the evolution on a smooth landscape given by the Fisher model and found that the rate of adaptation decreases with increasing complexity, as measured by the dimensionality of the adaptive landscape [30]. He derived analytically that the rate of change of fitness is given by

$$\frac{d\overline{w}}{dt} = -\frac{4N\mu r^2}{n} M \, \overline{w} \ln \overline{w}, \quad (1)$$

or, if time is measured in units of $(N\mu)^{-1}$ generations,

$$\frac{d\overline{w}}{dt} = -\frac{4r^2}{n} M \, \overline{w} \ln \overline{w}, \quad (2)$$

where

$$M = \frac{1}{\sqrt{2\pi}} \int_x^\infty (y-x)^2 \exp\left[-\frac{y^2}{2}\right] \quad (3)$$

and

$$x = \frac{r\sqrt{n}}{2\sqrt{-2\ln \overline{w}}}. \quad (4)$$

Here, $\bar{w}$ is the average fitness, N is the population size, μ is the mutation rate, n is the number of independent characters, r is the size of mutations in phenotype-space, $x = r\sqrt{n}/2z$ is a normalized measure of the size of mutations, and z is the initial distance from the optimum. This result shows that on a smooth fitness landscape described by Fisher's model, the rate of adaptation decreases with increasing complexity at least as fast as n^{-1}. This is a consequence of the fact that in complex organisms random mutations are less likely to be favorable, the probability of fixation is lower, and the increase in fitness is smaller in the event of fixation; this assumes that one considers mutations whose size is independent of complexity [30]. Thus, there appears to be a selective advantage in reducing complexity as measured by the number of independent characters.

Fisher's original model has a universal pleiotropy. Welch and Waxman investigated [31] how modularity, which they modeled by a modularly parceled pleiotropy, would affect Orr's results on the cost of complexity. Surprisingly they found that modular pleiotropy cannot eliminate the cost of complexity. The n^{-1} dependence observed by Orr is also observed if the degree of pleiotropy is restricted in a modular fashion. Modularity, as considered in their paper, increases the rate of adaptation if a single trait is maladapted, but may decrease the adaptation rate if all traits are equally maladapted. Therefore, modularity understood as a reduction in the number of traits that can be affected by a single mutation, does not necessarily provide a fitness advantage and cannot explain the increase in complexity observed over evolutionary time scales. On a smooth fitness landscape there always seems to be a cost associated with increasing complexity. Explanations for the emergence of modularity thus may have to involve a different set of assumptions than the Fischer model. For example, on a rugged fitness landscape complexity may increase the rate of evolution.

2.2.1. Selection for Stability or Robustness—Robustness is a generic property of biological systems [24]. It describes their ability to resist perturbations. A modular structure enhances the robustness of a system by decreasing the spread of a perturbation. Since robustness improves the fitness of an organism, and modularity contributes to fitness, modularity can be co-selected. This hypothesis has been confirmed in a study [32] of linear dynamics on a network. In this model evolving individuals are represented by matrices whose fitness is defined by the number of eigenvalues whose real part is negative. As the individuals evolve towards higher fitness they develop a hierarchical modularity which improves their robustness. Modularity is implicitly selected because of the selection for robustness.

2.2.2. Direct Selection for Modularity—It has also been proposed that modularity may be directly selected for rather than indirectly as a side effect of stability or robustness. Rainey and Cooper considered the following situation [22]. An environment, referred to as environment 1, contains an unexploited niche which requires a major evolutionary innovation to be exploited. This innovation happens to have evolved in a different lineage in a different location, referred to as environment 2. Organisms from environment 2 may be transported into environment 1, in which they cannot survive. Thus, they lyse and their DNA is released into environment 1. In such a situation, cells in environment 1 with the greatest ability to accommodate this DNA will benefit the most since they can now exploit the new

niche in environment 1. Thus, the ability to accommodate horizontally transferred genes confers a fitness advantage in the presence of unexploited niches and available DNA from individuals. This ability will be reduced by pleiotropic effects and enhanced by modular genome architectures. Therefore, modularity will be directly selected for in the presence of horizontal gene transfer and ecological opportunity because it facilitates the accommodation of foreign DNA, which may be beneficial. The authors emphasize that in their model modularity may increase evolvability but this is not the cause of the emergence of modularity.

2.2.3. Templated Modularity—A set of studies on the evolutionary emergence of modularity examined the evolutionary dynamics in a changing environment with goals that vary in a modular fashion such that each new goal shares subgoals with the previous goal. Such modularly varying goals have been found to lead to the spontaneous emergence of modularity.

In a first study [29], Kashtan and Alon studied the evolution of Boolean logic circuits and neural networks. When these systems were evolved under a fixed goal, they did not develop modularity. Even if the systems were started from a modular state, this modularity quickly decreased. On the other hand, if the systems were exposed to goals that periodically vary in a modular fashion, modularity did emerge and evolution proceeded faster. Randomly varying the environment did not lead to the emergence of modularity and resulted in a smaller increase in the rate of evolution than modularly varying goals.

An extension [33] of this work analyzed the nature of the environmental variation in more detail by comparing modularly varying goals to different types of randomly varying goals. Here it was found that modularly varying goals generally lead a to large speedup of evolution while randomly varying goals may or may not increase the rate of evolution. The advantage of goals that vary in a modular fashion is more pronounced for more complex goals. As an explanation for this observation, Kashtan *et al.* suggested that modularly varying goals can move populations away from local fitness maxima near which they may be stuck.

The results from the previous two paragraphs have been confirmed in an analytic model of evolution under modularly varying goals [4]. In this linear model, an evolving individual is represented by a matrix A which maps an input v to an output u by $Av = u$. The fitness of the individual is defined as

$$F(A) - F_0 = -\varepsilon \|A\|^2 - \|AV - U\|^2, \quad (5)$$

where the first term on the right-hand side ascribes a cost to the individual based on the magnitude of the elements of A and the second term represents a reward for correctly mapping a given set of inputs V to a given set of outputs U. If A is block-diagonal or almost so, it will be considered modular. In this analytic model, Kashtan *et al.* observed the same trends as before. Constant goals lead to non-modular structures and slow convergence, while modularly varying goals lead to modular solutions and fast convergence, especially for

harder goals. If goals stop varying, modularity decreases, and randomly changing goals generally result in evolutionary confusion.

Most recently, Kashtan *et al.* considered evolution in a spatially, rather than temporally, heterogeneous environment in the presence of extinctions [34]. In this case they observed that without extinctions networks evolve to be highly optimal but not modular, whereas with extinctions modularity emerges. Here they suggest that modularity is selected for because it enables individuals to rapidly adapt to free niches after an extinction event. They also realized that, in contrast to their previous studies, recombination is very important for rapid adaption to new niches.

2.2.4. Spontaneous Emergence of Modularity as a Phase Transition—In this section we will show in general how modularity emerges spontaneously under a small set of assumptions. The argument is motivated by two observations. First, evolvability is a selectable trait and will be selected for in a changing environment [35]. Second, modularity enhances the evolvability of organisms [21]. Consequently, the selection for evolvability leads to the emergence of modularity. We will first review an evolutionary model in which modularity emerges spontaneously. In this model, a population of individuals evolves on a rugged fitness landscape, the individuals engage in horizontal gene transfer, and the environment is changing in time. We will then review some results that this model may explain.

Evolvability evolves [35]. The evolvability of organisms is determined by the mutational processes acting on their genome and the rates at which these occur. Such processes may include point mutation, recombination, transposition, or horizontal gene transfer. The capability to perform these processes as well as the rates at which they occur are encoded in the genome and, thus, under selective pressure. In a changing environment, organisms with adaptable evolutionary frameworks will have a fitness advantage over their peers that are not as adaptable. This advantage imposes a selection for adaptable frameworks and hence evolvability.

As described in the previous sections, modularity confers many benefits to an evolving system. It allows for biological information to be stored in pieces or to be swapped in large chunks. It also enhances the robustness of a system and makes it more evolvable. However, do these benefits imply that modularity is inevitable? What is the probability for the emergence of modularity? Is modularity a typical or a special case? Here we explore the conjecture that modularity will spontaneously emerge in any evolving population which evolves on a rugged fitness landscape in a changing environment and undergoes horizontal gene transfer.

Lipson *et al.* were among the first to describe a quantitative model for the emergence of modularity [2]. In a simple linear algebra model they suggested that modularity arises spontaneously in response to variation. However, Gardner and Zuidema analyzed the model of Lipson *et al.* and came to the conclusion that it failed to establish a clear link between modularity and evolvability [36]. As we will see, one reason is that Lipson *et al.* did not consider horizontal gene transfer.

The spontaneous emergence of modularity has been observed in a generic evolutionary model described in [37] and extended in [38]. In this model evolution is assumed to occur on a rugged fitness landscape with many local optima. Such a rugged landscape, which imposes a pressure for efficient evolutionary structures, can be generically described by a spin glass Hamiltonian [38],

$$H^\alpha \left(s^{\alpha,l}\right) = \frac{1}{2\sqrt{N_D}} \sum_{i \neq j} \sigma_{i,j} \left(s_i^{\alpha,l}, s_j^{\alpha,l}\right) \Delta_{i,j}^\alpha. \quad (6)$$

Here $s_i^{\alpha,l}$ is a string by which "individual" l can be identified. The index i is in the range 1 i N, where N is the length of the string. This is a generic model of evolution in which $s_i^{\alpha,l}$ can, for example, represent an amino acid in a protein sequence or a protein in the genome. The variable l is a label for the different individuals in the population such that 1 l N_{size}, where N_{size} is the number of individuals. The interactions between the $s_i^{\alpha,l}$ are governed by a structure, or connection matrix $\Delta_{i,j}^\alpha$, whose possible forms are enumerated by a with 1 a D_{size}. Here D_{size} is the number of possible structures. The connection matrix generically represents the structure of interactions such as, for example, protein folds, or regulatory constraints. The matrix $\sigma_{i,j}(s_i, s_j)$ is symmetric in i and j and represents the interaction strength. Its values are chosen from a standard normal distribution. These random couplings encode the effect of the environment. They can take on both negative and positive values, which corresponds to having both ferromagnetic and anti-ferromagnetic interactions, leading to frustration. As a consequence, the fitness landscape is rugged with a large number of local extrema. Choosing the $\sigma_{i,j}(s_i, s_j)$ in a correlated way can lead to a less rugged, or even smooth if all are positive and equal, landscape [37]. The structure a is described by the matrix $\Delta_{i,j}^\alpha$, a binary symmetric contact matrix. The number of connections in each structure and hence the number of non-zero elements in $\Delta_{i,j}^\alpha$ is constrained to be a fixed number N_D. This ensures that modularity cannot emerge as a consequence of an increasing number of connections. Connections can only be redistributed.

Horizontal gene transfer (HGT) is restricted to predefined blocks of equal length and the rate of horizontal gene transfer and the mutation rate are approximately equal. The modularity M is defined to be the number of non-zero elements in blocks along the diagonal of $\Delta_{i,j}^\alpha$. The size of the blocks is equal to the length of the HGT segments. Note that this modularity will have a non-zero value, M_0, even for random distributions of connections.

This model allows one to study not only the evolution of individuals in a variable environment, but also the evolution of structural connections $\Delta_{i,j}^\alpha$. A simulation thus models the evolution of a population of D_{size} structures, and for each structure a population of N_{size} sequences.

There are three different levels of evolutionary change in this model. First, the sequences within each structure change most rapidly by point mutation and horizontal gene transfer. In each round, the 50% of the population with the highest fitness within each structure are randomly duplicated, where the fitness of each individual is given by the spin glass

Hamiltonian in equation (6). The value of the Hamiltonian is the energy, while the fitness is non-decreasing in the negative of the energy. Second, the environment changes after T_2 rounds of mutation, recombination, and selection. Environmental change is accomplished by assigning a new value to each of the elements of $\sigma_{i,j}$ with probability p. Thus, p represents the severity of environmental change and $1/T_2$ the frequency. Third, the evolution of the structures represents the slowest change. The $\Delta_{i,j}^{\alpha}$ undergo mutation and selection every T_3 rounds. The fitness of the structures is obtained by averaging the fitness of the sequences in each structure over the T_3/T_2 environmental changes. Of the structures only the top 5% of the population are selected for the next round, which are randomly amplified to maintain a constant population size and also mutated.

Simulations within this model show the emergence of modularity when the environment is changing and when horizontal gene transfer is present. Figure 1 shows the emergence of modularity for one instance of the model. $M_0 = 22$ is the modularity of the random state. The modularity grows linearly, because it is far from its steady-state value. This growth of modularity can be considered a symmetry-breaking event, where the order parameter is the excess modularity $M - M_0$. The symmetry being broken is the permutation symmetry of the connection matrix. It is the linear topology of the HGT event that allows this symmetry to be broken. The matrix moves from an initially uniform random distribution of entries to a modular distribution of entries clustered along the diagonal. Figure 2 shows that over large time scales this increase in modularity is associated with an increase in fitness, measured as a decrease in energy. Figure 3 shows the change of fitness for short times. During times of constant environment, the fitness increases rapidly, but every time the environment changes, the fitness decreases substantially because the sequences are not well adapted to the new environment. These results are not sensitive to the values chosen for the parameters [38].

The evolution of evolvability can also be observed in this model. Evolvability can be measured by the increase in fitness while the sequences evolve in one environment. Over long time scales, the average gain in fitness between two subsequent environmental changes is not constant. As figure 4 shows, there is a clear trend towards increasing evolvability over time. The evolved modular structure of the connection matrix allows the sequences to evolve faster in each new environment.

These results robustly persist if a different initial contact matrix is used. Choosing a scale-free network yields almost identical results [38]. Similarly, relaxing the biologically motivated constraint that horizontal gene transfer can only recombine predefined pieces of equal length does not hinder the emergence of modularity. For gene transfer that starts at any position with equal probability and swaps pieces of a Poisson random length, the evolution of modularity, shown in figure 5, robustly shows the expected trend.

Without environmental variability or without horizontal gene transfer, however, no emergence of modularity is observed. As figure 6 shows, the modularity remains near M_0, the value for random networks. It is important to note that neither environmental change, nor horizontal gene transfer explicitly favor modularity. Rather, the system adopts a modular state under these conditions because modularity allows the system to respond better to the continuously changing environment. Thus, there is an implicit selection for evolvability in a

variable environment, and horizontal gene transfer increases the evolvability of modular systems. In combination, horizontal gene transfer and environmental change implicitly select for modularity. One would therefore expect that the degree of modularity increases with increasing environmental change. Figure 7 shows such a trend for varying degrees of the severity of environmental change. From an initially modular state, modularity decreases if there is no environmental pressure, while it increases in the presence of environmental change. For larger values of p, modularity increases more rapidly. This trend can be seen more clearly in the derivative of modularity with respect to time, shown in figure 8 for different severity of environmental change.

A similar observation can be made when varying the frequency of environmental change rather than the intensity. For very high frequencies, modularity decreases with frequency because the environment is changing too fast for the system to evolve in response to it. But, as figure 9 shows, for moderate frequencies of environmental change, modularity increases with frequency just as it did with magnitude of environmental change. Figure 10 shows the rate of change of modularity versus frequency of environmental change.

The emergence of modularity is a response to the past variation in the environment of the system. Therefore, one can make the argument that in analogy to the fluctuation-dissipation theorem one would expect modularity to be proportional to the variance of the previously encountered environments [38]. As figures 8 and 10 show, this can indeed be observed in this model. Similarly, in analogy to the competition between energy and entropy, one would expect there to be a steady-state value of modularity below the maximum modularity which depends on the parameters of the system. At this steady state there will be a balance between the entropic forces of random mutations driving modularity to its baseline value, M_0, and the selective forces which seek to enhance evolvability by increasing modularity. This effect can also be observed in the model by starting the system in a highly modular state. As figure 11 shows, with time the modularity decreases from this very high value.

This model illustrates the spontaneous emergence of modularity in a population of evolving individuals under two conditions: the individuals can engage in horizontal gene transfer and the environment changes. This emergence of modularity is a symmetry-breaking event caused by the selection for evolvability in a changing environment [35] and the fact that modular systems can take advantage of horizontal gene transfer to adapt to a new environment. The rate of modularity growth increases with the amplitude and frequency of environmental change. A constant environment does not promote the emergence of modularity.

Other theoretical studies have been performed which support the results of this model. For example, Crombach and Hogeweg studied the evolution of simulated gene regulatory networks [39] and confirmed the observation that alternating environments lead to the evolution of evolvability. They found that even though mutations are random, their phenotypic effects become strongly biased by an evolving genotype-phenotype map. Martin and Wagner investigated the effects of recombination on models of transcriptional regulation circuits [40]. Their findings include that the presence of recombination leads to

the emergence of modular regulatory control, a reduction in the deleterious effects of mutations, and greater phenotypic diversity.

