Predicting attractors from spectral properties of stylized gene regulatory networks

Dzmitry Rumiantsau, Annick Lesne, Marc-Thorsten Hütt

Predicting attractors from spectral properties of stylized gene regulatory networks

Dzmitry Rumiantsau,[1,*] Annick Lesne,[2,3,†] and Marc-Thorsten Hütt[1,‡]
[1]*Department of Life Sciences and Chemistry, Constructor University, D-28759 Bremen, Germany*
[2]*Sorbonne Université, CNRS, Laboratoire de Physique Théorique de la Matière Condensée, LPTMC, F-75252 Paris, France*
[3]*Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, F-34293 Montpellier, France*

(Received 12 July 2022; revised 19 January 2023; accepted 7 April 2023; published 5 July 2023)

How the architecture of gene regulatory networks shapes gene expression patterns is an open question, which has been approached from a multitude of angles. The dominant strategy has been to identify nonrandom features in these networks and then argue for the function of these features using mechanistic modeling. Here we establish the foundation of an alternative approach by studying the correlation of network eigenvectors with synthetic gene expression data simulated with a basic and popular model of gene expression dynamics: Boolean threshold dynamics in signed directed graphs. We show that eigenvectors of the network adjacency matrix can predict collective states (attractors). However, the overall predictive power depends on details of the network architecture, namely the fraction of positive 3-cycles, in a predictable fashion. Our results are a set of statistical observations, providing a systematic step towards a further theoretical understanding of the role of network eigenvectors in dynamics on graphs.

DOI: 10.1103/PhysRevE.108.014402

I. INTRODUCTION

Understanding how collective dynamical states can be predicted just from network architecture remains a major challenge in network science. In a few cases exemplified by Turing patterns on graphs, collective states near the instability threshold are given by eigenvectors of the graph's Laplacian matrix [1,2]. Beyond such examples, the research is in an early phase where we still need systematic elements founding a broader theory [3,4].

In several disciplines, a common approach is to study relationships between structural and functional connectivity [5–9]. However, due to its inherent link-by-link comparison of structure and dynamics, it is not well adapted for the study of *collective* dynamical states.

We chose to consider Boolean attractors in threshold dynamics on signed directed graphs and to investigate numerically under which topological conditions (involving, e.g., the cycle composition of the graph) eigenvectors can also predict collective states.

Introduced in particular by Stuart Kauffman in 1969 [10,11] the random Boolean networks (RBN) model is a mathematical language to explore networks of interacting genes in a highly stylized form. In its original formulation with random Boolean update rules and random networks, where each node has the same number, K, of inputs, this model displayed remarkable statistical properties—such as the transition from stable to chaotic behavior as a function of K and the scaling of attractor numbers with network size [12].

By including a threshold on the input in the dynamical rules ("threshold update rules," see below Sec. II A), the predictive power of this stylized model for real-life biological systems became apparent [13,14] (see also Ref. [15]).

Similar discrete dynamics on graphs have been explored in the context of spin glasses [16], cellular automata [17], and excitation spreading [18]. Technically, the threshold dynamics from Refs. [13,19] belong to a category of totalistic cellular automata (see Refs. [20,21]).

For its biological application and biological interpretability, random Boolean networks require signed directed graphs (in contrast to other types of discrete dynamics on graphs, like cellular automata). An edge $(v_1, v_2) \in E$, $v_i \in V$ of a graph $G(V, E)$ denotes the regulatory effect of the expression of a gene v_1 on the expression of another gene v_2. It is therefore a condensed representation of an intricate set of biological processes (e.g., v_1 encoding a transcription factor, which has a binding site in the regulatory region of gene v_2). The formal language of Boolean dynamics is now firmly established as one modeling approach in systems biology, among for example ODE models and metabolic flux models [22–25]. The Cell Collective database [26] is devoted to this topic, as well as a segment of the BioModels database [27].

Threshold Boolean models have been criticized as models of biological systems [28], as they do not discriminate between functionally different nodes in the biological network, which are expected to follow distinct dynamical rules. While this criticism may also hold for gene regulatory networks (see, e.g., Ref. [29]), the case of uniform (threshold-type) update rules is still a meaningful setting to further our understanding of these types of systems—in particular the relationship

*d.rumiantsau@constructor.university
†annick.lesne@sorbonne-universite.fr
‡m.huett@constructor.university

