

Dynamics of network motifs

Thomas M. A. Fink, Emmanuel Barillot, and Sebastian E. Ahnert

Institut Curie, CNRS UMR 144, 75005 Paris, France and

Theory of Condensed Matter, Cavendish Laboratory, Cambridge CB3 0HE, UK

(Dated: February 1, 2006)

A biased distribution of local network structures, or network motifs, has been widely observed but remains poorly understood. By comprehensively studying their dynamics, we show that some network motifs are fundamentally more versatile—capable of executing a variety of tasks—than others. We calculate the value of a motif by comparing its versatility to its range of possible update rules. Our work suggests that network motifs are not optimised to perform a single dynamical task, but that some motifs are recurrent because, by a change of rules, they possess a broad range of functionality.

Complex dynamical networks [1] are found in the brain [2], genetic regulation and transcription [3], the World Wide Web [4], transport and traffic and ecological food webs [5]. Recently it was observed that some local network structures, or network motifs, are much more frequently observed in complex networks than would be expected by chance [6, 7]. Although this biased distribution of motifs appears to apply to a broad range of networks, it remains unclear why some motifs are ubiquitous and others are not. Understanding the comparative advantages of different small networks would explain why real dynamical networks make disparate use of motifs.

Recent work on network motifs includes the development of efficient methods of motif identification [8] and understanding the distribution of motifs imposed by underlying network geometry [9]. Certain motifs exhibiting dynamical behaviour have been identified as essential ingredients of specific biological processes [10–13]. The transition from inactivity to periodic and chaotic behaviour has been studied on small networks in the context of continuous neural networks [14].

In this Letter we comprehensively study the dynamics of network motifs by evaluating all possible binary update rules over all possible network connectivities. We introduce a formalism for classifying dynamical behaviour and show that some network motifs are fundamentally more versatile—capable of executing a variety of tasks—than others. Because it is more likely to be capable of an arbitrary task, a versatile motif would ostensibly occur more frequently in real networks, all else being equal. While versatility roughly increases with motif complexity, we find evidence of a critical complexity, after which increasing the number of bonds confers no advantage.

In addition to studying the range of tasks that can be executed by a given motif, we consider the range of network motifs which can perform a given task (the network design problem). We find that for a given task, there are typically many update rules running on different networks which can execute it, but with varying degrees of network versatility, simplicity, and resistance to

mutations. Plotting these networks/rules in the associated phase space, we find evidence of a trade-off between the quantities; in other words, there is limited freedom in choosing *how* a network performs a given task.

To understand the dynamical behaviour of network motifs, we consider the most general class of discrete, binary dynamics, boolean networks. We evaluate all possible update rules over each network for 2 and 3 nodes (Figure 1). Beyond this, the complexity of the problem explodes: there are 10^{19} update rules and 3044 networks for 4 nodes (Table I). Unlike others [6, 7, 9, 14], we include networks with self-interactions; these are important in many natural networks, particularly genetic networks, which rely heavily on autoregulation [3]. Moreover, we find that their presence dramatically increases the variety of tasks that a motif can achieve.

Boolean dynamics and endofunctions—A boolean network has two parts: a directed network of N nodes and from 0 to N^2 edges, in which each node can take the value 0 or 1 (the hardware); and a set of local update rules for the nodes (the software). An update rule assigns an output of 0 or 1 for each of the possible combinations of inputs. The state of the network is the vector of values

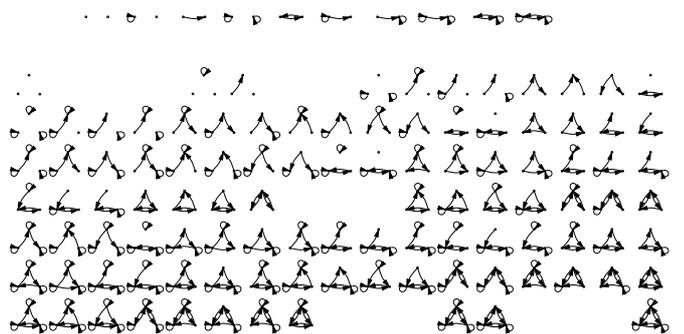


FIG. 1: Directed networks with self-interactions, ordered by number of edges. There are 10 distinct networks on 2-nodes and 104 networks on 3 nodes. For 10 nodes, the number of networks is of the order of Avogadro’s number.



