

# Fibration building blocks of information-processing networks

Flaviano Morone, Ian Leifer, Hernán A. Makse

*Levich Institute and Physics Department,*

*City College of New York, New York, NY 10031*

## Abstract

A major ambition of systems science is to uncover the building blocks of information-processing networks to decipher how these networks are put together to regulate their function. Here we introduce a graph representation of the information flow in the network as a set of universal input-trees, one for each node, which contains all pathways along which information can be transmitted from node to node. In this representation, we find remarkable symmetries in the universal input-trees that deconstruct the network into functional building blocks called fibers. Nodes in a fiber have isomorphic input-trees and thus process equivalent information and synchronize their activity. Each fiber can then be collapsed into a single representative base node through an information-preserving transformation called graph fibration, introduced by Grothendieck in the context of algebraic geometry. We exemplify the symmetry fibrations in gene regulatory networks and then show that apply **across all domains from biological, social and infrastructure networks.** The building blocks are classified into topological classes of input-trees characterized by integer branching ratios and fractal golden ratios of Fibonacci sequences representing cycles of information. Thus, fibration symmetries describe how complex networks are built from the bottom up to process information through the synchronization of their constitutive building blocks.

A recurring theme across all sciences is to find the fundamental building blocks of a complex system and then study the different ways to put them together [1–3] to guarantee the system’s function. For instance, every natural number can be written into a product of primes in a unique way. Thus, prime numbers are the building blocks of natural numbers. This mathematical notion of building blocks is extended to the more abstract notion of group theory and, under certain circumstances, finite groups can be also factored into “prime” groups [4]. The latter example, entirely abstract as it may be, has important implications for natural systems, due to the fundamental relationship between group theory and the notion of symmetry, that has led to the discovery of the fundamental building blocks of matter [3, 5]. Here we ask whether the same idea can provide the fundamental building blocks of information-processing networks [1, 2, 6, 7]; such as transcriptional regulatory networks that control gene expression in cells [2, 7, 8], as well as other networks like metabolic, cellular processes and pathways, neural and ecological networks, or even beyond biology to domains like social networks, internet and the web, economics, software and infrastructure networks. Previous studies identified ‘network motifs’ [2, 6, 8] as the building blocks of these networks. The crux of the matter is to test whether the building blocks of biological networks obey a predictive principle and whether such a principle can be expressed in the language of symmetries.

We introduce the concept of symmetry fibration by analyzing the gene regulatory network of bacterium *Escherichia coli* [8, 9], since it is a well-characterized biological network. We find that this network exhibits symmetries that are formally called **graph fibrations** [10–12]; first introduced by Grothendieck [10] in the context of algebraic geometry.

Fibrations identify clusters of synchronized genes (called **fibers**) which process equivalent information, such that genes in a fiber can be collapsed into a single representative gene called the **base**. The fibers are then the building blocks of the genetic network according to a symmetry principle that conserves the information flow in the genetic network. This symmetry principle classifies the building blocks into topological classes of universal input-trees characterized by integer branching ratios and more complex topologies with golden ratios of Fibonacci sequences representing network cycles. We then show that symmetry fibrations applies to a vast range of complex networks across different domains.

We search for symmetries in the *E. coli* transcriptional network (most updated compilation at RegulonDB [9]) where nodes are genes and a directed link represents a transcriptional

regulation (activator or repressor) mediated by a transcription factor (see SM Section VII and Ref. [8] for details of network construction).

We start our discussion by analyzing the symmetries in the sub-circuit regulated by gene *cpxR* which regulates its own expression (via an autoregulation positive loop) and regulates other genes as shown in Fig. 1a. Gene *cpxR* is not regulated by any other transcriptor factor so that it does not receive ‘information’ from the main network (we say that this gene forms its own ‘strongly connected component’, see below). Therefore, it is an ideal simple circuit to explain the concept of graph fibration.

A link in a transcriptional network represents regulatory ‘messages’ that are dynamically sent from a source gene to a target gene via the transcription factor. The transcription factor acts as a ‘messenger’ to repress or activate the transcription rate of the target gene. This information flow is not restricted to two interacting genes, but is transferred to different regions within the network that are accessible through the connecting pathways. The information arriving to a gene contains the entire history transmitted through all pathways that reach this gene. We formalize this process of communication between genes with the notion of **universal input-tree** of the gene. In a genetic network  $G$ , for every gene  $i$  there is a corresponding input-tree, denoted as  $T_i$ , which is the tree of all pathways of  $G$  ending at  $i$ . More precisely,  $T_i$  is a tree with a selected node  $i$  at the root, such that every other node  $j$  in the tree represents the initial node of a path in the network ending at  $i$ . **We note that the universal input-tree corresponds to the universal total space in traditional categorical or topological terminology [10], the universal total graph from [11], the view in the theory of distributed systems, or the unfolding of a nondeterministic automaton in concurrency theory [11].**

SM Section VIII A provides a formal mathematical definition. In practice, the input-tree of a gene is constructed as follows. Consider the circuit in Fig. 1a. The input-tree of *spy* (Fig. 1b) starts with *spy* at the root (first layer). Since this gene is upregulated by *baeR* and *cpxR*, then, the second layer of the input-tree contains these two pathways of length one starting at both genes. Gene *baeR* is further upregulated by *cpxR* and by itself through the autoregulation loop and *cpxR* is also autoregulated. Thus, the input-tree continues to the third layer taking into account these three possible pathways of length 2, one starting at *baeR* and two starting at *cpxR*. The procedure now continues, and since there are loops in the circuit, the input-tree has an infinite number of layers.

The universal input-tree formalism is a powerful framework to search for symmetries that preserve the information flow in the network, in that it replaces the canonical notion of a single trajectory with the set of all possible ‘histories’ from an initial to a final state of the network, and this makes, in practice, reasonably straightforward to ‘guess’ a type of symmetry which is not apparent in the classical network framework. Simply put, if two input-trees have the same ‘shape’ node by node and link by link, as in a local in-isomorphism, then the genes at the root of the input-trees receive the same information through the totality of all paths ending at the root, they share the same dynamical state and they synchronize their activity [11–20]. The informal notion of equivalence can be formalized by isomorphisms. An isomorphism between two input-trees is a bijective map that preserves the topology of the input-trees including their links. Specifically, a map  $\tau : T \rightarrow T'$  is an isomorphism iff for any pair of nodes  $a$  and  $b$  of  $T$  connected by a link, the pair of nodes  $\tau(a)$  and  $\tau(b)$  of  $T'$  is connected by the same type of link, too [11] (see SM Section VIII B for more details). In practice, this means that isomorphic input-trees are ‘the same’ except for the labeling of the nodes. Two input-trees related by an isomorphism are isomorphic or equivalent (understood as a local in-isomorphism, node to node and link to link) and genes with isomorphic input-trees are symmetric because they receive the same information and share their dynamical state [11, 12].

The set of all universal input-tree isomorphisms defines the symmetries of the genetic network, which can be described by a **Grothendieck fibration** [10]. The original Grothendieck definition of fibration is between categories [10], so the passage to a definition of fibrations between graphs requires to associate a category with a graph and rephrase Grothendieck’s definition in elementary terms. Different categories may be associated with a graph, giving rise to different notion of fibration between graphs. The notion of fibration that we use henceforth has been introduced in [11], and is called “graph fibration”.

A graph fibration of a network  $G$  is a morphism (formal definition in Section I and SM Section VIII C):

$$\psi : G \rightarrow B \tag{1}$$

that maps  $G$  to a network  $B$ , called the **base** of the fibration  $\psi$ . If two genes in  $G$  are mapped to the same node in  $B$ , they have the same universal input-tree. Also, if  $\psi$  is a minimal fibration in the sense of [11], then  $\psi$  maps genes with isomorphic input-trees to a common node in  $B$ . Thus, any two genes  $i$  and  $j$  with isomorphic input-trees are mapped

onto the same fiber of  $B$  by a minimal fibration  $\psi$ . The base  $B$  consists of a network where the genes in a fiber have been collapsed into one node by the fibration, leading to a dimensional reduction of the network into its irreducible components. Crucially, the fibration is a dimensional reduction that preserves the information in the network.

## I. FORMAL DEFINITION OF GRAPH FIBRATION

Following the notation of Ref. [11], a graph  $G$  is defined by a set  $N_G$  of nodes and a set  $A_G$  of directed links, and by two maps  $s_G, t_G : A_G \rightarrow N_G$  that define the source and the target of each link (following nomenclature in other fields, we also call such a graph, a 'network'). A graph morphism is a map  $\xi : G \rightarrow H$  that is given by two functions that map nodes to nodes  $\xi_N : N_G \rightarrow N_H$ , and links to links  $\xi_A : A_G \rightarrow A_H$ . An important property is that these maps commute with the source and target maps (see [11]):

$$s_H \circ \xi_A = \xi_N \circ s_G \quad \text{and} \quad t_H \circ \xi_A = \xi_N \circ t_G. \quad (2)$$

Therefore, a morphism maps nodes to nodes and links to links to preserve the incidence relation. The morphism becomes an epimorphism if and only if  $\xi_N$  and  $\xi_A$  are surjective.

We define a graph fibration between a graph  $G$  and  $B$  as a morphism  $\psi : G \rightarrow B$ , such that for each link  $a \in A_B$  and node  $x \in N_G$  satisfying  $\psi(x) = t(a)$  there is a unique link  $\tilde{a}^x \in A_G$  (called the lifting of  $a$  at  $x$ ) such that  $\psi(\tilde{a}^x) = a$  and  $t(\tilde{a}^x) = x$ . Then,  $G$  is called the total graph and  $B$  is the base of  $\psi$  and we say that  $G$  is fibered over the base  $B$ . The fiber over a node  $x \in N_B$  is the set of nodes of  $G$  that are mapped to  $x$ , denoted by  $\psi^{-1}(x)$ .

Ref. [11] further provides an intuitive definition of fibration in terms of local isomorphism. Let denote by  $G(-, x)$  the set of input links  $a \in A_G$  coming into node  $x$ , i. e.,  $t(a) = x$ . Then, an equivalent relation between two nodes  $x$  and  $y$  exists, i.e.  $x \simeq y$ , if there exists a bijection  $\varphi : G(-, x) \rightarrow G(-, y)$  such that  $s(a) \simeq s(\varphi(a))$ , for all  $a \in G(-, x)$ . A proposition proved in [11] shows that fibrations are equivalent to epimorphisms whose fibers satisfy this property.

A simpler and mathematically clear definition of fibration (following Grothendieck's legacy) is given in terms of a local lifting property from [11]: for a path  $\pi$  ending at a node  $x$  in the base, for each node  $y \in G$  in the fiber of  $x$  there is a unique path ending at  $y$  that is mapped to  $\pi$  by the fibration. This path is the lifting of  $\pi$  at  $y$ .

If two genes in  $G$  are mapped to the same node in  $B$ , they have the same universal input-tree. Also, if  $\psi$  is a minimal fibration in the sense of [11], then  $\psi$  maps genes with isomorphic input-trees to a common node in  $B$ . Thus, any two genes  $i$  and  $j$  with isomorphic input-trees are mapped onto the same fiber of  $B$  by a minimal fibration  $\psi$ .

The isomorphism involves not only mapping the nodes but also the types of links. Thus, it is not just the topology of the input-trees that determines the fibration, but also the type of links. The fibration formalism admits the case of different interactions among the genes, which implies that different genes could have different types of response functions, as long as the local in-isomorphism is approximately valid in the fiber to guarantee the isomorphism between genes that leads to synchronization in the fiber. The conditions for this validity and the relevance of heterogeneous response functions to the dynamical synchronization in the fiber is treated in SM Section IX.

The input-trees in the *cpxR* circuit are displayed in Fig. 1b. The input-trees of *baeR* and *spy* (Fig. 1c) are isomorphic and define the *baeR-spy* fiber. Clearly, the input-tree of *cpxR* is not isomorphic to either *baeR* or *spy*, and therefore *cpxR* is not symmetric with these genes, but it is isomorphic to *bacA*, *slt* and *yebE* forming another fiber. Likewise, *ung*, *tsr* and *psd* are all isomorphic forming their corresponding fiber (Fig. 1b). Figure 1d shows the fibration  $\psi : G \rightarrow B$  that collapses the genes in the fibers to the base  $B$ . Figure 1e shows another example (out of many) of a single connected component, *fadR*, and its corresponding isomorphic input-trees (Fig. 1f), fibers and base.

The input-tree encodes a sequence,  $a_i$ , defined as the number of genes in each  $i$ -th layer of the input-tree. The sequence  $a_i$  represents the number of paths of length  $i - 1$  that reach the gene at the root. This sequence is characterized by the branching ratio  $n$  of the input-tree defined as  $a_{i+1}/a_i \xrightarrow{i \rightarrow \infty} n$ , which represents the multiplicative growth of the number of paths across the network reaching the gene at the root. For instance, the input-trees of genes *baeR-spy* (Fig. 1b) encode a sequence  $a_i = i$  with branching ratio  $n = 1$  representing the single ( $n=1$ ) autoregulation loop inside the fiber.

We find that the network has other strongly connected components [in a strongly connected component, each gene is reachable from every other gene, SM Section X], three in total, which regulate more involved topologies of fibers: (i) a two-gene connected component composed of master regulators *crp-fis* (Fig. 2a, top), (ii) a five-gene connected component (SM Fig. 6), and (iii) the largest connected component at the core of the network which is

composed of genes involved in the pH-system that regulates hydrogen concentration (Fig. 2b). Each of the three connected components regulate a rich variety of fibers topologies shown in the figures which are collapsed by the application of fibration  $\psi : G \rightarrow B$  to produce their respective information-preserving bases.

We find that the full transcriptional network is organized into 91 different fibers. The algorithm to obtain the fibers in a network is based on a balanced coloring algorithm and was developed by Cardon and Crochemore in Ref. [21] (SM Section X). The complete list of fibers appears in SM Section XIII and SM-Table V with the statistics in SM Table IV. The plot of each fiber appears in Supplementary File 1. Despite their rich variety, the topologies of the input-trees present common features that allow us to classify all circuits into concise classes of 'fiber building blocks' (Figs. 3a and 3b). We define the 'fiber building block' as the genes in the fiber plus the minimum number of external regulator genes and loops needed to establish the local equivalence or isomorphism in the fiber.

We find that most of the input-tree topologies are fundamental and can be classified by two 'fiber numbers'  $|n, \ell\rangle$  reflecting two features: (a) infinite  $n$ -ary trees with branching ratio  $n$  representing the infinite pathways going through  $n$  loops in the fiber, and (b) finite trees representing finite pathways starting at  $\ell$  external regulators of the fiber. We find three main fiber numbers  $n$  in the regulatory network (Fig. 3a): (i)  $n = 0$  loops, Star Fiber (SF) class, (ii)  $n = 1$  loop, Chain Fiber (CF) class, and (iii)  $n = 2$  loops, Binary-Tree Fiber (BTF) class. This classification does not take into account the types of repressor or activator link inside the fiber, which could lead to further sub-classes of fibers.

Figure 3a shows examples of dissimilar circuits that can be concisely classified by  $|n, \ell\rangle$ . The full list of building blocks appears in Supplementary File 1 and more examples are displayed in SM Table VI (SF class), SM Table VII (CF class), and SM Table VIII (BTF class). For instance the  $n = 0$  SF class includes dissimilar circuits like  $|arcZ-ydeA\rangle = |0, 1\rangle$ ,  $|dcuC-ackA\rangle = |0, 2\rangle$ , and generalizations with  $\ell = 3$  regulators like  $|dcuR-aspA\rangle = |0, 3\rangle$  (Fig. 3a, top). The main feature of all these fiber building blocks is that they do not contain loops and therefore the input-trees are finite. The CF class contains  $n = 1$  loop in the fiber, and therefore an infinite chain in the input-tree, like the autoregulated loop in the fiber  $|ttdR\rangle = |1, 0\rangle$ . Adding an external regulator to this circuit, converts it to  $|purR\rangle = |1, 1\rangle$ . When another external regulator is added, we find  $|idnR\rangle = |1, 2\rangle$ . More elaborated circuits contain two autoregulated loops and feed-back loops featuring trees with branching ratio

$n = 2$ .

So far we have analyzed building blocks that receive information from the regulators in their respective connected components, but do not send back information to the regulators. These topologies are characterized by integer branching ratios,  $n = 0, 1, 2$ , as shown in Fig. 3a. We find, however, more interesting building blocks that also send information back to their regulators. These circuits contain additional cycles in the building block such that the input-trees become fractal. Notably, the building block of the fiber *uxuR-lgoR* that is regulated by the connected component *crp-fis* forms an intricate input-tree (see Fig. 3b, top) where the number of paths of length  $i - 1$  is encoded in a Fibonacci sequence:  $a_i = 1, 3, 4, 7, 11, 18, 29, \dots$  characterized by the Fibonacci recurring relation:  $a_1 = 1, a_2 = 3$ , and  $a_i = a_{i-1} + a_{i-2}$  for  $i > 2$ . This sequence leads to the non-integer branching ratio known as the golden ratio:  $a_{i+1}/a_i \xrightarrow{i \rightarrow \infty} \varphi = (1 + \sqrt{5})/2 = 1.6180\dots$

This remarkable topology arises in the genetic network due to the combination of two cycles of information flow. First, the autoregulation loop inside the fiber at *uxuR* creates a cycle of information of length  $d = 1$  which contributes to the input-tree with an infinite chain with branching ratio  $n = 1$ . This sequence is reflected in the Fibonacci series by the term  $a_i = a_{i-1}$ . The crucial addition to the building block is a second cycle of length  $d = 2$  between *uxuR* in the fiber and its regulator: *uxuR*  $\rightarrow$  *exuR*  $\rightarrow$  *uxuR*. This loop sends information from the fiber to the regulator and back to the fiber by traversing a path of length  $d = 2$  that creates a 'delay' of  $d = 2$  steps in the information that arrives back to the fiber (see Fig. 3b, top right). This short-term 'memory' effect is captured by the second term  $a_i = a_{i-2}$  in the Fibonacci sequence leading to the golden ratio. We call this topology the Fibonacci fiber (FF).

This argument implies that an autoregulated fiber that further regulates itself by connecting to its connected component via a cycle of length  $d$  encodes a generalized Fibonacci sequence defined as  $a_i = a_{i-1} + a_{i-d}$  (Fig. 3b, bottom). We find such a Fibonacci sequence in the *evgA-nhaR* fiber building block. This fiber contains an autoregulation cycle inside the fiber but also forms a cycle of length  $d = 4$  by sending information back to the pH-connected component: *evgA*  $\rightarrow$  *gadE*  $\rightarrow$  *gadX*  $\rightarrow$  *hns*  $\rightarrow$  *evgA* (Fig. 3b, third row). This topology forms a fractal input-tree with  $a_i = a_{i-1} + a_{i-4}$  and generalized golden ratio  $\varphi_4 = 1.38028\dots$  (see SM Section XIII to calculate generalized Fibonacci golden ratios). We call this topology a generalized 4-Fibonacci fiber, 4-FF. Generalized Fibonaccis can also appear inside the

strongly connected components, like the *rcsB-adiY* 3-FF in the pH system (Fig. 3b, second row).