Misevic *et al*. investigated how the reproductive mode shapes the genetic architecture of digital organisms [41]. In the theoretical model described in this section, modularity is expected to emerge in the presence of horizontal gene transfer, which enables the exchange of pieces of genetic information between different individuals. This leads to new genotypes that combine genetic material from two "parent" genotypes. Sexual reproduction also allows for large-scale exchange of genetic information and results in genotypes which are combinations of two parent genotypes. Hence, one would expect horizontal gene transfer and sexual reproduction to have similar effects. In their study Misevic *et al*. found that sexual organisms have more modular and longer genomes, are more robust, and have a higher fitness [41]. In addition they observed that the strength of epistatic interactions is weaker in sexual organisms than in asexual ones and that the reproductive mode has a significant effect on the evolution of the genetic architecture. In a follow-up study [42], they probed how changing environments influence the reproductive mode of these digital organisms. They report that in rapidly and strongly changing environments sexual reproduction becomes dominant regardless of the reproductive mode in which the population starts. Furthermore, in such environments predominantly sexual populations achieve a higher average fitness. Conversely, in slowly changing environments, asexual reproduction evolved to be predominant in most populations and sexual and asexual populations are equally fit on average.

Results very similar to the ones described in this section were obtained by Callahan *et al*. in a later study who observed the spontaneous emergence of a type of modularity called collinearity in a computational model of the evolution of polyketide synthases [43]. This emergence was observed for a wide range of parameters, despite the fact that modularity provides no direct fitness benefit. Modularity emerges only in the presence of continuous evolutionary pressure and horizontal gene transfer. This result was explained by a secondary selection effect because modularity increases the fitness benefits of recombination if the environment changes rapidly.

These studies provide evidence for the importance of environmental variation and horizontal gene transfer on the evolution of evolvability and modularity and are in agreement with the theory of spontaneous emergence of modularity described in this section.

3. Experimental Observation of Modularity

In this section we review experimental observations in various biological networks that support the theory described in the previous section. We will present evidence for the evolution of evolvability, for the presence of modularity and the enhanced evolvability provided by modularity, and for the connection of environmental variation and horizontal gene transfer to the emergence of modularity.

3.1. Modularity in Pathogens

Pathogens are exposed to extreme environmental pressure and engage in extensive horizontal gene transfer. Therefore, we would expect them to evolve substantial modularity. Studies show that pathogens not only are very modular but also that this modularity enhances their evolvability by allowing them to vary mutation rates between different parts of their genome.

Structural and evolutionary modules have been observed in viruses. For example, Karlin *et al.* found that *Paramyxovirinae* are composed of six modules [19]. Viral proteins have also been shown to be modular. Ferron *et al.* found modules by homology search in sequence data and ensured the validity of these modules by considering the results from other sources such as structure definition, biological data, and additional sequence properties [44]. The modular organization they discovered has helped to characterize virus domains by structure and function. In influenza, it has been observed that the antibody immune response is dominantly directed to the epitope regions of the hemagglutinin protein on the surface of the virus particle. While the mutation rate is assumed to be constant throughout the RNA of the virus, the observed substitution rate, or observed rate of evolution, in these five epitope regions is significantly greater [45].

All organisms balance the need for stability against the need of variability. Low rates of genetic change decrease deleterious mutations, but they also reduce the probability of beneficial mutations that may be necessary for adaptation. The optimal rates of genetic moves may vary with time or space. Radman *et al.* observed that bacteria and viruses have evolved the ability to respond to these variations by genetically controlling their mutation and recombination rates to adapt to changing environments [46]. The bacterial SOS response provides a prototypical example: under genotoxic or metabolic stress, bacteria will start to express mutator genes and upregulate several recombination genes [46]. This evidence that rates of genetic change are under selective control supports the idea that evolvability can evolve.

The hypermutation described in the previous paragraph can also be limited to select parts of the genome, making it particularly useful for pathogens, which can increase their mutation and recombination rates at sites encoding surface antigens to escape the host's immune system. This result has been confirmed by other studies reviewed by Massey and Buckling who proposed that the constantly changing environment of pathogens selects for mechanisms which can generate phenotypic variation [47]. A generalized hypermutation would lead to a significantly greater mutational load than the localized hypermutation of only those parts of the genome involved in interactions with the host environment [48]. Allocating mutation and recombination rates in a modular fashion across the genome enhances evolvability under excessive environmental pressure.

The positive effect of environmental variation on evolvability was also observed by Kepler and Perelson who showed in a differential equation model of virus dynamics that the presence of compartments with different drug concentrations increases the likelihood that a resistant strain emerges [49]. Especially for high drug concentrations, resistance was found

to emerge with a higher probability in the presence of spatial heterogeneity. Thus, environmental variability improves the evolvability of viruses.

3.2. Modularity in Metabolic Networks, Gene Networks, and Protein-Protein Interaction Networks

Since the seminal paper by Hartwell *et al*. [6], the concept of modularity has been firmly established in cell biology and with it the idea that modular structures may facilitate evolutionary change. Advances in genomics and proteomics have allowed the creation of large data sets from which networks can be constructed. Many of these networks have been shown to posses a hierarchical modular structure. Care needs to be taken when analyzing networks obtained from databases of biological interactions, such as protein-protein interactions. It is difficult to estimate the error rate of experimental techniques and sometimes the overlap of the results between interaction data obtained through two different methods can be small [50]. Promising attempts have been made to judge and improve the quality of protein-protein interaction data by combining the results from different methods [51]. This approach could also enhance other experimentally determined biological networks. In this section we review some of the evidence for the presence of modularity in metabolic networks, gene networks, and protein-protein interaction networks and the relationship between modularity and environmental variation and horizontal gene transfer.

3.2.1. Metabolic Networks—Ravasz *et al*. were among the first to investigate the structure of metabolic networks in detail [52]. They found that metabolic networks have a scale-free architecture and are nevertheless highly clustered. Furthermore, the clustering coefficient in metabolic networks is independent of size, which is in stark contrast to random scale-free networks in which the clustering coefficient decreases with size. To explain these surprising findings, Ravasz *et al*. suggested a new algorithm that can generate networks with topological properties in agreement with empirical metabolic networks. The resulting networks have a hierarchical modular structure [52]. In 2006, Spirin *et al*. extended the study of metabolic networks to include evolutionary information from genomic data that allowed them to analyze both functional and evolutionary modules [14]. In this metabolic-genomic network they also found modules on different scales, indicating hierarchical modularity. DaSilva *et al*. introduced a parameter called "core coefficient" to quantify hierarchical modularity in networks and found that the core coefficient in metabolic networks significantly exceeds that of random networks [53].

After it was recognized that metabolic networks exhibit modularity, the question arose whether metabolic networks are more modular in some organisms than in others. Parter *et al*. probed this question by analyzing the relationship between environmental variability and modularity in the metabolic networks of more than one hundred species of bacteria [54]. Their study showed that there is a positive correlation between environmental variation and modularity, shown in figure 12. This correlation remains significant even after the difference in network size is taken into account. In addition, modularity was shown to have a stronger correlation with environmental variability than with phylogenetic proximity [54]. Both of these results support the hypothesis that metabolic networks of organisms under greater environmental pressure evolve to be more modular.

Not only the importance of environmental change, but also the relation between horizontal gene transfer and modularity has been supported by empirical findings in metabolic networks. In 2008, a study by Kreimer *et al.* explored modularity in more than three hundred bacterial metabolic networks and found three main determinants of modularity: network size, the environment, and horizontal gene transfer [55]. The extent of horizontal gene transfer was obtained from [56]. There it was measured by computing the probability that a DNA segment is extrinsic using Bayesian inference. A gene segment was considered extrinsic to a recipient if its nucleotide composition was significantly different from the rest of the recipient's genome while matching the nucleotide composition of a donor.

Recently it has been pointed out that the dependence on network size may be a an artifact of the method used to compute modularity [57]. It was argued that modularity as conventionally defined will tend to increase with an increasing number of modules or nodes. This is a consequence of the null model implicit in the definition of modularity: a random graph that has the same degree sequence as the network under investigation. Since the probability that an edge will fall within a given module in this random network decreases with increasing network size, a larger network will tend to get a higher modularity score [57]. This can also be understood by realizing that stochastic noise will allow an algorithm to detect modules in any network, especially in sparse networks. For larger networks, there will be more such noise-induced modules, giving the impression of greater modularity. See figure 13 for an example of this size-dependent modularity in random networks. From this figure is clear, however, that the modularity observed in the natural network is significantly greater than the noise-induced value observed in the background model of a random network. Thus, any comparison of modularity in networks of different size or number of edges always needs to consider the values of modularity one would expect to obtain in a distribution of random networks with the same size and sparsity. However, the correlations observed between modularity and horizontal gene transfer and between modularity and environmental effects should not be affected by this methodological bias. In particular, the data show that even bacteria with small metabolic networks, in which modularity may be underestimated by the algorithm, exhibit high modularity if they live in highly variable environments [55]. It was also shown that bacteria occupying a limited number of niches have less modular metabolic networks than species occupying a greater variety of niches and pathogens that alternate between hosts have more modular metabolic networks than do single-host pathogens [55]. Thus, environmental variability and horizontal gene transfer seem to be closely related to the modular structure of bacterial metabolic networks.

3.2.2. Gene Networks—In gene networks, the advantages of a modular architecture become highly apparent. The possibility of using novel combinations of modules rather than evolving new genes from scratch greatly enhances evolvability. A genomic study has shown that such a modular rewiring contributed significantly to the evolution of new functionality, especially in the evolution of proteins [58]. Moreover, it was reported that the loss of intermodule introns inhibits further modular evolution.

Segré *et al.* constructed an epistatic interaction network from the results of all single and double knockouts of almost one thousand metabolic genes in *S. cerevisiae* [59]. They found that this network consists of modules of genes arranged in a hierarchy, where modules are

clusters of genes which interact monochromatically (either all aggravating or all buffering). Based on this observation they suggested to extend the concept of epistasis from genes to functional modules, which interact epistatically as a group. In other words, a second-order modularity of epistasis had emerged.

Bhattacharyya *et al*. reviewed the importance of modular interactions in cell signaling circuits [60]. They proposed that in such circuits, modularity may contribute to evolvability by making the evolution of new complex circuits and resulting phenotypes easier. This availability of new phenotypes would be especially beneficial in competitive and changing environments and may explain how modularity is maintained despite nonmodular systems often having a higher fitness in an unchanging environment. Modularity in gene networks may also contribute to an organism's robustness and the ability to maintain homeostasis [61].

Following the theory developed in section 2.2.4, we would expect modularity to emerge in organisms in response to environmental pressure. Similarly, we would anticipate that systems which are directly involved in interactions with the environment would evolve to be more modular than systems which have no external interactions. This prediction has been verified by Singh *et al.* in an evolutionary study of three bacterial stress response networks [13]. They observed that the regulatory network for chemotaxis, a process which allows an immediate response to the environment, has greater modularity than that for sporulation, which is more indirectly affected by the environment. Furthermore, the network regulating DNA uptake, which is hardly impacted by the environment, displays no significant modularity. These results illustrate the influence of environmental change on the emergence of modularity in stress response networks.

Gene regulation can be considered a higher order modularity, and it was mentioned at the end of the previous subsection on metabolic networks. Modularity in gene regulatory networks decreases the complexity of the circuitry required for complex responses to external stimuli [23]. Transcriptional regulation factors are often acquired through horizontal gene transfer [27] and it has been observed that regulatory circuits evolves faster than the genes they regulate [62, 63]. Thus, the higher order modularity displayed by gene regulation enhances evolvability.

3.2.3. Protein-Protein Interaction Networks—Modularity has also been observed in protein-protein interaction networks. An example of a clustered protein-protein interaction network with visually discernible modules is shown in figure 14. One of the first studies of protein-protein interaction networks on a mesoscale level was carried out by Spirin *et al*. in 2003 who discovered highly statistically significant modules [64]. Functional modules in protein interaction networks can also be found from sequence data alone and agree with modules found by other methods [65]. This indicates that the modularity in protein networks is encoded in the genome. A study that incorporated data from a variety of sources including gene expression, functional annotations, evolutionary conservation, and protein structure supports the observation of a modular topology in protein interaction networks [66].

Han *et al.* studied the modular structure of protein networks in more detail by considering their temporal changes [67]. They found two types of hubs which they named "party hubs" and "date hubs." While party hubs interact with most of their partners at the same time, date hubs bind their partners at different times or places. They investigated the topological role these different types of hubs play in the protein network. They observed that party hubs function mainly inside of modules while date hubs act as global connectors between modules. Consequently, the network is less resilient to the removal of date hubs than it is to the removal of party hubs. This hierarchy of hubs provides evidence for hierarchical modularity in protein-protein interaction networks.

The influence of the environment on modularity that was seen in metabolic networks and gene networks has also been noted in protein-protein interaction networks. A study by Cohen-Gihon *et al.* showed that a protein with a function that is common to all organisms exhibits a lower degree of structural modularity than a protein that can only be found in few cell types [68]. Campillos *et al.* obtained even more explicit evidence for the effects of changing environments and horizontal gene transfer on modularity by studying evolutionarily cohesive functional modules in protein networks [69]. This allowed them to compare modules by evolutionary age. They found that young modules are frequently horizontally transferred between species. These young modules are enriched in functions related to interactions with the environment. These young modules also play an important role in the adaptation to new environments of species. Ancient modules, on the other hand, are often very well conserved and enriched in core functions such as metabolism and information processing. Furthermore, bacteria living in competitive, varying, and stressful environments acquired the most modules [69]. These observations clearly demonstrate how important horizontal gene transfer and environmental heterogeneity in space or time are for the presence of modularity.

A recent piece of evidence for the hypothesis that modularity emerges spontaneously in the presence of environmental change and horizontal gene transfer comes from a quantitative study of the evolution of modularity across evolutionary time scales [38]. In this paper, a measure of protein divergence time was used to show that modularity in protein interaction networks has increased with time. The evolutionary age of proteins was quantified using the concept of compositional age. This method considers proteins to be older if they contain a larger fraction of older amino acids. The calibration of this measure using known divergence times enabled the construction of a mapping between compositional age and real age. To quantify modularity, topological overlap matrices were constructed from the interaction networks and reordered using average linkage hierarchical clustering. Modularity was computed using several different quantitative definitions and normalized by network size. For all definitions it was robustly observed that modularity has grown throughout evolutionary time in both organisms that were studied [38]. These results are consistent with the theory that environmental change and horizontal gene transfer naturally lead to an evolution of increased modularity.

In reference [38], the graphs displaying the growth of modularity as a function of evolutionary time show that modularity, after increasing initially, seems to saturate. Considering that the rate of evolution is believed to increase rather than saturate, this is at

first a surprising result. To understand this observation, we extended the analysis presented in [38] by computing a second order modularity. We constructed a new weighted network whose nodes are the projections of the modules of the original network, that is we carried out one-step of a renormalization group operation. If, in the original network, nodes in one module had connections to nodes in other modules, then the vertices corresponding to these two modules in the new network were connected. The edges in the new network were assigned weights equal to the number of such inter-module connections in the original network. We then quantified the modularity in this new network of modules by the same method described in [38]. We observed that this second order modularity also increases with evolutionary time, as shown in figure 15. Notably, it continues to rise even after the first order modularity seems to have saturated. This may indicate the emergence of a higher level of structure.

All the findings described above for metabolic networks, gene networks, and protein-protein interaction networks are in agreement with and can be explained by the theory of spontaneous emergence of modularity described in section 2.2.4. In the presence of horizontal gene transfer and environmental pressure, modularity will emerge spontaneously in a population of evolving individuals.

3.3. Modularity in Ecological Networks

Ecological networks summarize interactions among all species in an ecosystem, represented as nodes, and the biotic interactions between them, represented as edges between the nodes. There are typically two types of interactions that can exist between species: mutualistic or trophic. A trophic link between two species indicates that one eats the other, an antagonistic relationship, while a mutualistic link indicates a relationship in which both species benefit, such as between plant and their pollinators or seed dispersers. Most empirical studies of ecological networks focus on only one of these types of interaction which results in either trophic networks (food webs) or mutualistic networks (*e.g.* pollination networks).

Topological analysis has revealed that most food webs have a hierarchical structure and some can be decomposed into compartments or modules [9]. This modularity increases the stability of food webs [70] by localizing the impact of a disturbance within a single compartment and minimizing impact on other compartments [9].

We first describe why mutualistic networks are not expected to be modular. A recent study by Thebault and Fontaine found that the relation between network architecture and stability is fundamentally different between mutualistic networks and trophic networks [70]. While in the latter compartmentalization increases stability, it has the opposite effect in the former. The difference between trophic and mutalistic networks is the difference between a rugged and a smooth fitness landscape. Trophic interactions lead to frustration while mutualistic interactions do not. One model of evolution of ecological networks is the "tangled nature" model introduced in [71]. In this model, the evolution is governed by a replication rate, or microscopic fitness, which is very similar to the negative of the spin-glass Hamiltonian (6):

$$H(S^\alpha, t) = \frac{1}{N(t)} \sum_{S \in \mathscr{S}} n(S, t) \sum_{i=1}^{L} J_i(S^\alpha, S) S_i^\alpha S_i - \mu N(t). \quad (7)$$

Here t is the time, and the vector S^a, whose elements can only take on the values ± 1, describes an individual. The sum over S runs over the entire genome space $\mathscr{S}$, $N(t)$ is the population size at time t, and $n(S, t)$ is the occupancy of position S at time t (how many individuals have genotype S at time t). The interaction matrix $J^{ab} = J(S^a, S^b)$ does not change with time and describes the interactions (trophic, mutualistic, or competitive) between an individual with genotype a and and individual with genotype b. At any given time only a fraction of the entire genome space is occupied and hence contributes to H. The second term $\mu N(t)$ describes the limited availability of resources in the environment, where μ is the mean sustainable population size. The main differences between the tangled nature model and the spin-glass Hamiltonian (6) are that in the former the interactions are not symmetric — an interaction between two individuals can benefit one individual, but harm the other — and that the fitness of each individual depends on the occupation number of all positions in genotype space to which it is connected through a non-zero interaction term J_i.

