

PAPER • OPEN ACCESS

## Breaking reflection symmetry: evolving long dynamical cycles in Boolean systems

To cite this article: Mathieu Ouellet *et al* 2024 *New J. Phys.* **26** 023006

View the [article online](#) for updates and enhancements.

### You may also like

- [A review of - flavor symmetry in neutrino physics](#)

Zhi-Zhong Xing and Zhen-Hua Zhao

- [LOW-FREQUENCY OSCILLATIONS IN GLOBAL SIMULATIONS OF BLACK HOLE ACCRETION](#)

Sean M. O'Neill, Christopher S. Reynolds, M. Coleman Miller *et al.*

- [Enhancing variational quantum state diagonalization using reinforcement learning techniques](#)

Aakash Kundu, Przemysaw Bedeek, Mateusz Ostaszewski *et al.*

# New Journal of Physics

The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft 

IOP Institute of Physics

Published in partnership  
with: Deutsche Physikalische  
Gesellschaft and the Institute  
of Physics



OPEN ACCESS

PAPER

## Breaking reflection symmetry: evolving long dynamical cycles in Boolean systems

RECEIVED  
7 April 2023

REVISED  
29 November 2023

ACCEPTED FOR PUBLICATION  
2 January 2024

PUBLISHED  
6 February 2024

Original Content from  
this work may be used  
under the terms of the  
[Creative Commons  
Attribution 4.0 licence](#).

Any further distribution  
of this work must  
maintain attribution to  
the author(s) and the title  
of the work, journal  
citation and DOI.



Mathieu Ouellet<sup>1</sup> , Jason Z Kim<sup>2</sup> , Harmange Guillaume<sup>3,4</sup>, Sydney M Shaffer<sup>4,5,6</sup> , Lee C Bassett<sup>1,11,\*</sup>  and Dani S Bassett<sup>1,5,7,8,9,10,11,\*</sup> 

<sup>1</sup> Department of Electrical & Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA 19104, United States of America

<sup>2</sup> Department of Bioengineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104, United States of America

<sup>3</sup> Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States of America

<sup>4</sup> Cell and Molecular Biology Group, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States of America

<sup>5</sup> Department of Biological Engineering, School of Engineering & Applied Science, University of Pennsylvania, Philadelphia, PA 19104, United States of America

<sup>6</sup> Department of Pathology and Laboratory Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, United States of America

<sup>7</sup> Department of Physics & Astronomy, College of Arts & Sciences, University of Pennsylvania, Philadelphia, PA 19104, United States of America

<sup>8</sup> Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States of America

<sup>9</sup> Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, United States of America

<sup>10</sup> Santa Fe Institute, Santa Fe, NM 87501, United States of America

<sup>11</sup> These authors contributed equally.

\* Authors to whom any correspondence should be addressed.

E-mail: [lbassett@seas.upenn.edu](mailto:lbassett@seas.upenn.edu) and [dsb@seas.upenn.edu](mailto:dsb@seas.upenn.edu)

**Keywords:** symmetry, Boolean networks, structural motifs, symmetry breaking, dynamical system

### Abstract

In interacting dynamical systems, specific local interaction rules for system components give rise to diverse and complex global dynamics. Long dynamical cycles are a key feature of many natural interacting systems, especially in biology. Examples of dynamical cycles range from circadian rhythms regulating sleep to cell cycles regulating reproductive behavior. Despite the crucial role of cycles in nature, the properties of network structure that give rise to cycles still need to be better understood. Here, we use a Boolean interaction network model to study the relationships between network structure and cyclic dynamics. We identify particular structural motifs that support cycles, and other motifs that suppress them. More generally, we show that the presence of *dynamical reflection symmetry* in the interaction network enhances cyclic behavior. In simulating an artificial evolutionary process, we find that motifs that break reflection symmetry are discarded. We further show that dynamical reflection symmetries are over-represented in Boolean models of natural biological systems. Altogether, our results demonstrate a link between symmetry and functionality for interacting dynamical systems, and they provide evidence for symmetry's causal role in evolving dynamical functionality.

### 1. Introduction

One of the most intriguing characteristics of complex systems is that they evince emergent global functions from local interactions. Gene regulatory networks are a quintessential example, describing the complex network of short time-scale interactions between molecules to produce the long time-scale cycles of reactions that sustain life [1–4]. Such cyclic behaviors play a fundamental role in many processes, including cell cycles [5], biological clocks [6], cell fate [7], cancer regulation and DNA damage [8], and signaling [9]. It is known

that cyclic behavior in random Boolean models arises more often than fixed points and that such behavior is favored by evolution [10]. Despite their significance and prevalence, long cyclic reactions remain far from understood, in part because the process of determining their underlying mechanisms is made difficult by the nonlinear and heterogeneous distribution of interactions. Can we distill simple principles for how specific patterns of local interactions determine long and complex cycles of reactions?

Biological systems have been fruitfully modeled as Boolean networks to shed light on this question. In these models, the state of each component—a gene, protein, or RNA—is described by a binary value, and the interactions between components—binding, chemical reaction, and so on—are described by Boolean functions. Prior work has extensively studied the interaction functions [11–13] to model probabilistic [14] and multi-level [15] interactions or to stabilize existing sequences of reactions [16]. Other work has focused on the intensive study of specific network topologies [2, 3, 17–20] and local structures that are typically referred to as motifs [21]. However, we still lack a general understanding of how the local interaction topology determines long sequences of cycles, thereby limiting our ability to make principled predictions across different networks about the global effects of local structures.

Here, we provide such an understanding through the analytical and numerical study of Boolean network topology. First, we use an evolutionary algorithm to optimize for network motifs with long cycle lengths and discover the existence of *suppressed motifs* that are almost entirely absent in the evolved networks. Next, we discover that many such evolved networks display a *dynamical reflection symmetry*, such that if the network at state  $\vec{x}(t)$  transitions to state  $\vec{y} = \vec{x}(t+1)$ , then that same network at state  $\vec{1} - \vec{x}(t)$  transitions to state  $\vec{1} - \vec{y}$ . Moreover, we find that the suppressed motifs systematically break this symmetry.

To demonstrate the practical utility of our finding, we apply it to real biological systems and find that reflection symmetry appears naturally in networks that have evolved to support long dynamical cycles, whereas suppressed motifs decrement the length of dynamical cycles. Our findings demonstrate how dynamical symmetries play a crucial role in the observed complexity of biological systems.

## 2. Boolean network model

Our Boolean network model follows a typical construction [15] motivated by biologically meaningful functions [22], the notion of threshold logic [23], and multiple models of Boolean networks [10, 24, 25] that have been used successfully in biology. The system state  $\vec{x}(t)$  is represented by an  $n$ -dimensional Boolean vector in  $\mathbb{B}^n = \{\text{True, False}\}^n$ , a finite space of dimension  $2^n$  that we refer as the *state space*. For simplicity, we will often use an integer representation of True as 1 and False as 0. We consider networks where interactions between the nodes are either null, inhibitory, or excitatory. We represent those interactions by a weighted adjacency matrix,  $\mathbf{A}$  where  $A_{ij}$  is 1 for an excitatory edge from node  $i$  to node  $j$ ,  $-1$  for an inhibitory one, or 0 if there is no interaction (see figure 1(B)).

The states of all nodes are updated in discrete time steps; all states are updated at once. At every time step, each node sums the excitatory and inhibitory inputs, and if that sum is greater than 0, then the node becomes active (1); otherwise, it becomes inactive (0). The update rule can be formally specified as follows:

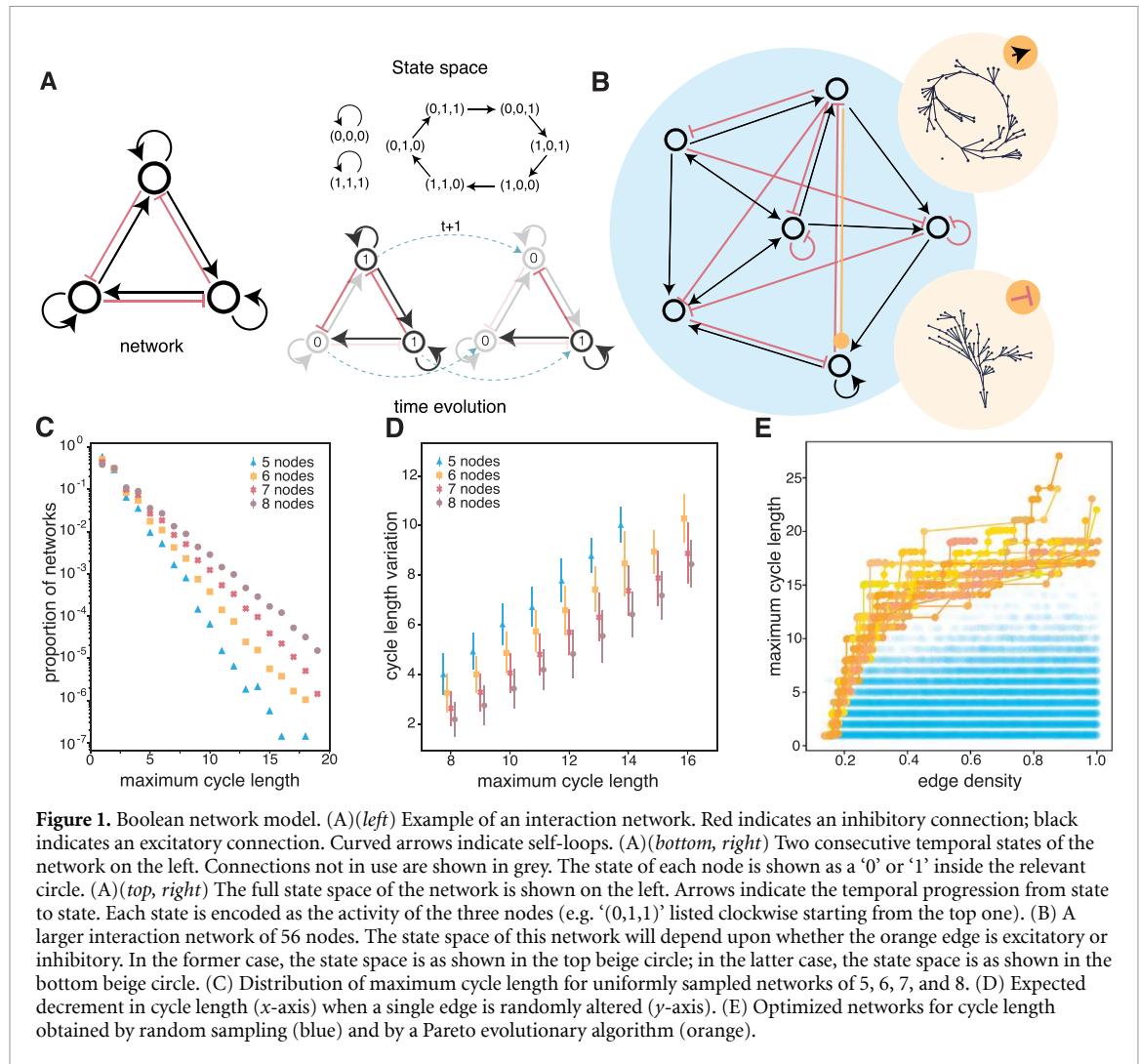
$$x_i(t+1) = \begin{cases} 1, & \text{if } \sum_j A_{ji} x_j(t) > 0 \\ 0, & \text{if } \sum_j A_{ji} x_j(t) \leq 0 \end{cases} \quad (1)$$