## II. CLASSIFICATION OF FIBERS

In summary, there are two types of fibers in the genetic network studied according to the topology of the input-tree. Input-trees with integer branching ratio  $n$  and input-trees with irrational branching ratio which are generalized Fibonacci sequences with golden ratios  $\varphi_d$ . The fiber numbers define a vector  $|n, \ell\rangle$  and  $|\varphi_d, \ell\rangle$  and classify fibers through topological classes of input-trees in these two general classes, respectively. Fiber number  $n$  defines the branching ratio of infinitely long perfect  $n$ -ary trees in the universal input-tree of the fiber. It directly quantifies the loops inside the fiber, since each loop can be traversed an infinite number of times, leading to an infinite chain or tree in the universal input-tree. For an integer-branched ratio input-tree, fiber number  $\ell$  defines the number of finite chains in the input set coming from the external regulators. The branching ratio  $n$  defines the fundamental class. The finite chain  $\ell$  fiber number defines the subclass.

Fibonacci input-trees are fractals and are characterized by the golden ratio  $\varphi$  and generalized golden ratios,  $\varphi_d$  for  $d \geq 3$ . The main feature of these structures is that beyond a cycle inside the fiber (an autorregulation loop), there is a cycle (of length  $d$ ) traversing the external regulators to the fiber going through the strongly connected component. We classify the Fibonacci fibers with a vector  $|\varphi_d, \ell\rangle$ , where the main fiber number is the fractal branching ratio  $\varphi_d$ ,  $d$  is the length of the cycle going through the strongly connected component external to the fiber, and  $\ell$  is the number of external regulators to the fiber. If there are more than one cycle, then the building block is defined using the shortest cycle. The Fibonacci golden ratio corresponds to the *exuR* circuit with  $d = 2$ . In this case, the cycle has length  $d = 2$ . It means that the fiber directly connects to the regulator and the regulator connects back to the fiber. This structure gives rise to a Fibonacci sequence with the golden ratio  $\varphi_2 = \varphi = (1 + \sqrt{5})/2$ . The other circuits are generalized Fibonacci circuits with  $d = 3$  and 4. In these circuits the fiber sends information to a strongly connected components to which the regulator of the fiber belongs. The path of length  $d$  is the shortest path through the connected component.

The isomorphism is defined only in terms of input-trees and not the output trees since

synchronization of nodes in a fiber depends only on the information received through the inputs of the nodes, but not on the output-trees. Thus, the output-tree does not contribute to the dynamical synchronization of nodes in the fiber. This is true unless that the output of a node is connected to the input through a path in the network. This case corresponds to the Fibonacci fibers. Symmetry permutations that preserve the input and output trees are automorphisms which form a permutation symmetry group [13–15] as opposed to isomorphisms in fibrations which form groupoids [12].

We notice that, by definition, the input-tree contains the information incoming from all nodes including the strongly connected component. In the circuits shown in Fig 1, for instance, the *BaeR-spy* fiber is connected to *cpxR* which does not receive any link from any component. Thus, we say that *cpxR* forms its own isolated connected component.

There are, however, fibers that receive information from strongly connected components larger than one gene, such as the Fibonacci fiber *evgA-nhaR* in Fig. 3b that forms a 4-FF because it receives (and also sends) information from the pH strongly connected component. In this case, the input-tree contains the paths that participate in the connected component as shown in Fig. 3b, on the right. In general, the Fibonacci fibers may receive information from multiple paths in the connected component as shown in the generic case of the d-FF in Fig. 3b, bottom. Furthermore, an important theorem proven by Norris [22] demonstrates that it suffices to find an isomorphism up to  $N - 1$  levels, where  $N$  is the number of nodes in the circuit. Thus, the proof of isomorphism between the nodes in the fibers needs only a finite number of levels in the input tree.

The fibration formalism is related to PageRank, Markov chains, and random walk processes [23], since if a graph  $G$  contains two nodes with the same PageRank, then the nodes are in the same fiber of some color-preserving fibration. Although the converse is not necessarily true. That is, two nodes in the same fiber, may not have the same PageRank. This is because PageRank takes into account only the immediate input set of nodes while fibers take into account isomorphisms between the entire input-tree. Therefore, fibers identified by symmetry fibrations can tell us more about the information processing of network nodes than the PageRank metric.

It is worth to discuss further the generalized golden ratio in the context of spectral analysis, and provide the rationale for its calculation as the maximum eigenvalue of the adjacency matrix of the base network, that is also called the network's spectral radius.

Given a directed network, whether representing signals or connectivities between nodes, we often would like to know how to quantify the information transmitted from a starting node  $i$  to a terminal node  $j$  as a function of the number of links traversed in sequence from  $i$  to  $j$ . This function is called the generating function of paths from  $i$  to  $j$ , denoted as  $F_{ij}(z) = \sum_{\ell=0}^{\infty} n_{\ell} z^{\ell}$ , and defined as

$$F_{ij}(z) = \frac{\text{cof}(I - zA)_{ij}}{\det(I - zA)} , \quad (3)$$

so  $n_{\ell}$  represents the total number of paths from node  $i$  to node  $j$  of length exactly  $\ell$ . The largest power  $L$  in the generating function  $F_{ij}(z) = \sum_{\ell=0}^L n_{\ell} z^{\ell}$  is finite if and only if there are no loops in the network (no loops  $\rightarrow L < \infty$ ); otherwise  $L = \infty$  if the network contains loops.

The function  $F_{ii}(z)$  represents the generating function of paths starting from and ending at  $i$ , and the function  $F(z) = \sum_{i=1}^N F_{ii}(z)$  represents the generating function of all the loops in the network, which is explicitly given by

$$F(z) = \text{tr}(I - zA)^{-1} = \sum_{i=1}^N \frac{1}{1 - z\lambda_i} , \quad (4)$$

where  $\lambda_i$  are the eigenvalues of the adjacency matrix. Function  $F(z)$  has poles at points  $z$  such that  $z_i = 1/\lambda_i$ , and the branching ratio of the network is given by the largest eigenvalue  $\lambda_{\max}$ , which also gives the generalized golden ratio of the universal input-tree of the Fibonacci fiber.

The largest eigenvalue of the adjacency matrix also characterizes certain properties of network dynamics. For example, in the particular case of the Kuramoto model,  $\lambda_{\max}$  marks the position of the phase transition separating the random state and the synchronized state. More precisely, the Kuramoto's critical interaction strength  $\gamma_c$  is inversely proportional to the spectral radius of the network, that is  $\gamma_c \propto 1/\lambda_{\max}$ .

### III. MULTI-LAYER COMPOSITE FIBERS

Building blocks can also be combined to make a composite (like prime numbers or quarks can be combined to form natural numbers or composite particles like protons and neutrons). The ability to assemble fibration building blocks to make larger composites is important in that it helps to investigate and understand systematically higher order functions of biological

systems composed of many genetic elements. We discovered a particular type of composite made up of two fibration building blocks, that we named multi-layer composite fiber. For instance, the double layer *add-oxyS* fiber in the *crp-fis* connected component (see Fig. 3b and ID# 7 in SM Table V and Supplementary File 1) is a composite  $|add-oxyS\rangle = |0, 1\rangle \oplus |1, 1\rangle$ . This composite is of interest since it allows for information to be shared between two genes, for instance *add* and *oxyS*, which are distant in the network.

This composite satisfies a simple engineering 'sum-rule': two elementary fibers (building blocks) are layered in series in a predefined order where the first layer is represented by an entry fiber (carrying transcription factors), and the second layer is formed by a terminator fiber of output genes, as shown in Fig. 3b. This multi-layer composite fiber is biologically significant because genes in the output layer may further regulate and synchronize otherwise uncoordinated genetic modules that implement different functions. Indeed, they resemble the building blocks of multilayered deep neural network where each subsequent gene in the layer synchronizes despite the fact that nodes can be distant in the network. More generally, composite fibers with multiple layers streamline the construction of larger aggregates of fibration building blocks performing more complex function in a coordinated fashion. These 'anomalous' topologies complete the classification of input-trees.

#### IV. FIBRATION LANDSCAPE ACROSS BIOLOGICAL NETWORKS, SPECIES AND SYSTEM DOMAINS

To study the applicability of fibration symmetries across domains of complex networks we have analyzed 383 publically available datasets. Full details of each network and results can be accessed at [https://docs.google.com/spreadsheets/d/1-RG5vR\\_EGNPqQcnJU8q3ky10pWi30jTh5Uo-Xa0Pj0c/edit#gid=1973910256](https://docs.google.com/spreadsheets/d/1-RG5vR_EGNPqQcnJU8q3ky10pWi30jTh5Uo-Xa0Pj0c/edit#gid=1973910256). The codes to reproduce this analysis are at [github.com/makselab](https://github.com/makselab) and the full datasets at [kcorelab.org](https://kcorelab.org) (see also SM Section XII). We first analyze biological networks spanning from transcriptional regulatory networks, metabolic networks, cellular processes networks and pathways, disease networks, and neural networks. The classification in terms of fiber numbers  $|n, \ell\rangle$  allows to systematically study the count of fibers across all datasets,  $N(n, \ell)$ . We find a varied set of fibers that characterize the biological landscape whose distributions are displayed in Fig. 4a. The fiber distributions across species are shown in Fig. 4b spanning *A. thaliana*, *E. coli*, *B.*

*subtilis*, *S. enterica* (*salmonella*), *M. tuberculosis*, *D. melanogaster*, *S. cerevisiae* (yeast), *M. musculus* (mouse), and *H. sapiens*. Our analysis allows to investigate the specific attributes and commonalities of the fiber building blocks inside and across domains. Certain features of the fiber number distribution are visible in the transcriptional networks. For instance, a tail with  $\ell$  is seen in the  $n = 0$  class as well as the features of the  $n = 1$  class, and a prevalence of more complex multi-layer and Fibonacci building blocks. Across species (Fig. 4b), bacteria like *E. coli* or *B. subtilis* display  $n = 0$  building blocks, while yeast, mouse and humans also display complex building blocks like multi-layers and Fibonaccis.

To test the existence of fibrations across other domains we extend our studies to complex networks ranging from social, infrastructure, internet, software, economic networks and ecosystems (SM Section XII). Figure 4c shows the obtained fiber distributions for each domain. A normalized comparison across domains is visualized in Fig. 4d showing the cumulative number of fibers over domains per network size of  $10^4$  nodes. The results support the applicability of the concept of fibration beyond biology to describe the building blocks of networks across different domains.

## V. RELATION BETWEEN THE FIBRATION LANDSCAPE AND THE STRUCTURE OF STRONGLY CONNECTED COMPONENTS

The global organization across different domains and species raises the question of the dependence of the fibration landscape of the network on their organization at the level of strongly connected components. It has been shown that the structure of gene regulatory networks have a peculiar organization at the global level and tend to show a multi-layer hierarchical structure with depressed feedback regulation at the transcription level [24, 25]. For instance, regulatory networks show strongly connected components much smaller than what would be expected by chance for a random network of the same size and degree distribution [26]. Then, we analyze networks across domains to understand how the reduction achieved by the fibration depends on the number and size of strongly connected components in  $G$ .

We characterize the reduction achieved by the fibers and the strongly connected component by defining associated reduction ratios. To characterize the dependence of the reduction achieved by the fibration on the number of strongly connected components we plot in Fig. 4e

the ratio  $r_f = 1 - N_f/N$  versus  $r_{\text{scc}} = 1 - N_{\text{scc}}/N$  over all studied networks, where  $N_f$  is the number of fibers and  $N_{\text{scc}}$  is the number of strongly connected components in each network (in both cases we include components and fibers of size one). The resulting phase diagram divides the networks in two domains. One domain where the reduction is more pronounced by collapsing the fibers ( $r_f > r_{\text{scc}}$ ). The other domain is where the reduction achieved by collapsing each strongly connected component into a single node is more pronounced ( $r_f < r_{\text{scc}}$ ).

Results shown in Fig. 4e indicate that the majority of networks (77% of them, specially, most of the transcriptional networks) exhibits more reduction by collapsing fibers into the base network as compared to the reduction by strongly connected components (23% of them). Most of the metabolic networks show more reduction by the strongly connected components, perhaps due to the large number of loops characteristic of these networks. Results for random network model (Erdős-Rényi) show that they have ( $r_{\text{scc}} = 1, r_f = 0$ ), since random networks are characterized by a single dominant giant connected component and no symmetries, i. e., no fibers

The dependence of the fiber reduction on the size of strongly connected components is characterized by plotting  $r'_f = n_f/N$  versus  $r_{\text{scc}} = n_{\text{scc}}/N$ , where  $n_f$  and  $n_{\text{scc}}$  are the number of nodes participating in fibers and strongly connected components larger than one, respectively. Figure 4f shows the phase diagram over all networks confirming the results of Fig. 4e. The results are consistent with a global network organization with a few and relatively small strongly connected components feeding information to a large number of fibers. This is the global architecture seen in the transcriptional network of *E. coli* with three strongly connected components across the network and a set of fibers receiving information from those components. This structure is also in agreement with the hierarchical organization found in Ref. [25, 26] for gene regulatory networks.

## VI. GENE CO-EXPRESSION ANALYSIS

The existence of isomorphic input-trees has important consequences for the dynamics in the network. Since symmetric genes in a fiber receive the same information through the pathways in the network, they synchronize their activity to produce gene co-expression levels that sustain cellular functions. This equivalence is understood as a local in-isomorphism

between nodes and links. Indeed, dynamical system theory shows that symmetries in the network leads to synchronization [12–20]. We corroborate this result numerically in Fig. 1g in the particular example of the *baeR-spy* fiber.

We also test the synchronization of gene expression levels inside the fibers with publically available transcription profile experiments available from the literature. We use gene expression data profiles in *E. coli* compiled at Ecomics <http://prokaryomics.com> [27]. This portal collects microarray and RNA-seq experiments from different sources such as the NCBI Gene Expression Omnibus (GEO) public database [28] and ArrayExpress [29] under different experimental growth conditions. The data is also compiled at the Colombos web portal [30]. The database contains transcriptome experiments measuring the expression profiles of 4,096 genes in *E. coli* strains over 3,579 experimental conditions which are described as: strain, medium, stress, and perturbation. Raw data is pre-processed to obtain expression levels by using noise reduction and bias correction normalized data across different platforms [27].

*E. coli* can adapt its growth to the different conditions that finds in the medium. This adaptation is made by sensing extra and intracellular molecules and using them as effectors for transcription factors. Thus, the different subcircuits and fibers are activated by specific experimental conditions. The Ecomics portal allows to obtain those experimental conditions where a set of genes has been significantly expressed. We perform standard gene expression analysis (see [colombos.net](http://colombos.net) and Ref. [30]) of the expression levels in *E. coli* obtained under these conditions.

For a given set of genes, we find the experimental conditions for which the genes have been significantly expressed by comparing the expression samples over the 4,096 different biological conditions. Following [30], the experimental conditions are ranked with the inverse coefficient of variation (ICV) defined as  $ICV_i = |\mu_i|/\sigma_i$ , when  $\mu_i$  is the average expression level of the genes in the condition  $i$  and  $\sigma_i$  is the standard deviation. Following [30], we select those conditions with  $ICV_i > 1$ , i.e., where the average expression levels in the particular condition  $i$  are higher than the standard deviation. This score reflects the fact that, in a relevant condition, the genes show an increment on their expression above the individual variations caused by random noise. Details on the expression analysis can be found at Ref. [30] and at <https://doi.org/10.1371/journal.pone.0020938.s001>. The obtained expression profiles are organized by the experimental conditions which are labeled according

to the GEO database [28]. From these data, we calculate the co-expression matrix using the Pearson correlation coefficient between the expression levels of two genes  $n$  and  $m$  in the relevant conditions:

$$C(n, m) = \frac{1}{p} \sum_{i=1}^p \left( \frac{x_{n,i} - \mu_n}{\sigma_n} \right) \left( \frac{x_{m,i} - \mu_m}{\sigma_m} \right),$$

where  $p$  is the number of conditions,  $x_{n,i}$  is the expression value of gene  $n$  at condition  $i$ , and  $\mu_n$  and  $\sigma_n$  are the respective mean and standard deviation of the expression levels of gene  $n$  over the  $p$  conditions.

Results for the correlation matrix are shown in Fig. 2a bottom for fibers regulated by the *crp-fis* strongly connected component. Gene expression is obtained for every gene, so we plot the correlation matrix calculated over each pair of genes. Genes that belong to the same operon are transcribed as a single unit, so these genes are expected to synchronize beyond the fact that they belong to a fiber, although with variations due to attenuators. Thus, we group together these genes as operons in the figure. The proper test of fiber synchronization is achieved by comparing gene co-expression belonging to different operons. We find that the expression levels of the genes that belong to the same fiber are highly correlated as predicted by the fibration symmetry. Genes that belong to different fibers show no significant correlations among them. In particular, there is no significant correlation between the expression of genes in a given fiber and the two master regulators *crp* and *fis*. This result is consistent with the fibration symmetry and occurs despite the fact that both, *crp* and *fis*, directly regulates all genes in the studied fibers.

Thus, our findings identify the fiber building blocks that predict synchronization in correlated co-expression profiles and identify functionally related genes from the symmetries of the genetic network. Fibrations make sure that genes are turned on and off at the right amount to assure the synchronization of expression levels in the fiber needed to execute cellular functions. Results shown elsewhere [31] indicate that symmetries also factorize according to function in simple connectomes. Thus, fibration symmetries may help to systematically organize the building blocks of biological networks in general from the invariances in the information flow encoded by the topologies of the constitutive input-trees that lead to the synchronization needed for cellular function.

### **Acknowledgments**

Research was sponsored by NIH-NIGMS R01EB022720, NIH-NCI U54CA137788/U54CA132378 and NSF-IIS 1515022. We thank L. Parra, A. Holodny and P. Casaccia for discussions. FM and HAM designed the research, developed the theory, performed the analysis and wrote the article, and IL performed the network analysis.