In a mutualistic network, all interactions are beneficial for both parties, and hence, the values of all J_i are positive. This is analogous to having only ferromagnetic couplings, in which case there is no frustration and it is easy to find the ground state. In a trophic network, however, interactions are always beneficial for one party and detrimental for the other. This is comparable to a spin-glass with ferromagnetic and anti-ferromagnetic couplings, which is characterized by frustration and slow dynamics. This difference in the dynamics between mutualistic and trophic networks has been observed numerically in variations of the tangled nature model. Rikvold and Sevim studied [72, 73] the distribution of the durations of quasi-steady states in mutualistic and trophic networks obtained from simulations and found that both follow a power law. The power-law exponent for the mutualistic case is more negative than that of the trophic case, which indicates that the mutualistic network is characterized by faster dynamics, as it would be expected for evolution on a smooth landscape. It was also observed that trophic interactions lead to hierarchically structured networks in the simulations while mutualistic interactions do not [72, 73]. As discussed in section 2.2.4, if evolution occurs slowly on a rugged fitness landscape, we expect the emergence of modularity. However, if evolution occurs rapidly, as it does on a smooth landscape, the discussion at the beginning of section 2.2 suggests that we should not expect emergence of modularity. By this independent line of reasoning, the results of [72, 73] also suggest that the dynamics in mutalistic networks are rapid. Thus, modularity is not expected to spontaneously arise in mutalistic networks.

On the other hand, based on the general theory proposed in section 2.2.4, we expect that food webs under greater environmental pressure might evolve to become more hierarchical. To test this hypothesis we investigated hierarchy in 22 empirical food webs from rivers in New Zealand. The data was assembled by Thompson and Townsend (*e.g.* [74, 75]) and provided on the Interaction Web Database[1]. We restricted our study to a limited geographical region to reduce the effect of potential confounding factors on food web

architecture such as latitude or biome. Including only river food webs also minimizes the influence different habitats may have on network topology. As a proxy for environmental pressure we considered the availability of energy from detritus (particulate organic matter) input. Detritus is central to many food webs [76] and particularly small rivers are fueled by detritus input from surrounding terrestrial plants [77]. Thompson and Townsend showed that the forest river food webs in this data set have a much greater concentration of both coarse and fine particulate organic matter than their counterparts flowing through grassland [75].

Our analysis proceeded as follows. We used the Euclidean commute time as a distance metric, defined for each pair of nodes as the expected time it takes a random walk to travel from one of the nodes to the other and back [78]. This Euclidean commute time between the nodes of a weighted graph decreases when the number of paths connecting two nodes increases. The commute time between two nodes also decreases when the length of any path connecting the nodes decreases. These properties make the Euclidean commute time well-suited for clustering tasks [79]. Let L denote the graph Laplacian, defined as $L = D - A$, where A is the adjacency matrix and $D = \mathrm{diag}(A_i)$ with $A_i = \Sigma_j A_{ij}$ is the degree matrix, a diagonal matrix whose elements are the degrees of the nodes. It is shown in [79] that the computation of the average commute time can be obtained from L^+, the Moore-Penrose pseudoinverse [80] of the graph Laplacian L, by

$$n(i,j) = V_G \, (\boldsymbol{e}_i - \boldsymbol{e}_j)^T \, L^+ \, (\boldsymbol{e}_i - \boldsymbol{e}_j). \quad (8)$$

Here $(\boldsymbol{e}_i)_j = \delta_{ij}$ and $V_G = \Sigma_{ij}\, a_{ij}$. Since it can be shown [79] that L^+ is symmetric and positive semidefinite, $T_{ij} = [n(i,j)]^{1/2}$ is a Euclidean distance metric, called the Euclidean commute time (ECT) distance.

After finding the commute distances, we performed average linkage hierarchical clustering on the matrices of commute distances to build a hierarchy of clusters. Finally, we quantified hierarchy by computing the cophenetic correlation coefficient (CCC) for each network. The CCC is a measure of how well the dendrogram distances correlate with the original commute distances. The CCC is defined as the Pearson correlation coefficient between the node-node distances in the original data and that in the tree-like representation:

$$\mathrm{CCC} = \frac{\sum_{i<j}(T_{ij}-T)\,(c_{ij}-c)}{\sqrt{\sum_{i<j}(T_{ij}-T)^2 \sum_{i<j}(c_{ij}-c)^2}}, \quad (9)$$

where T is the average of the commute distances, T_{ij}, and c is the average of the dendrogram distances, c_{ij}. Our results are shown in figure 16. The CCC values are significantly different between rivers surrounded by forest (pine, broadleaf) and rivers surrounded by grassland (tussock, pasture). Consistent with our hypothesis, food webs of rivers flowing through grassland, with little detritus input, and thus with substantial environmental pressure, are

---

[1] http://www.nceas.ucsb.edu/interactionweb/

more hierarchical than food webs of rivers flowing through forests, which provide greater detritus input.

In the particular data set under study, we observed a trend between CCC values and network size that is shown in figure 17. To determine whether this trend may bias our results we compared the CCC values of each food web to the distribution of CCC values of the largest connected components of 100 random networks of the same size and total number of edges as the food web. This allowed us, in analogy to the normalized modularity given in [54], to define a normalized CCC as

$$\mathrm{CCC_{norm}} = \frac{\mathrm{CCC - CCC_{rand}}}{1 - \mathrm{CCC_{rand}}}, \quad (10)$$

where CCC is the CCC value of the food web, and $\mathrm{CCC_{rand}}$ is the average CCC value of the random networks. To consider not only the mean of the CCC values of the random networks, but also their distribution, we also computed the standard score (z-score) of each food web CCC relative to the distribution of CCC values of the random networks of the same size and sparsity:

$$\mathrm{Z_{CCC}} = \frac{\mathrm{CCC - CCC_{rand}}}{\sigma}, \quad (11)$$

where σ is the standard deviation of the CCC values of the random networks. As shown in figure 18, the normalized CCC values show the same trend as the unnormalized values, indicating that network size is not the cause of the observed trend.

For most of the food webs that we analyzed, we were able to find explicit numbers for the amount of fine particulate organic matter present in the river from [75]. This allowed us to investigate how CCC varies explicitly with detritus input rather than with the surrounding habitat. The results in figure 19 show a clear trend of decreasing CCC with decreasing detritus availability. Not all fine particulate organic matter in a river is a consequence of detritus input but significant detritus input will lead to more particulate organic matter.

The analysis of food webs also revealed that modularity and hierarchy, despite being closely related concepts, do not have to be positively correlated. Indeed, we observed the opposite trend for this food web data set. A comparison of the CCC values and Newman's modularity, maximized using the spectral algorithm described in [81], is shown in figure 20. The data show a negative correlation between these measures of hierarchy and modularity. Note also, however, that the values of Newman's modularity are quite low for most of the food webs and may not indicate significant modularity. While we often speak of modularity as a short-hand, the fundamental NP $\rightarrow$ P transition is induced by multiple levels of modularity, *i.e.* hierarchy. Therefore, hierarchy is the more fundamental order parameter to consider. Our analysis of New Zealand river food webs shows an increase in hierarchy with increasing environmental pressure.

3.4. Modularity in Development

It has long been recognized that the body plans of higher organisms are modular: they consist of easily identifiable parts which serve a well defined function and are structurally separated from other parts [15]. It has also been observed that among metazoans, modularity increases with complexity and that modules may be hierarchically structured [7]. This modularity in phenotype is an immediate consequence of a developmental modularity which can be observed over many levels from parts of genes to the scale of organisms [82].

Modularity in development has been associated with evolvability. Raff and Sly pointed out that modularity enables the evolution of ontogeny because it makes three processes possible: the dissociation of developmental processes (*e.g.* heterochrony), the duplication and subsequent divergence of developmental modules, and the co-option of features into new functions [83]. Thus, developmental modularity improves evolvability and allows for the emergence of complex anatomies from genomes which need not be as complex. A study by Yang [84] provides empirical evidence for the hypothesis that modularity confers evolvability. He compared the taxonomic diversity of insect lineages with different degrees of modularity and found that lineages with greater life-stage modularity have greater rates of diversification [84]. Litvin *et al*. showed that environmental change and intrinsic genetic variation can alter the connectivity of the modules in gene regulatory networks [5] lending additional support to the idea that modularity confers evolvability by permitting a dynamic rewiring of network components in response to environmental perturbation. That is, modularity increases evolvability.

Meir *et al*. studied a computer model of the neurogenic network of *Drosophila melanogaster* and found it to be very robust to a change in parameters or initial conditions [85]. They also showed that, within their model, this robustness provides both functional and evolutionary flexibility. It allows a network performing one function to evolve the ability to achieve additional functions. In this case, robustness confers evolvability.

In 2006, Davidson and Erwin proposed that the hierarchical modular structure of gene regulatory networks leads to different rates of evolution between major aspects of body plan morphology and terminal properties of body plans [86]. They found a hierarchy with four types of modules in gene regulatory networks, "kernels," "plug-ins," "switches," and "batteries," each of which has a different function during development. Kernels shape the phylum- and superphylum-level characteristics, plug-ins and switches are associated with class, order, and family characteristics, and batteries are involved in speciation. The idea that the genetic framework upon which selection acts is not unstructured, as assumed in classic evolutionary theory, sparked a controversy [87, 88]. But a recent publication provides further evidence that the hierarchical structure of developmental regulatory networks provides an organizing structure for the evolution of the body plan [16]. An explicit calculation of the rate of evolution of genes in different types of modules in gene regulatory networks demonstrated the influence of hierarchical structure on evolvability.

3.5. Modularity in Physiology

Traditionally it has been believed that the healthy physiologic state is characterized by homeostasis and that maintenance of all physiologic variables in narrow ranges around optimal values is the key to good health. Pathology on the other hand was thought to result from a deviation of one or more physiologic variables from their healthy values. As a consequence, much of Western medicine is focused on restoring physiologic variables to their normal values [89].

More recently this view has been questioned, and it has been suggested that variation in physiologic variables may not be a detriment to, but rather a necessary component of, health [90]. The fluctuations observed in physiologic systems around mean values are just as important as the mean values themselves and aging and disease are characterized by a loss of variability [89]. The hypothesis that variability is associated with health is supported experimentally by Boker *et al.* who found that introducing noise into mechanical ventilators leads to an improvement in lung function [91].

However, not all noise is good noise. For example, atrial fibrillation, which leads to very irregular heartbeat intervals, is certainly not associated with health. Thus, the idea arose that "complex" physiologic time signals are indicative of healthy systems. Different measures of complexity have been suggested, most of which are based on entropy (*e.g.* [92]). Here we propose that complexity in physiology can be understood as modularity, and that modularity deteriorates in aging and disease. Healthy physiology is characterized by a modular partitioning phase space that facilitates transitions between states. This modularity promotes adaptability to external stimulus. A healthy state is described neither by a complete disconnect between modules nor by very strongly connected modules. Rather, there is an optimal amount of connectedness between modules.

3.5.1. Heart Rate—An early study showed that the decoupling of physiologic systems is associated with a decrease in variability and with disease [93]. Goldstein *et al.* studied the consequences of acute brain injury on heart rate variability. They found that neurological injury leads to a decoupling of the autonomic and cardiovascular systems and that this decoupling leads to a decrease in heart rate and blood pressure variability. It was also observed that a recoupling of cardiovascular signals is necessary for recovery.

A recent study provides empirical evidence that environmental stress can increase the modularity of a physiologic system, which in turn improves the system's ability to respond to the external stress. Anton Burykin and Timothy Buchman investigated the response of the human heart beat to exercise stress tests [94]. During such a test, subjects are exposed to increasing levels of speed and inclination on a treadmill. After some time for acclimatization in each level, the exercise load is increased. When the subject reaches maximal exertion, exercise load is decreased. Burykin and Buchman analyzed the obtained heart beat time series data by constructing an interbeat covariance matrix and computing the modularity of this matrix. They observed, as shown in figure 21, that the modularity increases abruptly in response to the external stress. It is interesting to note that the modularity already begins to increase 400 heart beats before a change in the heart beat frequency is observed [94].

Similarly, an experiment conducted by Carlsson *et al*. supports the hypothesis that disease is associated with a decrease in modularity [95]. The researchers compared heart interbeat interval time series data from three groups: healthy subjects, patients suffering from atrial fibrillation (AF), and patients suffering from congestive heart failure (CHF). They extracted motifs that occurred frequently in each patient's data set and performed a topological analysis of the space of these motifs. Their results, shown in figure 22, show the presence of two clear modules in the space of frequent motifs of healthy subjects that can be observed for all time scales. The motif space of AF patients does not separate into modules, while the motif space of CHF patients develops a separation only for longer time scales [95].

3.5.2. Postural Control—The benefits of explicitly introducing variability to restore the functionality of a physiologic system has been observed. In the context of postural control, Collins *et al*. showed that noise can enhance the detection of a subthreshold tactile stimulus [96], a phenomenon referred to as stochastic resonance. To test the concept of stochastic resonance in postural control, Priplata and coworkers applied subsensory noise to the feet of young and elderly subjects during quiet standing [97, 98]. This noise resulted in a reduction of postural sway in both groups, with a larger improvement in the elderly. In a follow-up study, they extended the group of subjects to include patients with diabetes and patients who had had a stroke. Subsensory noise led to improved balance in all groups; the improvement was greater for subjects with greater baseline sway (worse balance) [99]. Costa *et al*. quantified complexity by a measure called multiscale entropy and observed that the complexity of postural sway dynamics in elderly subjects with a history of falls is lower compared to that of both young subjects as well as elderly subjects without a history of falls [92]. Applying subsensory noise to the feet increased the complexity of sway fluctuations in the elderly. These results show that variability is an essential part of healthy physiology.

3.5.3. Brain Networks—A relation between modularity on the one hand and aging and pathology on the other, which is similar to the one in cardiovascular signals, has also been observed in brain networks. Meunier *et al*. studied the modular structure of human brain functional networks in young and older adults and found that both showed significant modularity and that the network structure of the human brain changes with age [100]. They also displayed the modularity of the two groups as a function of the number of edges, *i.e.* applied different threshold values, and observed that the modularity of the young group is consistently higher than the modularity of the older group. This difference was not statistically significant for any of the cutoff values used, but it is apparent in the graph that the difference becomes more significant as the number of edges decreases. We extended the analysis to networks with 130 edges, slightly below the lowest value of 150 used by Meunier *et al*., and observed a statistically significantly higher modularity in the younger group than in the older group. Our results are shown in figure 23.

To further support the hypothesis that brain networks of young adults are more modular than those of older adults, we used a completely different method to quantify modularity on the same networks. We computed a matrix of Euclidean commute time distances, as described in section 3.3 for each brain network and used average linkage hierarchical clustering to create a dendrogram of possible partitions for each network. Then we quantified modularity

as the ratio of intramodule weight over intramodule area in the adjacency matrix of the network [38]:

$$M = \frac{\sum\limits_{j,k \neq j} A_{jk} \delta(c_j, c_k)}{\sum\limits_{j,k \neq j} A_{jk}} \times \left(\frac{\sum\limits_{j,k \neq j} \delta(c_j, c_k)}{n(n-1)} \right)^{-1} , \quad (12)$$

where A_{jk} is the adjacency matrix of the network and $\delta(c_j, c_k) = 1$ if nodes j and k are in the same module and $\delta(c_j, c_k) = 0$ otherwise. The results, shown in figure 24, show that the modularity of brain networks from young adults is consistently higher than that of brain networks from older adults across the relevant part of the dendrogram of network divisions.

Empirical evidence also shows that the brain networks of diseased patients are less modular than those of healthy subjects. Chavez *et al.* analyzed the structure of brain networks from magnetoencephalographic signals in epileptic patients and compared it to that of healthy controls [101]. They observed that the patients suffering from epilepsy had brain networks with greater connectivity and lower modularity than the brain networks of the controls. They also found that in epilepsy patients, nodes have more connections to nodes in different functional modules.

Studies on very different physiological systems reveal a common trend: aging and disease lead to a decrease in the modularity of physiological systems, which reduces the ability of these systems to respond to external stress.

3.6. Social Networks

The emergence of hierarchical structure in response to environmental pressure has also been observed in social networks. Social networks are evolving systems. Thus, the insights gained from the study of general evolving systems may be applied to understand their behavior. In a very recent publication [102], the temporal evolution of hierarchical structure in the world trade network was analyzed over the last 40 years. It was shown that during recessions, which can be considered a form of environmental pressure, the world trade network tends to become more hierarchical, with a larger observed increase in hierarchy during more severe recessions. In addition it was found that globalization transforms the trade network into a less hierarchical state. This decreased hierarchy makes the trade network more sensitive to environmental shocks and leads to a slower recovery after recessions. These observations are again consistent with the theory presented in section 2.2.4. Increased environmental pressure leads to hierarchical structures which, in turn, improve a system's ability to respond to environmental perturbations.