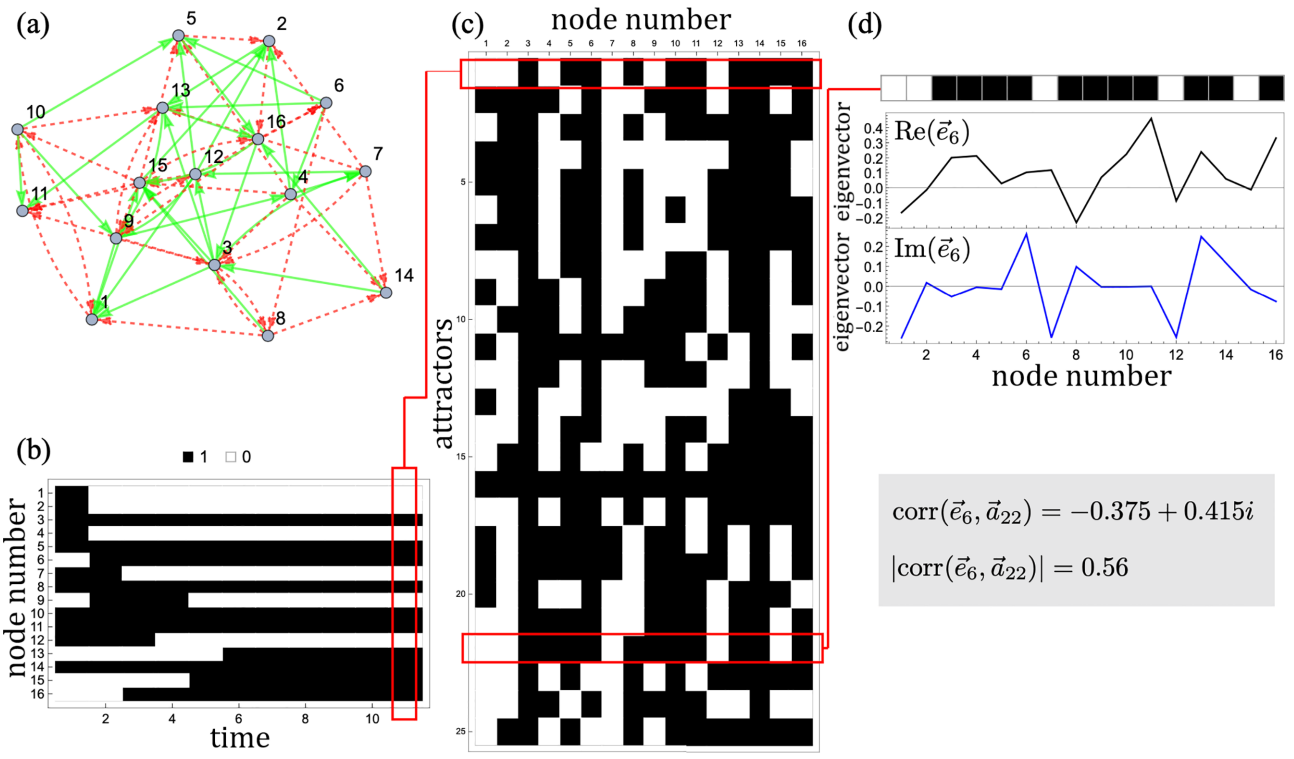


FIG. 1. General outline of our investigation. (a) Example of a small stylized gene regulatory network with $n = 16$ nodes and $m_+ = m_- = 32$ directed positive (activating) and negative (inhibitory) links shown in solid green lines and dashed red lines, respectively. (b) Example of a simulated time course starting from random initial conditions on the network from (a) with a binary state space and the threshold-based update rule from Ref. [13]. The asymptotic state (fixed-point attractor) is reached after a short transient. (c) Set of 25 examples of fixed-point attractors for the network from (a). The attractor shown in (b) is given as the first entry in this set. (d) Top: Visual comparison of an attractor [attractor $\vec{a}_{22}$ from (c)] with one left eigenvector, $\vec{e}_6$, of the adjacency matrix of the network shown in (a) with the real and imaginary parts, Re($\vec{e}_6$) and Im($\vec{e}_6$), shown in separate panels under the attractor pattern. Bottom: Numerical result for the Pearson correlation coefficient between the Boolean attractor $\vec{a}_{22}$ and the (complex) eigenvector $\vec{e}_6$, corr($\vec{e}_6, \vec{a}_{22}$), together with the modulus of this quantity, |corr($\vec{e}_6, \vec{a}_{22}$)|.

between network architecture and the dynamics of gene activity—as evidenced by the success of the approaches developed in Refs. [13,14]. An early example of a detailed model with biologically motivated update rules designed individually for each node is presented in Ref. [30] (see also Ref. [31]).

Beyond Boolean threshold dynamics the relationship between network architecture and stable attractors has been addressed in a range of works. In Ref. [32] the notion of an energy landscape known for continuous systems is reformulated for discrete systems, a work of relevance to cell type reprogramming. Adding to the previous understanding of the original random Boolean network (RBN) model [10,11], in Ref. [33] the role of subsets of update rules, canalyzing functions, for global stability has been discussed. Tools for studying RBNs on scalefree graphs have been developed in Ref. [34]. In Refs. [35,36] the evolvability of robust attractors has been discussed. The diverse research directions are also excellently summarized in Ref. [37]. However, the relationship between attractors and spectral characteristics of the network is not addressed in these works.

Here we show (1) that signed directed random graphs differ strongly from each other in the correlation between eigenvectors of the graph and attractors under threshold Boolean dynamics and (2) that the asymmetry of proportions of negative and positive cycles in the graph is a good predictor of this coupling of eigenvectors and dynamical attractors. In case of an over-representation of positive three-node cycles, a large percentage of attractors is structurally determined by the eigenvectors.

II. METHODS

A. Boolean network model

The general design of our investigation is summarized in Fig. 1.

Throughout the investigation the *adjacency matrix* A of a directed network G with nodes $\{1, 2, \ldots n\}$ is defined as follows: $A_{ij} = 0$ if there is no link from i to j in G, $A_{ij} = 1$ if the link from i to j is activating, and $A_{ij} = -1$ if it is inhibitory. We consider only networks without self-loops, so $A_{ii} = 0$, $i \in \{1, 2, \ldots, n\}$.

The dynamical model used in this study is Boolean threshold dynamics. Every node $i \in \{1, 2, \ldots n\}$ in the network G has two possible states, $S_i = 1$ and $S_i = 0$. At every time step t, future node states $S_i(t + 1)$ are determined from present states $S_i(t)$ via the following update rule [13]:

$$S_i(t+1) = \begin{cases} 1, & \sum_{j=1}^{n} a_{ji} S_j(t) > 0 \\ 0, & \sum_{j=1}^{n} a_{ji} S_j(t) < 0 \\ S_i(t), & \sum_{j=1}^{n} a_{ji} S_j(t) = 0 \end{cases}.$$

Since the network is directed, its adjacency matrix is not necessarily symmetric, and hence its right and left eigenvectors are different (and possibly complex). Given the expression $\sum_{j=1}^{n} A_{ji} S_j(t)$ in the above update rules, we are only considering left eigenvectors of A in our analysis.