FIG. 2: Examples of dynamics and their reduced representations, called endofunctions, for $N = 2$ and 3 nodes. The left dynamic can be written $a \rightarrow b, b \rightarrow c, c \rightarrow d, d \rightarrow a$, or $abcd \rightarrow bcda$, or simply $bcda$. Similarly, the center and right dynamics are $bbab$ and $eadhacc$.

of the N nodes, and thus there are 2^N possible states. The state is updated according to the combined set of local update rules, which is called a boolean function (BF). The number of possible update rules per node is 2^{2^k} , where k is the number of inputs (not outputs). For a fully connected network, there are $B = (2^N)^{2^N}$ BFs. The number of networks grows with N as 2, 10, 104, 3044, ...; this is the number of structures of finite relations (Sloane A000595). We call an instance of the hardware a network and an instance of the software a program.

In accordance with the BF, each state in a network is followed by another new state until the dynamics enters into an endless loop, or cycle. The 2 node network, for example, has 4 possible states: 00, 01, 10, 11. We label these a, b, c, d . Then the dynamic $00 \rightarrow 01, 01 \rightarrow 11, 10 \rightarrow 00, 11 \rightarrow 01$, say, can be written $a \rightarrow b, b \rightarrow d, c \rightarrow a, d \rightarrow b$ or $abcd \rightarrow bcda$, or simply $bcda$. The $4^4 = 256$ BFs on the fully connected network correspond to the 256 words that can be composed of the letters $abcd$. The behaviour of a BF, or word, is best appreciated by drawing its graph. When we leave the labels on the graph, we call the graph a dynamic, but when we drop the labels, we call the graph an endofunction (Figure 2). There is a one-to-one correspondence between BFs and dynamics, but the number of endofunctions is much smaller, since two different labelled graphs may have the same unlabelled graph. There are 19 endofunctions for 4 states: $\lambda \lambda \lambda$, where it is understood that all graphs contain 4 nodes and the circles $\circ \circ \circ \circ$ are cycles of length 1, 2, 3, 4.

An endofunction is a function from some set S to the same set S , where the elements of S are unlabeled. There is a one-to-one correspondence between endofunctions on

N	Net-works	B Boolean functions	E Endo-functions	CP Cycle patterns	CL Cycle lengths
1	2	2^2	3	3	2
2	10	4^4	19	11	4
3	104	8^8	951	66	8
4	3044	16^{16}	3799624	914	16
5	291968	32^{32}	9.06×10^{13}	43819	32

TABLE I: Properties of small boolean networks. The last 4 columns, which are the 2^N th terms of more natural series, apply to the fully connected network only. How they diminish with the cutting of bonds is the subject of this Letter.

j elements and directed graphs of j points and j edges, the only rule being that only one arrow can leave each node. The number of endofunctions grows with j as 1, 3, 7, 19, 47, 130, ... (Sloane A001372). When converting dynamics to endofunctions, some of the latter occur more frequently than others. The least frequent is the identity endofunction $abcd \dots$ (a set of N 1-cycles), which occurs only once; the most frequent is $aabcd \dots$ (a 1-cycle followed by a linear tail), with frequency $2^N!$.

Versatility—Versatility is the ability to do many different tasks with minimal reorganisation. In the case of network motifs, this means adopting many endofunctions by a change of software whilst keeping the hardware fixed.