4. Conclusion

We have presented the hypothesis that a changing environment selects for adaptable frameworks, and competition among different evolutionary frameworks leads to selection of structures with the most efficient dynamics, which are the modular ones. From a computer science point of view, by forming a hierarchy, the NP-complete problem of searching the

entire sequence space is replaced by a polynomial-time approximation. Many low-lying states are lost, but those that remain are found more quickly. From a physics point of view, the Hamiltonian is being made somewhat separable. Shorter modules are exponentially more easily evolved. Natural (and man-made) systems were shown in several examples to employ modularity to a non-zero extent.

We have defined a module to be a component that can operate relatively independently of the rest of the system. Modularity was said to have emerged when there are more intramodule connections than intermodule connections. We reviewed the hypothesis that modularity arises because there is a generic requirement for a system in a changing environment to be evolvable. This theory of spontaneous emergence of modularity states that systems become modular under three conditions: changing environments, information exchange, and slow evolution. These conditions appear to be met in much of biological evolution.

Mathematically, modularity measures the compartmentalization of biological organization. In the form of a linear expansion, the theory of spontaneous emergence of modularity would be stated as the rate of change of modularity is proportional to the environmental change:

$$p_E - p_0 = \frac{M'}{R} \quad (13)$$

where p_E is the environmental pressure, R is the resistance to evolution or ruggedness of the fitness landscape, M' is the rate of change of modularity, and p_0 is the initial value of environmental pressure for which the system had been in steady state. The theory reviewed here explains how environmental pressure, horizontal gene transfer rate, and ruggedness of the fitness landscape promote the emergence of modularity. This theory was shown to explain features on scales ranging from proteins to physiology to social networks.

Additional theoretical challenges lie in explaining how multiple levels of hierarchy form. One idea is that as the benefit from the development of modularity at one level saturates, an additional, higher-order level is nucleated, figure 25. Numerical support for this idea was shown in figure 15. Mathematical description of this hierarchical partitioning of biological space would seem to be an interesting, open research topic.

Acknowledgments

References

1. Simon HA. The architecture of complexity. P Am Philos Soc. 1962; 106:467–482. http://www.jstor.org/stable/985254.

2. Lipson H, Pollack JB, Suh NP. On the origin of modular variation. Evolution. 2002; 56:1549–1556. http://www3.interscience.wiley.com/journal/118941191/abstract. 10.1111/j.0014-3820.2002.tb01466.x [PubMed: 12353747]

3. Espinosa-Soto C, Wagner A. Specialization can drive the evolution of modularity. PLoS Comput Biol. 2010; 6:e1000719. http://dx.plos.org/10.1371/journal.pcbi.1000719. 10.1371/journal.pcbi. 1000719 [PubMed: 20360969]

4. Kashtan N, Mayo AE, Kalisky T, Alon U. An analytically solvable model for rapid evolution of modular structure. PLoS Comput Biol. 2009; 5:e1000355. http://dx.plos.org/10.1371/journal.pcbi. 1000355. 10.1371/journal.pcbi.1000355 [PubMed: 19360090]

5. Litvin O, Causton HC, Chen BJ, Pe'er D. Modularity and interactions in the genetics of gene expression. P Natl Acad Sci USA. 2009; 106:6441–6446. http://www.pnas.org/cgi/content/abstract/ 106/16/6441. 10.1073/pnas.0810208106

6. Hartwell LH, Hopfield JJ, Leibler S, Murray AW. From molecular to modular cell biology. Nature. 1999; 402:C47–52. http://www.ncbi.nlm.nih.gov/pubmed/10591225. 10.1038/35011540 [PubMed: 10591225]

7. Wagner GP. Homologues, natural kinds and the evolution of modularity. Integr Comp Biol. 1996; 36:36–43. http://icb.oxfordjournals.org/cgi/content/abstract/36/1/36. 10.1093/icb/36.1.36

8. Schlosser, G.; Wagner, GP., editors. Modularity in development and evolution. University of Chicago Press; Chicago: 2004.

9. Krause AE, Frank KA, Mason DM, Ulanowicz RE, Taylor WW. Compartments revealed in food-web structure. Nature. 2003; 426:282–285. http://www.ncbi.nlm.nih.gov/pubmed/14628050. 10.1038/nature02115 [PubMed: 14628050]

10. Montoya JM, Pimm SL, Solé RV. Ecological networks and their fragility. Nature. 2006; 442:259–264. http://dx.doi.org/10.1038/nature04927. 10.1038/nature04927 [PubMed: 16855581]

11. Baltimore D. Our genome unveiled. Nature. 2001; 409:814–816. http://dx.doi.org/ 10.1038/35057267. 10.1038/35057267 [PubMed: 11236992]

12. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of escherichia coli. Nat Genet. 2002; 31:64–68. http://www.ncbi.nlm.nih.gov/pubmed/11967538. 10.1038/ng881 [PubMed: 11967538]

13. Singh AH, Wolf DM, Wang P, Arkin AP. Modularity of stress response evolution. P Natl Acad Sci USA. 2008; 105:7500–7505. http://www.pnas.org/cgi/content/abstract/105/21/7500. 10.1073/pnas. 0709764105

14. Spirin V, Gelfand MS, Mironov AA, Mirny LA. A metabolic network in the evolutionary context: Multiscale structure and modularity. P Natl Acad Sci USA. 2006; 103:8774–8779. http:// www.pnas.org/cgi/content/abstract/103/23/8774. 10.1073/pnas.0510258103

15. Klingenberg CP. Morphological integration and developmental modularity. Annu Rev Ecol Evol S. 2008; 39:115–132. http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=35967311. 10.1146/annurev.ecolsys.37.091305.110054

16. He J, Deem MW. Hierarchical evolution of animal body plans. Dev Biol. 2010; 337:157–161. http://www.ncbi.nlm.nih.gov/pubmed/19799894. 10.1016/j.ydbio.2009.09.038 [PubMed: 19799894]

17. Callebaut, W.; Rasskin-Gutman, D., editors. Modularity: Understanding the development and evolution of natural complex systems. MIT Press; Cambridge, Massachusetts: 2005.

18. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. P Natl Acad Sci USA. 1998; 95:14863–14868. http://www.pnas.org/cgi/ content/abstract/95/25/14863. 10.1073/pnas.95.25.14863

19. Karlin D, Ferron F, Canard B, Longhi S. Structural disorder and modular organization in paramyxovirinae N and P. J Gen Virol. 2003; 84:3239–3252. http://www.ncbi.nlm.nih.gov/ pubmed/14645906. [PubMed: 14645906]

20. Solé RV, Valverde S. Spontaneous emergence of modularity in cellular networks. J Roy Soc Interface. 2008; 5:129–133. http://rsif.royalsocietypublishing.org/cgi/content/abstract/5/18/129. 10.1098/rsif.2007.1108 [PubMed: 17626003]

21. Bogarad LD, Deem MW. A hierarchical approach to protein molecular evolution. P Natl Acad Sci USA. 1999; 96:2591–2595. http://www.pubmedcentral.nih.gov/articlerender.fcgi? artid=15812&tool=pmcentrez&rend.

22. Rainey PB, Cooper TF. Evolution of bacterial diversity and the origins of modularity. Res Microbiol. 2004; 155:370–375. http://dx.doi.org/10.1016/j.resmic.2004.01.011. 10.1016/j.resmic.2004.01.011 [PubMed: 15207869]

23. McAdams HH, Srinivasan B, Arkin AP. The evolution of genetic regulatory systems in bacteria. Nat Rev Genet. 2004; 5:169–178. http://dx.doi.org/10.1038/nrg1292. 10.1038/nrg1292 [PubMed: 14970819]

24. Kitano H. Biological robustness. Nat Rev Genet. 2004; 5:826–837. http://dx.doi.org/10.1038/nrg1471. 10.1038/nrg1471 [PubMed: 15520792]

25. Hallinan JS. Gene duplication and hierarchical modularity in intracellular interaction networks. Biosystems. 2004; 74:51–62. http://dx.doi.org/10.1016/j.biosystems.2004.02.004. 10.1016/j.biosystems.2004.02.004 [PubMed: 15125992]

26. Soyer OS. Emergence and maintenance of functional modules in signaling pathways. BMC Evol Biol. 2007; 7:205. http://www.biomedcentral.com/1471-2148/7/205. 10.1186/1471-2148-7-205 [PubMed: 17974002]

27. Price MN, Dehal PS, Arkin AP. Horizontal gene transfer and the evolution of transcriptional regulation in escherichia coli. Genome Biol. 2008; 9:R4. http://genomebiology.com/2008/9/1/R4. 10.1186/gb-2008-9-1-r4 [PubMed: 18179685]

28. Force A, Cresko WA, Pickett FB, Proulx SR, Amemiya C, Lynch M. The origin of subfunctions and modular gene regulation. Genetics. 2005; 170:433–446. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1449736&tool=pmcentrez&re. 10.1534/genetics.104.027607 [PubMed: 15781713]

29. Kashtan N, Alon U. Spontaneous evolution of modularity and network motifs. P Natl Acad Sci USA. 2005; 102:13773–13778. http://www.pnas.org/cgi/content/abstract/102/39/13773. 10.1073/pnas.0503610102

30. Orr HA. Adaptation and the cost of complexity. Evolution. 2000; 54:13–20. http://www.ncbi.nlm.nih.gov/pubmed/10937178. [PubMed: 10937178]

31. Welch JJ, Waxman D. Modularity and the cost of complexity. Evolution. 2003; 57:1723. http://dx.doi.org/10.1554/02-673. 10.1554/02-673 [PubMed: 14503615]

32. Variano E, McCoy J, Lipson H. Networks, dynamics, and modularity. Phys Rev Lett. 2004; 92:188701. http://prl.aps.org/abstract/PRL/v92/i18/e188701. 10.1103/PhysRevLett.92.188701 [PubMed: 15169539]

33. Kashtan N, Noor E, Alon U. Varying environments can speed up evolution. P Natl Acad Sci USA. 2007; 104:13711–13716. http://www.jstor.org/stable/25436556.

34. Kashtan N, Parter M, Dekel E, Mayo AE, Alon U. Extinctions in heterogeneous environments and the evolution of modularity. Evolution. 2009; 63:1964–1975. http://www3.interscience.wiley.com/journal/122249924/abstract. 10.1111/j.1558-5646.2009.00684.x [PubMed: 19473401]

35. Earl DJ, Deem MW. Evolvability is a selectable trait. P Natl Acad Sci USA. 2004; 101:11531–11536. http://www.pnas.org/cgi/content/abstract/101/32/11531. 10.1073/pnas.0404656101

36. Gardner A, Zuidema W. Is evolvability involved in the origin of modular variation? Evolution. 2003; 57:1448. http://dx.doi.org/10.1554/03-056. 10.1554/03-056 [PubMed: 12894952]

37. Sun J, Deem M. Spontaneous emergence of modularity in a model of evolving individuals. Phys Rev Lett. 2007; 99:228107. http://link.aps.org/doi/10.1103/PhysRevLett.99.228107. 10.1103/PhysRevLett.99.228107 [PubMed: 18233336]

38. He J, Sun J, Deem MW. Spontaneous emergence of modularity in a model of evolving individuals and in real networks. Phys Rev E. 2009; 79:031907. http://link.aps.org/doi/10.1103/PhysRevE.79.031907. 10.1103/PhysRevE.79.031907

39. Crombach A, Hogeweg P. Evolution of evolvability in gene regulatory networks. PLoS Comput Biol. 2008; 4:e1000112. http://www.ncbi.nlm.nih.gov/pubmed/18617989. 10.1371/journal.pcbi.1000112 [PubMed: 18617989]

40. Martin OC, Wagner A. Effects of recombination on complex regulatory circuits. Genetics. 2009; 183:673–684. http://www.genetics.org/cgi/content/abstract/genetics.109.104174v1. 10.1534/genetics.109.104174 [PubMed: 19652184]

41. Misevic D, Ofria C, Lenski RE. Sexual reproduction reshapes the genetic architecture of digital organisms. P Roy Soc B. 2006; 273:457–464. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1560214&tool=pmcentrez&re. 10.1098/rspb.2005.3338

42. Misevic D, Ofria C, Lenski RE. Experiments with digital organisms on the origin and maintenance of sex in changing environments. J Hered. 2010; 101(Suppl):S46–S54. http://jhered.oxfordjournals.org/cgi/content/abstract/101/suppl_1/S46. 10.1093/jhered/esq017 [PubMed: 20200140]

43. Callahan B, Thattai M, Shraiman BI. Emergent gene order in a model of modular polyketide synthases. P Natl Acad Sci USA. 2009; 106:19410–19415. http://www.pnas.org/cgi/content/abstract/106/46/19410. 10.1073/pnas.0902364106

44. Ferron F, Rancurel C, Longhi S, Cambillau C, Henrissat B, Canard B. VaZyMolO: A tool to define and classify modularity in viral proteins. J Gen Virol. 2005; 86:743–749. http://vir.sgmjournals.org/cgi/content/abstract/86/3/743. 10.1099/vir.0.80590-0 [PubMed: 15722535]

45. Gupta V, Earl DJ, Deem MW. Quantifying influenza vaccine efficacy and antigenic distance. Vaccine. 2006; 24:3881–3888. http://dx.doi.org/10.1016/j.vaccine.2006.01.010. 10.1016/j.vaccine.2006.01.010 [PubMed: 16460844]

46. Radman M, Matic I, Taddei F. Evolution of evolvability. Ann NY Acad Sci. 1999; 870:146–155. http://www.blackwell-synergy.com/doi/abs/10.1111/j.1749-6632.1999.tb08874.x. 10.1111/j.1749-6632.1999.tb08874.x [PubMed: 10415480]

47. Massey RC, Buckling A. Environmental regulation of mutation rates at specific sites. Trends Microbiol. 2002; 10:580–584. http://www.ncbi.nlm.nih.gov/pubmed/12564995. [PubMed: 12564995]

48. Moxon R, Bayliss C, Hood D. Bacterial contingency loci: The role of simple sequence DNA repeats in bacterial adaptation. Annu Rev Genet. 2006; 40:307–333. http://www.ncbi.nlm.nih.gov/pubmed/17094739. 10.1146/annurev.genet.40.110405.090442 [PubMed: 17094739]

49. Kepler TB, Perelson AS. Drug concentration heterogeneity facilitates the evolution of drug resistance. P Natl Acad Sci USA. 1998; 95:11514–11519. http://www.pnas.org/cgi/content/abstract/95/20/11514. 10.1073/pnas.95.20.11514

50. Shoemaker BA, Panchenko AR. Deciphering protein-protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol. 2007; 3:e42. http://dx.plos.org/10.1371/journal.pcbi.0030042. 10.1371/journal.pcbi.0030042 [PubMed: 17397251]

51. Braun P, Tasan M, Dreze M, Barrios-Rodiles M, Lemmens I, Yu H, Sahalie JM, Murray RR, Roncari L, de Smet AS, Venkatesan K, Rual JF, Vandenhaute J, Cusick ME, Pawson T, Hill DE, Tavernier J, Wrana JL, Roth FP, Vidal M. An experimentally derived confidence score for binary protein-protein interactions. Nat Methods. 2009; 6:91–7. http://dx.doi.org/10.1038/nmeth.1281. 10.1038/nmeth.1281 [PubMed: 19060903]

52. Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabási AL. Hierarchical organization of modularity in metabolic networks. Science. 2002; 297:1551–1555. http://www.sciencemag.org/cgi/content/abstract/297/5586/1551. 10.1126/science.1073374 [PubMed: 12202830]

53. da Silva MR, Ma H, Zeng AP. Centrality, network capacity, and modularity as parameters to analyze the core-periphery structure in metabolic networks. Pr Inst Electr Elect. 2008; 96:1411–1420. http://ieeexplore.ieee.org/xpl/freeabsall.jsp?arnumber=4567408. 10.1109/JPROC.2008.925418

54. Parter M, Kashtan N, Alon U. Environmental variability and modularity of bacterial metabolic networks. BMC Evol Biol. 2007; 7:169. http://www.biomedcentral.com/1471-2148/7/169. 10.1186/1471-2148-7-169 [PubMed: 17888177]

55. Kreimer A, Borenstein E, Gophna U, Ruppin E. The evolution of modularity in bacterial metabolic networks. P Natl Acad Sci USA. 2008; 105:6976–6981. http://www.ncbi.nlm.nih.gov/pubmed/18460604. 10.1073/pnas.0712149105

56. Nakamura Y, Itoh T, Matsuda H, Gojobori T. Biased biological functions of horizontally transferred genes in prokaryotic genomes. Nat Genet. 2004; 36:760–766. http://dx.doi.org/10.1038/ng1381. 10.1038/ng1381 [PubMed: 15208628]

57. Good BH, de Montjoye YA, Clauset A. Performance of modularity maximization in practical contexts. Phys Rev E. 2010; 81:046106. http://www.ncbi.nlm.nih.gov/pubmed/20481785.