Note that in contrast to Ref. [13] we do not add inhibitory self-links to nodes, which only have positive input. In Ref. [13] this was done with the goal of a quantitative comparison of the model's dynamics with (discretized) gene expression time series. For our statistical investigation, it is more appropriate to fully control the number of edges in the random graphs investigated here. As pointed out in Ref. [28], the last row in the update rule given above already implies some level of self-regulation. This is not present in the original Kauffman model [10,11]. Also, in a typical network analysis of transcriptional regulatory networks (e.g., the analysis of network motifs as in Ref. [38]) self-links are not considered.

The main object of our investigation is a quantity that we call the *predictability* of a given attractor $\vec{a}_j$. It is defined as the maximal (Pearson) correlation of this attractor with a left eigenvector $\vec{e}_k$:

$$\pi(\vec{a}_j) = \max_k |\mathrm{corr}(\vec{e}_k, \vec{a}_j)|, \tag{1}$$

where the maximum is taken over all (left) eigenvectors of the adjacency matrix of the network.

In Fig. 1(d) eigenvector $\vec{e}_1$ has been selected for comparison, because for this eigenvector the correlation to the given attractor, $\vec{a}_{19}$, is maximal.

B. Attractor predictability

For the network shown in Fig. 1(a) we have obtained 67 distinct attractors by starting from all 65.536 possible initial conditions. The corresponding attractor predictabilities $\pi(\vec{a}_j)$ are shown as a histogram. We then define a global quantity, *Network-level attractor predictability* $\pi(G)$ of a network G, as the fraction of attractor predictabilities more than two standard deviations away from randomness, which is determined using the predictability distribution of randomized attractors. The randomization of attractors is performed by randomly shuffling each attractor vector (i.e., conserving the number of 0s and 1s in each attractor). For the sake of convenience, we abbreviate the name of this quantity as *network-level predictability* throughout the paper. For this particular example [the network shown in Fig. 1(a)] the network-level predictability is 0.29.

A summary of the pipeline for finding predictability distributions of observed and shuffled attractors is shown in Fig. 2(a). Figure 2(b) illustrates the definition of network-level predictability.

C. Network generation model for varying asymmetry in positive and negative directed cycles

In a stylized regulatory network a directed cycle is said to be *positive* if it contains an even number of inhibitory links (possibly zero), and is said to be *negative* otherwise. Adopting the standard definition of an asymmetry (see, e.g., Ref. [40]), we introduce a quantity $A_k(G)$, quantifying the difference between the numbers of positive and negative cycles of length k in network G:

$$A_k(G) = \frac{N_{k,+}(G) - N_{k,-}(G)}{N_{k,+}(G) + N_{k,-}(G)},$$

where $N_{k,+}(G)$ and $N_{k,-}(G)$ denote the numbers of positive and negative cycles of length k in G, respectively. We call this quantity k-cycle asymmetry or, for the case $k = 3$, dominantly considered in the following discussion, simply cycle asymmetry.

To create a directed network of n nodes with m_+ positive and m_- negative links, which has a value of 3-cycle asymmetry close to a_{target}, we first generate a random directed Erdös-Rényi network with exactly $m_+ + m_-$ unsigned links. Then we iterate through all of its 3-cycles in random order, allocating signs of links in those cycles in a way which brings 3-cycle asymmetry closer to a_{target} whenever possible. The algorithm also strives to maintain a balance between the number of different positive and negative 3-cycle compositions in the resulting network. As the final step, signs of links that are not part of any 3-cycle are chosen randomly, subject to m_+ and m_- constraints.

The sets of high- and low-asymmetry networks used in this study contain 10.000 samples each, with $n = 16$, $m_+ = m_- = 32$ and $a_{\mathrm{target}}^{\mathrm{high}} = 0.8$, $a_{\mathrm{target}}^{\mathrm{low}} = -0.8$, respectively. They have average 3-cycle asymmetries of 0.79308 and −0.80284, with the standard deviation of less than 0.05.

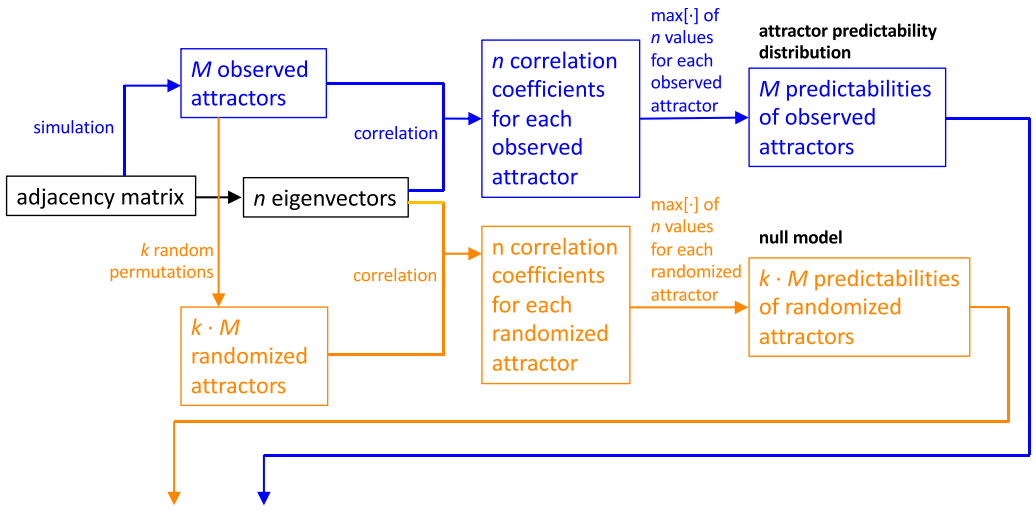
D. Generation of signed directed regular random networks