- 
- [1] Hartwell, L. H., Hopfield, J. J., Leibler, S. & Murray, A. W. From molecular to modular cell biology. *Nature* **402**, C47-C52 (1999).
  - [2] Alon, U. *An Introduction to Systems Biology: Design Principles of Biological Circuits* (CRC Press, Boca Raton, 2006).
  - [3] Gell-Mann, M. *The Quark and the Jaguar* (Holt Paperbacks, New York, 1994).
  - [4] Dixon, J. D. & Mortimer, B. *Permutation Groups*, Graduate Texts in Mathematics, 163 (Springer-Verlag, New York, 1996).
  - [5] Weinberg, S. *The Quantum Theory of Fields* (Cambridge University Press, Cambridge, 2005).
  - [6] Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. Network motifs: simple building blocks of complex networks. *Science* **298**, 824-827 (2002).
  - [7] Buchanan, M., Caldarelli, G., De Los Rios, P. Rao, F. & Vendruscolo M. (editors), *Networks in Cell Biology* (Cambridge University Press, Cambridge, 2010).
  - [8] Shen-Orr, S. S., Milo, R., Mangan, S. & Alon, U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genet.* **31**, 64-68 (2002).
  - [9] Gama-Castro, S. *et al.*, RegulonDB version 9.0: high-level integration of gene regulation, coexpression, motif clustering and beyond. *Nucleic Acids Res.* **44**, D133-D143 (2016). <http://regulondb.ccg.unam.mx>, last accessed September 15, 2018.
  - [10] Grothendieck, A. Technique de descente et théorèmes d'existence en géométrie algébrique, I. Généralités. Descente par morphismes fidélement plats. Séminaire N. Bourbaki **5**, Talk no. 190, p. 299-327 (1958-1960). [http://www.numdam.org/article/SB\\_1958-1960\\_\\_5\\_\\_299\\_0.pdf](http://www.numdam.org/article/SB_1958-1960__5__299_0.pdf), <https://ncatlab.org/nlab/show/Grothendieck+fibration>
  - [11] Boldi, P. & Vigna, S. Fibrations of graphs. *Discrete Mathematics* **243**, 21-66 (2001). <http://vigna.di.unimi.it/ftp/papers/FibrationsOfGraphs.pdf>, <http://vigna.di.unimi.it/fibrations>
  - [12] Golubitsky, M. & Stewart, I. Nonlinear dynamics of networks: the groupoid formalism. *Bull. Amer. Math. Soc.* **43**, 305-364 (2006). <http://www.ams.org/journals/bull/2006-43-03/S0273-0979-06-01108-6/S0273-0979-06-01108-6.pdf>
  - [13] Abrams, D. M., Pecora, L. M. & Motter, A. E. Focus issue: Patterns of network synchronization. *Chaos* **26**, 094601 (2016).

- [14] Pecora, L. M., Sorrentino, F., Hagerstrom, A. M., Murphy, T. E. & Roy, R. Cluster synchronization and isolated desynchronization in complex networks with symmetries. *Nature Comm.* **5**, 4079 (2014).
- [15] Sorrentino, F., Pecora, L. M., Hagerstrom, A. M., Murphy T. E. & Roy, R. Complete characterization of the stability of cluster synchronization in complex dynamical networks. *Sci. Adv.* **2**, e1501737 (2016).
- [16] Stewart, I., Golubitsky, M. & Pivato, M. Symmetry groupoids and patterns of synchrony in coupled cell networks. *SIAM J. Appl. Dyn. Syst.* **2**, 609-646 (2003).
- [17] Arenas, A., Díaz-Guilera, J. K. A., Moreno, Y. & Zhou, C. Synchronization in complex networks. *Phys. Rep.* **469**, 93-153 (2008).
- [18] Rodrigues, F. A., Peron, T. K., Ji, P. & Kurths, J. The Kuramoto model in complex networks. *Physics Reports* **610**, 1-98 (2016).
- [19] Strogatz, S. *Nonlinear Dynamics and Chaos: with Applications to Physics, Biology, Chemistry, and Engineering* (Westview Press, Boulder, 2000).
- [20] Kamei, H. & Cock, P. J. A. Computational of balanced relations and their lattice for a coupled cell network. *SIAM J. Appl. Dyn. Syst.* **12**, 352-382 (2013).
- [21] Cardon, A. & Crochemore, M. Partitioning a graph in  $O(|A| \log_2 |V|)$ . *Theoretical Computer Science* **19**, 85-98 (1982).
- [22] Norris, N. Universal covers of graphs: isomorphism to depth  $n - 1$  implies isomorphism to all depths. *Discrete Applied Mathematics* **56**, 61-74 (1995).
- [23] Boldi, P., Lonati, V., Santini, M. & Vigna, S. Graph fibrations, graph isomorphism, and PageRank. *RAIRO Inform. Théor.* **40**, 227-253 (2006).
- [24] Ma, H.-W., Buer, J. & Zeng, A.-P. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory revealed by a new top-down approach. *BMC Bioinformatics* **5**, 199 (2004).
- [25] Yu, H. & Gerstein, M. Genomic analysis of the hierarchical structure of regulatory networks. *Proc. Nat. Acad. Sci. US* **103**, 14724-14731 (2006).
- [26] Jothi, R., Balaji, S., Wuster, A. Grochow, J. A., Gsponer, J., Przytycka, T. M., Aravind, L. & Madan Babu, M. Genomic analysis reveals a tight link between transcription factor dynamics and regulatory network architecture. *Molecular Systems Biology* **5**, 294 (2009).
- [27] Kim M. *et al.* Multi-omics integration accurately predicts cellular state in unexplored condi-

- tions for Escherichia coli. *Nat. Commun.* **7**, 13090 (2016).
- [28] Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M. *et al.* NCBI GEO: archive for functional genomics data sets— update. *Nucleic Acids Res.* **41**, D991-D995 (2016).
- [29] Kolesnikov, N., Hastings, E., Keays, M., Melnichuk, O., Tang, Y. A., Williams, E., Dylag, M., Kurbatova, N., Brandizi, M., Burdett, T. *et al.* ArrayExpress update: simplifying data submissions. *Nucleic Acids Res.* **43**, D1113-D1116 (2015).
- [30] Moretto, M., Sonego, P., Dierckxsens, N., Brilli, M., Bianco, L., Ledezma-Tejeida, D., Gama-Castro, S., Galardini, G., Romualdi, C., Laukens, C., Collado-Vides, J., Meysman, P. & Engelen, K. COLOMBOS v3.0: leveraging gene expression compendia for cross-species analyses. *Nucleic Acids Res.* **44**, D620-D623 (2016). <http://colombos.net>
- [31] Morone, F. & Makse, H. A. Symmetry group factorization reveals the structure-function relation in the *Caenorhabditis elegans* connectome. Submitted (2018).
- [32] Boldi, P. & Vigna, S. An effective characterization of computability in anonymous networks. In J. L. Welch, editor, Distributed Computing. 15th International Conference, DISC 2001, number 2180 in *Lecture Notes in Computer Science*, pp 33-47 (SpringerVerlag, 2001).
- [33] Boldi, P. & Vigna, S. Universal dynamic synchronous self-stabilization. *Distr. Comput.* **15**, 137-153 (2002).
- [34] Brandt, H. Über eine Verallgemeinerung des Gruppenbegriffes. *H. Mathematische Annalen* **96**, 360-366 (1927). <http://eudml.org/doc/159172>
- [35] Weinstein, A. Groupoids: unifying internal and external symmetry. A tour through some examples. *Notices Amer. Math. Soc.* **43**, 744-752 (1996). <https://arxiv.org/abs/math/9602220>
- [36] Pradines, J. In Ehresmann’s footsteps: from group geometries to groupoid geometries (2007). <https://arxiv.org/pdf/0711.1608.pdf>
- [37] DeVille, L. & Lerman, E. Modular dynamical systems on networks. *J. European Mathematical Society* **17**, 2977-3013 (2015). <https://arxiv.org/abs/1303.3907>
- [38] Klipp, E., Liebermeister, W., Wierling, C. & Kowald A. Systems Biology (Wiley-VCH, Weinheim, 2016)
- [39] Medina-Rivera, A., Abreu-Goodger, C., Thomas-Chollier, M., Salgado, H., Collado-Vides, J. & van Helden, J. Theoretical and empirical quality assessment of transcription factor-binding

- motifs. *Nucleic Acids Res.* **39**, 808-824 (2011).
- [40] Hertz, G. Z., Hartzell, G. W. & Stormo, G. D. Identification of consensus patterns in unaligned DNA sequences known to be functionally related. *Comput. Appl. Biosci.* **6**, 81-92 (1990).
- [41] Kaplan, S., Bren, A., Zaslaver, A., Dekei, E. & Alon, U. *Diverse two-dimensional input-functions control bacterial sugar genes*. *Mol. Cell.* **29**, 786-792 (2008).
- [42] Coram, D. S. & Duvall, P. F. Approximate fibrations. *Rocky Mountain Journal of Mathematics* **7**, 275-288 (1977).

**FIG. 1. Definition of input-tree, fibration, fiber and base.** **a**, The circuit controlled by the *cpxR* gene regulates a series of fibers as shown by the different colored genes. The circuit regulates more genes represented by the dotted lines which are not displayed for simplicity and do not change the results. The full lists of genes and operons in this circuit are in SM Table V, ID=27, 28 and 54. **b**, The input-tree of representative genes involved in the *cpxR* circuit showing the isomorphisms that define the fibers. For each fiber, we show the number of paths of length  $i - 1$  at every layer of the input-tree,  $a_i$ , and its branching ratio  $n$ . **c**, Isomorphism between the input-trees of *baeR* and *spy*. The input-trees are composed of an infinite number of layers due to the autoregulation loop at *baeR* and *cpxR*. How to prove the equivalence of two input-trees when they have an infinite number of levels? An important theorem proven by Norris [22] demonstrates that it suffices to find an isomorphism up to  $N - 1$  levels, where  $N$  is the number of nodes in the circuit. Thus, in this case, 2 levels are sufficient to prove the isomorphism. **d**, Fibration  $\psi$  transforms the *cpxR* circuit  $G$  into its base  $B$  by collapsing the genes in the fibers into one. **e**, Fibration of the *fadR* circuit and **f**, its isomorphic input-trees. Full list of genes in this circuit appears in SM Table V, ID=3, 4, and 58. **g**, Symmetric genes in the fiber receive the same information through the pathways in the network and, therefore, synchronize their activity to produce activity levels leading to a common cellular function. We use the mathematical model of gene regulatory dynamics of Boolean kinetics from Ref. [8] (sigmoidal interactions lead to qualitatively similar results) to show the synchronization inside the fiber *baeR-spy* when the fiber is activated by its regulator *cpxR*. Notice that *cpxR* does not synchronize with the fiber. The result that fibers lead to synchronization has been proven in [12–16, 20].

**FIG. 2. Strongly connected components of the genetic network and synchronization of gene co-expression in the fibers.** **a**, We find three main connected components, beside a number of single gene components like the *cpxR* circuit in Fig. 1a and *fadR* in Fig. 1e. A (strongly) connected component is formed by genes that can be reached by a path from any other gene in the component (irrespective of the type of links, either repressor or activator). **a, Top**, Two-gene connected component of *crp-fis*. This component controls a set of fibers as shown, but the fibers do not belong to the connected component since they do not send back information to the connected component. We highlight the fiber *uxuR-lgoR* which sends information to its regulator *exuR* and forms a 2-Fibonacci fiber  $|\varphi_2 = 1.6180.., \ell = 2\rangle$ , as well as the double-layer composite  $|add - oxyS\rangle = |0, 1\rangle \oplus |1, 1\rangle$ .

**a, Bottom.** Co-expression (correlation) matrix calculated from the Pearson coefficient between the expression levels of each pair of genes. Synchronization of the genes in the respective fibers appears as the block structure of the matrix. **b,** The core of the network is the connected component formed by genes involved in the pH system as shown. This component supports two Fibonacci fibers: 3-FF and 4-FF.

FIG. 3. **Classification of building blocks.** **a, Typical fiber building blocks.** These building blocks are characterized by a fiber that does not send back information to its regulator. They are characterized by two integer fiber numbers:  $|n, \ell\rangle$ . We show selected examples of circuits and input-trees. The full list of fibers appears in SM Table V and Supplementary File 1. The statistical count of every class is in SM Table IV and more examples of each class in SM Tables VI, VII and VIII. The last example shows a generic building block for a general n-ary tree  $|n, \ell\rangle$  with  $\ell$  regulators. **b, Fibonacci building blocks.** These building blocks are more complex and characterized by an autoregulated fiber that sends back information to its regulator. This creates a fractal input-tree that encodes a Fibonacci sequence with golden branching ratio in the number of paths  $a_i$  versus path length,  $i - 1$ . When the information is sent to the connected component that includes the regulator, then a cycle of length  $d$  is formed and the topology is a generalized Fibonacci block with golden ratio  $\varphi_d$  as indicated. We find three such building blocks: 2-FF, 3-FF and 4-FF.

FIG. 4. **Fiber landscape across domains and species.** **a, Fiber landscape for biological networks.** Total number of fiber building blocks across 5 types of biological networks analyzed in the present work. The count includes the total number of fibers in the networks of each type. **b, Fiber landscape across species.** Count of fibers across species. **c, Fiber landscape across domains.** Count of fibers across the major domains studied in the main text. The biological domain panel includes all networks in **a**. **d, Global fiber landscape.** Cumulative count of fibers in all domains in **c**. The cumulative count represents the total number of fibers per network of  $10^4$  nodes. Specifically, the quantity is calculated as the total number of fibers divided by the total number of nodes in all networks per domain multiplied by  $10^4$ . **e, Reduction by number of fibers vs strongly connected components (SCC).** Reduction ratios in terms of number of fibers and number of SCC, and **f**, in terms of size of fibers and size of SCC as defined in the text.

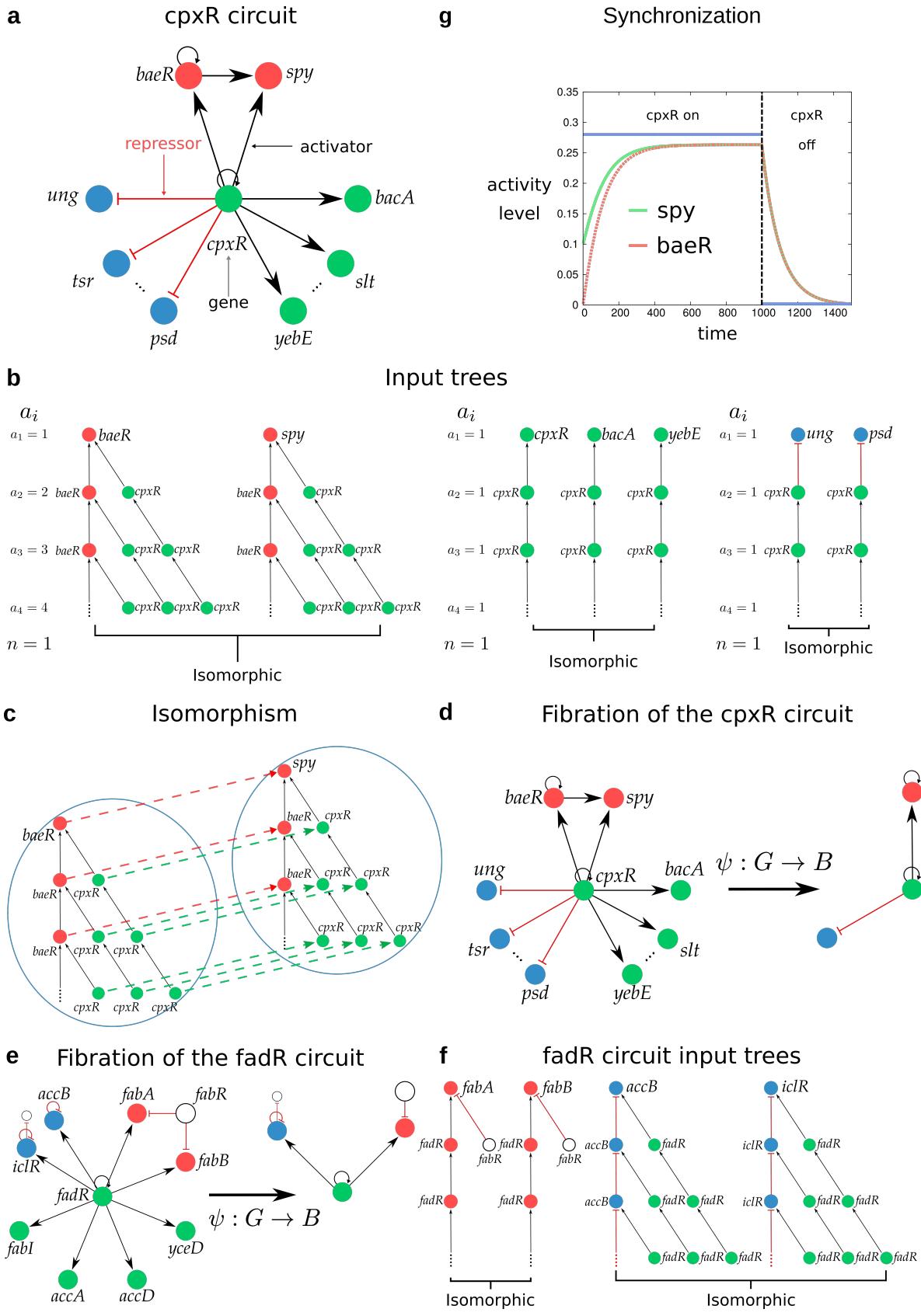


FIG. 1:

## 2a cpr-fis connected component

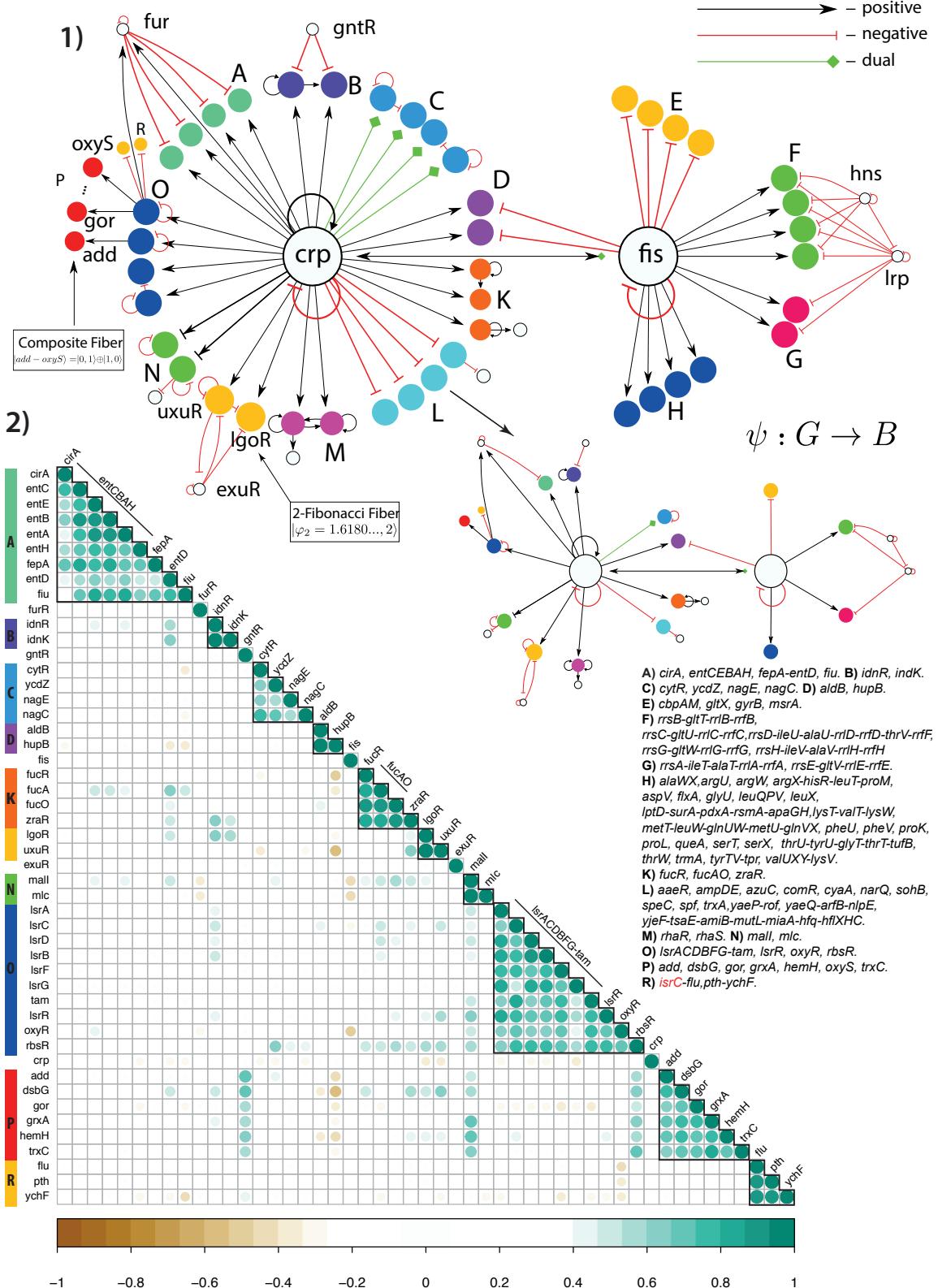


FIG. 2: a

## 2b pH connected component

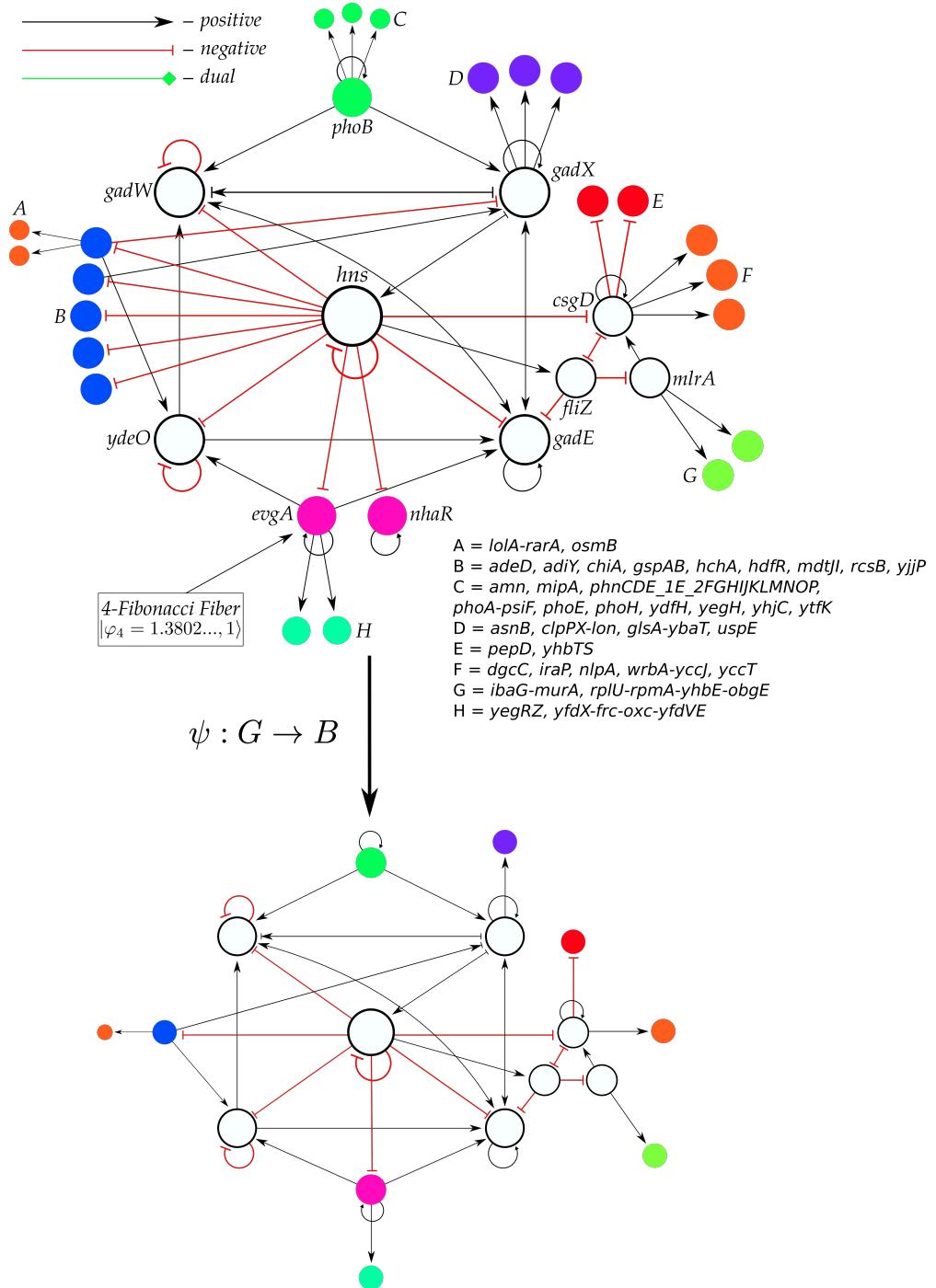


FIG. 2: b

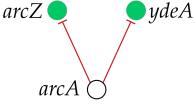
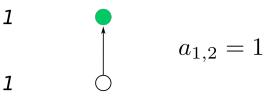
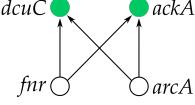
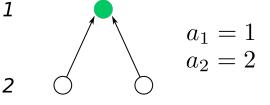
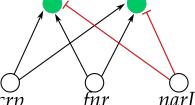
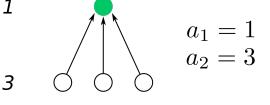
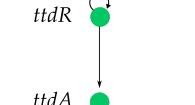
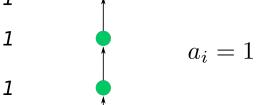
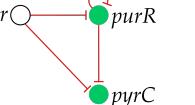
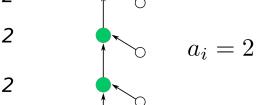
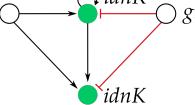
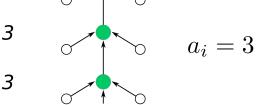
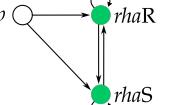
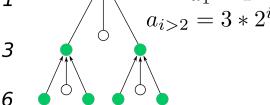
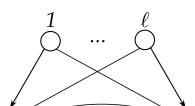
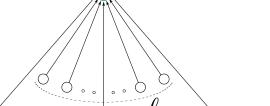
$ n, \ell\rangle$	Genetic circuit	Input tree
$ 0, 1\rangle$		 $a_i$ $1$ $1$ $a_{1,2} = 1$
$ 0, 2\rangle$		 $a_i$ $1$ $2$ $a_1 = 1$ $a_2 = 2$
$ 0, 3\rangle$		 $a_i$ $1$ $3$ $a_1 = 1$ $a_2 = 3$
$ 1, 0\rangle$		 $a_i$ $1$ $1$ $1$ $a_i = 1$
$ 1, 1\rangle$		 $a_i$ $2$ $2$ $2$ $a_i = 2$
$ 1, 2\rangle$		 $a_i$ $3$ $3$ $3$ $a_i = 3$
$ 2, 1\rangle$		 $a_i$ $1$ $3$ $6$ $a_1 = 1$ $a_{i>2} = 3 * 2^{i-1}$
$ n \geq 3, \ell\rangle$		 $a_1 = 1$ $a_{i>2} = (n + l) * n^{i-1}$

FIG. 3: a

$ \varphi_d, \ell\rangle$	Genetic circuit	Input tree
$ 1.6180..., 2\rangle$	<p>2-Fibonacci Fiber (2-FF)</p>	<p><math>a_i</math></p> $a_i = a_{i-1} + a_{i-2}$ <p>A = <math>uxuR</math>, B = <math>lgoR</math>, C = <math>exuR</math>, D = <math>crp</math></p>
$ 1.4655..., 1\rangle$	<p>3-Fibonacci Fiber (3-FF)</p>	<p><math>a_i</math></p> $a_i = a_{i-1} + a_{i-3}$ <p>A = <math>rcsB</math>, B = <math>adiY</math>, C = <math>gadX</math>, D = <math>hns</math></p>
$ 1.3802..., 1\rangle$	<p>4-Fibonacci Fiber (4-FF)</p>	<p><math>a_i</math></p> $a_i = a_{i-1} + a_{i-4}$ <p>A = <math>evgA</math>, B = <math>nhaR</math>, C = <math>gadE</math>, D = <math>gadX</math>, E = <math>hns</math></p>
$ \varphi_d, \ell\rangle$	<p>d-Fibonacci Fiber (d-FF)</p>	<p><math>a_i</math></p> $a_i = a_{i-1} + a_{i-d}$

FIG. 3: b

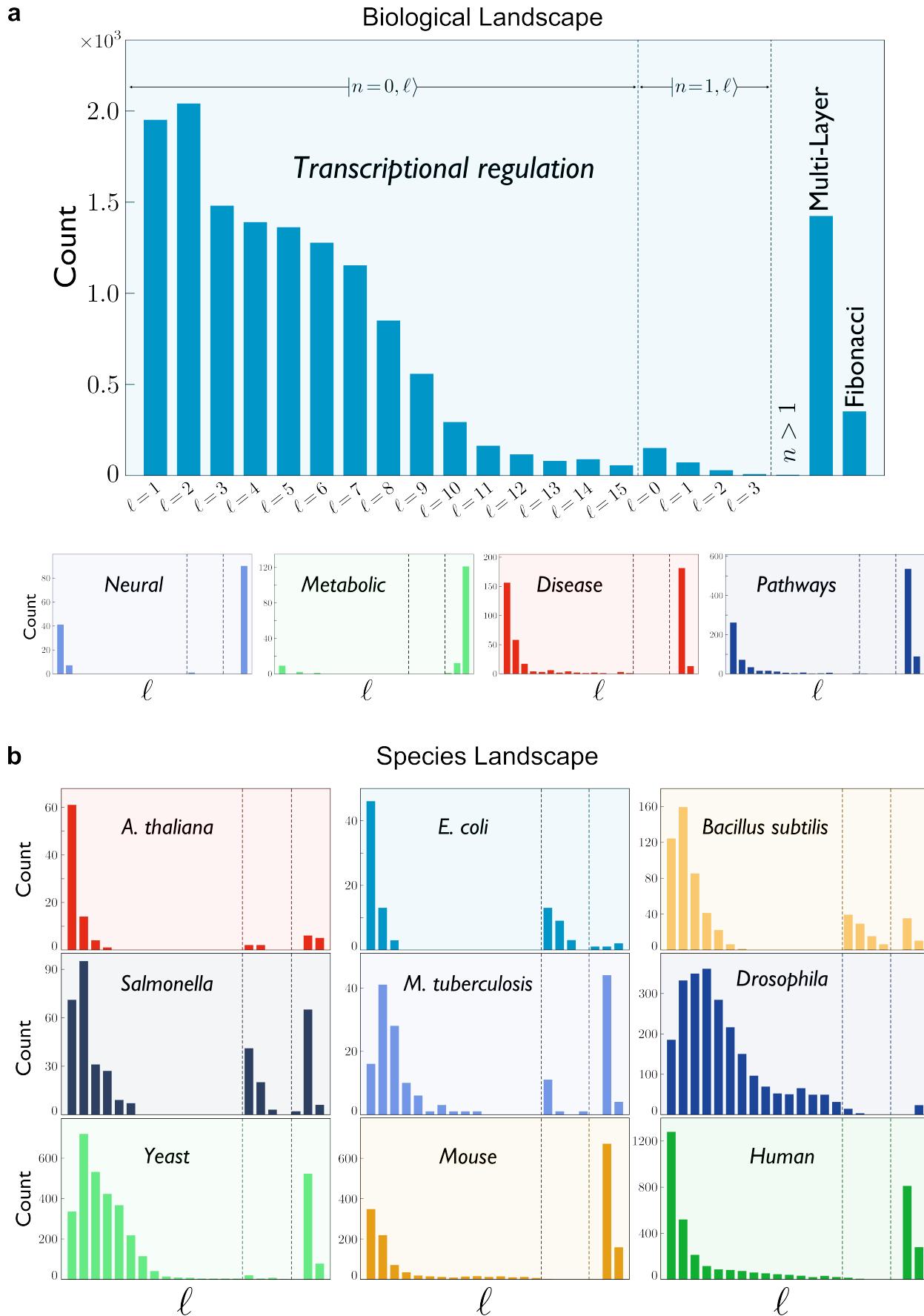


FIG29: a, b

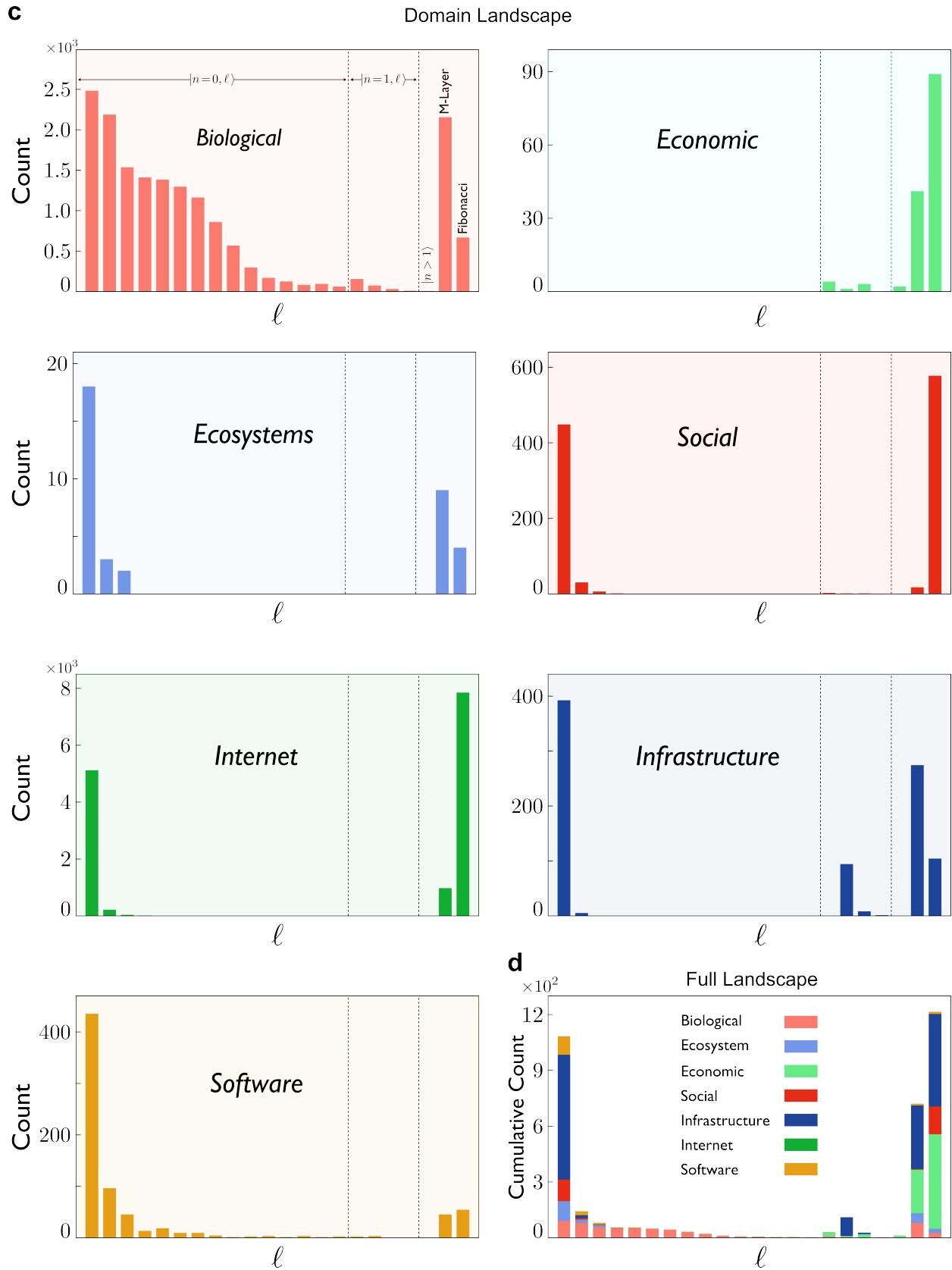
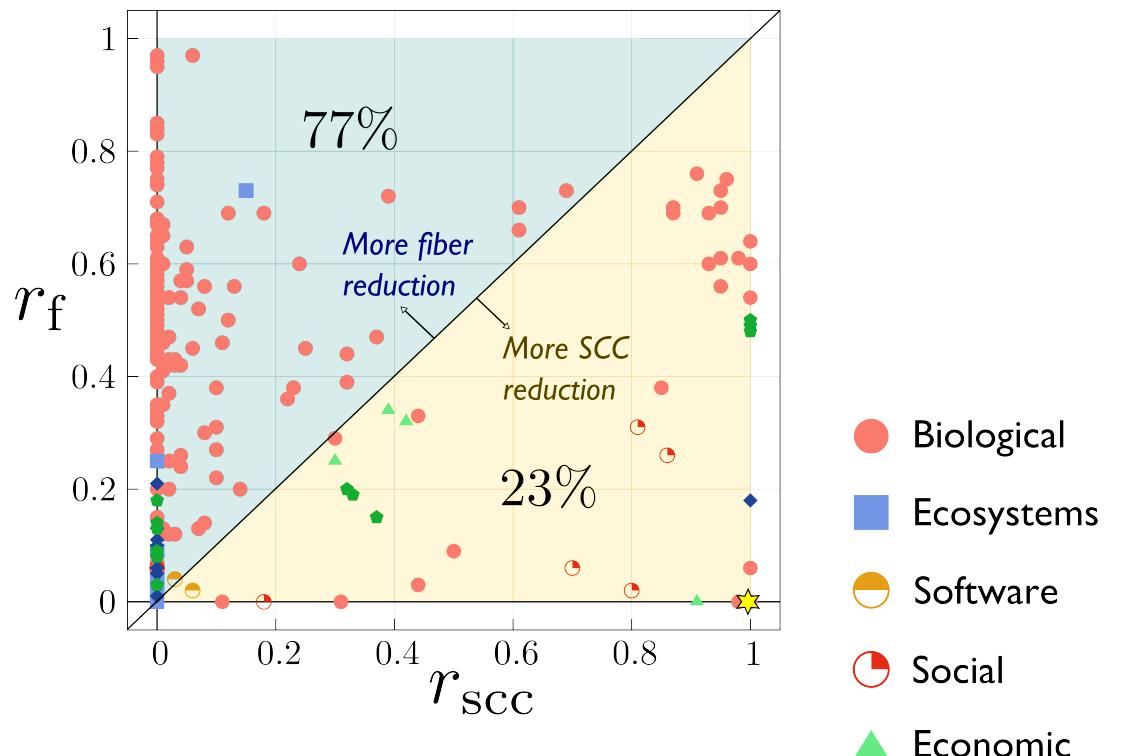