58. Patthy L. Modular assembly of genes and the evolution of new functions. Genetica. 2003; 118:217–231. http://www.springerlink.com/content/n3300425302h06t6. 10.1023/A:1024182432483 [PubMed: 12868611]

59. Segrè D, Deluna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. Nat Genet. 2005; 37:77–83. http://dx.doi.org/10.1038/ng1489. 10.1038/ng1489 [PubMed: 15592468]

60. Bhattacharyya RP, Reményi A, Yeh BJ, Lim WA. Domains, motifs, and scaffolds: The role of modular interactions in the evolution and wiring of cell signaling circuits. Annu Rev Biochem. 2006; 75:655–680. http://www.ncbi.nlm.nih.gov/pubmed/16756506. 10.1146/annurev.biochem.75.103004.142710 [PubMed: 16756506]

61. Wang X, Dalkic E, Wu M, Chan C. Gene module level analysis: Identification to networks and dynamics. Curr Opin Biotech. 2008; 19:482–491. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2615490&tool=pmcentrez&re. 10.1016/j.copbio.2008.07.011 [PubMed: 18725293]

62. Lozada-Chávez I, Janga SC, Collado-Vides J. Bacterial regulatory networks are extremely flexible in evolution. Nucleic Acids Res. 2006; 34:3434–45. http://nar.oxfordjournals.org/cgi/content/abstract/34/12/3434. 10.1093/nar/gkl423 [PubMed: 16840530]

63. Maslov S, Sneppen K, Eriksen KA, Yan KK. Upstream plasticity and downstream robustness in evolution of molecular networks. BMC Evol Biol. 2004; 4:9. http://www.biomedcentral.com/1471-2148/4/9. 10.1186/1471-2148-4-9 [PubMed: 15070432]

64. Spirin V, Mirny LA. Protein complexes and functional modules in molecular networks. P Natl Acad Sci USA. 2003; 100:12123–12128. http://www.pnas.org/cgi/content/abstract/100/21/12123. 10.1073/pnas.2032324100

65. von Mering C, Zdobnov EM, Tsoka S, Ciccarelli FD, Pereira-Leal JB, Ouzounis CA, Bork P. Genome evolution reveals biochemical networks and functional modules. P Natl Acad Sci USA. 2003; 100:15428–15433. http://www.pnas.org/cgi/content/abstract/100/26/15428. 10.1073/pnas.2136809100

66. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, Jensen LJ, Bastuck S, Dümpelfeld B, Edelmann A, Heurtier MA, Hoffman V, Hoefert C, Klein K, Hudak M, Michon AM, Schelder M, Schirle M, Remor M, Rudi T, Hooper S, Bauer A, Bouwmeester T, Casari G, Drewes G, Neubauer G, Rick JM, Kuster B, Bork P, Russell RB, Superti-Furga G. Proteome survey reveals modularity of the yeast cell machinery. Nature. 2006; 440:631–636. http://www.ncbi.nlm.nih.gov/pubmed/16429126. 10.1038/nature04532 [PubMed: 16429126]

67. Han JDJ, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJM, Cusick ME, Roth FP, Vidal M. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. Nature. 2004; 430:88–93. http://dx.doi.org/10.1038/nature02555. 10.1038/nature02555 [PubMed: 15190252]

68. Cohen-Gihon I, Lancet D, Yanai I. Modular genes with metazoan-specific domains have increased tissue specificity. Trends Genet. 2005; 21:210–213. http://www.ncbi.nlm.nih.gov/pubmed/15797615. 10.1016/j.tig.2005.02.008 [PubMed: 15797615]

69. Campillos M, von Mering C, Jensen LJ, Bork P. Identification and analysis of evolutionarily cohesive functional modules in protein networks. Genome Res. 2006; 16:374–382. http://genome.cshlp.org/cgi/content/abstract/16/3/374. 10.1101/gr.4336406 [PubMed: 16449501]

70. Thebault E, Fontaine C. Stability of ecological communities and the architecture of mutualistic and trophic networks. Science. 2010; 329:853–856. http://www.sciencemag.org/cgi/content/abstract/329/5993/853. 10.1126/science.1188321 [PubMed: 20705861]

71. Christensen K, di Collobiano SA, Hall M, Jensen HJ. Tangled nature: A model of evolutionary ecology. J Theor Biol. 2002; 216:73–84. http://dx.doi.org/10.1006/jtbi.2002.2530. 10.1006/jtbi.2002.2530 [PubMed: 12076129]

72. Rikvold PA. Self-optimization, community stability, and fluctuations in two individual-based models of biological coevolution. J Math Biol. 2007; 55:653–77. http://www.springerlink.com/content/c50567n71681442u/. 10.1007/s00285-007-0101-y [PubMed: 17534620]

73. Rikvold PA, Sevim V. Individual-based predator-prey model for biological coevolution: Fluctuations, stability, and community structure. Phys Rev E. 2007; 75:051920. http://link.aps.org/doi/10.1103/PhysRevE.75.051920. 10.1103/PhysRevE.75.051920

74. Thompson RM, Townsend CR. Impacts on stream food webs of native and exotic forest: An intercontinental comparison. Ecology. 2003; 84:145–161. http://www.esajournals.org/doi/abs/10.1890/0012-9658%282003%29084%5B0145%3AIOSFWO%5. 10.1890/0012-9658(2003)084[0145:IOSFWO]2.0.CO;2

75. Thompson RM, Townsend CR. Energy availability, spatial heterogeneity and ecosystem size predict food-web structure in streams. Oikos. 2005; 108:137–148. http://doi.wiley.com/10.1111/j.0030-1299.2005.11600.x. 10.1111/j.0030-1299.2005.11600.x

76. Sabo, JL.; Soykan, CU.; Keller, A. Functional roles of leaf litter detritus in terrestrial food webs. In: de Ruiter, P.; Wolters, V.; Moore, JC., editors. Dynamic food webs: Multispecies assemblages, ecosystem development and environmental change. Elsevier; Amsterdam: 2005. p. 211-222.

77. Vannote RL, Minshall GW, Cummins KW, Sedell JR, Cushing CE. The river continuum concept. Can J Fish Aquad Sci. 1980; 37:130–137. http://www.citeulike.org/user/iwagner/article/3662110.

78. Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007; 17:395–416. http://www.springerlink.com/content/jq1g17785n783661. 10.1007/s11222-007-9033-z

79. Saerens, M.; Fouss, F.; Yen, L.; Dupont, P. The principal components analysis of a graph, and its relationships to spectral clustering. Proceedings of the 15th European conference on machine learning (ECML); Berlin: Springer Verlag; 2004. p. 371-383.

80. Barnett, S. Matrices: Methods and applications. Oxford University Press; USA: 1990.

81. Newman MEJ. Modularity and community structure in networks. P Natl Acad Sci USA. 2006; 103:8577–8582. http://www.pnas.org/cgi/content/abstract/103/23/8577. 10.1073/pnas.0601602103

82. Raff EC, Raff RA. Dissociability, modularity, evolvability. Evol Dev. 2000; 2:235–237. http://www3.interscience.wiley.com/journal/119052048/abstract. 10.1046/j.1525-142x.2000.00069.x [PubMed: 11252551]

83. Raff RA, Sly BJ. Modularity and dissociation in the evolution of gene expression territories in development. Evol Dev. 2000; 2:102–113. http://www.ncbi.nlm.nih.gov/pubmed/11258388. [PubMed: 11258388]

84. Yang AS. Modularity, evolvability, and adaptive radiations: A comparison of the hemi- and holometabolous insects. Evol Dev. 2001; 3:59–72. http://www.ncbi.nlm.nih.gov/pubmed/11341675. [PubMed: 11341675]

85. Meir E, von Dassow G, Munro E, Odell GM. Robustness, flexibility, and the role of lateral inhibition in the neurogenic network. Curr Biol. 2002; 12:778–786. http://www.ncbi.nlm.nih.gov/pubmed/12015114. [PubMed: 12015114]

86. Davidson EH, Erwin DH. Gene regulatory networks and the evolution of animal body plans. Science. 2006; 311:796–800. http://www.sciencemag.org/cgi/content/abstract/311/5762/796. 10.1126/science.1113832 [PubMed: 16469913]

87. Coyne JA. Comment on "gene regulatory networks and the evolution of animal body plans". Science. 2006; 313:761. author reply 761http://www.sciencemag.org/cgi/content/abstract/313/5788/761b. 10.1126/science.1126454 [PubMed: 16902111]

88. Erwin DH. Response to comment on "gene regulatory networks and the evolution of animal body plans". Science. 2006; 313:761c–761c. http://www.sciencemag.org/cgi/content/abstract/313/5788/761c. 10.1126/science.1126765 [PubMed: 16902111]

89. Buchman TG. The community of the self. Nature. 2002; 420:246–251. http://dx.doi.org/10.1038/nature01260. 10.1038/nature01260 [PubMed: 12432410]

90. West, BJ. Where medicine went wrong: Rediscovering the path to complexity. World Scientific; 2006.

91. Boker A, Haberman CJ, Girling L, Guzman RP, Louridas G, Tanner JR, Cheang M, Maycher BW, Bell DD, Doak GJ. Variable ventilation improves perioperative lung function in patients undergoing abdominal aortic aneurysmectomy. Anesthesiology. 2004; 100:608–616. http://www.ncbi.nlm.nih.gov/pubmed/15108976. [PubMed: 15108976]

92. Costa MD, Priplata AA, Lipsitz LA, Wu Z, Huang NE, Goldberger AL, Peng CK. Noise and poise: Enhancement of postural complexity in the elderly with a stochastic-resonance–based therapy.

Europhys Lett. 2007; 77:68008. http://stacks.iop.org/0295-5075/77/i=6/a=68008. [PubMed: 17710211]

93. Goldstein B, Toweill D, Lai S, Sonnenthal K, Kimberly B. Uncoupling of the autonomic and cardiovascular systems in acute brain injury. Am J Physiol Reg I. 1998; 275:R1287–R1292. http://ajpregu.physiology.org/cgi/content/abstract/275/4/R1287.

94. Buchman, TG.; Burykin, A. Personal communication. 2010.

95. Carlsson, G.; Danciger, J.; Morton, J. Mapping geometry in heart rate data. personal communication. 2010.

96. Collins JJ, Imhoff TT, Grigg P. Noise-enhanced tactile sensation. Nature. 1996; 383:770. http://dx.doi.org/10.1038/383770a0. 10.1038/383770a0 [PubMed: 8893000]

97. Priplata AA, Niemi JB, Salen M, Harry JD, Lipsitz LA, Collins JJ. Noise-enhanced human balance control. Phys Rev Lett. 2002; 89:238101. http://link.aps.org/doi/10.1103/PhysRevLett.89.238101. 10.1103/PhysRevLett.89.238101 [PubMed: 12485044]

98. Priplata AA, Niemi JB, Harry JD, Lipsitz LA, Collins JJ. Vibrating insoles and balance control in elderly people. Lancet. 2003; 362:1123–1124. http://www.ncbi.nlm.nih.gov/pubmed/14550702. 10.1016/S0140-6736(03)14470-4 [PubMed: 14550702]

99. Priplata AA, Patritti BL, Niemi JB, Hughes R, Gravelle DC, Lipsitz LA, Veves A, Stein J, Bonato P, Collins JJ. Noise-enhanced balance control in patients with diabetes and patients with stroke. Ann Neurol. 2006; 59:4–12. http://www.ncbi.nlm.nih.gov/pubmed/16287079. 10.1002/ana.20670 [PubMed: 16287079]

100. Meunier D, Achard S, Morcom A, Bullmore E. Age-related changes in modular organization of human brain functional networks. NeuroImage. 2009; 44:715–723. http://www.ncbi.nlm.nih.gov/pubmed/19027073. 10.1016/j.neuroimage.2008.09.062 [PubMed: 19027073]

101. Chavez M, Valencia M, Navarro V, Latora V, Martinerie J. Functional modularity of background activities in normal and epileptic brain networks. Phys Rev Lett. 2010; 104:118701. http://prl.aps.org/abstract/PRL/v104/i11/e118701. 10.1103/PhysRevLett.104.118701 [PubMed: 20366507]

102. He J, Deem MW. Structure and response in the world trade network. Phys Rev Lett. 2010; 105:198701. http://prl.aps.org/abstract/PRL/v105/i19/e198701. 10.1103/PhysRevLett. 105.198701 [PubMed: 21231202]

103. Fortunato S. Community detection in graphs. Phys Rep. 2010; 486:75–174. http://dx.doi.org/10.1016/j.physrep.2009.11.002. 10.1016/j.physrep.2009.11.002

104. Barber M. Modularity and community detection in bipartite networks. Phys Rev E. 76 http://pre.aps.org/abstract/PRE/v76/i6/e066102. 10.1103/PhysRevE.76.066102

105. Leicht EA, Newman MEJ. Community structure in directed networks. Phys Rev Lett. 100 http://prl.aps.org/abstract/PRL/v100/i11/e118703. 10.1103/PhysRevLett.100.118703

Appendix A. Quantitative Definitions of Modularity and Hierarchy

In this appendix we will summarize the quantitative definitions of modularity and hierarchy used throughout this review paper. A recent and thorough exposition of definitions of modularity and algorithms for the detection of modules can be found in [103].

Appendix A.1. Newman's Modularity

Newman's modularity is one of the most widely used quantitative measures of modularity. Consider an undirected graph with adjacency matrix A_{ij} and a partition of this graph into clusters or modules defined by $\{c_i\}$, where c_i describes which module node i belongs to. Then, Newman's modularity is defined as [81]

$$Q = \frac{1}{2m} \sum_{i,j} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j), \quad \text{(A.1)}$$

where $k_i = \Sigma_j A_{ij}$ is the degree of each node, $m = \frac{1}{2}\sum_i k_i = \frac{1}{2}\sum_{ij} A_{ij}$ is the total number of edges and $\delta(c_i, c_j)$ is defined as

$$\delta(c_i, c_j) = \begin{cases} 1 & \text{if nodes } i \text{ and } j \text{ belong to the same module} \\ 0 & \text{otherwise.} \end{cases} \quad \text{(A.2)}$$

Conceptually, Newman's modularity compares the fraction of within-module edges in the graph to the expected fraction of within-module edges in a random graph with the same degree sequence as given by the configuration model. The value of Q is normalized such that it always lies in the interval $(-1, 1)$. The definition of Q has also been adapted to bipartite networks [104] and directed networks [105] by choosing a different null model.

The value of Q depends not only on the graph under investigation but also on the chosen partition. Thus, it is not a property of the graph itself. However, numerous methods, usually for the purpose of community detection, have been developed to search for the partition of the graph that will maximize Q. This maximal Q is a property of the graph and can be considered as a measure of modularity.

Although the method of maximizing Q is widely used for community detection, some shortcomings have been identified [57, 103]. First, it has been recognized that modularity maximization can fail to detect modules which are very small relative to the size of the network even if these modules are clearly defined. This effect is referred to as the resolution limit. Second, there are exponentially many partitions of the network, which can greatly differ from one another, with modularity scores that are similar to the maximum modularity. Good *et al.* call this phenomenon the degeneracy problem. Finally, the maximum modularity of a network can have a strong dependence on the size of the network and the number of modules, which complicates the comparison of modularity values between different networks. Some of these problems are more severe in networks which are sparse or hierarchical as many biological networks are. While the first two of these three problems can make the identification of modules more difficult, they do not seriously affect efforts to quantify modularity in networks. The third of these problems, however, has to be addressed when studying the effect a variable may have on modularity in biological networks. One approach is to normalize the modularity of a network by comparing it to a distribution of random networks which share the same topological features such as network size and degree distribution, as it was done in [54]:

$$Q_m = \frac{Q_{\text{real}} - Q_{\text{rand}}}{Q_{\text{max}} - Q_{\text{rand}}}. \quad \text{(A.3)}$$

Here Q_m is the normalized modularity, Q_{real} is the raw modularity, Q_{rand} is the average modularity of the random networks, and Q_{max} is the upper bound of the modularity, which

can either be estimated [54] or taken to be the largest value from the distribution of random networks.

Appendix A.2. Other Measures of Modularity

A further quantitative measure of modularity given a network with adjacency matrix A_{jk} and a partition of this network into modules is the ratio of the fraction of the weight within modules (*coverage*) over the fraction of the off-diagonal area within modules [38]:

$$M = \frac{\sum\limits_{j,k \neq j} A_{jk}\delta(c_j, c_k)}{\sum\limits_{j,k \neq j} A_{jk}} \times \left(\frac{\sum\limits_{j,k \neq j} \delta(c_j, c_k)}{n(n-1)} \right)^{-1}. \quad \text{(A.4)}$$

As above, $\delta(c_j, c_k) = 1$ if nodes j and k are in the same module and $\delta(c_j, c_k) = 0$ if they are not. The first term in the product, the coverage, measures what fraction of the edges lies within modules, while the second term normalizes by the size of the modules. In other words, M is the ratio of the density of the subgraph formed by the modules over the density of the original graph. Unlike Newman's modularity, maximizing this measure of modularity will not yield meaningful partitions of the network, because it greatly favors small modules. However, if a partition of a network is obtained by other means, such as hierarchical clustering, M quantifies how much the density of the clusters exceeds that of the entire network — a measure of modular structure.