For a specific network node, we denote by k_{out}^{+}, k_{out}^{-}, k_{in}^{+}, and k_{in}^{-} the number of links of corresponding sign and direction. A signed directed regular network of degree k is a directed signed network where each node has $k_{\mathrm{out}}^{+} = k_{\mathrm{out}}^{-} = k_{\mathrm{in}}^{+} = k_{\mathrm{in}}^{-} = k$.

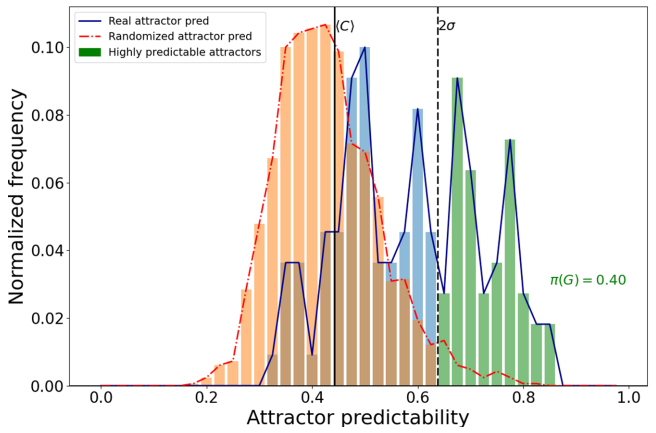
In order to generate a signed directed random regular network of degree k on n nodes, we first create an undirected unsigned connected random regular network of degree $4k$. Next, we iterate through the nodes in random order and distribute link directions randomly, compliant with degree constraints $k_{\mathrm{out}} = k_{\mathrm{out}}^{+} + k_{\mathrm{out}}^{-} = 2k$ and $k_{\mathrm{in}} = k_{\mathrm{in}}^{+} + k_{\mathrm{in}}^{-} = 2k$. Afterwards, subsets of k positive outgoing links and k positive incoming links are randomly chosen for each node, which fully determines signs of the remaining links. During those two steps, only random assignments of directions and signs, which do not lead to immediate conflicts (such as making a link whose sign is fully determined by the process of elimination with respect to one node violate degree constraints of another node) in the network, are performed. Finally, in order to avoid biasing the output towards specific topological configurations, direction- and sign-preserving pairwise switch randomization is run on the network for a large number of steps.

In this study, we use a set of 10 000 connected signed directed regular random networks of degree 2 with 16 nodes. The numbers are chosen for consistency with the ER dataset (see Sec. III B), since resulting networks also have 64 links in total.

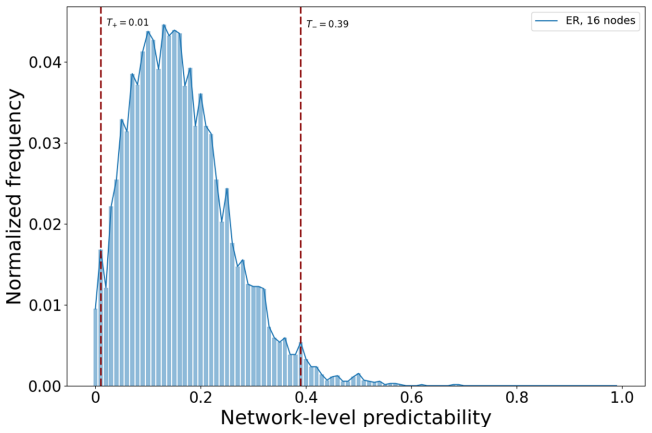
We define sets of regular networks with high and low network-level predictability by using the 2% quantile-motivated thresholds $T_+ : \pi(G) \geqslant 0.66$, $T_- : \pi(G) \leqslant 0.13$.



(a)



(b)



(c)

FIG. 2. (a) A schematic illustration of the computation of observed and shuffled attractor predictability distributions for a single network. Distinct attractors appear from many dynamical runs starting from different initial conditions. In practice, the number of dynamical runs is usually much larger than the number of observed attractors M, since distinct initial conditions often yield the same attractor. Random permutations of observed attractors which produce randomized attractors are chosen independently for each observed attractor. (b) Quantification of the network-level predictability $\pi(G)$ of an Erdös-Rényi (ER) network G with $n = 16$ nodes and $m_+ = m_- = 32$ directed positive and negative links (see Table ST1 [39]). Histogram of attractor predictabilities shown in blue (solid line), together with the corresponding histogram for randomized attractors (see Methods) shown in orange (dash-dotted line). Cases contributing positively to $\pi(G)$ (i.e., those with $\pi(\vec{a}_j) > \langle C \rangle + 2\sigma$, with the average predictability of random attractors, $\langle C \rangle$, and its standard deviation σ) are highlighted in green (to the right of vertical dashed line). (c) Distribution of network-level predictability, together with the thresholds $T_- = 0.01$ and $T_+ = 0.39$ for networks with high and low network-level predictability, respectively. T_+ was chosen as the largest value (up to two decimal paces) that labels at least 2% of all synthetic ER networks described in Sec. III B as those with high network-level predictability, and T_- was chosen similarly with the same 2% threshold in mind. Only networks with at least 50 distinct attractors enter this histogram.