FIG. 4: c, d

**e** Reduction by number of fibers vs SCC



**f** Reduction by size of fibers vs SCC

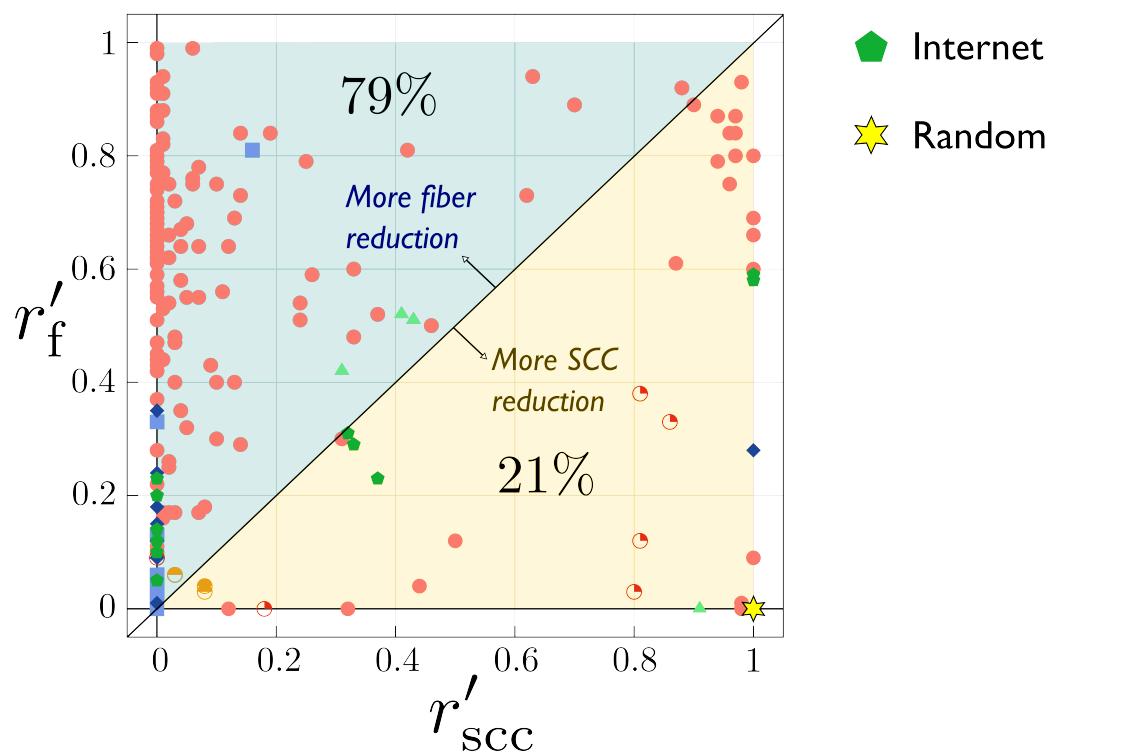


FIG. 4: e, f

## Supplementary Materials

### Fibration building blocks of information-processing networks

**Flaviano Morone, Ian Leifer, Hernán A. Makse**

#### **Contents**

<b>I. Formal definition of graph fibration</b>	5
<b>II. Classification of fibers</b>	9
<b>III. Multi-layer composite fibers</b>	11
<b>IV. Fibration landscape across biological networks, species and system domains</b>	12
<b>V. Relation between the fibration landscape and the structure of strongly connected components</b>	13
<b>VI. Gene co-expression analysis</b>	14
<b>References</b>	18
<b>VII. Transcriptional regulatory network of <i>E. coli</i></b>	34
<b>VIII. Minimal fibration</b>	34
A. Universal input-tree	37
B. Isomorphic universal input-trees	39
C. From isomorphic input-trees to fibrations	41
<b>IX. Influence of heterogeneities and randomness in fibrations in gene regulatory networks</b>	42
<b>X. Strongly connected component</b>	46
<b>XI. Algorithm to find fibers of minimal graph fibrations</b>	46
<b>XII. Datasets of biological and non-biological networks</b>	50

<b>XIII. Genetic building blocks and their fiber numbers</b>	52
A. Fibers statistics	54
B. Full list of fibers	54
C. Examples of fibers with integer fiber numbers	55

## VII. TRANSCRIPTIONAL REGULATORY NETWORK OF *E. COLI*

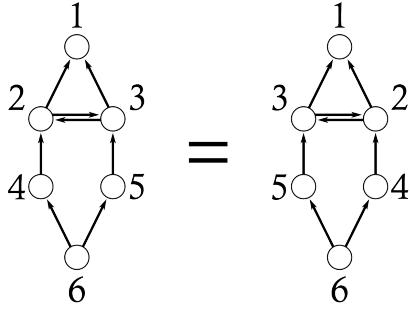
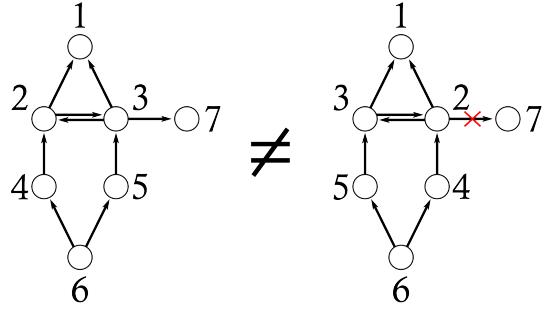
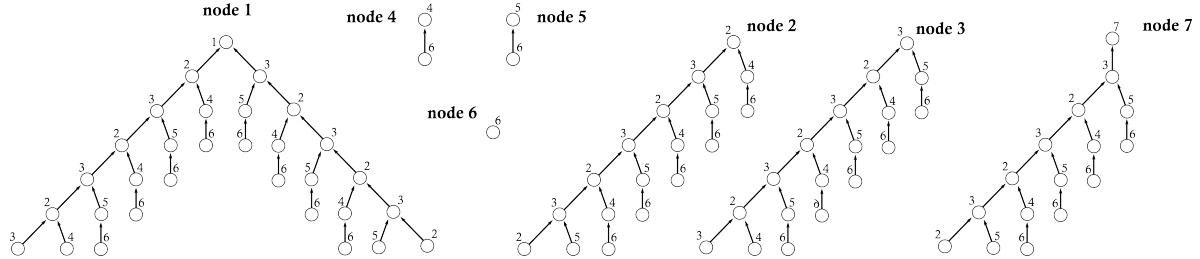
We use the transcription factor-gene target network in *Escherichia coli* K-12 obtained from the RegulonDB data source (<http://regulondb.ccg.unam.mx>) that manually curates all transcriptional regulations from literature searches [9]. We download all transcriptional regulatory interactions catalogued in RegulonDB version 9.0 from [http://regulondb.ccg.unam.mx/menu/download/datasets/files/network\\_tf\\_gene.txt](http://regulondb.ccg.unam.mx/menu/download/datasets/files/network_tf_gene.txt), last accessed September 15, 2018.

The database downloaded from RegulonDB is composed of a bipartite transcription factor - gene target network. A transcription factor (TF) is a protein encoded by the source gene that binds to a binding site of the target gene to regulate its rate of transcription. A TF can either promote (activator) or inhibit (repressor) the transcription from DNA to mRNA by the target gene. A TF can also regulate the transcription of its own gene, forming an autoregulation loop. Each TF is encoded by a gene. Therefore, a gene-gene regulatory network can be constructed from the bipartite transcription factor-gene target network where a directed link from gene 1 → gene 2 implies that gene 1 encodes for a TF that controls the rate of transcription of gene 2. Thus, we say that gene 1 sends a genetic 'message' to gene 2. The histories of all messages passing in the network defines the information flow in the network.

In the transcriptional regulatory network from RegulonDB there are 2014 nodes (genes). There are 192 genes that encode for TFs. Some of the genes are organized into operons. An operon is a set of genes that are transcribed as a single unit [9]. We group genes into operons. To simplify notation, in the circuits of Figs. 1, 3a and 3b, we call each operon by the name of one of the genes that belongs to the operon. Full names appear in SM Table V. In Fig. 2, the legends indicate the full name of the operons that includes all genes in the operon. There are 313 operons in the dataset. After grouping the genes into operons, the network is reduced to 880 nodes with 1764 edges.

## VIII. MINIMAL FIBRATION

Below we provide formal definitions of the main concepts developed in the main paper: input-trees, isomorphisms, fibrations, minimal fibration, fibers and bases. We start with a

**A** Network with a Symmetry Group**B** Network with NO Symmetry Group**C** Input trees

**FIG. 5: Group symmetries and fibrations with their input-tree.** **a,** Example of a network with a symmetry group. The automorphism shown maps the network into another network leaving invariant the connectivity of every nodes in the network [4, 12–14]. **b,** A network without automorphisms but with a fibration. The addition of a single out-link from  $3 \rightarrow 7$  breaks the whole group symmetry. However, since fibrations are defined according only to the input-tree, then the network still have a symmetry, a fibration arising from the fact that the input-trees of nodes 2 and 3 are isomorphic, as well as between the input-trees of nodes 4 and 5 as shown in **(c)**. There are no more isomorphisms as shown by the rest of the input-trees. Therefore, nodes 2 and 3 form a fiber. Nodes 4 and 5 also form another fiber, yet independently of the other fiber. The fibration is a morphism that maps the network into a base which is formed by collapsing the isomorphic nodes into one, i.e., collapsing node 2 and 3 together, and node 4 and 5 together. The resulting base is also called a quotient graph.

review of the literature (not exhaustive).

The literature on fibrations and groupoids crosses the fields of mathematics, computer science and dynamical systems theory. The notion of fibration was first introduced by Grothendieck as fibrations between categories [10]. This original formulation

of Grothendieck is highly abstract and differs from our present work which refers to the notion of graph fibration. The original paper of Grothendieck has been published as a part of the Séminaire N. Bourbaki in 1958 and can be found at [http://www.numdam.org/article/SB\\_1958-1960\\_\\_5\\_\\_299\\_0.pdf](http://www.numdam.org/article/SB_1958-1960__5__299_0.pdf). A mathematical account of Grothendieck fibrations in the context of category theory appears in <https://ncatlab.org/nlab/show/Grothendieck+fibration>. The work of Boldi and Vigna [11] on graphs fibrations is one of the works that is the closest to our formulation. See <http://vigna.di.unimi.it/ftp/papers/FibrationsOfGraphs.pdf>. For a review of the history of graph fibrations from Grothendieck to modern studies, see the blog of Vigna at <http://vigna.di.unimi.it/fibrations/>. It should be noted that all these works on fibrations pertain to a highly abstract mathematical level which, in turn, provides the concept of fibration with a quite broad applicability. For a more accessible reading on fibrations within the particular context application to biological networks, the reader is recommended to follow our paper and supplementary sections.

Graph fibrations have been also applied in computer science to understand the ranking algorithm PageRank [23], and to understand the state of synchrony of processors in computing distributed systems [32, 33]. In parallel, the work of Golubitsky and Stewart [12, 16] and collaborators in dynamical systems theory considers the equivalent formalism of symmetry groupoids of coupled cell networks. A review of the groupoid formalism in dynamical systems appears in [12]. An axiomatic abstract account of groupoids is found in [34–36]. DeVille and Lerman [37] also discuss the relation between graph fibrations and the groupoid formalism.

To elaborate on the definition of network fibrations, we can compare fibrations to well-known symmetry automorphisms which form symmetry groups [4, 12–14] using the example networks of Figs. 5a and 5b. The network of Fig. 5a contains the automorphism defined by:

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 3 & 2 & 5 & 4 & 6 \end{pmatrix}, \quad (5)$$

because the nodes are connected exactly to the same nodes before and after the application of the permutation. An automorphism thus preserves the full connectivity of the network. That is the automorphism preserves the input and output tree, such that automorphisms are more difficult to realized. This might explain by group symmetries is not the framework to describe biology. Instead, the fibration symmetries, with its emphasis in the preservation

of the input-trees only, might be the symmetry that characterize information-processing systems, while automorphisms are more appropriate for inanimate matter.

Next, consider the slightly modified network depicted in Fig. 5b left, which differs from the network in Fig. 5a by one extra out-going link from node 3 to 7. In this network, the permutation of nodes  $2 \leftrightarrow 3$  and  $4 \leftrightarrow 5$ , Eq. (5), is not an automorphism anymore, because it does not preserve the in and out connectivities of all nodes, e.g., node 3 is connected with 7 but loses this connection after the permutation (Fig. 5b right). It is interesting to see how fragile group symmetries are: if we connect just one extra node to the network as shown in Fig. 5b, the symmetry (i.e. the network automorphism group) is broken. This occurs because automorphisms require very strict arrangements of nodes and links to preserve, rigidly, the global structure of the network.

This example raises the following question: are there extra symmetries in the network shown in Fig. 5b beyond its automorphisms? The answer to this question is, indeed, yes: there are extra symmetries in the network of Fig. 5b, called fibrations [10, 11], which do not form a group [4] but a groupoid. A groupoid is a set of transformations satisfying the axioms of invertibility, identity and associativity but not closure [12, 16], while in a group, transformations satisfy the four axioms. For this reason, groupoids are fundamentally different algebraic structures compared with traditional group symmetries.

### A. Universal input-tree

Roughly speaking, fibrations take into account only the input-trees of the nodes, but not the output-trees. Thus, node 3 is connected to node 7 via an out-going link, and this link destroys the symmetry group, but this node is still symmetric with 2 via a fibration, since the input-trees of nodes 2 and 3 are isomorphic, even though node 3 is connected with 7. This is because the connection  $3 \rightarrow 7$  is an out-going link of node 3 and, therefore, is not part of its input-tree. Simply put, fibrations preserve input-trees only, while automorphisms preserve both input and output-trees, since they preserve the full connectivity of the network, and thus, they represent more stringent symmetries than fibrations. We formalize this idea shortly, after introducing some definitions.

The basic ingredient to define fibrations of a network  $G$  is the **universal input-tree**, which contains the full information received by a given node through the totality of all the

possible paths ending in that node and starting from every other node in the network. Thus, for every node  $i$  in the network  $G$  there is a corresponding input-tree, called  $T_i$ , which is defined as a tree with a selected node  $r_i$ , called the root, and such that every other node is a path  $\mathcal{P}_{j \rightarrow i}$  of  $G$  starting from  $j$  and ending in  $i$  [11, 37]. A link from node  $\mathcal{P}_{j \rightarrow i}$  to node  $\mathcal{P}_{k \rightarrow i}$  exists if  $\mathcal{P}_{j \rightarrow i} = e_{j \rightarrow k} \mathcal{P}_{k \rightarrow i} =$ , where  $e_{j \rightarrow k}$  is an edge of  $G$ .

For example, let us construct the input-tree  $T_2$  of node 2 in the network on the left of Fig. 5b. The root is the node  $r_2$  at the uppermost level of the tree. Every other node of the input-tree of node 2 is a path  $\mathcal{P}_{j \rightarrow 2}$  ending in 2. There are two paths of length 1:  $\mathcal{P}_{3 \rightarrow 2}^{(1)}$  and  $\mathcal{P}_{4 \rightarrow 2}^{(1)}$ ; three paths of length 2:  $\mathcal{P}_{2 \rightarrow 2}^{(2)}$ ,  $\mathcal{P}_{5 \rightarrow 2}^{(2)}$ , and  $\mathcal{P}_{6 \rightarrow 2}^{(2)}$ ; and so on. Since  $\mathcal{P}_{2 \rightarrow 2}^{(2)} = e_{2 \rightarrow 3} \mathcal{P}_{3 \rightarrow 2}^{(1)}$ , we put a link in the input-tree from  $\mathcal{P}_{2 \rightarrow 2}^{(2)}$  to  $\mathcal{P}_{3 \rightarrow 2}^{(1)}$  because  $\mathcal{P}_{2 \rightarrow 2}^{(2)} = e_{2 \rightarrow 3} \mathcal{P}_{3 \rightarrow 2}^{(1)}$ . We then add all other links in the input-tree using the same criterion. The resulting input-tree  $T_2$  is shown in Fig. 5c, together with the input-trees of all other nodes in the network in Fig. 5b.

In general, we label each node of  $T_i$  using the starting point of the corresponding path  $\mathcal{P}_{j \rightarrow i}$ . For example, in  $T_2$  nodes  $\mathcal{P}_{3 \rightarrow 2}^{(1)}$  and  $\mathcal{P}_{4 \rightarrow 2}^{(1)}$  are labeled 3 and 4 respectively, and the length of the path is equal to the depth of the node in the input-tree.

Thus, in practice, we arrive at the following way to construct the input-tree: we start with the node at the root, lets say node 2. We label every node  $\mathcal{P}_{j \rightarrow 2}$  in the input-tree by node  $j$  where the path starts. The first layer of the input-tree consists of all the nodes that are at a distance one from the root. In this case, nodes 3 and 4. Thus we add two links to 2 from 3 and 4 in the input-tree.

The second layer of the input-tree is obtained applying the same procedure to each node in the first layer, 3 and 4. For instance, node 3 receives a link from 2 and 5. Therefore the second layer of the input-tree contains nodes 2 and 5 connected to node 3. We repeat the procedure with the other node in layer 2: node 4. Node 4 receives a link only from node 6, and node 6 from no one. So, we add a link from 6 to 4 and this path does not propagate further. The third layer of the input-tree is obtained iteratively applying the same procedure, and so on.

We note that the input-trees of nodes 1, 2, 3 and 7 are infinite since the network contains a cycle (or loop) between nodes  $2 \rightleftharpoons 3$ . For instance,  $T_1$  is infinite because there are paths crossing the loop infinite times. On the other hand, the input-trees of nodes 4, 5 and 6 are finite since they do not cross the loop.