If additional information is available about a network, the definition of modularity can be adapted to accommodate this information. For example, in the simulation described in section 2.2.4 a natural partitioning of the network arises from the horizontal gene transfer segments. Since horizontal gene transfer is restricted to predefined blocks, a sensible measure of modularity is given by the number of non-zero entries in the connection matrix within these blocks. For example, if there are 12 blocks of length 10 each, the quantity

$$\sum_{k=0}^{11}\sum_{j=1}^{10}\sum_{j'=j+1}^{10} \Delta_{10k+j,10k+j'}, \quad \text{(A.5)}$$

where   is the connection matrix, quantifies how many of the interactions take place within the predefined blocks. Because the total number of interactions is constant, if this number is large, then more interactions occur within the blocks than between them, indicating a modular structure.

The bandedness in figures 13 and 15 is a proxy for modularity measuring the locality of interactions in the network. To measure bandedness, the adjacency matrix of the network has to be reordered to concentrate interactions along the diagonal. This can be done using hierarchical clustering [38]. The bandedness is then defined as the ratio of the fraction of the interactions within a band along the diagonal over the fraction of the area within the band:

$$\frac{\displaystyle\sum_{0<|i-j|<W}^{D} A_{ij}}{\displaystyle\sum_{i\neq j} A_{ij}} \times \left(\frac{\displaystyle\sum_{0<|i-j|<W} 1}{\displaystyle\sum_{i\neq j} 1}\right)^{-1} \quad \text{(A.6)}$$

Here, W is the width of the band, and A_{ij} are the elements of the adjacency matrix.

Appendix A.3. Cophenetic Correlation Coefficient

For a given network one can define a distance between any pair of nodes using, for example, the commute distance described in section 3.3. Once distances are defined, one can construct a hierarchical tree, or dendrogram, of the network using hierarchical clustering. A dendrogram can be created for any network regardless of whether it exhibits a hierarchical structure or not. From this dendrogram a new pair-wise distance between nodes can be obtained given by the height at which two nodes are joined in the dendrogram. If the network is hierarchical, the distances obtained from the dendrogram will faithfully represent the original distances, but if the network does not have a hierarchical structure, the pair-wise distances from the dendrogram will differ greatly from the original distances. The condition for a set of pair-wise distances to be tree-like is that for any triple of nodes

$$T_{ij} \leq \max\left[T_{ik}, T_{jk}\right], \quad \text{(A.7)}$$

where T_{ij} is the distance between nodes i and j etc. The cophenetic correlation coefficient (CCC) quantifies how well the tree-like representation describes the network from which it was constructed and is thus a quantitative measure of the hierarchy in a network. It is defined as the Pearson correlation coefficient between the node-node distances in the original network and those in the dendrogram:

$$\text{CCC} = \frac{\displaystyle\sum_{i<j}(T_{ij}-T)\,(c_{ij}-c)}{\sqrt{\displaystyle\sum_{i<j}(T_{ij}-T)^2 \sum_{i<j}(c_{ij}-c)^2}}. \quad \text{(A.8)}$$