E. Numerical experiments

1. *Standard settings*

Our main example will be stylized gene regulatory networks—random Erdös-Rényi networks with $n = 16$ nodes and $m_+ = 32$ directed positive and $m_- = 32$ directed negative links. We created a database by running simulations on 50 000 such networks of this size and for each network computing all distinct attractors from the complete set of 65 536 possible initial conditions.

2. *Filtering criteria for biological networks*

We analyze network-level attractor predictability for the 78 networks of models from the Cell Collective database [26], as well as the nine-node network from Ref. [41] and six biological networks from Ref. [42]. In order to have meaningful structural information in these networks, we apply filtering criteria on connectivity (weak connectivity in [0.15,1]), number of observed fixed-point attractors (at least 10), number of cycles (at least 1 cycle of length 3 or 4), and the ratio of

positive and negative links (between 0.3 and 0.65). This leads to 15 networks included in our analysis.

3. GeneNetWeaver experiments

Simulations with GeneNetWeaver [43] were performed in the following way: We considered the averaging GeneNetWeaver model (see Ref. [39]) with measurement noise, as well as a stochastic version with both intrinsic and measurement noise using GeneNetWeaver default parameters. We used 400 stylized gene regulatory networks from the database described in standard settings, and calculated GeneNetWeaver attractors for each of them from 250 random initial conditions.

4. Analysis of small networks

We define in-degree asymmetry of a node k in a network G as

$$a_G(k) = \frac{n_+^k - n_-^k}{n_+^k + n_-^k},$$

where n_+^k and n_-^k denote the numbers of positive and negative links incoming to node k, respectively.

In Fig. 5(b) a threshold of ± 0.6 is used to define 3-cycle asymmetry extremes. The other quantity studied in that experiment is the number of network nodes with extreme in-degree asymmetry, with node k considered extreme if $|a_G(k)| > 0.6$. Due to network architecture, the number of such nodes has three possible values: 3, 4, or 5. All 128 link sign assignments are analyzed, with no constraints on the fraction of positive links.

The algorithm employed in Fig. 5(d) performs a single link sign swap on every iteration. To provide a smoother trajectory of network-level predictability, changes 3-cycle asymmetry as gradually as possible in the desired direction, while also trying to affect the smallest number of cycles. Among candidates that satisfy those two costraints, selection of link sign swap is random, to eliminate possible bias.

III. RESULTS

A. Attractor predictability as a network-level property

We investigate network-level predictability in three ways: (1) We create a database of ER networks with high and low network-level predictability values and discover that the two populations differ systematically in cycle asymmetry. To ensure that this effect is not a consequence of node degree, we generate a large set of signed directed *regular* random networks and use it to validate our findings (Sec. III B). (2) Using an algorithm which allows us to systematically vary the cycle asymmetry of a network, we validate the statistical significance of cycle asymmetry's ability to discriminate high and low network-level predictability (Sec. III C). (3) We study small graphs of only few cycles to gather some analytical insight about the way attractor-eigenvector relationship emerges and how this relationship changes as a function of cycle sign (Appendix B).

We checked that our results are not qualitatively affected by changes in network size, connectivity, or the nonexhaustive (random) sampling of initial conditions (see the various cases listed in the supplementary material [39]).

B. Topological properties of networks with high and low attractor predictability

With the basic outline of our approach given in Fig. 1 and the key quantity we analyze (network-level predictability of a network) illustrated in Fig. 2(b), we can now turn to the quantitative investigation of a larger ensemble of graphs. The settings for those numerical experiments are described as "standard settings" in Methods. Figure 2(c) shows the distribution of network-level predictability together with the thresholds we used to define the subsets of networks with high and low network-level predictability, respectively. These two sets now allow us to investigate topological differences between networks with extreme values of network-level predictability.

While most of the usual topological quantities (degree distributions, centrality measures, assortativity, etc.) were surprisingly uninformative in discriminating these two sets of networks (see Figs. S11–S13), striking topological differences become visible on the level of small directed cycles, if the signs of links which form the cycle are taken into account. Figure 3(a) shows the asymmetry of positive and negative 3-cycles (see Methods) for the two sets of graphs. Those sets display a difference in observed cycle asymmetry that is significant with a p value of 6.53×10^{-53} according to a two-sample Kolmogorov-Smirnov test [see Fig. 3(a)]. Furthermore, while using a 2-standard-deviation network-level predictability threshold [see Fig. 2(b)], we see that networks with high network-level predictability assume relatively high $[A_3(G) \geqslant 0.4]$ positive values of 3-cycle asymmetry 662% more often than those with low network-level predictability. Conversely, networks from the latter set exhibit a high $[A_3(G) \leqslant -0.4]$ fraction of negative 3-cycles 462% more often than those from the former one.

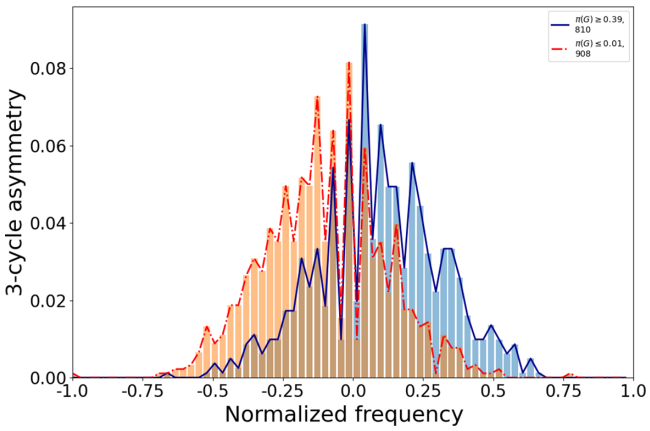
To ensure that observed topological difference between sets of networks with high and low network-level predictability is not a consequence of the properties of degree distribution, it is necessary to eliminate it as an intervening factor. We approach the problem by constructing signed directed regular random networks of the same size and connectivity (see Methods) and studying this new dataset with the same methods that we describe for ER networks in this section.