## B. Isomorphic universal input-trees

The universal input-tree  $T_i$  at node  $i$  can be interpreted as the collection of all possible ‘histories’ starting at some node and ending in node  $i$ . The important result is the following: if two input-trees  $T_i$  and  $T_j$  are isomorphic, then the corresponding nodes  $i$  and  $j$  in the network  $G$  receive the same information and their current dynamical state is the same [11]. This equivalence is understood in terms of a local in-isomorphism as defined in Boldi & Vigna [11] (see Section I) that maps nodes to nodes and links to links, so it formalizes the fact that the dynamical interactions represented by a directed link from gene to gene could be in principle different across genes, as long as the links are the same (or similar, in case that the produced synchronization is approximate) inside the fiber. We elaborate on this point in SM Section IX.

An isomorphism between  $T_i$  and  $T_j$  is defined as a bijective map  $\tau : T_i \rightarrow T_j$ , which maps one-to-one the nodes and edges of  $T_i$  to nodes and edges of  $T_j$ .

A minimal condition for the existence of an isomorphism between the input-trees is that the two input-trees have the same number of nodes (we could also add a condition of the same degree sequence). Thus, it is clear that there could be no isomorphism between the input-trees of nodes 2 and 4, since the former contains an infinite number of nodes and the latter just two. Thus, a minimal condition for an isomorphism to exist is that it should be a mapping between two input trees with the same number of nodes, since the mapping needs to be bijective, i.e., with an inverse. By inspection it is then clear that there is an isomorphism between the input trees of nodes 4 and 5. This isomorphism is the map  $\tau_{4 \rightarrow 5} : T_4 \rightarrow T_5$ , and it is written as a transformation following the notation:

$$\tau_{4 \rightarrow 5} = \begin{pmatrix} 4 & 6 \\ \downarrow & \downarrow \\ 5 & 6 \end{pmatrix}, \quad (\text{isomorphism between input-trees of nodes 4 and 5}). \quad (6)$$

which maps the root of  $T_4$  to the root of  $T_5$  as  $\tau_{4 \rightarrow 5}(4) = 5$ , and node  $6 \in T_4$  to node  $6 \in T_5$  as  $\tau_{4 \rightarrow 5}(6) = 6$ . In this particular example the links are of the same type, so there is no need to specify the type of links in the isomorphism, but in general the local equivalence require that nodes are map to nodes and also links to the same type of link.

The map in Eq. (6) is one of the simplest isomorphism since the input-tree contains only one level. In this particular case, to see that nodes  $T_4$  and  $T_5$  are isomorphic, it is thus

enough to see that both nodes 4 and 5 connect to one and the same node, which is node 6 in this case. That is, both input-trees of nodes 4 and 5 are isomorphic because they are made up of just two nodes and one edge, and this isomorphism implies that 4 and 5 receive the same information. This is the simplest form of an isomorphism between input-trees. In this case, we say that node 4 and 5 have the same *input-set*, which is an input-tree of only one level, that is the set of incoming links. The input-set formalism is developed in Ref. [12].

Next, we consider the input-trees of nodes 2 and 3. By visual inspection, both input-trees have the same ‘shape’. However, these trees are infinite in the number of levels. How do we decide if two input-trees are isomorphic when they have an infinite number of levels? Remarkably, to determine if two input-trees are isomorphic, it suffices to check that they are isomorphic up to the  $N - 1$  level, thanks to a theorem by Norris [22], where  $N$  is the total number of nodes in the network  $G$ . Since  $G$  has  $N = 7$ , we use six levels in the input trees and, indeed, determine that there is an isomorphism between  $T_2$  and  $T_3$  which corresponds to the following map:

$$\tau_{2 \rightarrow 3} = \begin{pmatrix} 2 : 3 & 4 & 2 & 5 & 6 & 3 & 4 & 6 & \dots \\ \downarrow & \downarrow \\ 3 : 2 & 5 & 3 & 4 & 6 & 2 & 5 & 6 & \dots \end{pmatrix}, \quad (\text{isomorphism between input-trees of 2 and 3}). \quad (7)$$

Here, the notation starts with the root of the tree and then we write nodes in each level from top to bottom starting from left to right in each level. There are no other isomorphism between the other input-trees. Notice that  $T_7$  is not isomorphic to  $T_3$  by just one link to the root.

The existence of isomorphic input-trees implies that the genes receive the same information in the network leading to synchronization in the nodes within the fiber. This result also appear in a number of studies in Refs. [11, 23, 32, 33, 37] and has been applied to the case of processors in distributed systems. DeVille and Lerman has also shown that graph fibrations lead to synchrony subspaces in networks [37]. Analogous work in the dynamical systems community by Golubitsky, Stewart and collaborators [12, 16] and reference therein, shows that synchronization [17–19] is a consequence of groupoid symmetries. These works have shown that symmetry groupoids are analogous to fibers. These results extend analogous work using automorphisms and group theory in dynamical systems showing how automorphisms lead to synchronized nodes in orbits. See for example Refs. [12–16, 20] and references

therein for extensive work on how automorphisms forming symmetry groups lead to synchronization of nodes in orbits. The orbits from automorphisms are the fibers in fibrations. For a review on symmetry groupoids and synchronization see [12].

In general, the existence of an isomorphism  $\tau$  from the input-tree of node  $i$  to the input-tree of node  $j$  implies the synchronization of  $x_i$  and  $x_j$  [12]. This, again, implies the local equivalence at the node and link level. Simply put, two nodes are synchronized if their input-set are synchronized, too. Synchronous nodes in a fiber induced by the fibrations in the notation of Ref. [11] correspond to a coloring of nodes that are balanced in the notation of Ref. [12]. The balanced coloring assigns two nodes  $i$  and  $j$  the same color only if their inputs, self-consistently, receives the same content of colored nodes, whence the term ‘balanced’. The nodes with the same color in the balanced coloring procedure in Ref. [12] are the same as the fibers in the fibration notation used here. A balanced coloring procedure in a network is obtained using the algorithm explained in SM Section X proposed in the paper by Cardon and Crochemore [21].

### C. From isomorphic input-trees to fibrations

If a network  $G$  has at least one pair of isomorphic input-trees, then there exists a network  $B$ , called the **base** of  $G$ , such that  $G$  can be ‘fibered’ over  $B$ , as explained next. The base  $B$  is defined as follows:

- a node  $I \in B$  is a representative of the set of nodes  $\{i \in G\}$  whose input-trees are isomorphic;
- an edge  $e_{I \rightarrow J}$  where  $I, J \in B$  is defined as  $e_{I \rightarrow J} = \sum_{i \in I} e_{i \rightarrow j}$ , where  $e_{i \rightarrow j} \in G$ .

Having defined the base network  $B$ , we say that  $G$  is fibered over  $B$  if there exists a surjective morphism  $\psi : G \rightarrow B$ , called **graph fibration** [10, 11], that maps nodes and edges of  $G$  to nodes and edges of  $B$  as:  $\psi(i) = I$  for all  $i \in I$ , and  $\psi(e_{i \rightarrow j}) = e_{I \rightarrow J}$ . A surjective morphism is a map between two sets (the domain and codomain) where each element of the codomain (in this case  $B$ ) is mapped to, at least, by one element of the domain (in this case  $G$ ). The set of nodes  $i \in G$  that are mapped to the same node  $I \in B$ , and denoted by  $\phi^{-1}(I)$ , is called the **fiber** of  $G$  over node  $I$ . We notice that all input-trees of nodes which belong to the same fiber are pairwise isomorphic.

While the fibers of every minimal fibration are the same, minimal fibrations might differ in the way they map links, which in turns provide several different ways to map isomorphically input-trees. This lack of uniqueness of the isomorphism could have consequences to the biological interpretation of the fibers.

In practice, the Grothendieck fibration is the morphism  $\psi$  of a network  $G$  that maps  $G$  to a simplified network, the base  $B$  (also called the quotient), that consists of the following steps: *(i)* consider all the nodes in a fiber (which have isomorphic input-trees) and choose one as the representative  $I$ , *(ii)* collapse the nodes in the fiber into one single node in  $B$  and call it by the name of the representative node  $I$ , *(iii)* for every link of a node  $j$  in  $G$  directed to the node  $I$  in  $G$ , add a link in  $B$  from  $j$  to  $I$ . If the node  $j$  belongs to the fiber, then the corresponding link in  $B$  is an autoregulation loop in  $B$ , *(iv)* repeat for every fiber in  $G$ . This procedure has been applied to the circuits in the main text.

## IX. INFLUENCE OF HETEROGENEITIES AND RANDOMNESS IN FIBRATIONS IN GENE REGULATORY NETWORKS

Cellular interactions are characterized by noise, randomness and are heterogeneous. Below we address the existence of symmetries in genetic network in light of disorder in genetic interactions. In particular, in the binding of transcription factors (TF) to DNA binding sites that determines the expression levels of genes through their response functions.

Heterogeneities in TF-DNA interactions arises due to the fact that the response functions of genes have different amplitude, different half-saturation constants and they could be linear or nonlinear. This implies that the links in the network are not all of the same type. This situation is captured by the formalism of fibration since the fibration involves a morphism that maps nodes to nodes and links to links to preserve the incidence relation or the local in-isomorphism. Thus, the isomorphism is between pairs of nodes and only isomorphic links needs to be the same to satisfy the symmetry fibration. Thus, fibrations impose a less stringent condition between the type of dynamics of each pair of genes than a global equivalence between all genes in the network. Genes need to be symmetric in pairs in the fiber, being only locally equivalent but they can be globally heterogeneous. We next discuss how the requirement of local equivalence in the regulation functions of pair of genes can be approximately fulfilled in gene regulatory networks.

The gene regulation function is effectively determined by the binding properties of the TF to the DNA and, in turn, it is determined by the promoters nucleotide sequence [38]. That is, for a giving TF, the regulation functions depend on the promoter sequence to which the TF binds. A given TF will bind preferentially to promoter sequences that contain well-defined TF-binding motifs. These motifs are used in bioinformatics analysis to determine the binding probability of a TF to DNA through position-specific scoring matrices (PSSM). Such matrices are built by statistics over bindings sites DNA sequences experimentally characterized for a given TF [39, 40].

The fact that a given TF binds to specific TF-binding motifs implies, in the language of graph fibrations, that the outgoing links of a given TF are approximately of the same type. This property implies that the local in-isomorphism can be applied by mapping genes to genes and links to links in the fibers found in our study.

Evidence is seen in the sequence-logo motif analysis of TFs (see for instance, the TrpR repressor for tryptophan biosynthesis in *E. coli* logo analysis in Ref. [39]) showing enriched sequence-logo binding site DNA sequences. Well-defined DNA sequence motifs are seen in the logo patterns for the TrpR binding sites across all the binding site sequences where TrpR binds. The common DNA motif indicates that the response functions for the outgoing links from a TF can be approximately considered to be of the same type. This implies that the fibers can synchronize their activity according to the graph fibration, through the local in-isomorphism between regulated genes.

Further evidence is found in direct measurements of the response functions of TF across different binding sites across regulated genes. In Ref. [41], the response functions of different genes in the sugar catabolism system in *E. coli* have been measured. The response functions in this well-described system are quite diverse, however, this diversity does not affect the existence of fibrations. Synchronization in the fiber requires that the response function of a given TF or a pair of TF that regulate a particular set of genes mainly depend only on the TFs and are less dependent on the target gene. That is, the input-functions over a set of genes that are regulated by a pair of TFs, for instance, are expected to be approximately similar, even though other input-functions of other TFs might be diverse as found in [41]. This is corroborated in Ref. [41] (Fig. 2). For instance, the input-functions in the sugar genes are similar for a particular TF activated by a given external sugar, like arabinose

that activates TF AraC together with the activation of CRP by cAMP. For instance, as seen in Fig. 2 in Ref. [41], the input-functions for the genes *araBAD*, *araE*, *araFGH*, and *araJ* are quite similar; the input functions of the target genes are determined mainly by the TF AraC and CRP. The target genes then receive the same information approximately from the incoming genes *crp* and *araC* through the local in-isomorphism, and, if there is no other gene sending information to these genes, then the genes form a fiber satisfying the symmetry fibration. We note that in this particular case there are other links to the genes and therefore they do not belong form a fiber.

Next, we further address how dynamical heterogeneities can be handled theoretically by perturbation expansion around the symmetric state even if symmetries might be broken by heterogeneous randomness in the links inside the fiber. We show how this case is typical in theoretical physics discoveries, and we show that it can be analytically and computationally treated by perturbation expansion theory around the ideal symmetric case.

Variations from ideal symmetries are ubiquitous in nature. It is the case of all theories that are computable and were discovered in physics including the theory of the Higgs boson. Many computations in physics starts with an ideal symmetry and then proceeds with perturbation theory given by diagrammatic loop expansions around the exact symmetric solution to take account of heterogeneous interactions in real systems. Perturbative techniques are how the majority of real phenomena in physics was discovered. The most venerable case of this approach is the discovery of new hadrons by Murray Gell-Mann.

Indeed, there is no symmetry that is exactly realized in Nature or it is not broken. Even in the SU(3) flavor symmetry of chromodynamics in the Standard Model of Particle Physics, the constituent masses of quarks up, down and strange are all different and are not perfectly symmetric ( $m_u = 336$  MeV and  $m_d = 340$  MeV,  $m_s = 486$  MeV). Despite this asymmetry between the masses, by assuming the perfect symmetry (meaning all the masses are the same) one can put the three quarks in a triplet (which assumes that they can be interchanged by an SU(3) transformation) and predict the existence of all the composites hadrons in the octet of Gell-Mann, like protons, neutrons, and the heavier barions with high precision. The remaining small physical differences in quantum behavior from the asymmetries are then treated with perturbation theory from the ideal symmetric case. This perturbative theoretical approach which is at the root of all physical phenomena to account

for departures from symmetries, from the Standard Model to all quantum mechanics, can be readily translated to describe the cell's departure from ideal symmetry.

The symmetries of real systems are departures of exact symmetries and are treated as perturbations in the interactions between particles, in analogy to the dynamical interactions between genes, which can be quantified and made computable by perturbation theory. Thus, this situation can be translated to the symmetries in the cell which are observed experimentally and therefore point to the efficacy of the perturbation regime.

Thus, variability is an intrinsic property of all physical and biological systems, where it is the main source of phenotypic change among organisms and species. If all links in a biological network were the same, the regularities of living system functions can be traced back in the symmetries of the interaction network. The biological variation that can happen by symmetry breaking from this ideal symmetric state can explain, in the weak symmetry breaking regime from the ideal symmetric state, the biological phenotypic variations. Ideal symmetries must be broken, for evolution to exist, a weakly broken symmetry state would allow for the speciation needed for new functions and species to emerge.

Fibration theory considers this situation in the theory of perturbation for approximate fibrations which has been developed in [42]. The perturbative expansion can be rigorously controlled and the loop corrections can be analytically computed, in principle, at any order (see theorem 2.6 from [42]). Thus, approximate fibrations are computable by rigorous mathematical results in the framework of category theory.

Lastly, the verdict on the applicability of graph fibration theory, as on every physical theory, should be returned, ultimately, on the goodness of its predictions when compared with experimental data and practical applications. Figure 2a shows experimental evidence that the symmetry in the gene regulatory network in *E. coli* is well-satisfied, or the symmetry breaking must be weak because a weak symmetry breaking from the response functions implies a weak breaking of synchronicity. In this regards, we observe in Fig. 2a that fibration theory does predict consistently the gene synchronization patterns in a multitude of different cases.

## X. STRONGLY CONNECTED COMPONENT

In a directed network, the strongly connected component is composed of nodes that are reachable from every other node in the component. That is, there is a directed path from every node to any other node in the strongly connected component. A weakly connected component is obtained when we ignore the directionality of the links. Strongly connected components are relevant to genetic fibers since they contain loops that control the state of the genes. We find four types of strongly connected components. Single-gene component like *cpxR* and *fadR* in Figs. 1a and 1e. The other type of components are those in Fig. 2a and Fig. 2b and also a five-gene connected component shown in SM Fig. 6. We note that the fibers regulated by this component, for instance, do not belong to the connected component. This is because they receive information but do not send information back to the connected component. These fibers are characterized by integer fiber numbers. When the fiber receives and sends back information, that is, when the fiber belongs to the strongly connected component, then it becomes a Fibonacci fiber. The largest strongly connected component controls the pH system shown in Fig. 2b.

## XI. ALGORITHM TO FIND FIBERS OF MINIMAL GRAPH FIBRATIONS

The algorithm to obtain the fibers in a network is based on a balanced coloring algorithm and was developed by Cardon and Crochemore in Ref. [21].

Here we explain the algorithm proposed by Cardon and Crochemore [21] to construct a *minimal* balanced coloring of a network, namely a coloring that employs the least possible number of colors, which is associated with minimal fibrations.

The algorithm to find the fibers computes only minimal balanced colorings, which are associated with minimal fibrations. The algorithm's runtime scales as  $O(M \log_2 N)$  ( $N$  is the number of nodes and  $M$  the number of links), which implies that it is essentially linear with the network size, specially for sparse networks.