and we will write  $\tau(\vec{x}(t)) = \vec{x}(t+1)$ .

## 3. Numerical results

### 3.1. Sampling networks

To relate cycling dynamics to the topology of the interaction network, we began by considering random networks. We discovered that most random networks have a short cycle length (figure 1(C)), which is reflected in the exponential tail of the cycle-length distribution. The identification of cycles is an NP-hard problem [26]. Hence we limit our study of cycle lengths to networks containing eight or fewer nodes. Long cycles are also extremely rare. For instance, a random network of seven nodes has a cycle of length 19 with an approximate probability of one in a million (see figure 1(C)). We have also found that edge density is a determining factor in maximal cycle length, as denser networks are more likely to exhibit long cycles. Furthermore, altering the nature of a single edge (e.g. from excitatory to inhibitory) has a strong destructive effect on the maximal cycle length (see figure 1). We did not observe a correlation between the degree of asymmetry in the connection matrix and the length of cycles; in fact, we found that both the best and worst cycles often exhibit high degrees of symmetry, suggesting that the specific nature of the asymmetry is crucial in determining cycle length.

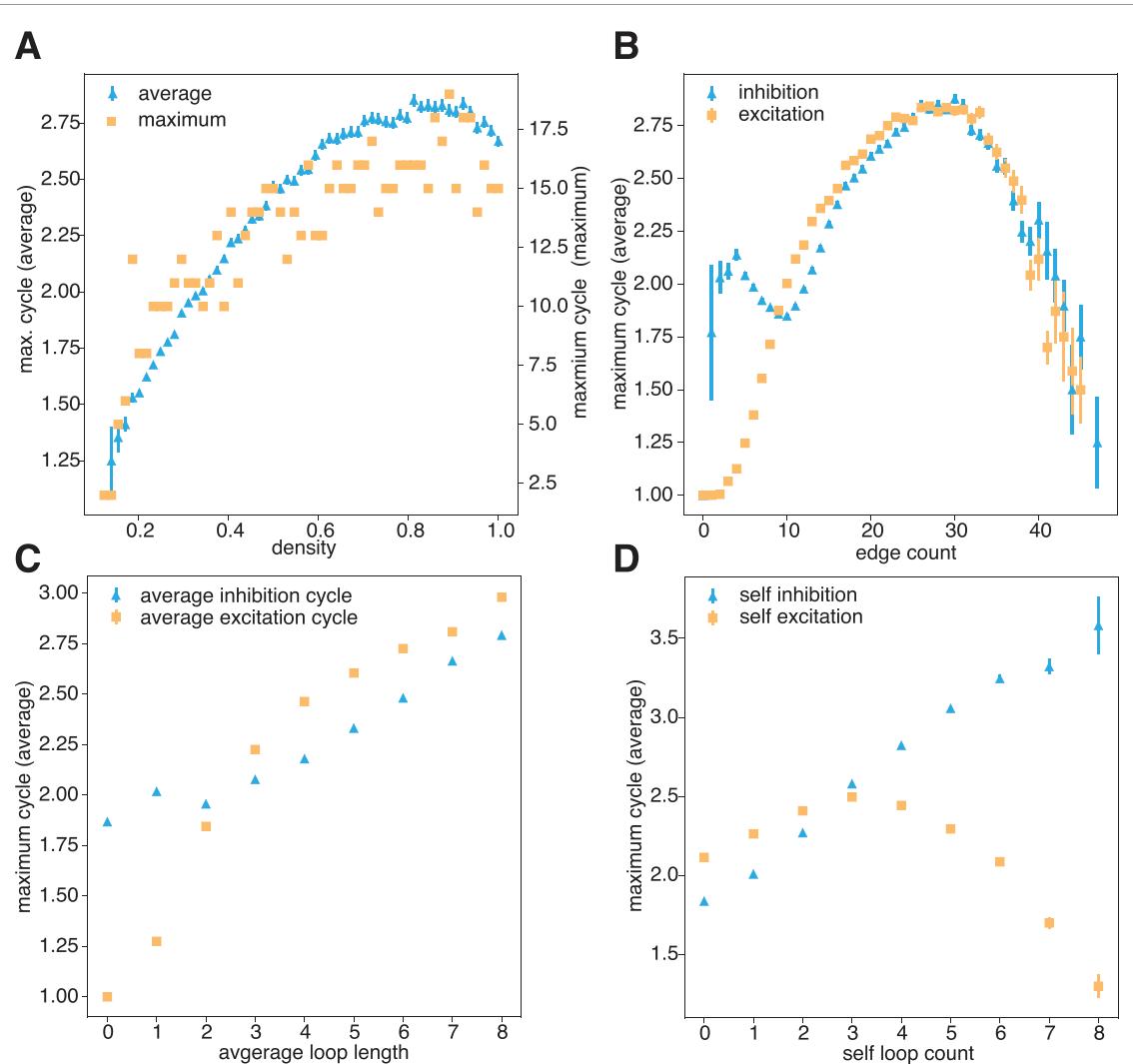


The apparent simplicity of this Boolean model belies surprising complexity. Unlike Hopfield networks, the lack of symmetry in the interaction matrix ( $a_{ij} \neq a_{ji}$ ) of Boolean networks implies the non-existence of a Lyapunov function, making them difficult to study analytically [27, 28]. Further, the inclusion of self-loops has been shown to increase the number and robustness of attractor states, thereby increasing the complexity of our model's dynamics [29]. For these reasons, such models are remarkably expressive and useful in explaining real biological observations such as cell differentiation [30].

We investigated the association between a network's maximum cycle length and its edge density to identify factors influencing cycle length. Our findings indicate that higher edge density leads to an increase in both average and maximum cycle lengths (see figure 2(A)). We observe that in random networks, a combination of inhibitory and excitatory edges results in longer cycles (see figure 2(B)). Furthermore, we found that the average maximum cycle length increases with the presence of excitatory or inhibitory circuits in the interaction network (see figure 2(C)). We also found that the specific distribution of excitatory and inhibitory connections, particularly in self-connections, differentially affected cycle length. Self-inhibition, which we defined as an inhibition circuit of length 1, was positively correlated with a high maximum cycle length. By contrast, self-excitation, which we defined as an excitatory circuit of length 1, was negatively correlated with a high maximum cycle length (see figure 2(D)).

To identify and study highly cyclic networks, we utilized an objective function that balances the maximization of cycle length and the minimization of network density. This trade-off is of significance in various domains. Biological systems often face energetic constraints on interactions or the physical pathways on which they rely [31, 32], while the boundaries of their design limit engineered systems. The goal of minimizing interactions is countered by the goal of facilitating a diverse dynamical repertoire [33–36].

We used a genetic Pareto algorithm [37] that encodes each network's genetic representation as a string of  $n^2$  characters in the set  $-1, 0, 1$  (see also figures A2(A)–(C)). This algorithm finds Pareto efficient solutions without the need to define one objective function encapsulating the two objectives, allowing us to analyze the



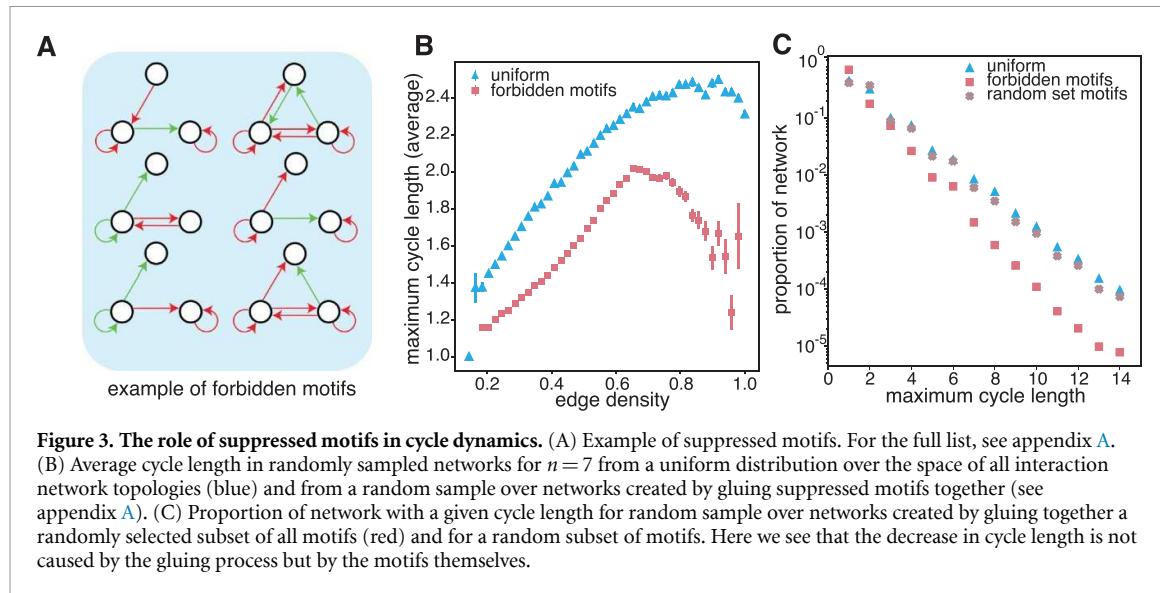
**Figure 2.** Topological properties of random networks. (A) The average and largest maximal cycle length as a function of density for an eight-node network. (B) The average maximal cycle length as a function of the number of edges of each type: excitatory and inhibitory. (C) The average maximal cycle length as a function of the average size of physical cycles within the interaction network. Note that the former is a dynamical property and the latter is an interaction property. (D) The average maximal cycle length as a function of the number of self-loops, as given by the diagonal entries in the adjacency matrix. All plots use a sample of 400 000 random networks.