We discover that asymmetry of positive and negative 3-cycles discriminates between networks with high and low network-level predictability (see Methods) as well (see Fig. S7 in Ref. [39]).

C. Network-level predictability for networks with varying cycle asymmetry

The main topological difference between highly predictable and poorly predictable networks, so far, has been observed on a statistical level using a large set of random networks.

Next, in order to assess whether indeed this topological difference can also be causally linked to network-level



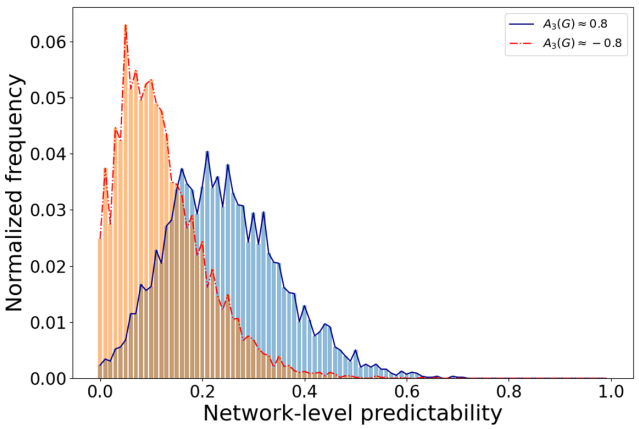


FIG. 3. (a) Histogram of 3-cycle asymmetry distributions for sets of networks with high (solid blue line) and low (dash-dotted red line) network-level predictability, defined using 2 standard deviation threshold [see Fig. 2(b)]. (b) Histograms of network-level predictability for networks with many negative cycles (negative 3-cycle asymmetry; orange dash-dotted line) and many cycles (positive 3-cycle asymmetry; blue solid line).

predictability, we create networks with controlled asymmetry of positive and negative cycles (see Methods) and analyze the network-level predictability for sets of networks with high and low cycle asymmetry, respectively. We discover that networks where the majority of cycles are positive exhibit predictability values in [0.3, 0.6] 614% more often than their counterparts with low cycle asymmetry. Conversely, networks with low cycle asymmetry have network-level predictability in [0, 0.15] 243% more often than those with high cycle asymmetry [see Fig. 3(b)].

D. Continuous dynamics and noise

An important question is whether network-level attractor predictability goes beyond the specific Boolean dynamical model and extends to, for example, continuous gene expression dynamics simulated via differential equations. In order to address this point, we use the GeneNetWeaver tool [43], a model based on ordinary differential equations (in the case of no noise or just measurement noise) or stochastic differential

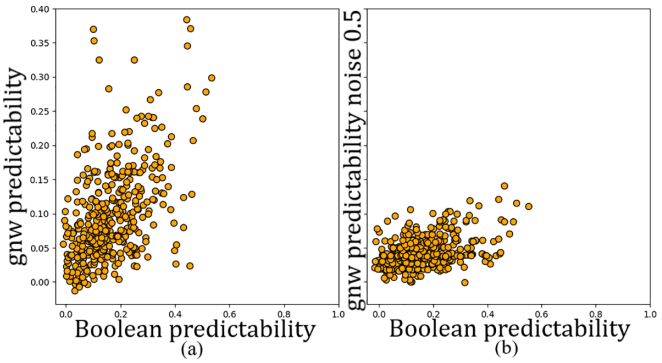


FIG. 4. Comparison of network-level attractor predictabilities derived from the Boolean model with those derived from the GeneNetWeaver data simulator. (a) GeneNetWeaver simulation with measurement noise but no intrinsic noise. Pearson correlation between the quantities is 0.51 (b) GeneNetWeaver simulation with both measurement noise and intrinsic noise. Default values are used for measurement noise, while intrinsic noise corresponds to noiseCoefficientSDE = 0.5 in GeneNetWeaver settings (10 times the default value). Pearson correlation between the quantities is 0.37.

equations (in the case of intrinsic noise). GeneNetWeaver has served as a generator of synthetic gene expression data for several DREAM gene network inference competitions [44,45]. Note that GeneNetWeaver has by default a multiplicative activation term [43]. For our simulations we rewrote this as an additive activation function to be comparable to the threshold dynamics used in our Boolean model. A summary of these activation functions and results for the original GeneNetWeaver model are given as supplementary material (see Fig. S10 and relevant text in Ref. [39]). We simulate steady states for several initial conditions (see Methods) and we compute, as in the case of Boolean dynamics, the correlation of these asymptotic states with the eigenvectors of the graph and thus obtain network-level predictabilities.

Figure 4(a) shows that in spite of the markedly different nature of the two models the predictabilities show a clear positive correlation. This positive correlation persists even in the presence of intrinsic noise [Fig. 4(b)].