As far as only minimal fibrations are considered, the algorithm will return always the same fibers containing the same nodes, for any initial condition and realization. However, while the fibers of every minimal fibration are the same, minimal fibrations might differ in

### soxR-soxS-fnr-fur-arcA connected component

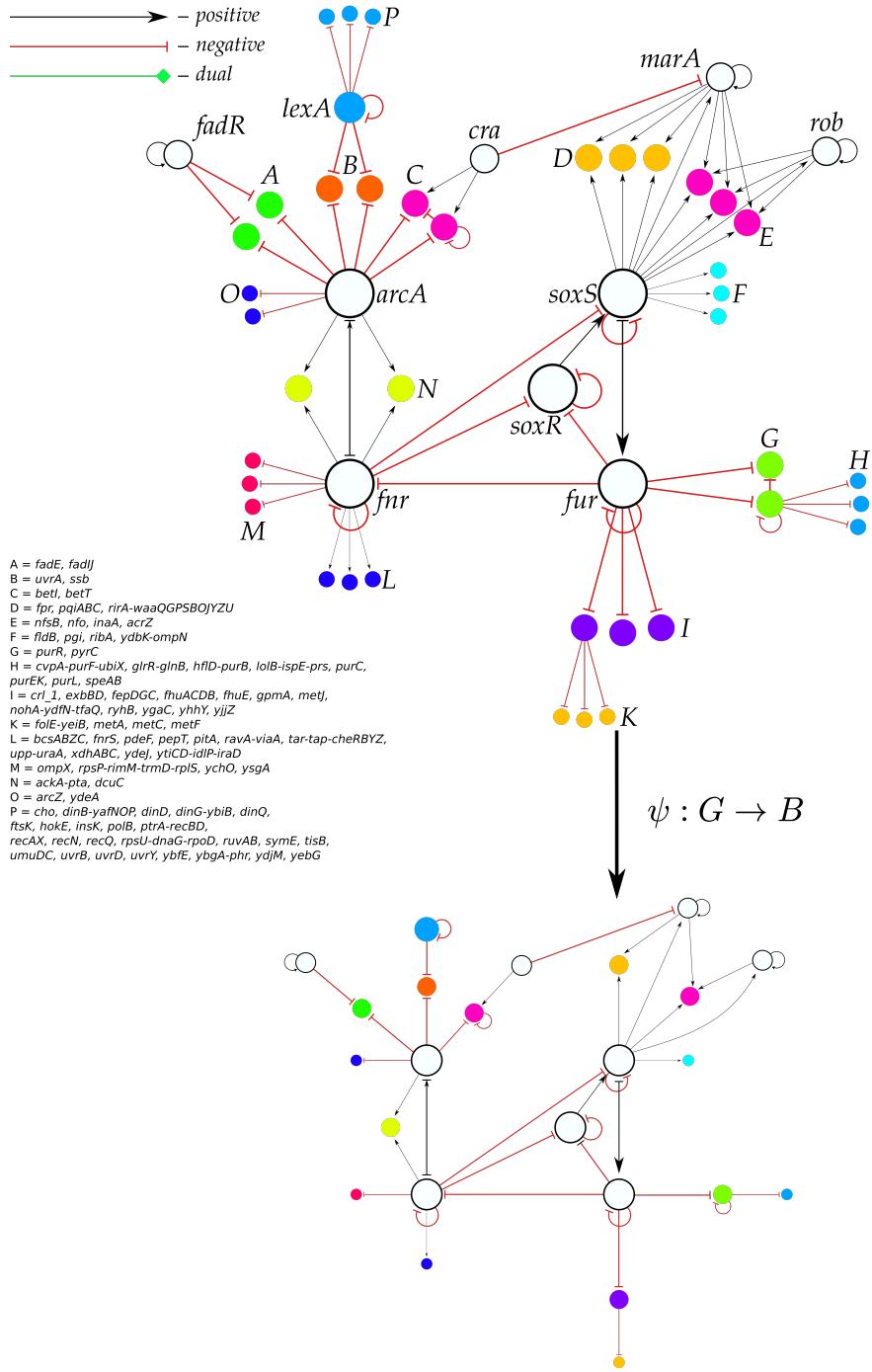


FIG. 6: A five-gene connected component of *soxR*, *soxS*, *fnr*, *fur*, and *arcA* with its regulated fibers.

the way they map links, which in turns provides different way to map isomorphically input trees. However, these different link mappings do not affect the graph partition into fibers.

The flow of information arriving to genes in a fiber is analogous to a process of assigning a color to each gene such that each gene ‘receives’ the colors from adjacent genes via its incoming links and ‘sends’ its color to the adjacent genes via its outgoing links. The nodes in a fiber have the same color symbolizing the fact that they receive the same information. This is called a *balanced coloring* in the framework of Ref. [12]. This process leads to the coloring of nodes in, for instance, the circuits in the main text as well as in the example shown in Fig. 7c explained in the next section. For instance, the genes *baeR* and *spy* (Fig. 1) have the same color and are in the same fiber since they receive the same colors: both *baeR* and *spy* receive one red color via the activator link from one red node (*baeR* from itself and *spy* from *baeR*) and one green activator link each from the green node *cpxR*.

This coloring creates a partition of nodes of  $G$  into disjoint sets such that each node in one set receives the same number of colors from nodes within other sets [12, 16]. A coloring of  $G$  with this property is the *balanced coloring* and represents an *equitable partition* of the network, see [12, 16]. The sets identified by a *balanced coloring* are exactly the fibers of  $G$  identified by minimal graph fibrations  $\psi$  [11, 12, 37].

The algorithm to obtain the fibers in a network is based on a balanced coloring algorithm and was developed by Cardon and Crochemore in Ref. [21]. See also the follow up in [20]. We have to construct a coloring of the nodes that is balanced. A coloring is balanced if two identically colored nodes are connected to identically colored nodes via their inbound links. Here we explain the algorithm proposed by Cardon and Crochemore [21] to construct a *minimal* balanced coloring of a network, namely a coloring that employs the least possible number of colors, which is associated with minimal fibrations. Each balanced colored cluster is a fiber in the network. The flow of the algorithm is exemplified with the example network of Fig. 7.

- **Step 1** - We start by assigning the same color to all nodes. In Fig. 7a all nodes are initially colored in blue. In addition, we assign to each link the same color of the node from where it emanates. To update the coloring (or, equivalently, to generate a new partition) of nodes, we construct the table shown in the right panel of Fig. 7a, as explained next. In the top row of this table we put the network nodes colored with their current color. In the leftmost column we put each type of colored link. In this

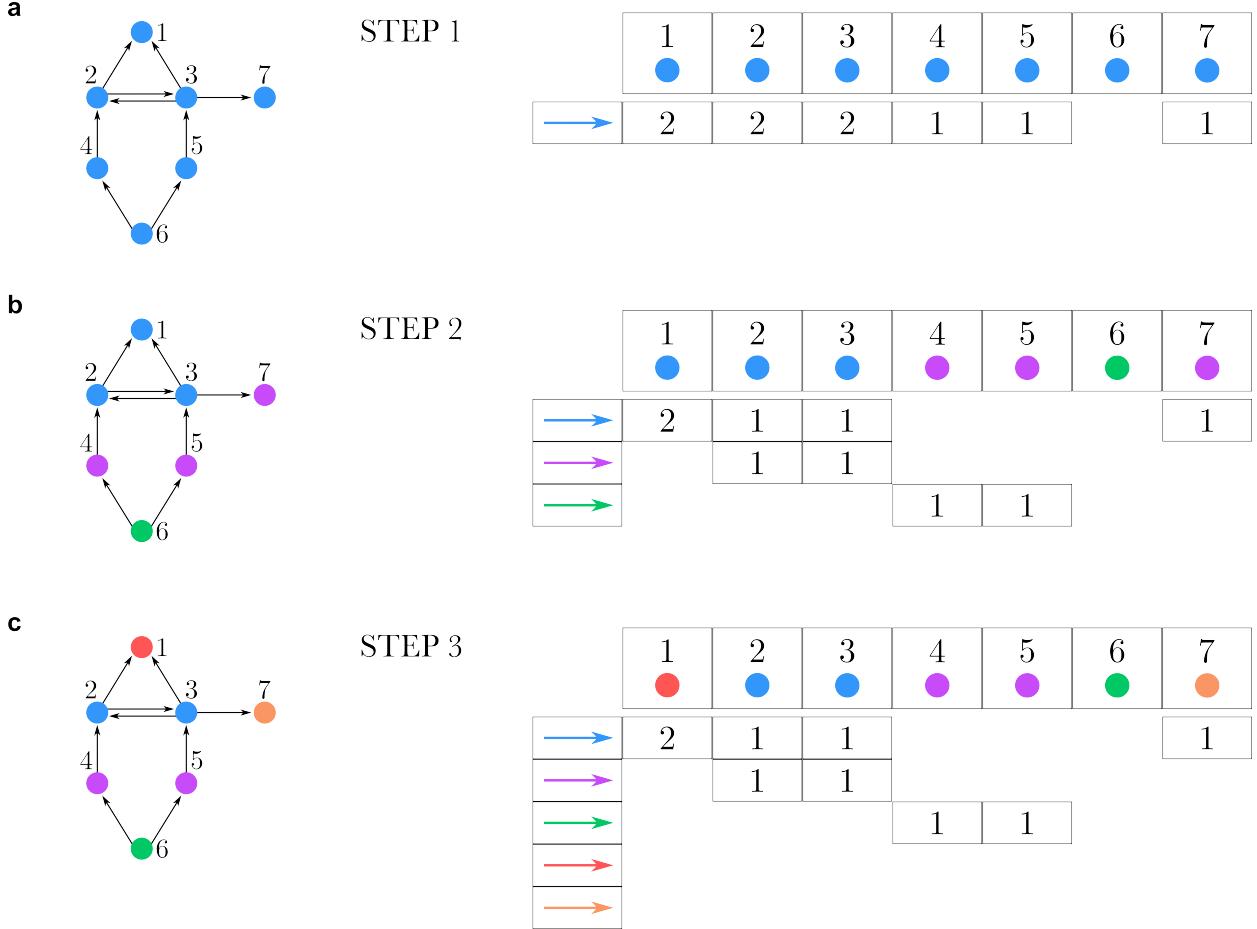


FIG. 7: **Algorithm to find the fibers of a network through a minimal balanced coloring.**

The goal of the algorithm is to find a minimal balanced coloring of the network, so that two nodes have the same color only if they are connected to the same number of identically colored nodes via inbound links. The colors represent the fibers in the network.

initial stage of the algorithm we only have a blue link for all the nodes. Then, we fill the entries of the table with the number of colored links of this blue type that are received by the corresponding node. For example, node 1 receives two 2 blue links as well as nodes 2 and 3. Nodes 4, 5 and 7 receive one blue link each, and node 6 nothing. The structure of this table determines the new coloring as explained in the next step.

- **Step 2** - Using the table in Fig. 7a we update the coloring of nodes as follows. We assign the same color to all nodes that receive the same number of colored links of each type. Specifically, nodes 1, 2 and 3 receive two blue links, so we assign them the same (blue) color. Analogously, nodes 4, 5 and 7 receive one blue link, so we assign them

the same color, but different from blue. We assign them a purple color. Similarly, we assign another color to node 6 (green). We then obtain the colored network in the left of Fig. 7b. Applying the counting of receiving coloring links to this network, we obtain the new coloring table shown in Fig. 7b, where each link has the color of the node from where it emanates. Thus, we update the table to generate the new coloring, as shown in the right panel of Fig. 7b.

- **Step 3** - Using the same criterion as in Step 2, we update the coloring of nodes, comprising now five different colors, and then we generate the new table, as shown in Fig. 7c. At this point the algorithm stops, because we do not need to introduce more colors, since each color is balanced. Each color corresponds to a fiber, and each node in each colored fiber receives the same colors from other fibers or from nodes in the same fiber. Therefore, the coloring shown in the network of Fig. 7c is the minimal balanced coloring of the network, and the colors indicate the fibers in the network.

## XII. DATASETS OF BIOLOGICAL AND NON-BIOLOGICAL NETWORKS

To investigate the applicability of graph fibrations in a broader context, we performed an extensive analysis of different complex networks from diverse domains in systems science.

Full details of each network analyzed can be accessed at [https://docs.google.com/spreadsheets/d/1-RG5vR\\_EGNPqQcnJU8q3ky10pWi30jTh5Uo-Xa0Pj0c/edit#gid=1973910256](https://docs.google.com/spreadsheets/d/1-RG5vR_EGNPqQcnJU8q3ky10pWi30jTh5Uo-Xa0Pj0c/edit#gid=1973910256). The codes to reproduce this analysis are at [github.com/makselab](https://github.com/makselab) and the full datasets appear at [kcorelab.org](https://kcorelab.org). See also tables below with information about the networks.

We first show the symmetry fibrations in biological networks and species. See Section IV. We characterize biological networks spanning from:

- **Biological networks:** transcriptional regulatory networks, metabolic networks, cellular processes networks and pathways, disease networks, neural networks.

We study the following species:

- Species: *A. thaliana*, *E. coli*, *B. subtilis*, *S. enterica* (salmonella), *M. tuberculosis*, *D. melanogaster*, *S. cerevisiae* (yeast), *M. musculus* (mouse), and *H. sapiens* (human).

We also study non-biological networks in Section IV:

- Social Networks: online social networks, Facebook, Twitter, Wikipedia, Youtube, email networks, communication networks, citation networks, collaboration networks, bloggers
- Internet: routers, autonomous systems, web graphs, hyperlinks, peer-to-peer
- Infrastructure Networks: power grid, airport, roads, flights
- Economic Networks
- Software Networks: Linux, jdk
- Ecosystems

Network Domain	Total No. of nodes	Total No. of edges	No. of networks
Biological	287390	4211856	289
Economic	1752	108639	5
Ecosystems	1879	5378	14
Infrastructure	24511	82534	16
Internet	244634	835565	27
Social	104909	1261009	15
Software	43391	503645	3

TABLE I: Features of the networks across domains. We report the total numbers for each domain summed over all the networks in the domain.

Network Subdomain	Total No. of nodes	Total No. of edges	No. of networks
Autonomous systems graphs	141842	481415	14
Bitcoin	9664	59777	2
Collaboration networks	50260	504897	4
Disease	4309	15254	66
Facebook	4039	88234	1
Youtube subscriptions	13723	76765	1
Internet peer-to-peer networks	31978	110154	4
Jazz	198	5484	1
Linux	30837	213954	1
Metabolic	4273	33829	50
Networks with ground-truth communities	1005	25571	1
Neural networks	3694	129812	8
Cellular processes and Pathways	9825	54712	127
Plant-Pollinator	1631	2719	11
Plant-Seed-Disperser	65	165	2
Power grid	4941	6594	1
Sentiment	99	278	2
Transcriptional regulatory	260258	3908769	32

TABLE II: Subtypes of networks belonging to the different domains.

### XIII. GENETIC BUILDING BLOCKS AND THEIR FIBER NUMBERS

There are two types of fibers in the genetic network studied according to the topology of the input-tree. Input-trees with integer branching ratio  $n$  and input-trees with irrational branching ratio which are generalized Fibonacci sequences with golden ratios  $\varphi_d$ . We introduce fiber numbers as a vector  $|n, \ell\rangle$  and  $|\varphi_d, \ell\rangle$  to classify fibers through topological classes of input-trees in these two general classes, respectively. Fiber number  $n$  defines the branching ratio of infinitely long perfect  $n$ -ary trees in the universal input-tree of the fiber. It directly quantifies the loops **inside** the fiber, since each loop can be traversed an infinite number of times, leading to an infinite chain or tree in the universal input-tree. For

Species	Total No. of nodes	Total No. of edges	No. networks
Yeast	55932	1392926	11
Arabidopsis Thaliana	790	1431	1
Bacillus subtilis	5602	11417	3
Drosophila	39549	321734	5
Escherichia coli	879	1835	1
Human	72587	1198712	248
Micobacterium Tuberculosis	1624	3212	1
Mouse	64709	987424	7
Salmonella	8293	15589	6

TABLE III: Number of networks per species.

an integer-branching ratio input-tree, fiber number  $\ell$  defines the number of finite chains in the input set coming from the external regulators. The branching ratio  $n$  defines the fundamental class. The finite chain  $\ell$  fiber number defines the subclass.

For Fibonacci-type universal input-trees, the branching ratio becomes fractal and are characterized by the golden ratio  $\varphi$  and other generalized golden ratios,  $\varphi_d$  for  $d \geq 3$ . The main feature of these structures is that beyond the cycles inside the fiber, there is a cycle traversing the external regulators to the fiber. This cycle goes through a connected component of genes and has a length  $d$ . In this case, we classify the Fibonacci fibers with a vector  $|\varphi_d, \ell\rangle$ , where the main fiber number is the fractal branching ratio  $\varphi_d$ ,  $d$  is the length of the cycle going through the connected component external to the fiber, and  $\ell$  is the number of external regulators to the fiber. If there are more than one cycle, then the building block is defined using the shortest cycle. The well-known Fibonacci golden ratio corresponds to the *exuR* circuit with  $d = 2$ . In this case, the cycle has length  $d = 2$ . It means that the fiber directly connects to the regulator and the regulator connects back to the fiber. This structure gives rise to a Fibonacci sequence with the golden ratio  $\varphi_2 = \varphi = (1 + \sqrt{5})/2$ . The other circuits are generalized Fibonacci circuits with  $d = 3$  and 4. In these circuits the fiber sends information to a connected components to which a regulator of the fiber belongs. The path of length  $d$  is the shortest path through the connected component.

The calculation of the generalized golden ratio proceeds as follows. We build the adj-

cency matrix of the base of the building block. For instance, for the 4-FF in Fig. 3b we have (by convention  $A_{ij}$  implies a link from  $i \rightarrow j$ , i.e.,  $A_{i \rightarrow j}$ ):

$$A = \begin{matrix} evgA \\ gadE \\ gadX \\ hns \end{matrix} \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}. \quad (8)$$

The generalized golden ratio is the largest eigenvalue of this matrix. In this case:  $\varphi_4 = 1.38028\dots$

### A. Fibers statistics

SM Table IV shows the counts in the *E. coli* network of each building block. For instance the most abundant building blocks are the following:

$|n = 0, \ell = 1\rangle$ : 45

$|n = 1, \ell = 0\rangle$ : 13

$|n = 0, \ell = 2\rangle$ : 13

$|n = 1, \ell = 1\rangle$ : 8

The list is completed with the fractal building blocks of Fibonacci sequences which are less numerous but more complex in their structure:

$|\varphi_2 = 1.6180\dots, \ell = 2\rangle$ : 1

$|\varphi_3 = 1.4655\dots, \ell = 1\rangle$ : 1

$|\varphi_4 = 1.3802\dots, \ell = 1\rangle$ : 1

### B. Full list of fibers

SM Table V shows the complete list of the 91 fibers building blocks found in the genetic network of *E. coli*. We list the genes in the fiber plus their external regulators. Supplementary File 1 shows the plot of the circuit of every fiber and the fiber building block.

Structure type	Amount in E-coli
$ n = 0, l = 1\rangle$	45
$ n = 0, l = 2\rangle$	13
$ n = 0, l = 3\rangle$	3
$ n = 1, l = 0\rangle$	13
$ n = 1, l = 1\rangle$	8
$ n = 1, l = 2\rangle$	3
$ n = 2, l = 0\rangle$	1
$ n = 2, l = 1\rangle$	1
$ \varphi_d = 1.3802.., l = 1\rangle$	1
$ \varphi_d = 1.4655.., l = 1\rangle$	1
$ \varphi_d = 1.6180.., l = 2\rangle$	1
Composite Fiber	1
<b>Total number of building blocks</b>	<b>91</b>

TABLE IV: Building block statistics. We show the count of every building block defined by the fiber numbers.

The first column in SM Table V is the ID of the fiber. This ID refers to the plot of the fiber building block in Supplementary File 1. The second column lists the genes in the fiber, the third column lists the external regulators. The last column specifies the fiber number associated with each fiber as  $|n, \ell\rangle$  or  $|\varphi_d, \ell\rangle$ .

### C. Examples of fibers with integer fiber numbers

- SM Table VI shows concrete examples of fiber building blocks in the Star Fiber Class with  $n = 0$ .
- SM Table VII shows concrete examples of fiber building blocks in the Chain Fiber Class with  $n = 1$ .
- SM Table VIII shows concrete examples of fiber building blocks in the Binary Tree Fiber Class with  $n = 2$ .