For $N = 2$, there are 10 possible networks (Figure 1 top). For the fully connected network, the distribution of endofunctions given a random program is given by (10). What happens to the distribution when we cut bonds from the fully connected network? This is described by the so-called master equations:

$$P_E(\cdot \cdot) = \lambda \quad (1)$$

$$P_E(\circ \cdot) = \frac{1}{4}[2\lambda + \lambda\lambda + \phi] \quad (2)$$

$$P_E(\cdot \rightarrow) = \frac{1}{4}[2\lambda + 2\lambda] \quad (3)$$

$$P_E(\circ \rightarrow) = \frac{1}{16}[4\lambda + 4\lambda\lambda + 4\phi + 3\phi\phi + 1\phi\phi\phi] \quad (4)$$

$$P_E(\cdot \leftrightarrow) = \frac{1}{8}[4\lambda + 2\lambda + \phi + \phi] \quad (5)$$

$$P_E(\circ \leftrightarrow) = \frac{1}{4}[\lambda + \lambda + \lambda\lambda + \phi] \quad (6)$$

$$P_E(\cdot \rightarrow \cdot) = \frac{1}{8}[2\lambda + \lambda + \lambda + \lambda + \phi + \lambda\phi + \lambda\lambda] \quad (7)$$

$$P_E(\circ \rightarrow \cdot) = \frac{1}{64}[8(\lambda + \phi + \lambda\lambda + \lambda\lambda) + 2(\phi\phi + \phi\phi) + 4(\lambda + \lambda + \lambda + \lambda + \lambda\phi + \lambda\phi) + 3\phi\phi + \phi\phi] \quad (8)$$

$$P_E(\cdot \leftrightarrow \cdot) = \frac{1}{32}[4(\lambda + \lambda) + 3(\lambda\lambda + \phi) + 2(\lambda + \lambda + \lambda + \lambda + \lambda\phi + \lambda\phi) + \phi\phi + \phi\phi] \quad (9)$$

$$P_E(\circ \leftrightarrow \cdot) = \frac{1}{256}[24(\lambda + \lambda + \lambda + \lambda + \lambda + \lambda + \lambda + \lambda) + 12(\lambda + \lambda + \lambda + \lambda + \lambda + \lambda + \lambda + \lambda) + 8\phi + 6(\phi\phi + \phi\phi) + 4\lambda + 3\phi + \phi\phi]. \quad (10)$$

These equations can be inverted to solve for the distri-

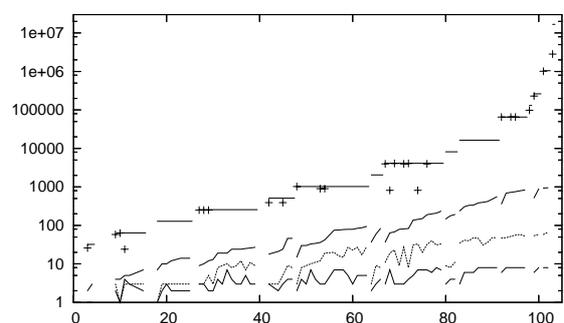


FIG. 3: The 104 3-node networks (we do not plot disjoint networks). From top, the numbers of boolean functions B (lines), distinct BFs \bar{B} (+s, if different), endofunctions E (lines), cycle patterns CP (dotted lines) and cycle lengths CL (lines).

bution of networks for a given endofunction.

A coarser picture, which only considers the long term behaviour, results from grouping together dynamics which have the same set of cycle lengths. For $N = 2$, these *cycle patterns* are $\circ \circ \circ \circ \circ \circ \circ \circ \circ \circ$. The number of cycle patterns CP grows with N as 3, 11, 66, 914... (the 2^N th terms of Sloane A026905); the N th term is the sum of the first 2^N partition numbers.

The coarsest classification associates all dynamics that contain at least one cycle of a given length. The possible cycle lengths (CL s) are the circles of size 1 to 2^N ; for $N = 2$, $\circ \circ \circ \circ$. Note that one dynamic can be a member of more than one cycle length class.