entire landscape of optimal solutions. Using this genetic algorithm, we evolved random networks along the Pareto front and confirmed that we produced networks with larger maximum cycle lengths than equi-dense random networks (figure 1(E)); for further information regarding the structure of Pareto networks, see figures A2(D) and (E).

### 3.2. Suppressed motifs

We then investigated the factors contributing to the variation in the set of Pareto optimal networks compared to the general population of networks. We found significant differences in the networks' local topology, which differs markedly between evolved and random networks. We observed that the Pareto front networks' global properties—specifically, their density and average degree—were similar to those of random networks. Yet, their local topology was dramatically different. When considering 3-motifs, i.e. subsets of three nodes in the graph with their connections, we discovered a subgroup of around thirty (out of a possible 3284) 3-motifs that were almost completely absent in the evolved networks (see figures 3(A) and A1(A)). The existence of these *suppressed motifs* suggests a condition on the network's local connectivity that affects its evolved functionality.

To evaluate the impact of suppressed motifs on cycle length, we artificially created networks containing a high density of suppressed motifs (see the ). We found that the average cycle length was significantly lower than expected in random networks for all network densities ( $p < 0.0001$ ) (figure 3(B)). We also found that the density of networks with a given maximum cycle length was lower than expected in random networks



**Figure 3. The role of suppressed motifs in cycle dynamics.** (A) Example of suppressed motifs. For the full list, see appendix A. (B) Average cycle length in randomly sampled networks for  $n = 7$  from a uniform distribution over the space of all interaction network topologies (blue) and from a random sample over networks created by gluing suppressed motifs together (see appendix A). (C) Proportion of network with a given cycle length for random sample over networks created by gluing together a randomly selected subset of all motifs (red) and for a random subset of motifs. Here we see that the decrease in cycle length is not caused by the gluing process but by the motifs themselves.

(figure 3(C)). To determine the specificity of the observed behavior, we next constructed networks containing a high density of randomly selected 3-node motifs. In this new population, we observed only a 20% decrease in the density of networks with a long cycle length; this is in comparison to a more than 95% decrease for suppressed motifs networks (figure 3(C)); see figure A4 for  $p$ -values and confidence intervals). These findings suggest that a decremented cycle length is specific to suppressed motifs and is not purely an artifact of sampling over a limited number of motifs.

### 3.3. Dynamical reflection symmetry drives long cycles

To better understand what drives the existence of suppressed motifs, we exhaustively enumerated all networks for  $n < 5$  (figure 4(A)). This becomes rapidly pointless for larger  $n$  since the number of networks is approximately given by  $\frac{3^{n^2}}{n!}$ . For each  $n < 5$ , we identified the networks with the maximal cycle length (see appendix A). What do all of these networks have in common? We might naively posit that symmetry in the structure of the interaction is important, and indeed the optimized networks with  $1 < n < 4$  exhibit structural symmetry. However, we observed that the 4-node networks with the maximum cycle length were not structurally symmetric, motivating the need for a different explanation.

As an alternative, we considered a dynamical reflection symmetry that manifests in the network's state-space representation. Such a reflection symmetry permits the inverse of a sequence of states as another sequence. The inverse state is given by the  $\neg$  operator or the standard NOT gate. Under this operator, the state of four Boolean nodes  $\vec{x} = (0110)$  becomes the state  $\neg\vec{x} = \vec{1} - \vec{x} = (1001)$ . Then, when we say *dynamical* reflection symmetry, we mean that if the system's dynamics evolve to map  $x(t)$  to  $x(t+1)$ , then the system's dynamics also evolve to map  $\neg x(t)$  to  $\neg x(t+1)$  as follows:

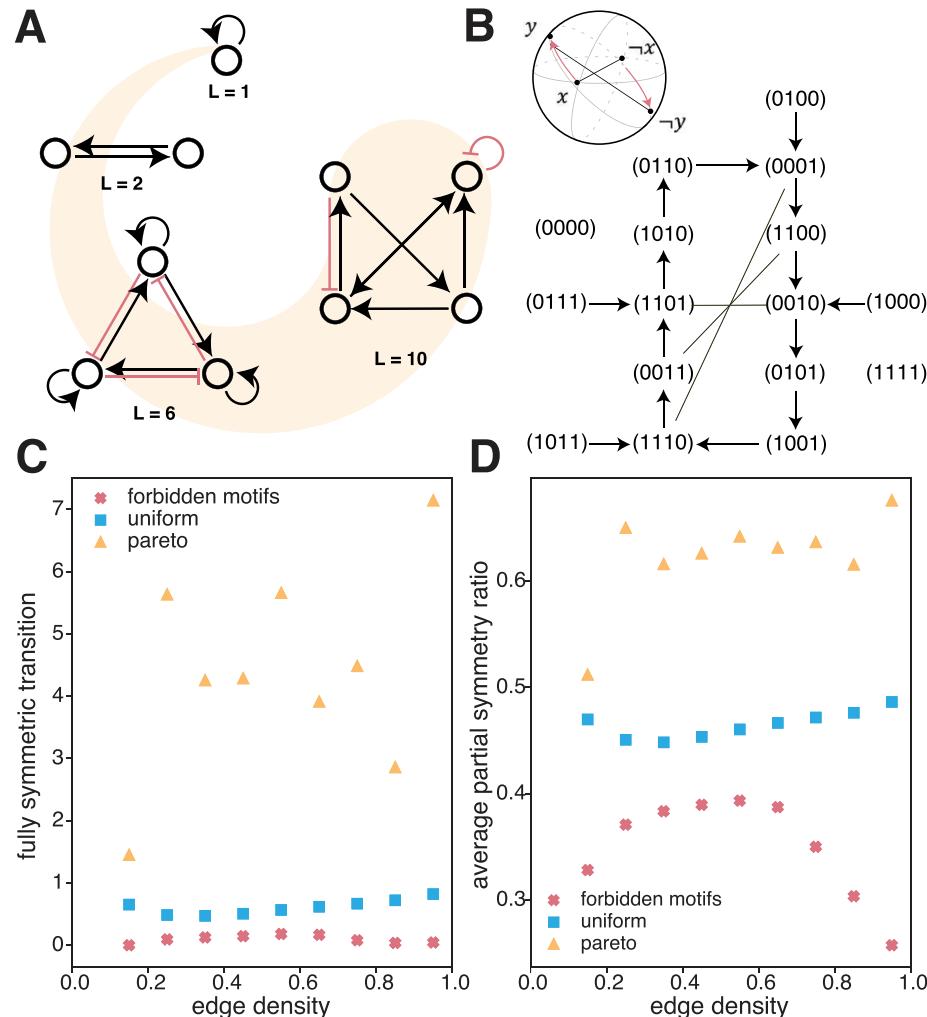
$$\tau(x) = y \iff \tau(\neg x) = \neg y. \quad (2)$$

We observed dynamical reflection symmetry in the network's state transition diagram in all of the  $1 < n < 5$  networks found to have maximal cycle lengths (figures 4(A) and (B)).

This observation motivates the question: Might reflection symmetry relate to suppressed motifs, and if so how? We found that the maximum cycle length was significantly lower (figure 4(C)) in the suppressed motif networks than in the random networks (pairwise two-sided  $z$ -test,  $p < 0.0001$ ). In contrast, our evolved networks—built to optimize the maximum cycle length—displayed a 3- to 7-fold increase in reflection symmetric transitions compared to random networks. These findings link reflection symmetry with the presence of suppressed motifs and a network's ability to display cyclic behaviors.