For each fiber number  $n$  we show several typical examples of building blocks with different values of  $\ell$ . We show typical cases of integer fiber numbers. Despite the fact that the circuits are characterized by a few topological classes of input-trees, the circuits in the network have a rich variety of different topologies.

Id	Fiber	Regulators	Fiber Number
1	aaeR, ampDE, azuC, comR, cyaA, narQ, sohB, speC, spf, trxR, yaeP-rof, yaeQ-arfB-nlpE, yjeF-tsaE-amiB-mutL-miaA-hfq-hflXKC	crp	$ n = 0, l = 1\rangle$
2	aaeXAB, agp, cpdB, cstA, glgS, glpR, grpE, hofMNOP, ivbL-ilvBN-uhpABC, lacI, mcaS, mhpR, nadC, ompA, ppdD-hofBC, preTA, raiA, rmf, rpsF-priB-rpsR-rplI, sfsA-dksA-gluQ, sxy, ubiG, ychH, yeiP, yeiW, yfiP-patZ, yibN-grxC-secB-gpsA, ykgR	crp	$ n = 0, l = 1\rangle$
3	accA, accD, fabI, fadR, yceD-rpmF-plsX-fabHDG-acpP-fabF		$ n = 1, l = 0\rangle$
4	accB, iclR	fadR	$ n = 1, l = 1\rangle$
5	ackA-pta, dcuC	arcA, fnr	$ n = 0, l = 2\rangle$
6	acrZ, inaA, nfo, nfsB	marA, rob, soxS	$ n = 0, l = 3\rangle$
7	add, dsbG, gor, grxA, hemH, oxyS, trxC	crp, oxyR, rbsR	$ n = 0, l = 1\rangle \oplus  n = 1, l = 1\rangle$
8	adeD, adiY, chiA, gspAB, hchA, hdfR, mdtJI, rcsB, yjjP	hns	$ \varphi_d = 1.4655.., l = 1\rangle$
9	agaR, agaS-kbaY-agaBCDI		$ n = 1, l = 0\rangle$
10	alaA-yfbR, avtA, leuE, livJ, livKHMGF, lysU, sdaA	lrp	$ n = 0, l = 1\rangle$
11	alaE, kbl-tdh, yojI	lrp	$ n = 0, l = 1\rangle$
12	alaWX, argU, argW, argX-hisR-leuT-proM, aspV, flxA, glyU, leuQPV, leuX, lptD-surA-pdxA-rsmA-apaGH, lysT-valT-lysW, metT-leuW-glnUW-metU-glnVX, pheU, pheV, proK, proL, queA, serT, serX, thrU-tyrU-glyT-thrT-tufB, thrW, trmA, tyrTV-tpr, valUXY-lysV	fis	$ n = 0, l = 1\rangle$
13	aldB, hupB	crp, fis	$ n = 0, l = 2\rangle$
14	allA, allS, gcl-hyi-glxE-ybbW-allB-ybbY-glxE	allR	$ n = 0, l = 1\rangle$
15	alsR, rpiB		$ n = 1, l = 0\rangle$
16	amiA-hemF, cmk-rpsA-ihfB, uspB	IHF	$ n = 0, l = 1\rangle$

17	amn, mipA, phnCDE_1E_2FGHIJKLMNOP, phoA-psif, phoB, phoE, phoH, ydfH, yegH, yhjC, ytfK		$ n = 1, l = 0\rangle$
18	ampC, dacC	bolA	$ n = 0, l = 1\rangle$
19	araE-ygeA, araFGH	araC, crp	$ n = 0, l = 2\rangle$
20	arcZ, ydeA	arcA	$ n = 0, l = 1\rangle$
21	argA, argCBH, argE, argF, argI, argR, artJ, artPIQM, lysO		$ n = 1, l = 0\rangle$
22	argO, lysP	argP, lrp	$ n = 0, l = 2\rangle$
23	aroF-tyrA, tyrB	tyrR	$ n = 0, l = 1\rangle$
24	aroH, trpLEDCBA, trpR		$ n = 1, l = 0\rangle$
25	asnB, clpPX-lon, glsA-ybaT, uspE	gadX	$ n = 0, l = 1\rangle$
26	aspA-dcuA, dcuR	crp, fnr,	$ n = 0, l = 3\rangle$
		narL	
27	bacA, cpxPQ, cpxR, ftnB, ldtC, ldtD, ppiD, sbmA-yaiW, slt, srkA-dsbA, xerD-dsbC-recJ-prfB-lysS, yccA, yebE, yidQ, yqaE-kbp, yqjA-mzrA		$ n = 1, l = 0\rangle$
28	baeR, spy	cpxR	$ n = 1, l = 1\rangle$
29	bcsABZC, fnrS, pdeF, pepT, pitA, ravA-viaA, tar-tap- cheRBYZ, upp-uraA, xdhABC, ydeJ, ytiCD-idlP-iraD	fnr	$ n = 0, l = 1\rangle$
30	bdcA, dkgB, grxD, mepH, mhpT, pgpC-tadA, rfe-wzzE- wecBC-rffGHC-wecE-wzxE-rffT-wzyE-rffM, rybB, tehAB, tsgA, ydbD, yeaE	nsrR	$ n = 0, l = 1\rangle$
31	betI, betT	arcA, cra	$ n = 1, l = 2\rangle$
32	bioA, bioBFCD	birA	$ n = 0, l = 1\rangle$
33	bluF, ydeI	rcdA	$ n = 0, l = 1\rangle$
34	borD, envY-ompT, mgrB, mgrR, mgtLA, mgtS, pagP, rstA, phoP ybjG		$ n = 0, l = 1\rangle$
35	cbpAM, gltX, gyrB, msrA	fis	$ n = 0, l = 1\rangle$
36	cdaR, garD, gudPXD		$ n = 1, l = 0\rangle$

37	cho, dinB-yafNOP, dinD, ding-ybib, dinQ, ftsK, hokE, insK, lexA, polB, ptrA-recBD, recAX, recN, recQ, rpsU-dnaG-rpoD, ruvAB, symE, tisB, umuDC, uvrB, uvrD, uvrY, ybfE, ybgA-phr, ydjM, yebG		$ n = 1, l = 0\rangle$
38	cirA, entCEBAH, fepA-entD, fiu	crp, fur	$ n = 0, l = 2\rangle$
39	copA, cueO	cueR	$ n = 0, l = 1\rangle$
40	cra, pitB, sbcDC	phoB	$ n = 0, l = 1\rangle$
41	crl_1, exbBD, fepDGC, fhuACDB, fhuE, gpmA, metJ, nohA- ydfN-tfaQ, ryhB, ygaC, yhhY, yjjZ	fur	$ n = 0, l = 1\rangle$
42	cusCFBA, cusR, yedX	hprR, phoB	$ n = 1, l = 2\rangle$
43	cvpA-purF-ubiX, glrR-glnB, hflD-purB, lolB-ispE-prs, purC, purEK, purL, speAB	purR	$ n = 0, l = 1\rangle$
44	cysDNC, cysK, tcyP, yciW, ygeH, yoac	cysB	$ n = 0, l = 1\rangle$
45	cytR, nagC, nagE, ycdZ	crp	$ n = 1, l = 1\rangle$
46	dapB, lysC	argP	$ n = 0, l = 1\rangle$
47	ddpXABCDF, patA, potFGHI, yeaGH, yhdWXYZ	ntrC	$ n = 0, l = 1\rangle$
48	decR, mlaFEDCB, yncE	marA	$ n = 0, l = 1\rangle$
49	dgcC, iraP, nlpA, wrbA-yccJ, yccT	csgD	$ n = 0, l = 1\rangle$
50	dicB-ydfDE-insD-7-intQ, dicC-ydfXW	dicA	$ n = 0, l = 1\rangle$
51	dsdC, norR	nsrR	$ n = 1, l = 1\rangle$
52	dtpA, omrA, omrB	ompR	$ n = 0, l = 1\rangle$
53	ecpA, ecpR	matA	$ n = 0, l = 1\rangle$
54	efeU_1U_2, motAB-cheAW, psd-mscM, tsr, ung	cpxR	$ n = 0, l = 1\rangle$
55	epd-pgk-fbaA, gapA-yeaD, mpl	cra, crp	$ n = 0, l = 2\rangle$
56	erpA, iscR, rnlAB		$ n = 1, l = 0\rangle$
57	evgA, nhaR	hns	$ \varphi_d = 1.3802.., l = 1\rangle$
58	fabA, fabB	fabR, fadR	$ n = 0, l = 2\rangle$
59	fadE, fadIJ	arcA, fadR	$ n = 0, l = 2\rangle$
60	fbaB, fruBKA, glk, gpmM-envC-yibQ, pfkA, ppc, pykF, pyrG-eno, tpiA	cra	$ n = 0, l = 1\rangle$

61	fldB, pgi, ribA, ydbK-ompN		soxS	$ n = 0, l = 1\rangle$
62	folE-yeiB, metA, metC, metF		metJ	$ n = 0, l = 1\rangle$
63	fpr, pqiABC, rirA-waaQGPSBOYZU		marA, soxS	$ n = 0, l = 2\rangle$
64	fucAO, fucR, zraR		crp	$ n = 1, l = 1\rangle$
65	gfcA, ybhL, yfiR-dgcN-yfB, ymiA-yciX		yjjQ	$ n = 0, l = 1\rangle$
66	hupA, trg		crp, fis	$ n = 0, l = 2\rangle$
67	ibaG-murA, rplU-rpmA-yhbE-obgE		mlrA	$ n = 0, l = 1\rangle$
68	ibpAB, yadV-htrE		IHF	$ n = 0, l = 1\rangle$
69	idnK, idnR		crp, gntR	$ n = 1, l = 2\rangle$
70	isrC-flu, pth-ychF		oxyR	$ n = 0, l = 1\rangle$
71	lgoR, uxuR		crp, exuR	$ \varphi_d = 1.6180.., l = 2\rangle$
72	lolA-rarA, osmB		rcsB	$ n = 0, l = 1\rangle$
73	lsrACDBFG-tam, lsrR, oxyR, rbsR		crp	$ n = 1, l = 1\rangle$
74	malI, mlc		crp	$ n = 1, l = 1\rangle$
75	manA, yhfA		crp	$ n = 0, l = 1\rangle$
76	mngAB, mngR			$ n = 1, l = 0\rangle$
77	nadA-pnuC, nadB		nadR	$ n = 0, l = 1\rangle$
78	nimR, nimT			$ n = 1, l = 0\rangle$
79	ompX, rpsP-rimM-trmD-rplS, ychO, ysgA		fnr	$ n = 0, l = 1\rangle$
80	pepD, yhbTS		csgD	$ n = 0, l = 1\rangle$
81	phoP, slyB			$ n = 2, l = 0\rangle$
82	pspABCDE, pspG		IHF, pspF	$ n = 0, l = 2\rangle$
83	purR, pyrC		fur	$ n = 1, l = 1\rangle$
84	rhaR, rhaS		crp	$ n = 2, l = 1\rangle$
85	rrsA-ileT-alaT-rrlA-rrfA, rrsE-gltV-rrlE-rrfE		fis, lrp	$ n = 0, l = 2\rangle$
86	rrsB-gltT-rrlB-rrfB, rrsC-gltU-rrlC-rrfC, rrsD-ileU-alaU-rrlD-rrfD-thrV-rrfF, rrsG-gltW-rrlG-rrfG, rrsH-ileV-alaV-rrlH-rrfH		fis, hns, lrp	$ n = 0, l = 3\rangle$
87	ssb, uvrA		arcA, lexA	$ n = 0, l = 2\rangle$
88	ttdABT, ttdR			$ n = 1, l = 0\rangle$

89	ycjG, ycjY-ymjDC-mpaA	pgrR	$ n = 0, l = 1\rangle$
90	yegRZ, yfdX-frc-oxc-yfdVE	evgA	$ n = 0, l = 1\rangle$
91	ykgMO, znuA, znuCB	zur	$ n = 0, l = 1\rangle$

TABLE V: List of fiber building blocks with ID, genes in the fiber, external regulators of the fiber and fiber numbers. We provide a Supplementary File 1 which plots every building block using the same IDs.

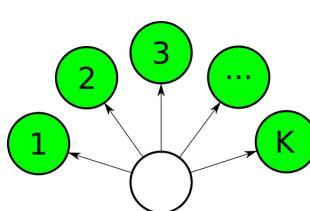
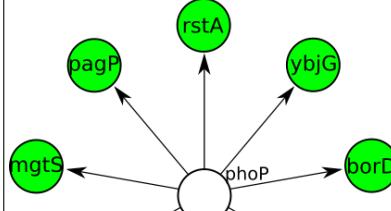
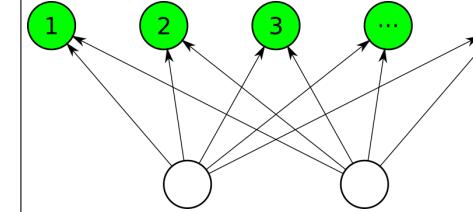
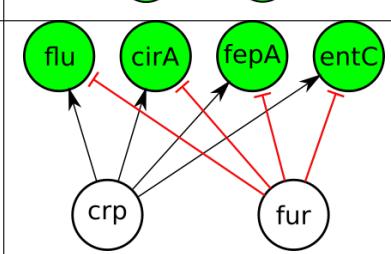
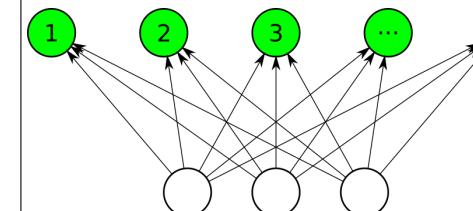
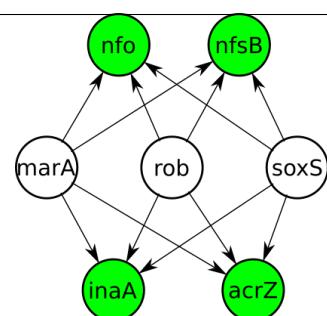
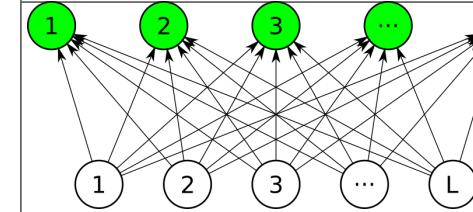
Class	Building block topology in the network	Count	Circuit example in <i>E. coli</i>
L-Star Fiber (L-SF) $ n = 0, \ell = L\rangle$		45	
		13	
		3	
		61 total	

TABLE VI: Examples of Star Fibers

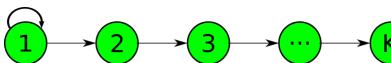
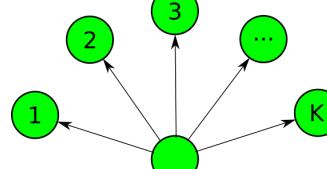
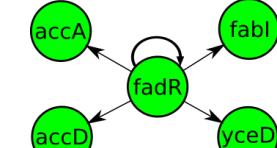
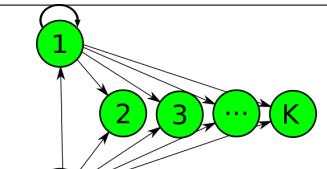
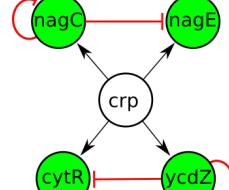
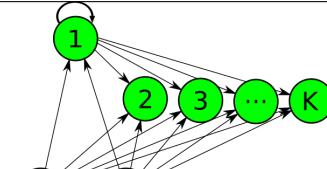
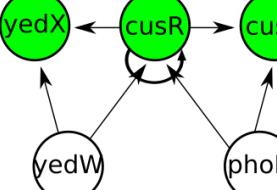
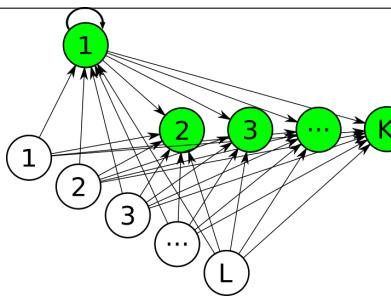
Class	Building block topology in the network	Count	Circuit example in <i>E. coli</i>
L-Chain Fiber (L-CF) $ n = 1, \ell = L\rangle$		5	
		8	
		8	
		3	
		24 total	

TABLE VII: Examples of Chain Fibers

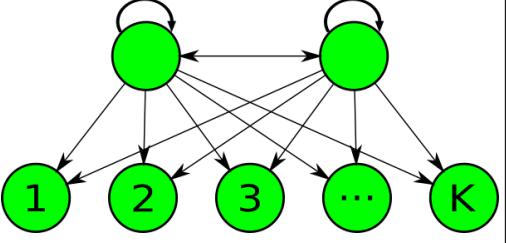
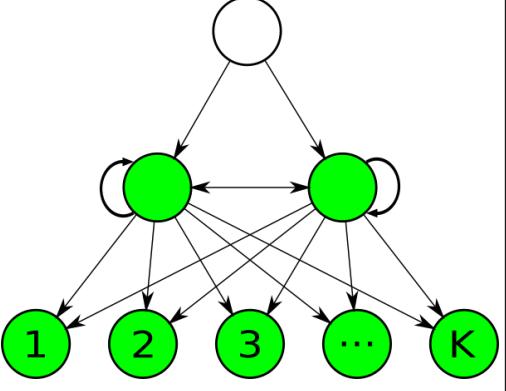
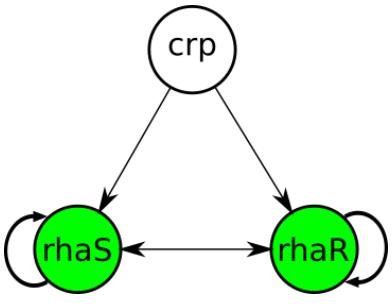
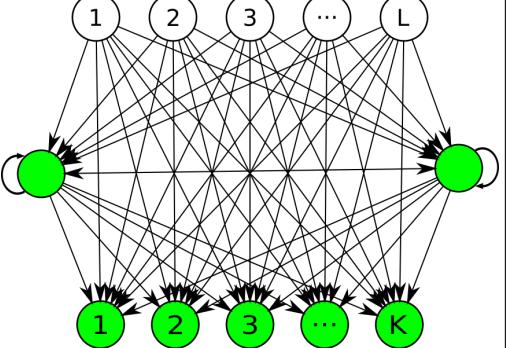
Class	Building block topology in the network	Count	Circuit example in <i>E. coli</i>
L-Binary Tree Fiber (L-BTF) $ n = 2, \ell = L\rangle$		1	
		1	
		2 total	

TABLE VIII: Examples of Binary Tree Fibers