The distributions of cycle patterns and cycle lengths can be deduced directly from a network's master equation. For the fully connected network (\leftrightarrow), they are

$$P_{CP}(\leftrightarrow) = \frac{1}{256} [64^\circ + 48(\circ^\circ + \circ) + 36\% + 24\circ + 12\% + 8\circ + 6(\% + \circ) + 3\% + \%] \quad (11)$$

$$P_{CL}(\leftrightarrow) = \frac{1}{256} [191^\circ + 93\circ + 32\circ + 6\circ]. \quad (12)$$

For $N = 3$, there are 104 networks (Figure 1 bottom) and $8^8 \simeq 1.7 \times 10^7$ boolean functions, which generate 951 endofunctions. However, not all of the 104 networks can yield all possible endofunctions. The less connected networks exhibit a surprisingly limited range of behaviour. For example, the master equations for the feedback (\triangleleft) and feed-forward (\triangle) networks are

$$P_E(\triangleleft) = 1/16 (6 \text{ } \begin{array}{c} \circ \\ \diagup \\ \diagdown \end{array} + 6 \text{ } \begin{array}{c} \circ \\ \diagdown \\ \diagup \end{array} + 2 \text{ } \begin{array}{c} \circ \\ \diagup \\ \diagup \end{array} + \text{ } \begin{array}{c} \circ \\ \diagup \\ \circ \\ \diagdown \end{array} + \text{ } \begin{array}{c} \circ \\ \diagdown \\ \circ \\ \diagup \end{array}) \quad (13)$$

$$P_E(\triangle) = 1/16 (5 \text{ } \begin{array}{c} \circ \\ \diagup \\ \diagdown \end{array} + 2 \text{ } \begin{array}{c} \circ \\ \diagdown \\ \diagup \end{array} + 3 \text{ } \begin{array}{c} \circ \\ \diagup \\ \diagup \end{array} + 2 \text{ } \begin{array}{c} \circ \\ \diagdown \\ \diagdown \end{array} + 2 \text{ } \begin{array}{c} \circ \\ \diagup \\ \circ \\ \diagdown \end{array} + \text{ } \begin{array}{c} \circ \\ \diagdown \\ \circ \\ \diagup \end{array}). \quad (14)$$

The feedback network exhibits a broad range of cycle lengths, whereas the feed-forward shows a variety of basins of attraction but only around 1-cycles (the only connected network with ≥ 3 bonds to do so). By contrast, the superposition $\triangleleft + \triangle = \triangleleft$, which we call feed-feedback, has 19 endofunctions, including all cycle lengths but 8. Interestingly, adding a self-interaction to a network can increase its versatility dramatically: \triangleleft , \triangleleft and \triangleleft can generate 5.3, 5 and 11 times as many endofunctions as \triangleleft , \triangle and \triangle (Table II). Networks without self-interactions can generate at most 6.7% and 29% of the possible endofunctions and cycle patterns.

The feed-forward network \triangle has been identified as an important ingredient in several biological processes [10–13]. An example is a pulse generator in synthetic biology [11] which can be modeled by \triangle with a switch (left node), an inhibitor (top node) and an output (right node). The pulse generation itself is represented by the following sequence of active (1) and inactive (0) states of switch/inhibitor/output: $000 \Rightarrow 100 \rightarrow 111 \rightarrow 110$,

where \Rightarrow denotes switching on the generator (*i.e.*, changing the constant rule on the switch node from 0 to 1). This dynamic is represented by the fourth term of the RHS of (14). Another example using \triangle is a sign-sensitive delay element in transcription networks [13]. This is a circuit that responds slowly to the off-on and rapidly to the on-off switching of a signal.

Feedback is also an essential part of various biological mechanisms. Examples include (i) a cellular digital clock, which generates series of pulses of protein expression [10], and (ii) negative autoregulation [12]. The first can be modeled using the 2-node feedback network (\leftrightarrow) while the second requires an additional self-interaction (\leftrightarrow). The dynamics are represented by \circ and $\%$.