To better understand how reflection symmetry might relate to suppressed motifs, we returned to our artificially constructed networks containing a high density of suppressed motifs and measured the number of reflection-symmetric transitions. The number of perfectly symmetric transitions is small; hence, it is of value to also estimate partial symmetries. Specifically, we estimate the fraction of bits that respects symmetry. Specifically, given a Boolean network of size  $n$ , the partial symmetry ratio  $p_{sr}$  for this network is given by:

$$p_{sr} = \frac{1}{2^n \cdot n} \sum_x \sum_i \theta(|f(x)_i - f(\neg x)_i|),$$



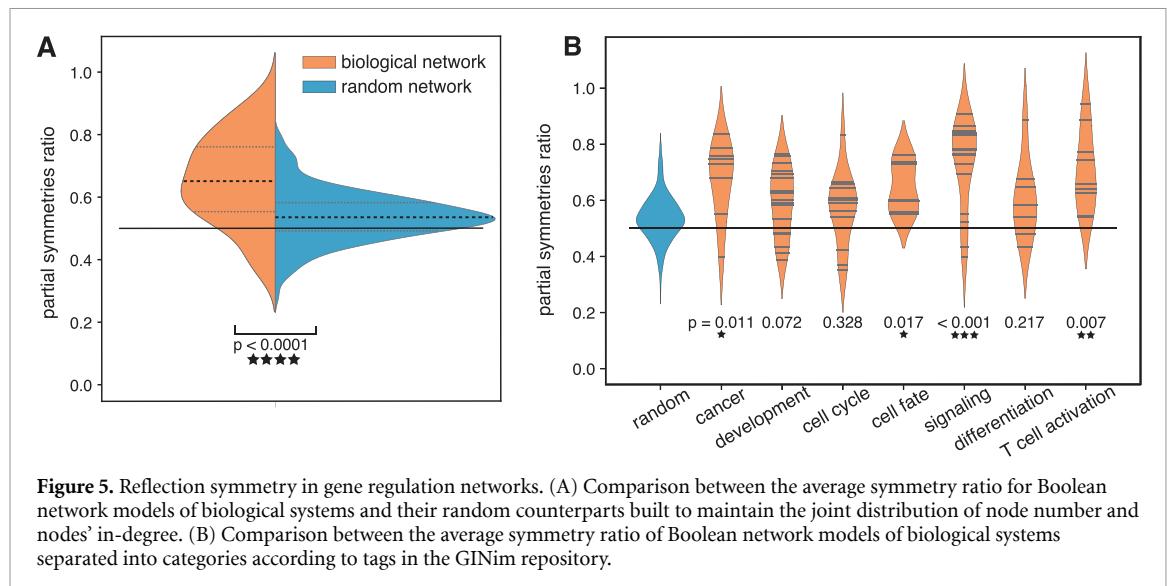
**Figure 4.** The role of symmetries in cycle dynamics. (A) Example networks of up to 4 nodes, with the highest cycle length possible in that network size. (B) The state space of the  $n = 4$  network shown in panel (A). The lines without arrowheads represent the states linked under the reflection symmetry (e.g. '(0,0,0,1)' is linked to '(1,1,1,0)'). The top right schematic shows conceptually how the reflection symmetry affects the system's dynamics where a state  $x$  is mapped under time evolution to the state  $y$ . (C) Average number of fully symmetric transitions. (D) We sample the average partial symmetry ratio, the fraction of bits that transition symmetrically, for random networks (blue), evolved networks (orange), and suppressed motif networks (red).

where  $\theta(x)$  takes the value 1 if  $x > 0$  and is zero elsewhere. With this broader definition of partial symmetry, we observed a similar trend in which evolved networks showed an increase in the partial symmetry ratio, whereas suppressed motif networks showed a decrease (figure 4(D)). These findings motivate further investigations into the causes that might drive correlations between dynamical reflection symmetry, suppressed motifs, and long cycles.

### 3.4. Real Boolean biological networks support reflection symmetries

Evolved networks that displayed a long maximum cycle length tended to express reflection-symmetric transitions more frequently than random networks. Interestingly, those transitions were not exclusively present inside the cycles but were also found in the non-cyclic part of the state space. This observation suggests that dynamical symmetries may play an even more profound role in evolved networks, and motivates investigation of their presence in broader categories of dynamical networks, both synthetic and natural. We turned to gene regulatory networks to test our hypothesis regarding the presence of reflection symmetries in biology. We used two repositories: the GINsim software [38] and the PyBoolNet python package [39].

The set of Boolean biological networks used contains 70 networks. Networks containing non-binary states were transformed to a binary representation using GINSIM. They have between 3 and 218 nodes, with an average of 33.7 nodes. The average connectivity is in the range of [1.18, 4.82] with an average of 2.55 connections. The 129 random Boolean networks were generated using the BoolNet package with uniform function generation, and uniform linkage based on Kauffman's method [40]. The number of genes was



**Figure 5.** Reflection symmetry in gene regulation networks. (A) Comparison between the average symmetry ratio for Boolean network models of biological systems and their random counterparts built to maintain the joint distribution of node number and nodes' in-degree. (B) Comparison between the average symmetry ratio of Boolean network models of biological systems separated into categories according to tags in the GINsim repository.

selected to reproduce the distribution found in the real network, and the function generation was done randomly using the homogeneous policy. The average number of inputs was selected to reproduce the average number found in the real networks with the given number of nodes.

Using these data, we observed that Boolean models of biological systems showed markedly more symmetries than random networks (figure 5(A)). Specifically, the average number of symmetric transitions in the biological model networks was 64.4%, whereas the average number in random networks was 53.6%. The slight deviation from 50% is due to the finite size of the analyzed networks. Further, the reflection symmetry ratio was significantly greater in the biological networks than in the random networks (two-sided *t*-test,  $t = 7.068$ ,  $p = 4.6e - 10$ ). In addition to this overall effect, we noted a marked variation across the different models: for 510 of the biological networks, less than 50% of transitions were symmetric; for 29 of them, more than 70% were symmetric; and for 12 of them, more than 80% were symmetric. To better understand this variability, we divided the models into biologically relevant categories using the tags provided in the GINsim repository. The seven most populated categories were retained for analysis (figure 5(B)). We observed that the most symmetrical categories were cell signaling, cell fate, cell activation, and cancer, with only a few networks lying under the 50% line.

By considering symmetries in Boolean networks, we illustrate a simple method to determine reflection symmetry and examine the overall development of the system. Notably, this concept of reflection symmetry can also be applied to non-binary state systems and even to systems where only part of the system's evolution is known. As an example, we have included a basic case study using continuously valued scRNA-seq data of cancer cells, demonstrating how these symmetries can be utilized to classify cells that respond to drugs versus those that are drug resistant (see appendix B).

## 4. Discussion

### 4.1. Functional role for reflection symmetry

In this study, we uncovered a new link between the expression of specific motifs and the existence of cycling behavior in Boolean networks. Potential links between these two properties have been reported previously for some gene regulatory networks [41]. For example, prior studies report a link between bi-fan motifs and cycling behavior [41] and find that chaotic motifs are linked to cycling [42]. Building upon these observations, we have shown that abnormally underrepresented motifs have a specific function as reflection-symmetry breakers. This relationship sheds light on how an interaction network with these motifs can maintain long cycles.

### 4.2. Diversity of dynamical symmetries

Here we investigated the problem of dynamical symmetries in Boolean networks. Many other types of symmetries exist. For example, one might seek to understand symmetries in the output function on each node and consider that the function  $f_1(\vec{x}) = (x_1 \text{ or } x_2)$  is invariant under permutation of its argument. In fact, output function symmetries at the node level are a powerful way to characterize complex network dynamics [43, 44]. Our work extends these observations by showing that symmetries at the global level can explain

some properties of networks with complex dynamics. Another common approach is the study of symmetric properties inside the interaction network, such as fibration symmetry [45, 46]. Our approach differs from these studies in evaluating symmetry in the state space [47], but nevertheless provides insights regarding a specific property of the interaction network. This property does not appear as a symmetry in the structure but as a balancing equation on each node. By taking a different perspective from prior studies, our approach sheds light on new mechanisms of dynamical symmetries and the function of the systems that support them.

#### 4.3. Methodological considerations

Several methodological considerations are pertinent to our work. First, there remains a conceptual and formal distinction between a system and its Boolean network model. Here, we have shown that biologically inspired Boolean networks display a high level of reflection symmetry and that not all biological processes have the same amount of symmetry. Yet, it remains unknown whether reflection symmetry is intrinsic to the system or a result of the map between a set of experiments and its Boolean network model. Future theoretical work could examine the impact of the mapping process on our findings. Further experimental work could examine the evolution of a simple biological system (e.g. yeast) and confirm the existence of reflection-symmetric states.

Boolean models of biological systems can include more general types of interactions and update rules than the ones we considered. Reflection symmetry can be easily identified in both thresholded and un-thresholded models. Furthermore, since every Boolean model maps binary states to binary states, our reflection symmetry definition is broadly applicable and does not depend on the nature of the updating scheme. The difference in models does affect the ability to analyze motifs since the model determines the possible types of interactions. Further work should examine this potential dependence and the broader impact of the updating schemes and the model on the condition necessary for mirror symmetry. Future work could also seek to generalize and test the theory of dynamical reflection symmetry for more general Boolean networks and multiple updating schemes and to random Boolean networks [19, 48]. Finally, various other deterministic network-based systems exhibit cyclical behavior, and it would be intriguing to investigate whether subclasses of motifs also exist in systems such as coupled maps on networks [49, 50] or evolutionary games on networks [51].

### 5. Conclusion

In identifying the important role of dynamical symmetry in Boolean networks, this work suggests strategies that can be used to engineer complex dynamical systems with particular dynamical features, or to modify an existing system's architecture to influence its properties. In particular, we have shown is possible to enhance or suppress cycling behaviors by altering only a small subset of edges in a system's interaction network topology. As applied to complex biophysical systems in pharmacology and microbiology, this understanding may aid the design of targeted diagnostics and therapeutics.

Our findings can also improve Boolean network modeling of real systems. In situations where cycling behavior is a defining system characteristic, the search for a suitable Boolean network model can be dramatically simplified by considering the symmetric subspace. This is crucial since it is impractical to explore the complete state space experimentally, even for small networks. Similarly, reflection-symmetric counterparts of observed trajectories can be added to training data sets, doubling the information available to machine learning models at no extra cost, analogously to mirroring images for neural network training.

In elucidating the formal relationship between reflection symmetry and cyclic behaviors, our work raises intriguing new questions for the scientific community. For example: What causes and regulates dynamical symmetries? Why does the degree of dynamical symmetry vary across different cellular and molecular processes? Which diseases or other perturbations to the system impact this symmetry? Is symmetry explained by evolution or caused by the physics of the system? Answers to these questions will profoundly impact our fundamental understanding of natural dynamical systems.

### Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/ouelletmathieu/BMS>.