Here T is the average of the distances in the original network, T_{ij}, and c is the average of the dendrogram distances, c_{ij}. Unlike measures of modularity which depend on how the network is partitioned, the CCC is a property of the network. It can, however, be affected by the chosen distance measure on the original data and hierarchical clustering algorithm.
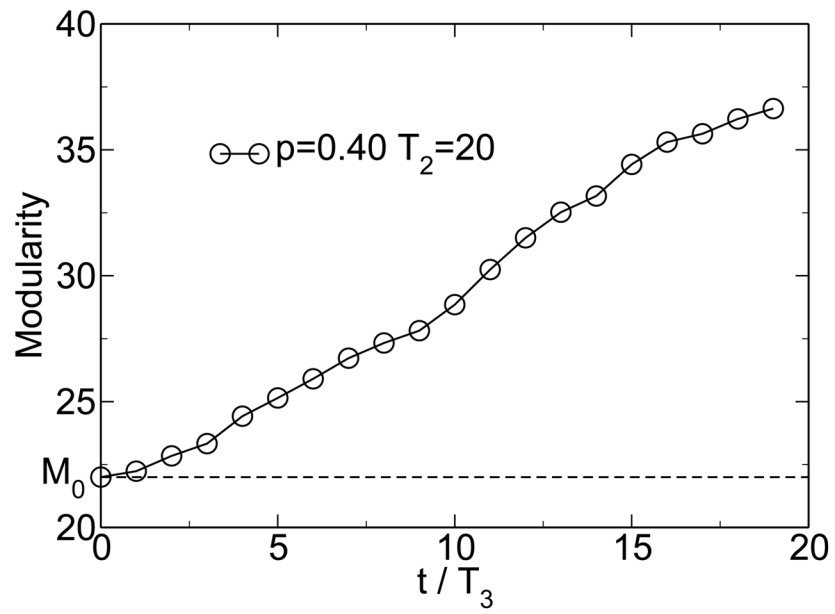
Figure 1.
Spontaneous emergence of modularity. $M_0 = 22$ is the baseline modularity for a random distribution of connections. From [38].
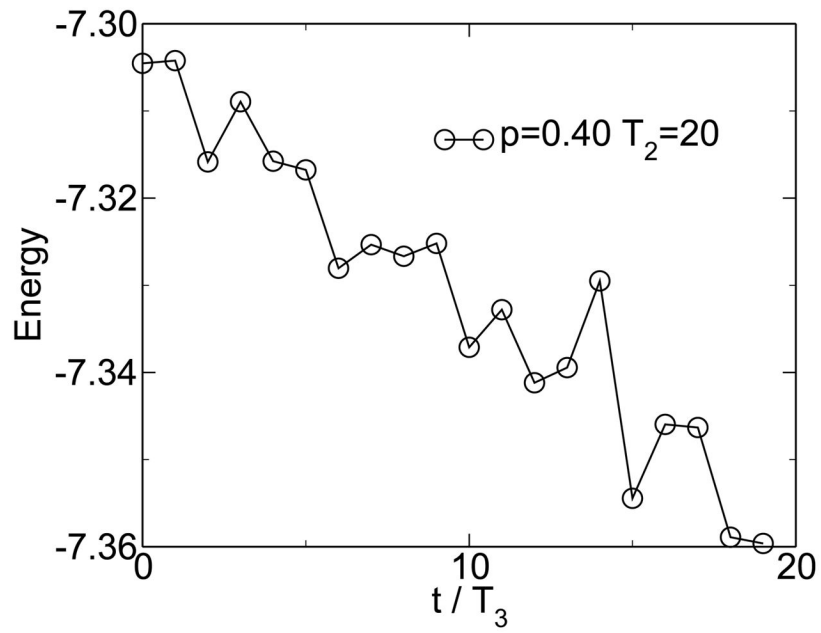
Figure 2.
The fitness (negative energy) increases with time and increasing modularity (see figure 1).
From [38].
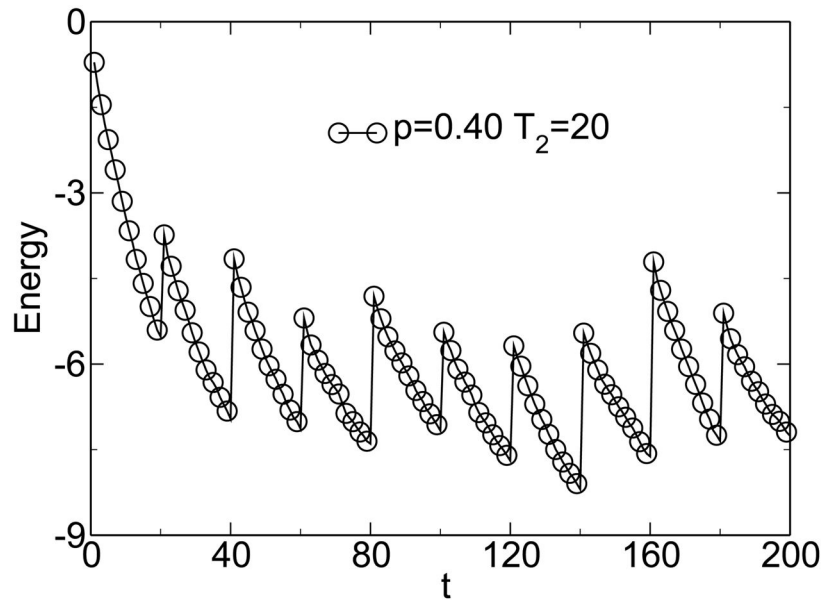
Figure 3.
Over short times, the fitness (negative energy) increases rapidly with time, but decreases significantly during environmental changes. From [38].
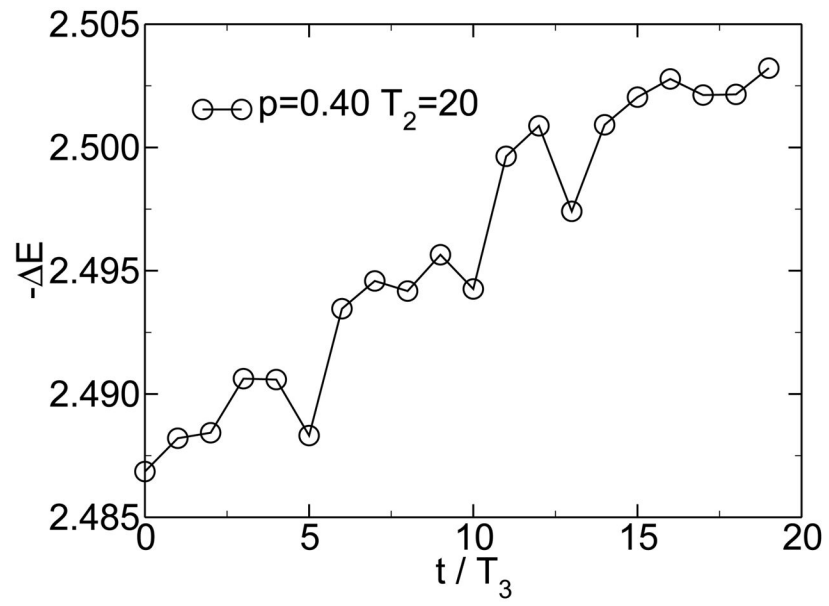
Figure 4.
The evolvability, as measured by the increase in fitness during evolution in one environment, increases with time as modularity increases. From [38].
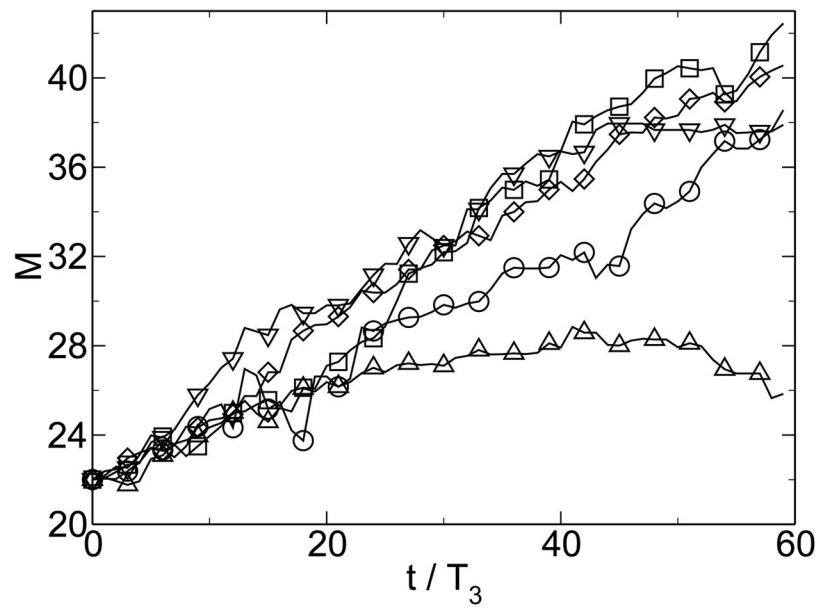
Figure 5.
Modularity growth for gene swaps of Poisson random length at a uniformly random position. The average swap lengths for the results shown are 10 ($\bigcirc$), 20 ($\square$), 20 ($\diamond$), 5 ($\triangle$), and 40 ($\triangledown$). The number of attempted swaps are 12, 6, 12, 24, and 3, respectively. From [38].
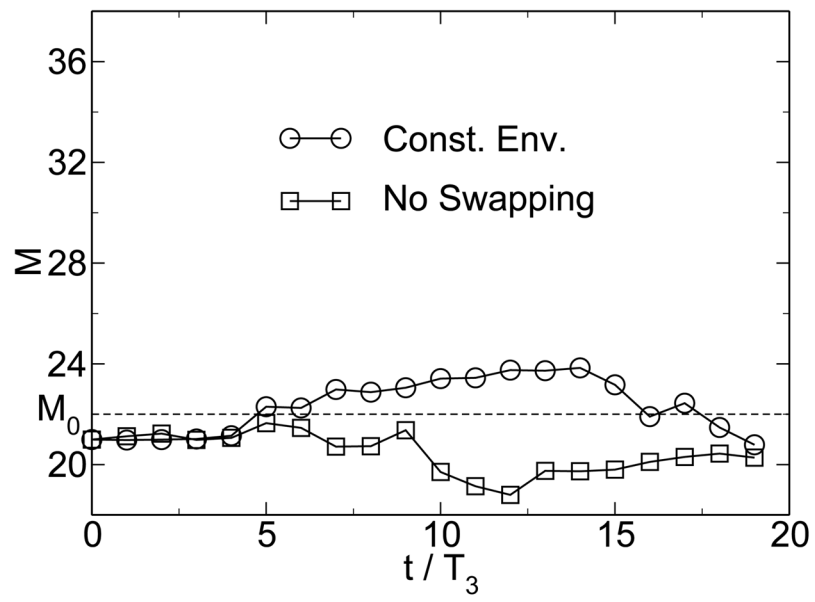
Figure 6.
If either environmental change or horizontal gene transfer is removed from the model, modularity does not emerge. From [38].
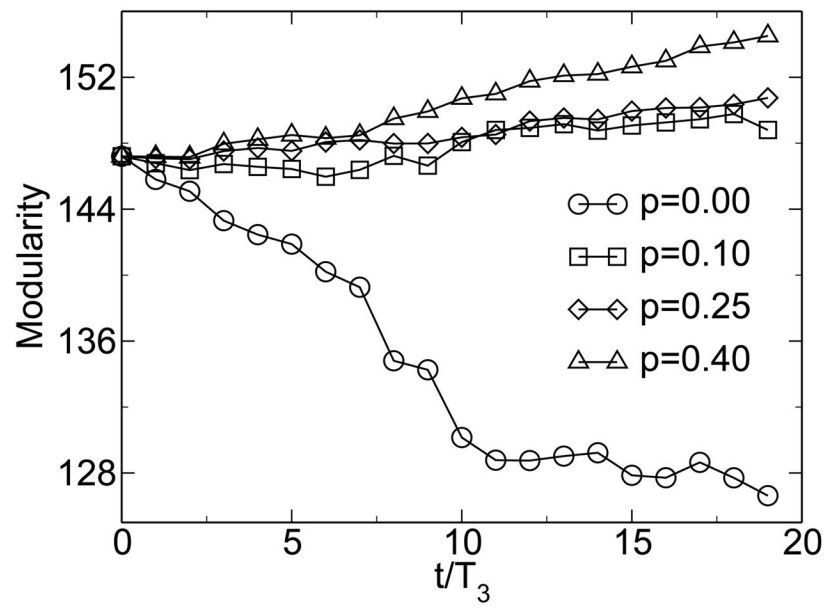
Figure 7.
Larger magnitudes of environmental change lead to a faster increase in modularity. From [38].
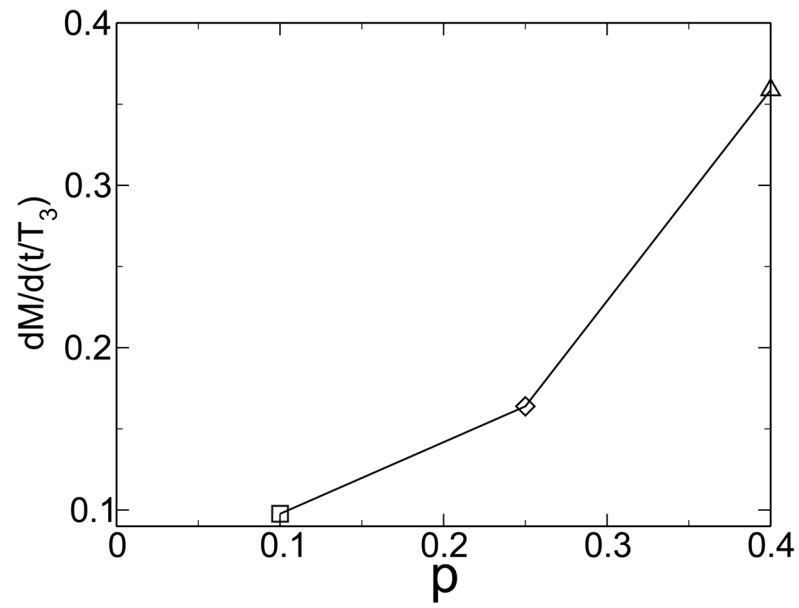
Figure 8.
The rate of change of modularity increases with increasing magnitude of environmental change. From [38].
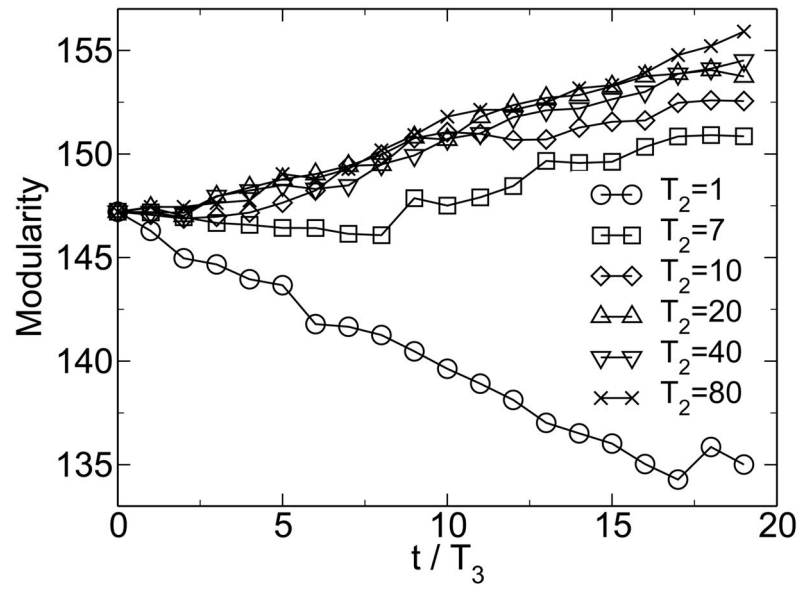
Figure 9.
More frequent environmental change leads to a faster increase in modularity. From [38].
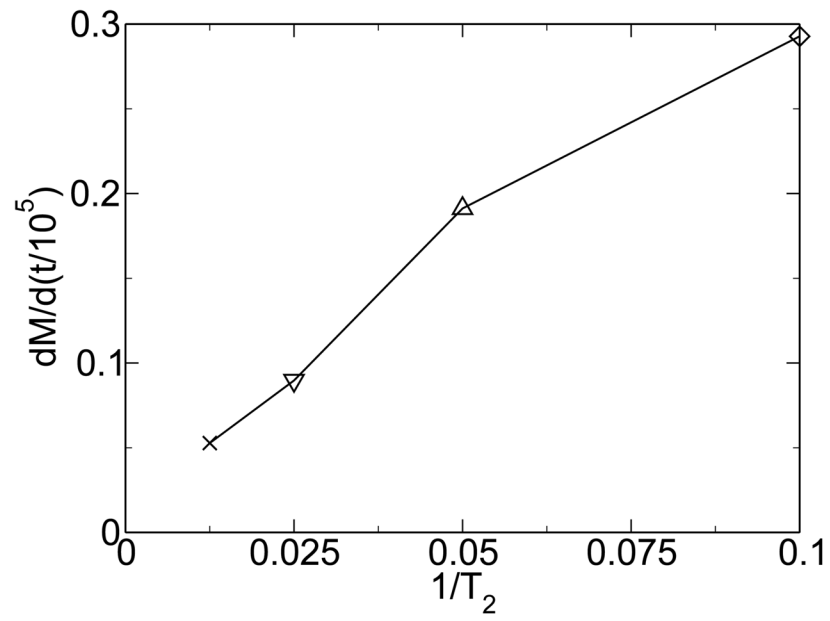
Figure 10.
The rate of change of modularity is approximately proportional to frequency of environmental change. From [38].
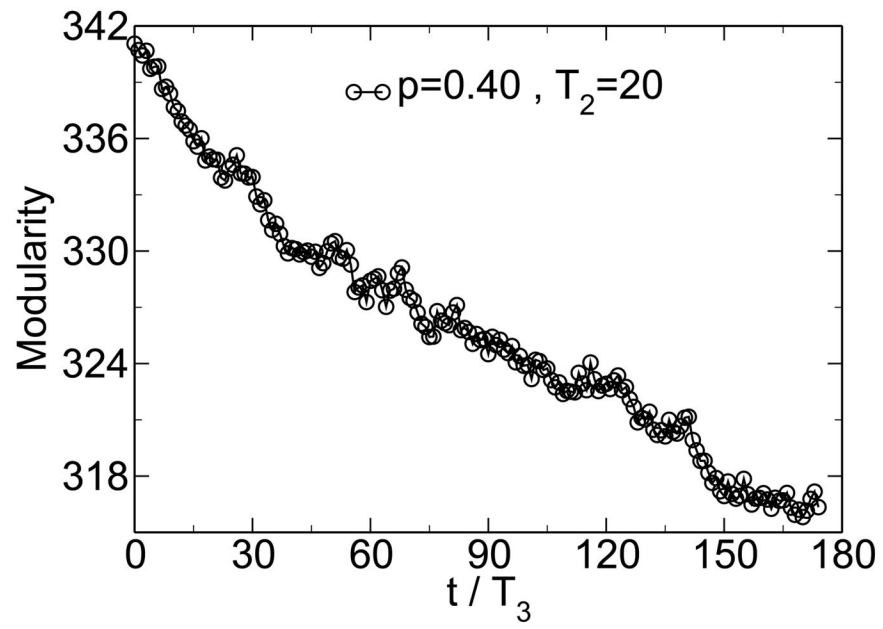
Figure 11.
If the modularity is initially above its steady-state value, it decreases with time because of the entropic effects of random mutations. From [38].
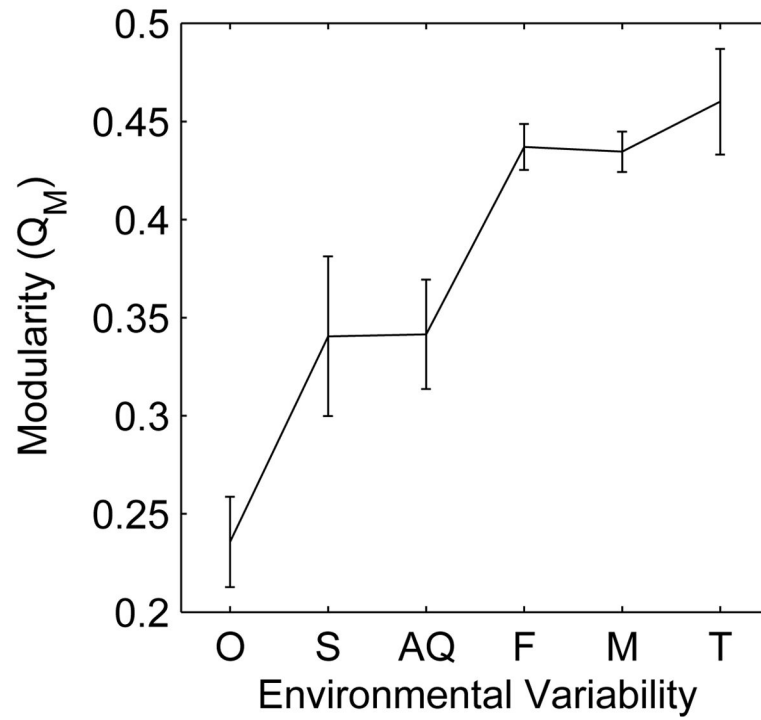
Figure 12.
Following [54]: Modularity increases with environmental variability. Bacteria are grouped by lifestyle and ordered from lowest to highest environmental variability. O stands for obligate, S for specialized, AQ for aquatic, F for facultative, M for multiple environments, and T for terrestrial.
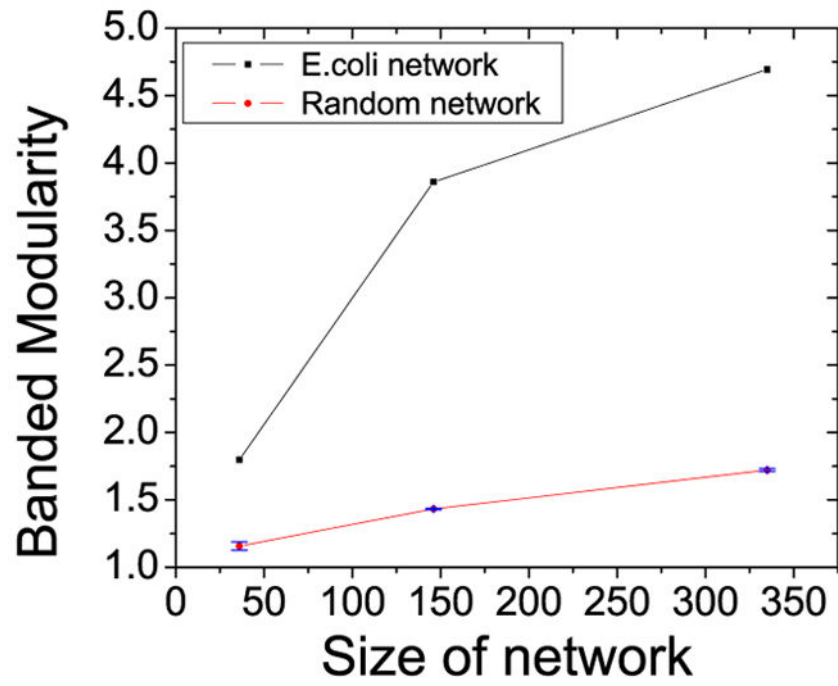
Figure 13.
The modularity of random networks increases with size. From [38].
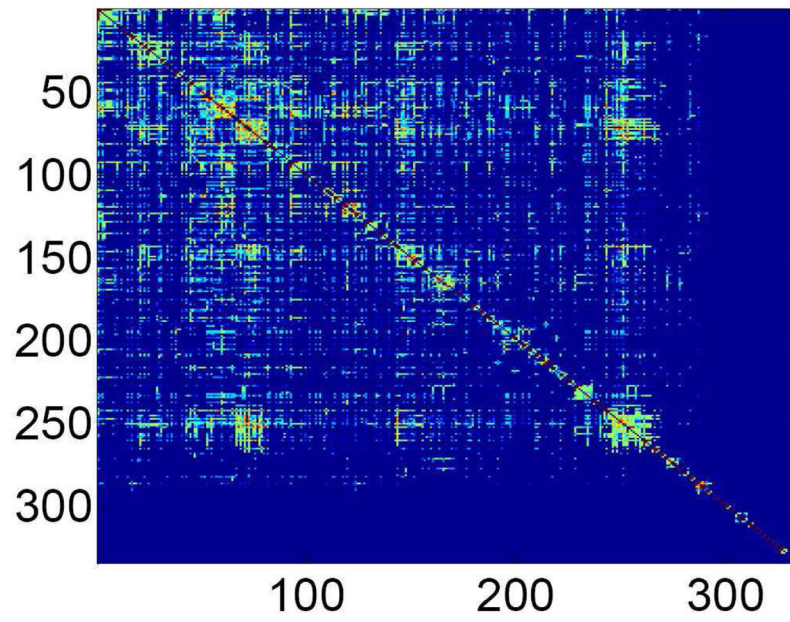
Figure 14.
Topological overlap matrix of the *E. coli* protein interaction network. From [38].
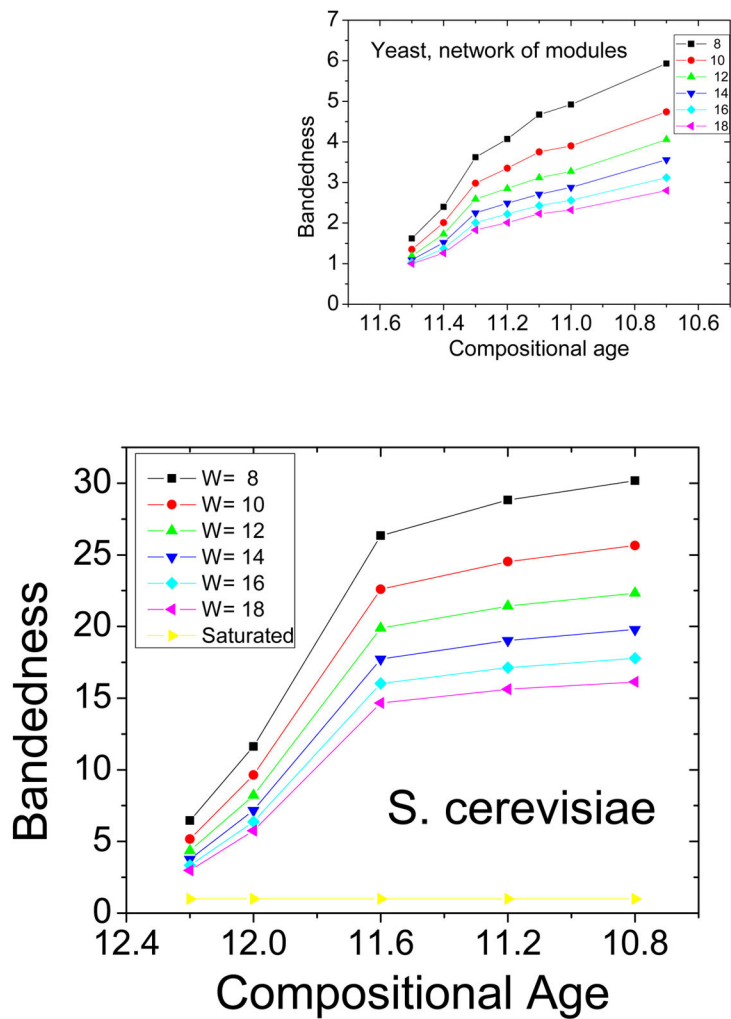
Figure 15.
Upper graph: second order modularity, lower graph: first order modularity from [38]. The second order modularity continues to increase after the first order modularity saturates. The compositional age axes have been aligned.
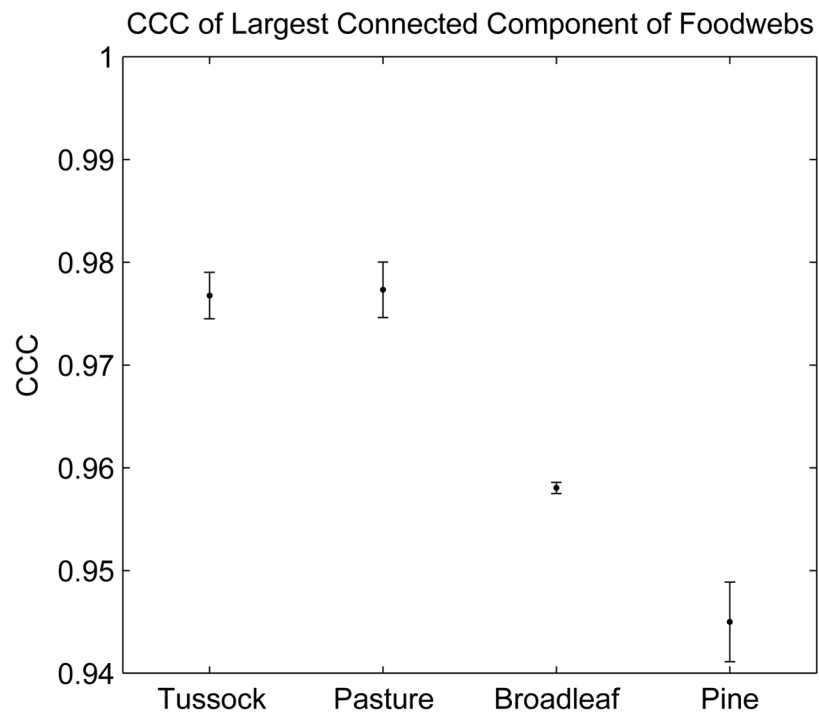
Figure 16.
Cophenetic correlation coefficient for river food webs surrounded by different terrestrial habitats. Food webs of rivers flowing through habitats with low detritus input are more hierarchical. For each habitat, the error bars are one standard error. Data from [74, 75].
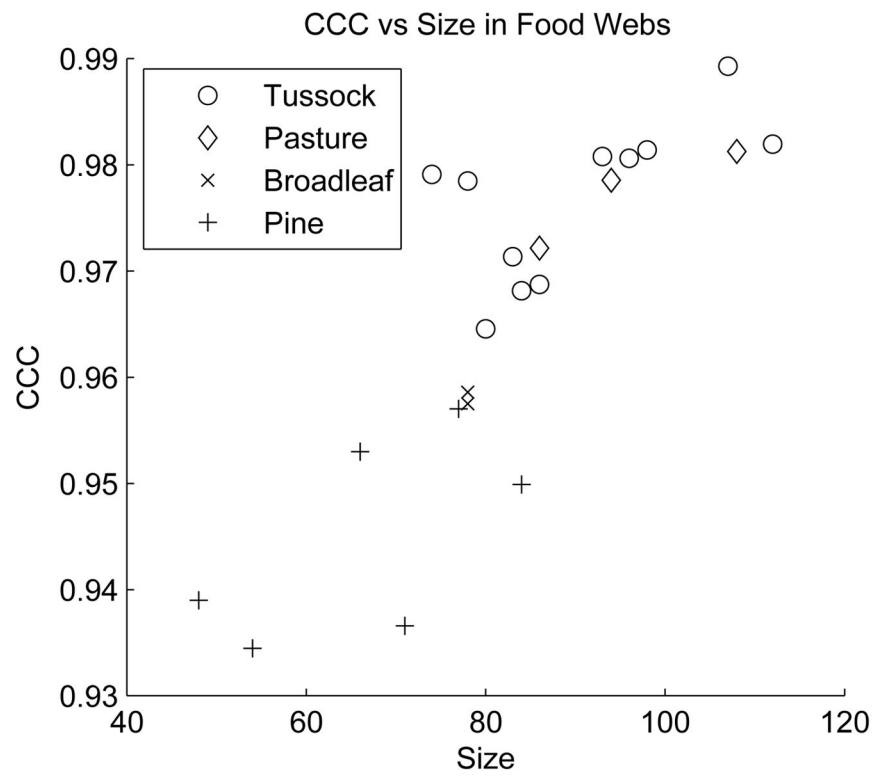
Figure 17.
The studied food webs show a trend between network size and CCC. Data from [74, 75].
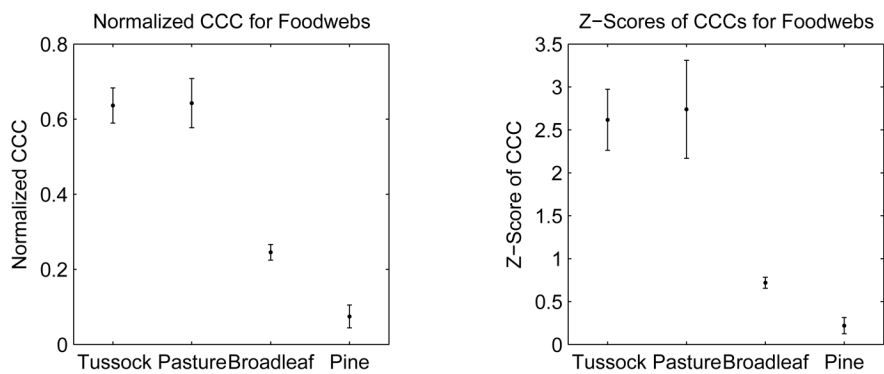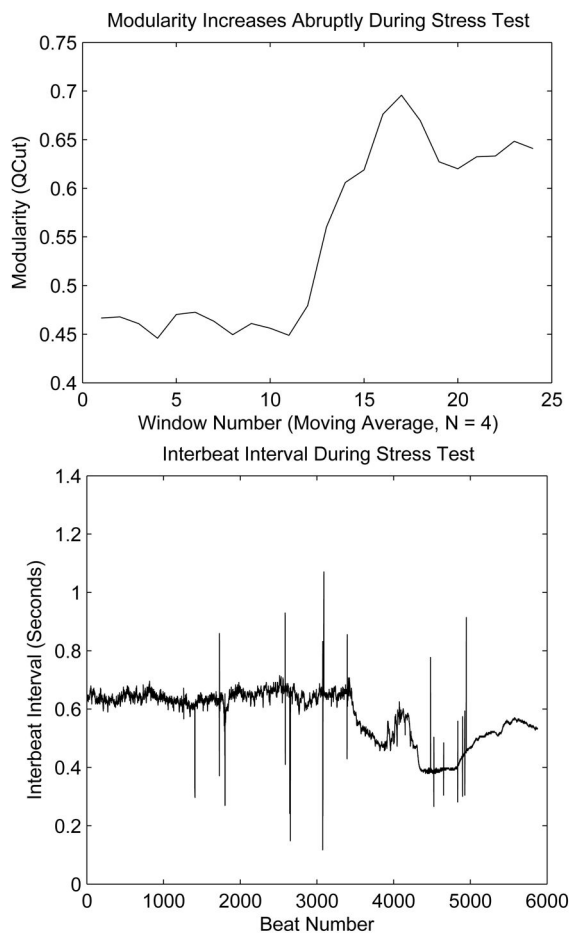
Figure 18.
Normalized CCC values and standard scores (z-scores) of river food webs flowing through different habitats relative to the distribution of CCC values obtained from random networks of equal size and total number of edges. The graphs show the same trend as the unnormalized CCC values, demonstrating that the trend is not an artifact caused by variations in network size. The error bars are one standard error. Data from [74, 75].
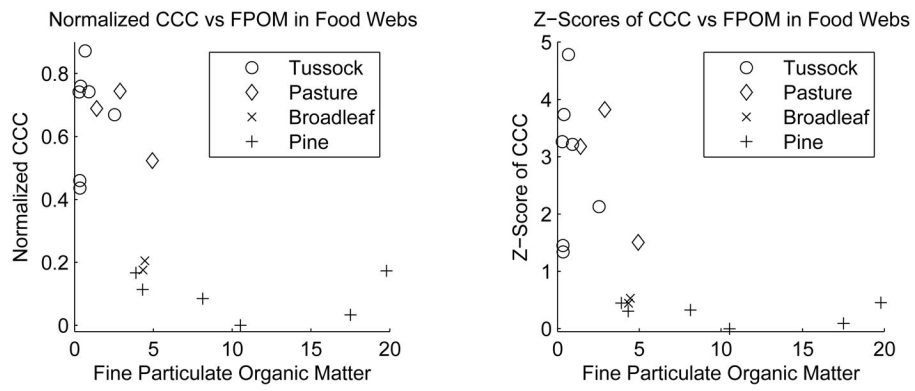
Figure 19.
Normalized CCC values and standard scores (z-scores) of river food webs as a function of fine particulate organic matter. Both normalized CCC measures decrease with decreasing particulate organic matter. Data from [75].
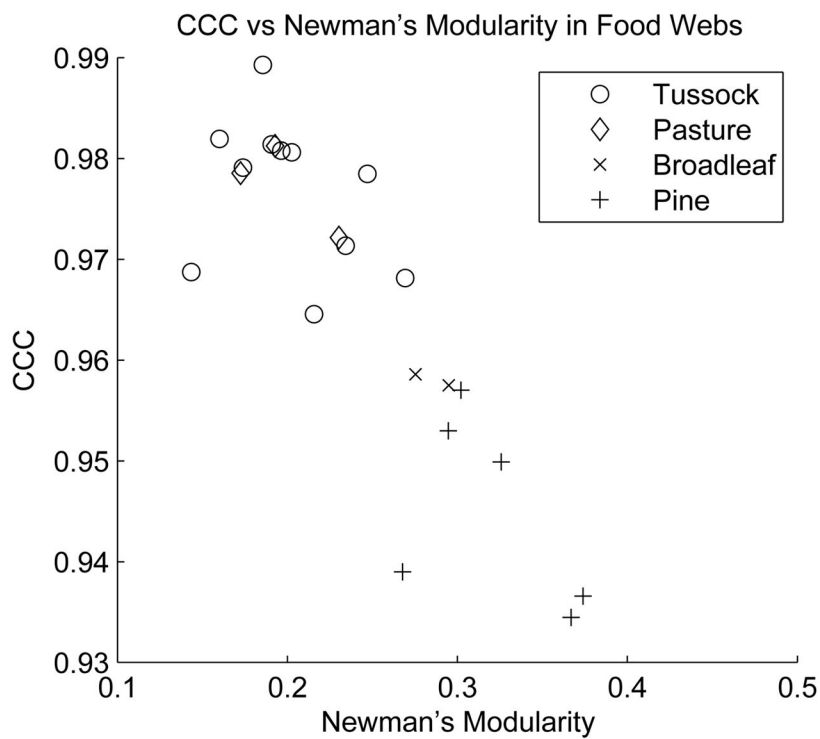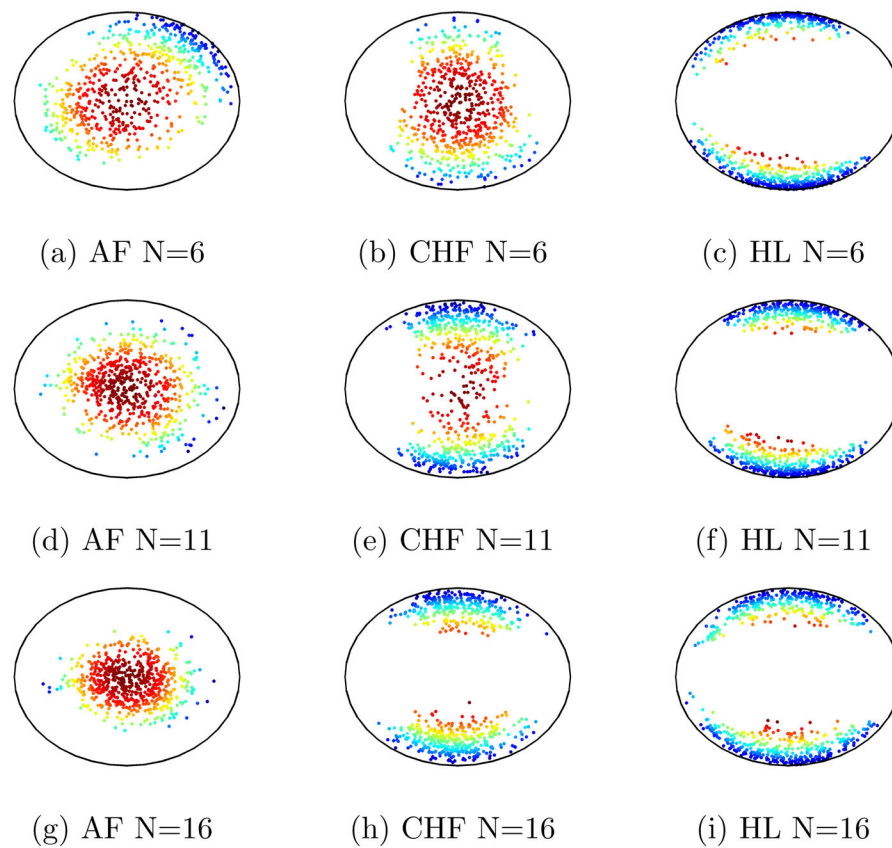
Figure 20.
Cophenetic correlation coefficient versus Newman's modularity for the largest connected component of river food webs. There is a negative correlation between CCC and Newman's modularity. However, the value for Newman's modularity is quite low for most of the food webs. Data from [74, 75].
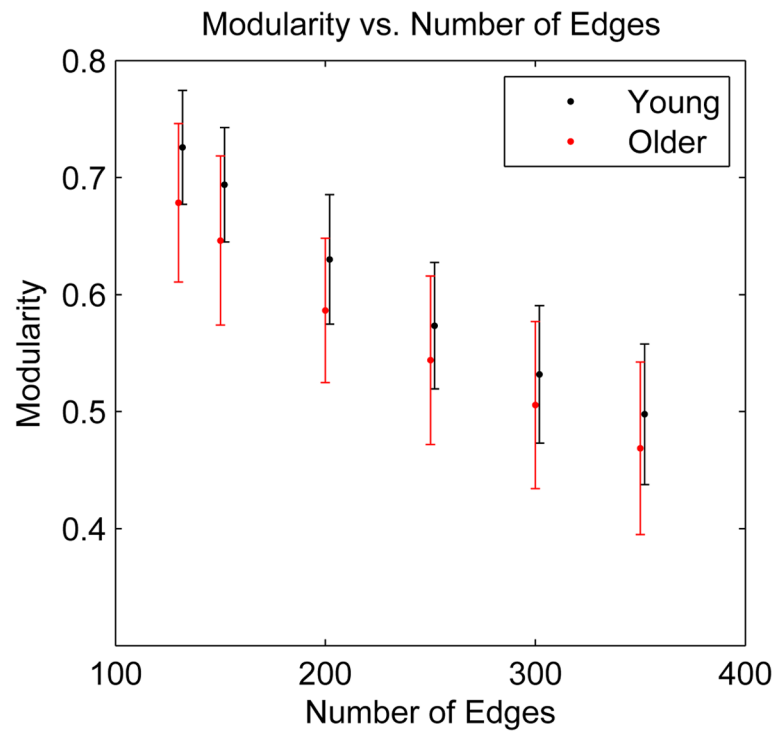
Figure 21.
From [94]. Upper graph: the modularity of the interbeat covariance matrix increases abruptly during a stress test on a treadmill. Lower graph: unprocessed record of interbeat intervals during the same test. The abscissæ of the two graphs correspond to the same time interval. Note that the increase in modularity precedes the rise in heart rate by 400 heart beats.