Value—The fully connected network can, with the appropriate software, emulate anything a simpler network can do. But a complicated network is more costly than a simple one. The number of edges is one measure of network complexity. A more natural definition is the amount of memory M necessary to specify the boolean function: $M = 2^{k_1} + 2^{k_2} + \dots$, where k_1, k_2, \dots are the number of inputs at each node. Note that $M = \log_2 B$ and that M is a function of the network hardware only.

The number of endofunctions over the number of programs is the extent to which a network can convert the space of programs into a variety of tasks. It is a measure of the value of a network. While for most networks the number of distinct programs is equal to 2^M , this is not the case for networks with symmetries. For instance, \triangleleft will give the same behaviour for a given BF under any permutation of the points, whereas this is not true for \triangle . In other words, the 128 programs on \triangle are all dis-

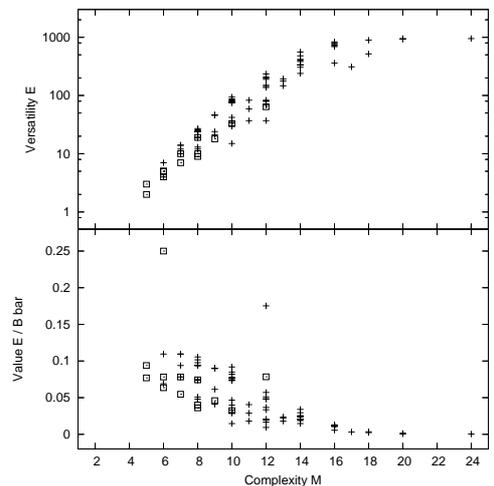


FIG. 4: The 104 3-node networks, ordered by memory M . The 13 motifs without self-interactions are highlighted as squares; disjoint networks are not shown. TOP: Versatility E (endofunctions). After 16 bits, increasing network complexity confers little advantage. BOTTOM: Value E/\bar{B} . Again, the most valuable networks tend to be less complex.

Net.	B	\bar{B}	E	CP	CL	Net.	B	\bar{B}	E	CP	CL
	32	26	2	1	1		256		19	10	7
	64	63	4	1	1		1024		33	10	7
	32		3	1	1		4096	816	64	19	7
	128		10	3	3		256		13	3	2
	64		5	3	3		2048		37	3	2
	256	250	10	3	3		256		25	7	5
	128		7	1	1		4096	1376	143	33	7
	64	24	5	3	4		1024		78	15	7
	512	392	18	3	3		4096		209	31	8
	256	250	9	3	3		65536		834	53	8

TABLE II: The 13 connected 3-node motifs without self-interactions (studied in [6, 7]) and 7 other networks of interest. We show the network diagram and the number of boolean functions B , distinct BFs \bar{B} (if different), endofunctions E , cycle patterns CP and cycle lengths CL .

tinct, whereas only 24 of the 64 programs are distinct. The reduction of programs B to distinct programs \bar{B} can be determined by the symmetry of the network:

$$B - \bar{B} = 2^{2^k} (2^{2^k} - 1) (\alpha 2^{2^k} + \beta), \quad (15)$$

where $\alpha, \beta = 0, 1/2$ for 2-fold symmetry and $\alpha, \beta = 5/6, 1/3$ for 6-fold symmetry, and k is the number of inputs on the symmetric nodes. For 3 nodes, we plot B and \bar{B} in Figure 3 and E/\bar{B} in Figure 4 (bottom). The most valuable networks tend to have lower complexity M , with feedback with 0 and 3 self-interactions being exceptional. **Stability**— If we make a mutation to a program (BF) by flipping a 0 to a 1 or 1 to 0, we may or may not cause the resulting endofunction or cycle pattern or cycle length to change. The fraction of the M possible mutations to the program (genotype) which do not change the behaviour (phenotype) is the stability S . There are times when it is important to have a high stability, such as when the environment is constant. In other circumstances there