E. Small networks

In order to develop some intuition about the observed statistical association between network-level predictability and network architecture, we study a minimal example consisting of five nodes and seven edges arranged to form three overlapping triangles, as depicted in Fig. 5(a). Starting from this template we create all possible signed graphs and study the two topological quantities, cycle asymmetry and the number of nodes with extreme in-degree asymmetry (see Methods for definitions) as candidates for explaining network-level attractor predictability. Structurally, this procedure leads to four values of 3-cycle asymmetries (all three cycles negative, $A = -1$, two cycles negative, one positive, $A = -1/3$, etc.) and three values of extreme in-degree asymmetry node count (see Methods for details). For all signed graphs we compute attractor predictability as previously described and assign categories of high and low predictability via thresholds.

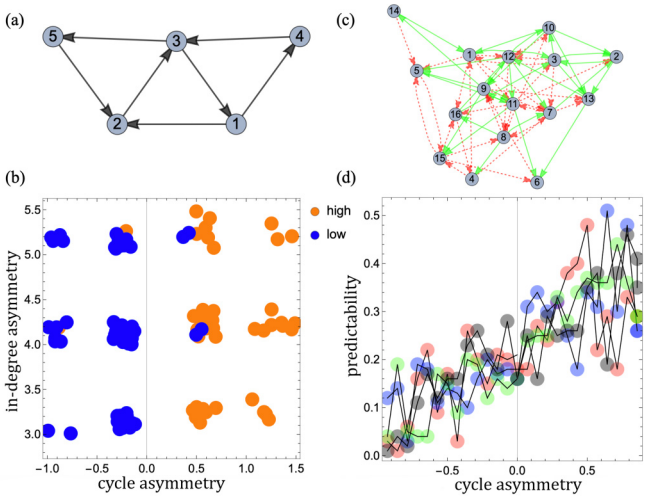


FIG. 5. Case study on a small graph with three 3-cycles and design of highly predictable networks. (a) Graph with $N = 5$ nodes and $M = 7$ edges, arranged to form three 3-cycles. (b) Distribution of high (orange, predominantly on the right) and low (blue, predominantly on the left) predictability in the plane of cycle asymmetry and extreme in-degree asymmetry node count. For defining such extreme nodes, in-degree asymmetry threshold of 0.6 has been employed. Predictability thresholds >0.2 and <0.1 have been used to define high and low predictability, respectively. Note that a small random shift of points has been applied for visual clarity. (c) Initial graph used for the design of highly predictable networks ($N = 16, M = 64$, number of 3-cycles $N_C = 28$. Positive and negative edges are shown with solid green and dashed red lines, respectively. (d) Changes of predictability under stepwise cycle sign switches. Starting from the graph displayed in (c) we gradually increase the number of positive cycles (to the right) or the number of negative cycles (to the left) and observe the corresponding change in network-level predictability. Four independent trajectories are shown in different colors.

Figure 5(b) shows the distribution of these two categories in the plane spanned by the two topological quantities. Cases of high and low predictability are almost perfectly separated along the cycle asymmetry axis. The few remaining exceptions are situated at high in-degree asymmetry. We verified that this general picture does not change qualitatively when varying these choices of thresholds used for predictability categories within reasonable ranges. The supplementary material shows a distribution of network-level predictabilities (see Fig. S16 in Ref. [39]).

Summarizing, we see that even in these small stylized graphs, high positive cycle asymmetry leads to high predictability, with a slight secondary influence from in-degree asymmetry.

As a next step, we can attempt to *design* highly predictable networks by adjusting the cycle content. In order to do this, we create a random directed ER graph, randomly distribute an equal number of positive and negative signs on the edges (with the subsidiary constraint that the initial 3-cycle asymmetry is zero), and then switch signs (conserving the balance of positive and negative edges) such that 3-cycles are iteratively turned into positive and into negative cycles, respectively. The initial graph is given in Fig. 5(c). Examples of the corresponding trajectories of predictability as a function of cycle asymmetry are shown in Fig. 5(d).

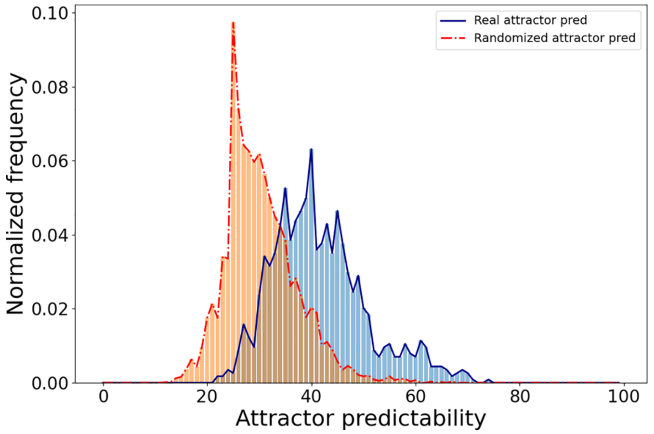


FIG. 6. Normalized predictability histograms of tumor cell invasion and migration model network attractors and their shuffled versions. For each attractor (except the constant zero vector) five random permutations are considered. The network has 3-cycle asymmetry of $A_3(G) = 0.53$, and network-level predictability of $\pi(G) = 0.36$ when the 2-standard deviation threshold is used [see Fig. 2(b)].

F. Application to biological examples

To assess how relevant the notion of predictability is for real-life biological networks we apply our methods to networks of models from Cell Collective database [26], as well as networks studied in Ref. [41] and Ref. [42] (see Methods). Figure 6 shows an example of attractor predictability distribution for a tumor cell invasion and migration network.

In line with studying attractor predictabiltiy as a network level property, we analyze its connection to 3-cycle asymmetry for a subset of networks obtained after applying a set of filtering criteria to the collection described above (see Methods). Table I shows that this topological property still discriminates between networks with high and poor network-level predictability.