### Acknowledgments

M O acknowledges the support of the Natural Sciences and Engineering Research Council of Canada (NSERC). J Z K acknowledges support from the NIH T32-EB020087. L C B and M O acknowledge support

from the National Science Foundation through the University of Pennsylvania Materials Research Science and Engineering Center (MRSEC) (DMR-1720530). D S B acknowledges support from the National Science Foundation (IIS-1926757, DMR-1420530), the Paul G Allen Family Foundation, and the Army Research Office (W911NF-16-1-0474, W011MF-191-244). The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

## Citation diversity statement

Recent work in several fields of science—including physics and biology, where our work here is situated [52–55]—has identified a bias in citation practices such that papers from women and other minority scholars are under-cited relative to the number of such papers in the field [56–60]. Here we sought to proactively consider choosing references that reflect the diversity of the field in thought, form of contribution, gender, race, ethnicity, and other factors. First, we obtained the predicted gender of the first and last author of each reference by using databases that store the probability of a first name being carried by a woman [54, 61]. By this measure (and excluding self-citations to the first and last authors of our current paper), our references contain 5.64% woman(first)/woman(last), 13.33% man/woman, 11.4% woman/man, and 69.64% man/man. This method is limited in that a) names, pronouns, and social media profiles used to construct the databases may not, in every case, be indicative of gender identity and b) it cannot account for intersex, non-binary, or transgender people. Second, we obtained predicted racial/ethnic category of the first and last author of each reference by databases that store the probability of a first and last name being carried by an author of color [62]. By this measure (and excluding self-citations), our references contain 29.6% author of color (first)/author of color(last), 11.7% white author/author of color, 16.78% author of color/white author, and 41.93% white author/white author. This method is limited in that a) names and Florida Voter Data to make the predictions may not be indicative of racial/ethnic identity, and b) it cannot account for Indigenous and mixed-race authors or those who may face differential biases due to the ambiguous racialization or ethnicization of their names. We look forward to future work that could help us to better understand how to support equitable practices in science.

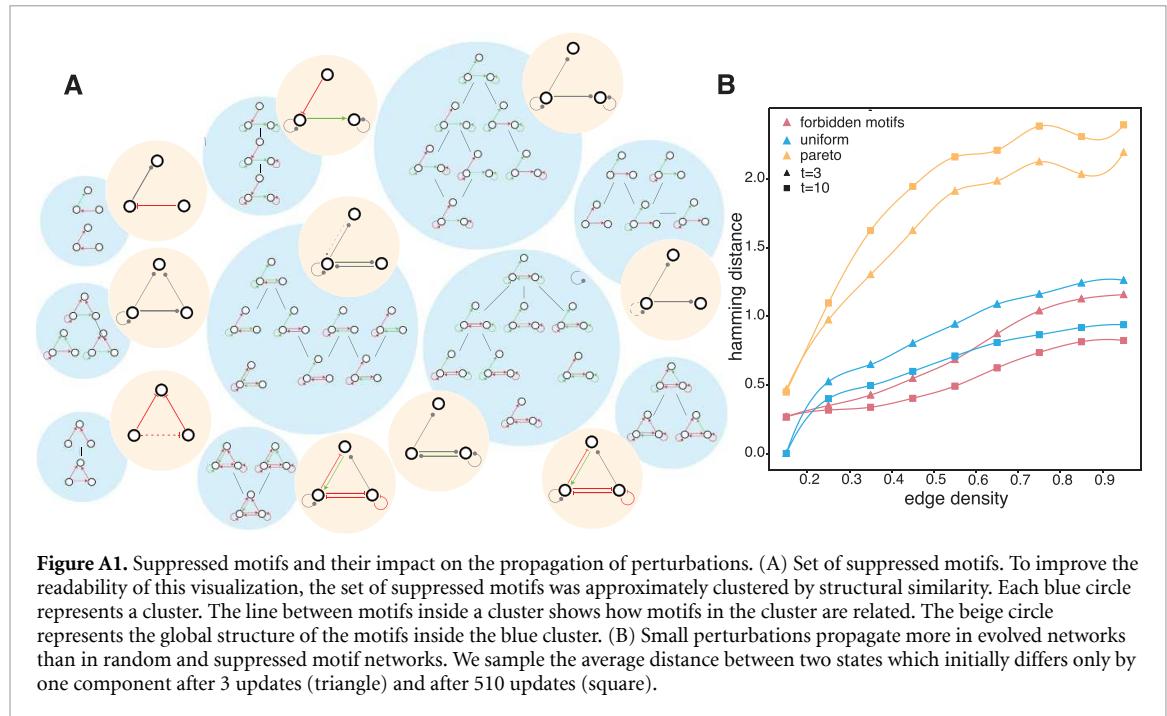
## Appendix A. Sampling

We sample the space of Boolean networks in two steps. First, a density is chosen uniformly at random from the range 0.2 to 1. Second, we select elements in the matrix and fill those elements with either a  $-1$  or  $+1$ —with equal probability—until we reach the target density. For the biased case, we again begin by choosing a density randomly from 0.2 to 1. Then, we select valid motifs randomly from the set of desired motifs and place them randomly in the interaction network. If we cannot find a position for a given motif, then we discard that motif and randomly draw another from the same set. Once the desired ratio of motifs is achieved, the network is filled with random edges to achieve the global density. Edges used for a motif (empty edges included) are protected and cannot be filled in this part of the process.

### A.1. Evolved networks

In our evolutionary algorithm, we set the population size to 600 interaction networks with a density taken from a uniform distribution ranging from 0.2 to 1.0. The selection for mating is handled by the NSGA2 algorithm [37]. In each step, 400 new offspring are created using two mating operators (see next section for details). We mutate each edge of the offspring's interaction networks with a probability of  $P = 0.02$ . When a component is selected for mutation, its value is changed to one of the two other values in  $\{-1, 0, 1\}$ . The generation process is then repeated until convergence. The whole process is repeated 30 times.

We use two mating operators that reflect the exploratory and convergent mating strategies, respectively. Both operators are used with equal probability each time we call the mating procedure. The first mating operator generates two offspring. For each index of the matrix, each offspring gets assigned either the first or second parent's component. If the first offspring receives its component from parent two, then the second offspring receives its component from parent one. Then both matrices are tested for validity; if one is found to be disconnected when considering the interaction matrix as an undirected graph, the process is restarted. The second mating operator generates one offspring at a time. We list the cycle basis of both parents' interaction networks. Then, random cycle structures are picked from the cycle basis to build a new interaction network containing a random mixture of cycles from both parents. Only cycles that agree for all of their common inhibition or excitation edges are selected. If two selected cycles cannot be joined because of disagreeing edges, the process is restarted. We use the first mating operator if the process fails more than 1000 times. Such failures happen but are rare and represent a negligible percentage of the offspring production.



Interestingly, each node in the evolved networks had approximately 20% more incoming excitatory edges than inhibitory edges. The evolved networks also exhibited nearly twice as many physical excitation circuits as the random networks (see figure A2(D)). These features were emergent properties, as the evolutionary algorithm did not explicitly optimize for them. In addition to comparing all evolved networks to all random networks, we separately examined the dynamics of evolved networks that implemented the exploratory versus the convergent mating strategy. We found that convergent mating consistently produced networks with greater maximum cycle length than exploratory mating (see suppl. figure A2(C)). In general, the structure of the two parents does not generate offspring with long cyclic behavior. Even for the convergent mating strategy, most cycles from both parents do not act cooperatively, generating important attractors leading to fixed points.

#### A.2. Motif characterization

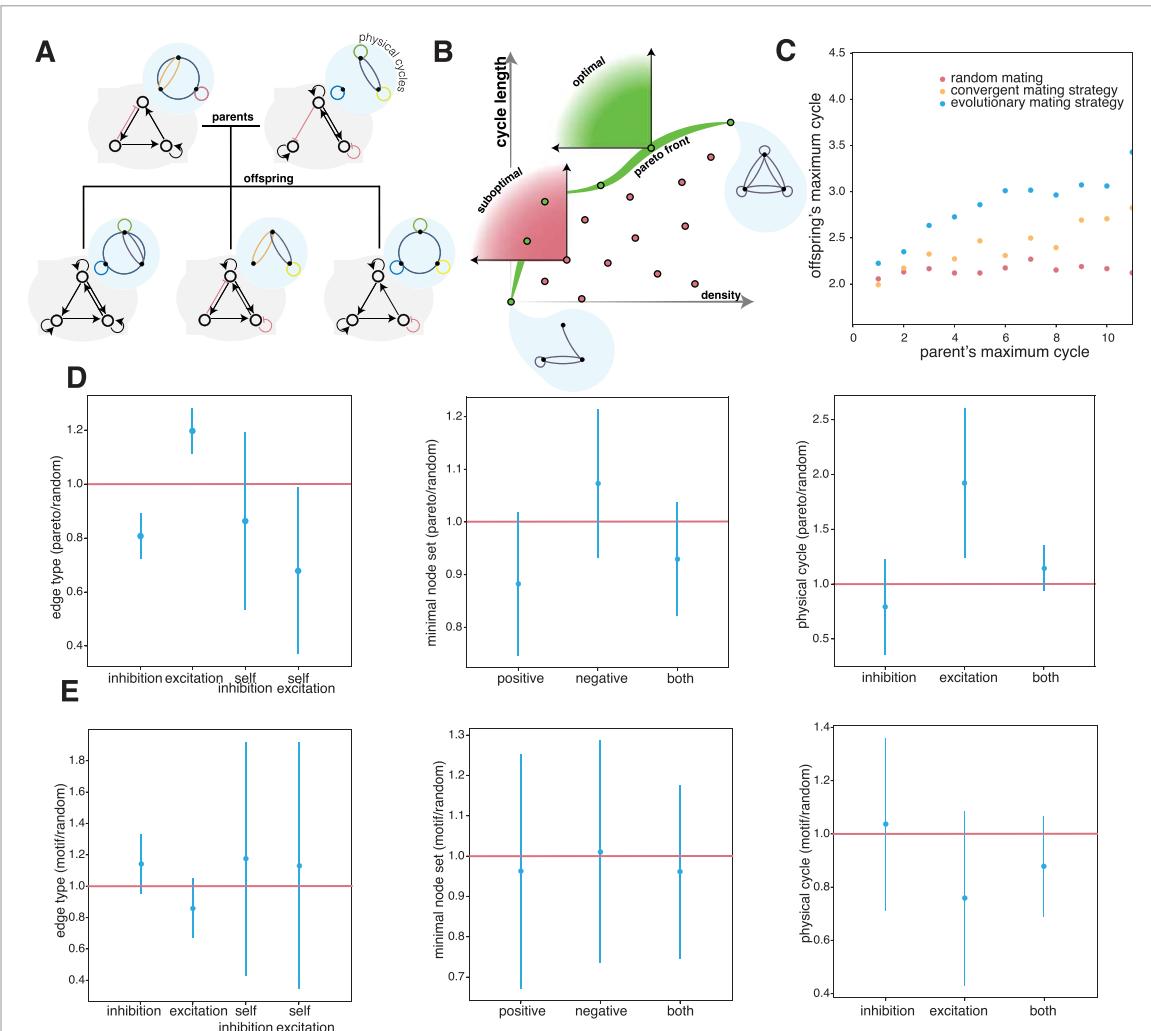
To calculate the  $z$ -value for the number of expected motifs, we used sampled data to obtain each motif's average expected number and its expected standard deviation. Since the expected number for 3-motifs is always one order of magnitude less than the standard deviation, we consider motifs with an individual  $z$ -score of less than  $-0.1$  to be suppressed. This repression typically corresponds to the absence of the motif in the graph. We then consider a given motif to be suppressed in the population if it is tagged as suppressed in at least 50% of the population's individual graphs.