(a) AF N=6   (b) CHF N=6   (c) HL N=6

(d) AF N=11   (e) CHF N=11   (f) HL N=11

(g) AF N=16   (h) CHF N=16   (i) HL N=16

Figure 22.
Results of a topological analysis of frequent-motif spaces performed by Carlsson *et al.* [95] for different time lengths, *N*. The motifs of healthy patients (HL) always separate into two components while the motifs of patients suffering from atrial fibrillation (AF) never show a separation. For congestive heart failure (CHF) patients, the motif space shows no separation for small *N*, but a separation develops with increasing *N*.
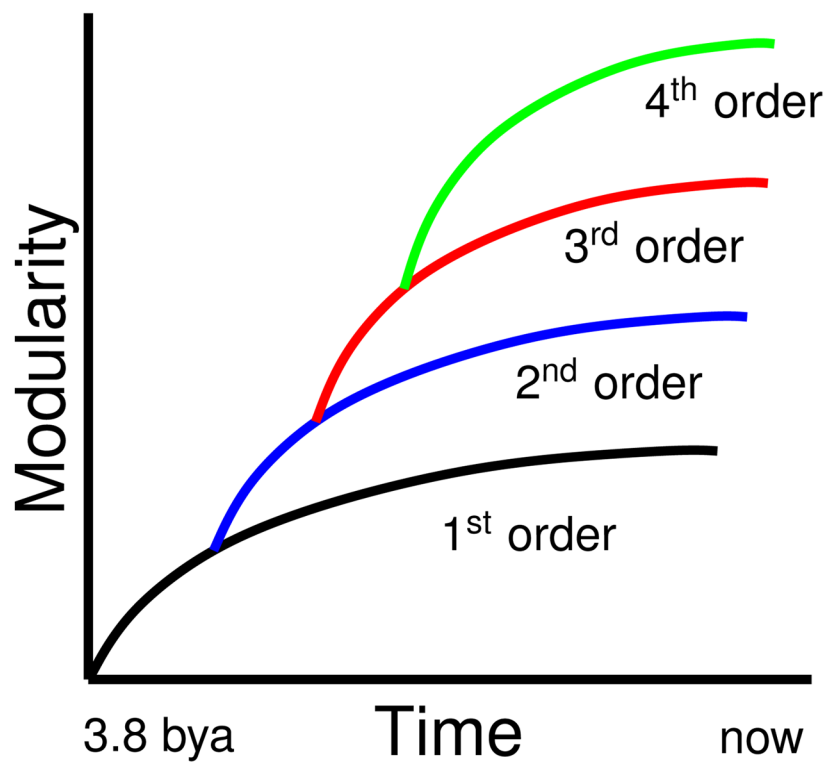
Figure 23.
Newman's modularity for different threshold values, following [100]. The modularity in the younger group is consistently larger than in the older group. The difference becomes more significant for larger threshold values, *i.e.* sparser networks. The modularities shown here are slightly larger that in [100] because we used the spectral algorithm described in [81] to maximize modularity rather than a greedy algorithm.
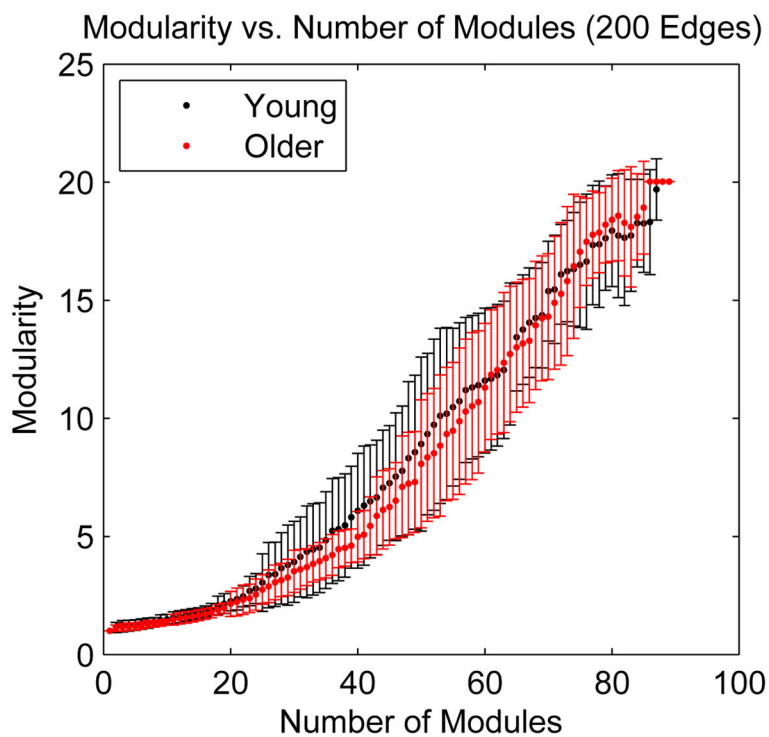
Figure 24.
The modularity as defined in equation (12) across the dendrogram obtained from average linkage hierarchical clustering on commute distance matrices of network divisions for 200 edges. The modularity of brain networks from young adults is higher than that of networks from older adults for all relevant parts of the dendrogram. Data from [100].

Figure 25.
The emergence of hierarchical modularity. Initially, modularity increases as first-order modules are formed. However, modularity is bounded, and so there is a time at which it saturates. We expect that before first-order modularity saturates, a second-order modularity emerges by combining the first-order modules into second-order modules. This process repeats to create a hierarchy of modules.