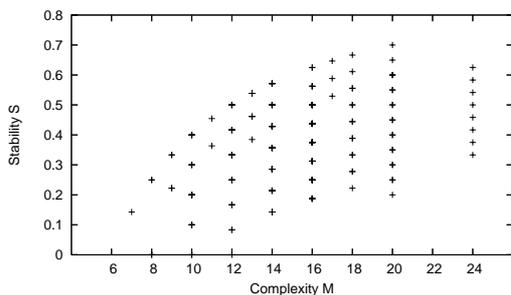


FIG. 5: Boolean networks which generate the cycle pattern , plotted in M - S phase space (many of the points are degenerate). Stability ranges from 0.08 to 0.7; the most stable BFs do not correspond to the simplest networks.

may be a need to swap to different tasks by changing single bits. We calculated the stability S with respect to cycle patterns (long-term behaviour) for all boolean functions/networks. We find that, within a set of programs which produce the same cycle pattern, there is a broad range in stability.

Stability S can be combined with versatility E and complexity M to create a unified picture in the context of network design. Given a particular behaviour, consider the set of boolean functions/networks which generate it. Typically this set is very large, and we plot the ensemble in E - M - S phase space. We find that there is limited freedom in specifying E , M and S , because optimising one tends to be in conflict with optimising another.

The conflict between program stability and network simplicity is illustrated by projecting the E - M - S phase space onto the M - S plane for a particular cycle pattern (Figure 5). There is a clear trade-off between low M and high S , and this is typical of other cycle patterns. Similarly, we have already seen a trade-off between versatility and simplicity in Figure 4 (top). The emerging picture suggests that in biological and other naturally occurring networks, robustness and versatility may come at the price of increased connectivity.

Our work suggests that elementary network circuits are not optimised to perform a single dynamical task, but are typically capable of generating a host of different dynamics via a change of update rules. This supports the evolutionary mechanism of co-opting a network used for one task to perform another; in other words, do not reinvent, borrow and modify. Boolean motifs provide an explicit set of dynamical building blocks for hierarchically constructing more complex task-specific networks, such as a clock of arbitrary period.

-
- [1] R. Albert and A. L. Barabasi, Rev. Mod. Phys. **74**, 47 (2002)
 - [2] C. Koch and G. Laurent, Science **284**, 96 (1999)
 - [3] N. Guelzim, S. Bottani, P. Bourguine, and F. Képès, Nature Genetics **31**, 60 (2002)
 - [4] J. Eckmann and E. Moses, Proc. Natl. Acad. Sci. U.S.A. **99**, 5825 (2002)
 - [5] D. Garlaschelli, G. Caldarelli, and L. Pietronero, Nature **423** 165 (2003)
 - [6] R. Milo *et al.*, Science **298**, 824 (2002).
 - [7] R. Milo *et al.*, Science **303**, 1538 (2004).
 - [8] E. Ziv, R. Koytcheff, M. Middendorf, and C. Wiggins, Phys. Rev. E **71**, 016110 (2005)
 - [9] S. Itzkovitz and U. Alon, Phys. Rev. E **71**, 026117 (2005)
 - [10] G. Lahav *et al.*, Nature Genetics **36**, 147 (2004)
 - [11] S. Basu *et al.*, Proc. Natl. Acad. Sci. U.S.A. **101**, 6355 (2004)
 - [12] N. Rosenfeld, M. B. Elowitz, and U. Alon, J. Mol. Biol. **323**, 785 (2002)
 - [13] S. Mangan, A. Zaslaver, and U. Alon, J. Mol. Biol. **334**, 197 (2003)
 - [14] V. P. Zhigulin, Phys. Rev. Lett. **92**, 238701 (2004)

- [15] N. G. de Bruijn, *J. Combinat. Theory (A)* **12**, 14 (1972).
- [16] M. Aldana-Gonzlez, S. Coppersmith and L. P. Kadanoff *Perspectives and Problems in Nonlinear Sci-*

ence (Springer, 2003).