#### A.3. Reflection symmetry

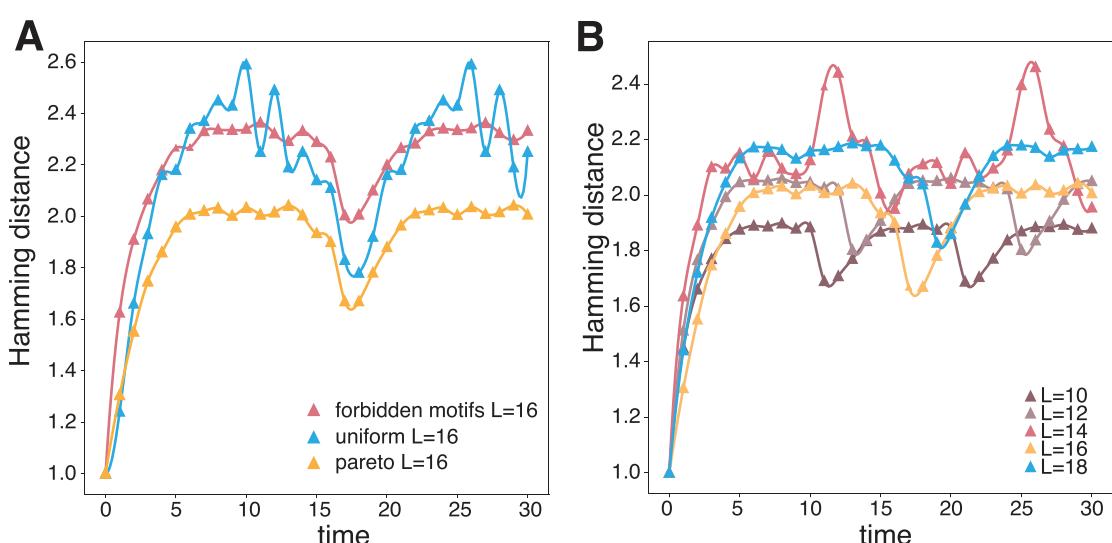
Since for each  $n \leq 4$ , there was at least one network with the maximal possible cycle length with this symmetry, it is quite tempting to conjecture that there always exists a maximal network with this property. This, of course, cannot be confirmed computationally and could only be realistically verified up to  $n = 5$  without developing a more sophisticated search algorithm.

#### A.4. Robustness

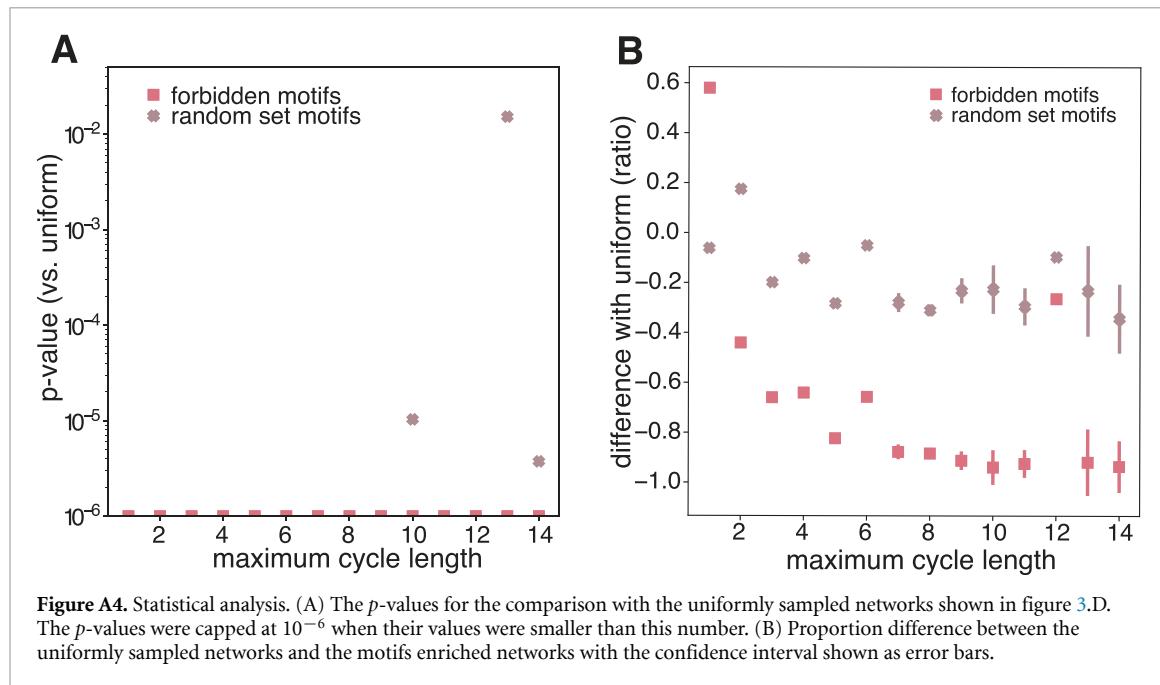
Another frequently cited property of the evolved networks is their robustness to perturbation. We sought to determine if the evolved networks and the suppressed motif networks had distinct levels of robustness. The notion of robustness depends upon a notion of response to perturbation. The distance between the non-perturbed and perturbed states is measured at each time step by the Hamming distance, which only counts the number of non-matching bits. A robust network is one where the distance stays constant or decreases over time. We expect evolved networks to show an increase in robustness, and we expect suppressed motif networks to show a decrease in robustness. The results are surprising: for a large range of edge densities, perturbations to suppressed motif networks do not drive a large divergence in state dynamics in comparison to uniform random networks (figure A1).



**Figure A2.** Offspring generation and properties of suppressed motifs. (A) Offspring generation is based on cycle preservation. (B) Example of the approximated Pareto front where green points are optimal in the Pareto sense, and where red points are sub-optimal. (C) Evaluation of the different mating strategies or policies that generate offspring. (D) Ratio of the number of different physical constituents of the Pareto front networks versus random networks. (E) Ratio of the number of different physical constituents for the suppressed motif networks versus the random networks.



**Figure A3.** Time evolution of perturbations. (A) Average Hamming distance from the 1 bit perturbed network to the unperturbed network as a function of the time for the Pareto front networks binned by cycle length. (B) Same as in panel (A) but for a maximal cycle length of 16 and plotted separately for Pareto front networks, random networks, and suppressed motif networks.



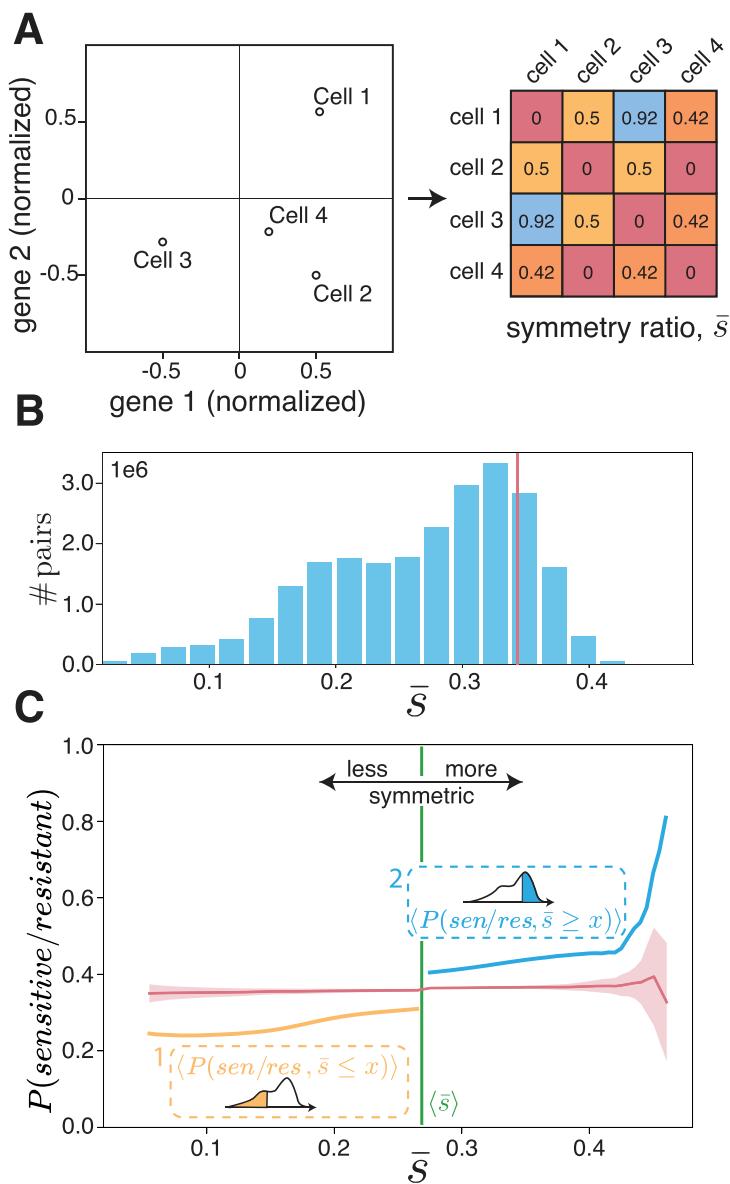
Intuitively, one might imagine that a network's sensitivity to perturbation could depend upon the maximum cycle length, such that the longer the maximum cycle the greater the tendency for perturbations to drive divergent dynamics. In contrast to this intuition, however, random and suppressed motif networks have quite different average cycle lengths. The difference could then be entirely caused by the presence of long cycles in the evolved population in comparison with the two other populations. To test this possibility, we have compared networks from each category with the same cycle length. For a given maximum cycle length, evolved networks have a slightly lower distance than the random and the suppressed networks for a given cycle length (see figure A3(A)). The difference however is quite small and is cyclic with the period equal to the cycle length, as expected (see figure A3(B)). Therefore, suppressed motifs and the optimization process have, for a given maximum cycle length, a rather low impact on how the networks handle perturbations.

## Appendix B. Deriving reflection symmetry directly from data

Boolean networks permit a simple, general definition of dynamical symmetry. However, Boolean models are not available for many natural dynamical systems, and creating them requires extensive work. Hence, here we analyze symmetry directly from data, circumventing the challenges of Boolean model construction.

We considered the problem of drug resistance in melanoma at the single-cell level. The state of gene expression can predict which cells will resist therapy and which ones will be sensitive and responsive to therapy. Notably, this prediction is possible before treating the cells with any drug. Thus, we define the state of a cell as a 'primed' resistant state if that cell would be resistant upon the application of a drug [63]. For this experiment, we profiled the gene expression of these cells using scRNA-seq. We then divided the cells into two categories: the drug-sensitive cells and the cells primed for drug resistance [63].

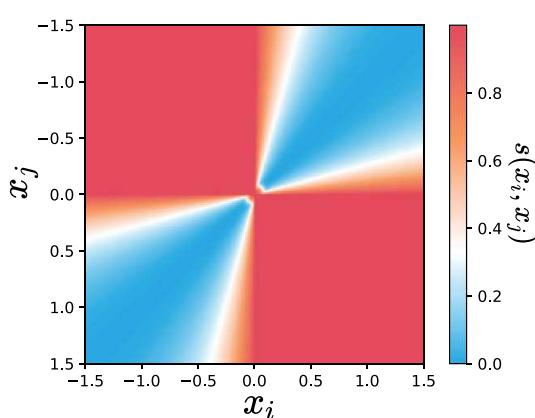
To apply the idea of symmetry to these data, we normalized the expression of each gene. Specifically, negative (positive) normalized quantities represented a lower (higher) expression level than the average expression found for a given gene in all the cells tested. Genes that are highly expressed in more than 70% of the cell are removed as these are typically housekeeping genes unrelated to the process of interest. We also remove genes that are not expressed in at least 25% of the cells since it becomes too hard to find two cells having the same gene expressed. Genes with zero expression are changed to NaN. Finally, genes with a decreasing expression distribution, i.e. where most expression values are close to 0, are removed because of their lack of suitability for defining a symmetry axis. We then defined a parity metric to determine whether two cells had a symmetric expression at the gene level. This metric quantifies how opposite the same gene expressions are for two different cells. For each gene, the maximum of the distribution of each gene's expression over all cells is defined as the zero expression level. Redefining the zero expression allows us to think of the expression of each gene as being more (positive value) or less (negative value) than the most common quantity found in the cells. The maximum of the distribution was preferred over the average since the distributions have highly zero-inflated counts due to the experimental technique used. The expression



**Figure B5.** Reflection symmetry in cancerous cells. (A) An illustrative example of the reflection symmetry ratio for four cells that express two genes. (Left) The normalized level of expression for all four cells and both genes. (Right) A symmetric matrix showing the computed reflection symmetry ratio between every pair of cells. In this example, cells 1 and 3 are highly reflection-symmetric, whereas cells 2 and 4 are not reflection-symmetric (asymmetric). (B) The distribution of the reflection symmetry ratio  $\bar{s}$  for all pair of cells. The red line indicates the expectation for uncorrelated, normally distributed genes. (C) The probability for a randomly selected pair of cells to have different states (drug-sensitive state/drug-resistant state) as a function of the reflection symmetry ratio  $\bar{s}$ . The orange curve shows that the probability for pair of cells with a reflection symmetry ratio  $\bar{s}$  less than  $x$  as shown in Inset 1. The blue curve shows the probability for pair of cells with a reflection symmetry ratio  $\bar{s}$  higher than  $x$  as shown in Inset 2. The red curve shows the probability obtained by random permutations of the labels for both cases. The green line shows the average reflection symmetry ratio over all the pairs of cells.

value for each gene is then scaled by the standard deviation. Let  $x$  be a gene,  $i, j$  be a pair of cells, and  $s(i, j)$  be the reflection symmetry ratio. We ask for the function  $s(i, j)$  to be 1 for a perfectly symmetric expression, i.e.  $x_i = -x_j$ . We also desire that the function will output a 0 if  $|x_i| \approx 0$  and  $|x_j| \approx 0$  even if  $x_i = -x_j$ . This condition allows us to avoid attributing some symmetry properties to noise by requiring a difference of at least 20% of the standard deviation between the two expressions  $x_i$  and  $x_j$ . Finally, we want the function to output a 0 when  $x_i$  and  $x_j$  are of the same sign. One possible function with these properties is given by

$$s(i, j) = 1 - \max \left\{ \left| \frac{x_i + x_j}{|x_i| + |x_j| + \epsilon} \right|^{h_1}, \frac{1 - \tanh h_2 (r_{x,ij} - r)}{2} \right\},$$



**Figure B6.** Reflection symmetry ratio  $s(i,j)$  given two gene expression values:  $x_i$  and  $x_j$ . For symmetric values, the ratio is 1 and collapses rapidly to 0 when the two values have the same sign.

where  $h_1, h_2$  are parameters that select the smoothness of the transition between the values of 0 and 1, and  $r_{x,ij} = \sqrt{x_i^2 + x_j^2}$ . The parameters used for the indicator are  $h_1 = 1.7$ ,  $r = 0.1$ ,  $h_2 = 25$ , and  $\epsilon = 1 \times 10^{-7}$ . Figure B6 shows the reflection symmetry ratio for these parameters.

We observed that the distribution of the parity metric spanned from 0 (non-symmetric pairs) to 0.4 (symmetric pairs). The distribution was also bi-modal, suggesting the existence of two pair types (see figure B5(B)). The expectation value of the parity metric for two normally distributed gene expression measurements is 0.343; by contrast, we find that the observed gene expression measurements are skewed toward non-symmetry. With these data and metrics in hand, we studied the symmetrical opposition of the two different states (drug-sensitive or primed). Specifically, we computed the probability that a pair of cells respecting a given criterion were of different states. When considering all pairs independently from their reflection symmetry, we observed a probability of approximately 35% for a pair to have different states. Therefore, we have approximately a one-in-three chance of picking a drug-sensitive and a primed cell when picking a pair of cells at random in our data set. We expected pairs with a high (low) reflection symmetry ratio to have a higher (lower) probability of residing in different paths (drug-sensitive vs. primed). Our expectation was confirmed: pairs with a low reflection symmetry ratio had a probability of around 25% to be of different states versus 80% for the high reflection symmetry ratio pairs (see figure B5(C)). Broadly, these data demonstrate that reflection symmetry correlates with drug resistance in cancer and could be used to classify a cell as drug-resistant or drug-responsive.

The application of symmetry analysis to gene expression inside melanoma scRNA-seq data is a simple example of the potential of considering symmetry in biology. However, the drug-sensitive and drug-resistant paths were not directly observed in the cells where the gene expression was measured. The labeling was inferred by principal component analysis of the cells' gene expression, where the variance was found to be principally located on one axis. This axis was later found to correlate highly with sensitivity and drug resistance. Additional studies could be conducted on data sets that are not linearly separable using one axis and on data expressing more than a binary category.

## ORCID iDs

Mathieu Ouellet  <https://orcid.org/0000-0002-0651-4817>

Jason Z Kim  <https://orcid.org/0000-0002-3970-4561>

Sydney M Shaffer  <https://orcid.org/0003-3539-9886>

Lee C Bassett  <https://orcid.org/0001-8729-1530>

Dani S Bassett  <https://orcid.org/0000-0002-6183-4493>

## References

- [1] Chen H, Wang G, Simha R, Du C and Zeng C 2016 Boolean models of biological processes explain cascade-like behavior *Sci. Rep.* **6** 20067
- [2] Berdahl A, Shreim A, Sood V, Paczuski M and Davidsen J 2009 Random sampling versus exact enumeration of attractors in random Boolean networks *New J. Phys.* **11** 043024
- [3] Iguchi K, Kinoshita S-i and Yamada H S 2007 Boolean dynamics of Kauffman models with a scale-free network *J. Theor. Biol.* **247** 138

[4] Sinha S, Jones B M, Traniello I M, Bukhari S A, Halfon M S, Hofmann H A, Huang S, Katz P S, Keagy J and Lynch V J *et al* 2020 Behavior-related gene regulatory networks: a new level of organization in the brain *Proc. Natl Acad. Sci.* **117** 23270

[5] Klemm K and Bornholdt S 2005 Topology of biological networks and reliability of information processing *Proc. Natl Acad. Sci.* **102** 18414 10.1073/pnas.0509132102

[6] Forger D B 2017 *Biological Clocks, Rhythms and Oscillations: The Theory of Biological Timekeeping* (MIT Press)

[7] Kobayashi T, Mizuno H, Imayoshi I, Furusawa C, Shirahige K and Kageyama R 2009 The cyclic gene Hes1 contributes to diverse differentiation responses of embryonic stem cells *Gen. Dev.* **23** 1870

[8] Wang D-G, S Wang, Huang B and Liu F 2019 Roles of cellular heterogeneity, intrinsic and extrinsic noise in variability of p53 oscillation *Sci. Rep.* **9** 5883

[9] Nash M S, Young K W, Challiss R J and Nahorski S R 2001 Receptor-specific messenger oscillations *Nature* **413** 381

[10] Pinho R, Borenstein E and Feldman M W 2012 Most networks in Wagner's model are cycling *PLoS One* **7** e34285

[11] Aracena J 2008 Maximum number of fixed points in regulatory Boolean networks *Bull. Math. Biol.* **70** 1398

[12] Mori F and Mochizuki A 2017 Expected number of fixed points in Boolean networks with arbitrary topology *Phys. Rev. Lett.* **119** 028301

[13] Aracena J, Richard A and Salinas L 2017 Number of fixed points and disjoint cycles in monotone Boolean networks *SIAM J. Discrete Math.* **31** 1702

[14] Raj A and Van Oudenaarden A 2008 Nature, nurture, or chance: stochastic gene expression and its consequences *Cell* **135** 216

[15] Chen L, Kulasiri D and Samarasinghe S 2018 A novel data-driven Boolean model for genetic regulatory networks *Front. Physiol.* **9** 1328

[16] Campbell C and Albert R 2014 Stabilization of perturbed Boolean network attractors through compensatory interactions *BMC Syst. Biol.* **8** 1

[17] Coppersmith S, Kadanoff L P and Zhang Z 2001 Reversible Boolean networks I: distribution of cycle lengths *Physica D* **149** 11

[18] Ouma W Z, Pogacar K and Grotewold E 2018 Topological and statistical analyses of gene regulatory networks reveal unifying yet quantitatively different emergent properties *PLoS Comput. Biol.* **14** e1006098

[19] Drossel B, Mihaljev T and Greil F 2005 Number and length of attractors in a critical Kauffman model with connectivity one *Phys. Rev. Lett.* **94** 088701

[20] Somogyvari Z and Payrits S 2000 Length of state cycles of random Boolean networks: an analytic study *J. Phys. A: Math. Gen.* **33** 6699

[21] Mangan S and Alon U 2003 Structure and function of the feed-forward loop network motif *Proc. Natl Acad. Sci.* **100** 11980

[22] Raeymaekers L 2002 Dynamics of Boolean networks controlled by biologically meaningful functions *J. Theor. Biol.* **218** 331

[23] Hu S-T 1965 *Threshold Logic*, in *Threshold Logic* (University of California Press)

[24] Greil F and Drossel B 2007 Kauffman networks with threshold functions *Eur. Phys. J. B* **57** 109

[25] Wacker B W, Velcovs M T and Rogers J A 2020 Boolean network topologies and the determinative power of nodes *J. Complex Netw.* **8** cnaa003

[26] Bridoux F, Gaze-Maillot C, Perrot K and Sené S 2021 Complexity of limit-cycle problems in Boolean networks *Int. Conf. on Current Trends in Theory and Practice of Informatics* (Springer) pp 135–46

[27] Hopfield J J 1982 Neural networks and physical systems with emergent collective computational abilities *Proc. Natl Acad. Sci.* **79** 2554

[28] Hopfield J J 2007 Hopfield network *Scholarpedia* **2** 1977

[29] Montagna S, Braccini M and Roli A 2020 The impact of self-loops on Boolean networks attractor landscape and implications for cell differentiation modelling *IEEE/ACM Trans. Comput. Biol. Bioinform.* **1** 2702–2713

[30] Braccini M, Montagna S and Roli A 2018 Self-loops favour diversification and asymmetric transitions between attractors in Boolean network models *Italian Workshop on Artificial Life and Evolutionary Computation* vol 1 (Springer) pp 30–41

[31] Bullmore E and Sporns O 2012 The economy of brain network organization *Nat. Rev. Neurosci.* **13** 336

[32] Wu G, Yan Q, Jones J A, Tang Y J, Fong S S and Koffas M A G 2016 Metabolic burden: cornerstones in synthetic biology and metabolic engineering applications *Trends Biotechnol.* **34** 652

[33] Zañudo J G and Albert R 2013 An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks *Chaos* **23** 025111

[34] Egger R, Tupikov Y, Elmalem M, Katlowitz K A, Benzeira S E, Picardo M A, Moll F, Kornfeld J, Jin D Z and Long M A 2020 Local axonal conduction shapes the spatiotemporal properties of neural sequences *Cell* **183** 537

[35] Jaeger J and Crombach A 2012 Life's attractors: understanding developmental systems through reverse engineering and in silico evolution *Adv. Exp. Med. Biol.* **751** 93

[36] Yamamoto H *et al* 2018 Impact of modular organization on dynamical richness in cortical networks *Sci. Adv.* **4** eaau4914

[37] Deb K, Pratap A, Agarwal S and Meyarivan T 2002 A fast and elitist multiobjective genetic algorithm: NSGA-II *IEEE Trans. Evol. Comput.* **6** 182

[38] Naldi A, Hernandez C, Abou-Jaoudé W, Monteiro P T, Chaouiya C and Thieffry D 2018 Logical modeling and analysis of cellular regulatory networks with GINsim 3.0 *Front. Physiol.* **9** 646

[39] Klarner H, Streck A and Siebert H 2017 PyBoolNet: a python package for the generation, analysis and visualization of boolean networks *Bioinformatics* **33** 770

[40] Kauffman S A 1969 Metabolic stability and epigenesis in randomly constructed genetic nets *J. Theor. Biol.* **22** 437

[41] Burda Z, Krzywicki A, Martin O C and Zagorski M 2011 Motifs emerge from function in model gene regulatory networks *Proc. Natl Acad. Sci.* **108** 17263

[42] Zhang Z, Ye W, Qian Y, Zheng Z, Huang X, Hu G and Kurths J 2012 Chaotic motifs in gene regulatory networks *PLoS One* **7** e39355

[43] Hosseini S, Reichl M D and Bassler K E 2014 Symmetry in critical random Boolean network dynamics *Phys. Rev. E* **89** 042808

[44] Stearns R E, Ravi S, Marathe M V and Rosenkrantz D J 2019 Symmetry properties of nested canalizing functions *Discrete Math. Theor. Comput. Sci.* **21** 1–17

[45] Morone F, Leifer I and Makse H A 2020 Fibration symmetries uncover the building blocks of biological networks *Proc. Natl Acad. Sci.* **117** 8306

[46] Leifer I, Morone F, Reis S D, Andrade J S Jr, Sigman M and Makse H A 2020 Circuits with broken fibration symmetries perform core logic computations in biological networks *PLoS Comput. Biol.* **16** e1007776

[47] Cantor Y, Khan B and Dombrowski K 2014 Towards a formal understanding of Bateson's rule: chromatic symmetry in cyclic Boolean networks and its relationship to organism growth and cell differentiation *Proc. Comput. Sci.* **36** 476

[48] Liu M and Bassler K E 2006 Emergent criticality from coevolution in random Boolean networks *Phys. Rev. E* **74** 041910

[49] Kaneko K 1992 Overview of coupled map lattices *Chaos* **2** 279

[50] Killingback T, Loftus G and Sundaram B 2013 Competitively coupled maps and spatial pattern formation *Phys. Rev. E* **87** 022902

[51] Killingback T and Doebeli M 1998 Self-organized criticality in spatial evolutionary game theory *J. Theor. Biol.* **191** 335

[52] Chatterjee P and Werner R M 2021 Gender disparity in citations in high-impact journal articles *JAMA Netw. Open* **4** e2114509

[53] Fulvio J M, Akinnola I and Postle B R 2021 Gender (im)balance in citation practices in cognitive neuroscience *J. Cogn. Neurosci.* **33** 3

[54] Dworkin J D, Linn K A, Teich E G, Zurn P, Shinohara R T and Bassett D S 2020 The extent and drivers of gender imbalance in neuroscience reference lists *Nat. Neurosci.* **23** 918

[55] Teich E G *et al* 2021 Citation inequity and gendered citation practices in contemporary physics (arXiv:2112.09047)

[56] Wang X, Dworkin J D, Zhou D, Stiso J, Falk E B, Bassett D S, Zurn P and Lydon-Staley D M 2021 Gendered citation practices in the field of communication *Ann. Int. Communun. Assoc.* **1** 134–153

[57] Dion M L, Sumner J L and Mitchell S M 2018 Gendered citation patterns across political science and social science methodology fields *Polit. Anal.* **26** 312

[58] Mitchell S M, Lange S and Brus H 2013 Gendered citation patterns in international relations journals *Int. Stud. Perspect.* **14** 485

[59] Maliniak D, Powers R and Walter B F 2013 The gender citation gap in international relations *Int. Organ.* **67** 889

[60] Caplar N, Tacchella S and Birrer S 2017 Quantitative evaluation of gender bias in astronomical publications from citation counts *Nat. Astron.* **1** 0141

[61] Zhou D, Cornblath E J, Stiso J, Teich E G, Dworkin J D, Blevins A S and Bassett D S Gender diversity statement and code notebook v1.0 2020

[62] Sood G and Laohaprapanon S 2018 Predicting race and ethnicity from the sequence of characters in a name (arXiv:1805.02109)

[63] Shaffer S M, Dunagin M C, Torborg S R, Torre E A, Emert B, Krepler C, Beqiri M, Sproesser K, Brafford P A and Xiao M 2017 Rare cell variability and drug-induced reprogramming as a mode of cancer drug resistance *Nature* **546** 431