Even though the overall agreement between cycle content and network-level predictability is not as clear as in our investigation of random graphs, the association is still quite visible in Table I. Note that in spite of the selection criteria we impose (see Methods), the real networks tend to be rather sparse. As a consequence, the numerical values of cycle asymmetry may not always be reliable, due to the small number of cycles in the graph (cycle statistics are provided in Table ST2 in Ref. [39]) and, in contrast to random graphs, where the cycle asymmetry derived from 3-cycles was by far the most dominant topological feature, here both 3-cycles and 4-cycles need to be taken into account.

IV. DISCUSSION

As with any systematic investigation of this type, our analysis contains several tunable parameters and one needs to understand, how the observations made here depend on these parameters.

TABLE I. Predictability $\pi(G)$, asymmetries $A_3(G)$ and $A_4(G)$ and agreement with theory for biological networks. Thresholds for high and low cycle asymmetry are ± 0.25. Low and high network-level predictability thresholds are $\pi(G) < 0.1$ (upper solid line) and $\pi(G) > 0.3$ (lower solid line). Agreement is assigned as follows: If both $A_3(G)$ and $A_4(G)$ are below -0.25 for low or above 0.25 for high network-level predictability, then the results is $++$. If the previous condition holds for either $A_3(G)$ or $A_4(G)$, and the other asymmetry is between the thresholds, then the result is a $+$. If network-level predictability and one of asymmetry values are extreme in the opposite sense, and the other asymmetry value is between the thresholds, then the result is a $-$. Results of $-+$ and $--$ are defined similarly, and are not present in the table.

Model name	$\pi(G)$	$A_3(G)$	$A_4(G)$	Agreement
Mammalian Cell Cycle 2006	0.01	0.20	−0.25	
Budding Yeast Cell Cycle 2009	0.07	−0.33	−0.76	++
CD4+ T Cell Differentiation and Plasticity	0.07	0.33	0.04	−
Oxidative Stress Pathway	0.08	−1.00	−1.00	++
Budding Yeast Cell Cycle	0.11	1.00	−0.27	
Arabidopsis thaliana Cell Cycle	0.15	−0.52	0.00	
emt26network.csv	0.16	0.80	0.93	
Lac Operon	0.21	nan	1.00	
T-LGL Survival Network 2011 Reduced Network	0.31	−0.33	0.00	−
B cell differentiation	0.32	1.00	1.00	++
gonadalsexdet.csv	0.34	0.64	0.62	++
Tumour Cell Invasion and Migration	0.36	0.53	0.27	++
sclcnetwork.csv	0.78	0.50	0.37	++
gastricnetwork.csv	0.79	0.42	0.30	++
Aurora Kinase A in Neuroblastoma	0.95	0.20	0.43	+

The main parameters are as follows:

(1) The connectivity of the network. There the factors limiting the range of this parameter are connectedness of the network (distinguish between strongly and weakly connected graphs) at the lower end and the rapid decline of the number of distinct attractors on the upper end (see Fig. S2 in Ref. [39]). Nevertheless, we found that our results remain valid for a range of weak connectivity values in [0.15, 0.6] (see Fig. S3 in Ref. [39]).

(2) The asymmetry of positive and negative links. The main part of our investigation has been performed with equal numbers of positive and negative links. In order to gain some insight in how the results depend on the positive-negative link asymmetry, we first look at an estimate of the average number of attractors as a function of this asymmetry (see Fig. S1 in Ref. [39]) As expected, the number of attractors decreases when the fraction of positive links is very high. This does not happen for very low fraction of positive links, but those values are excluded from consideration to eliminate possible bias they introduce into the dynamical behavior of the system.

This delineates a range of asymmetries for which predictability can be meaningfully investigated.

Next we can look at the average predictability as a function of asymmetry between positive and negative links. This can be done in two ways: (i) on the level of attractors, studying the distribution of the attractor-eigenvector correlations (in comparision with a null model of shuffled attractors) as a function of the positive-negative link asymmetry, and (ii) on the level of networks, where we can look at the network-level predictability (i.e., the fraction of attractors with value of predictability more than two standard deviations way from the average of random predictions). Based on our analysis, results of this study do not change qualitatively when asymmetry of positive and negative links lies within this range (see Fig. S4 in Ref. [39]).

(3) The last parameter to be addressed is the most challenging one: network size. A full enumeration of all initial conditions quickly becomes unfeasible when we depart from network sizes explored here. As this parameter increases, random sampling covers an ever smaller percentage of initial conditions, thus biasing the investigation towards attractors with large basins.

In order to see, whether this bias towards attractors with a large basin has already a substantial effect in the case of the network sizes under investigation here, we also conduct an investigation of larger networks with attractor sets drawn from randomly sampled initial conditions. Supplementary experiments indicate that our results remain valid for networks of size 40 with 50 000 random initial conditions each (see Fig. S6 in Ref. [39]), and although sampling of large basin attractors tends to somewhat amplify extreme cases of network-level predictability, results of this study are not affected qualitatively (see Fig. S9 in Ref. [39]). In addition, we verified that even if complete enumeration of initial conditions is impossible, network-level predictability is robust with respect to different samplings of initial conditions performed on the same network (see Fig. S8 in Ref. [39]). We furthermore checked that our key results do not depend on the exact choice of the minor technical parameters of our investigation, namely the number of sampled initial conditions (see Fig. S5 in Ref. [39]), the attractor predictability threshold (i.e., two standard deviations away from randomness, or quantile-based definitions; see Figs. S14 and S15 in Ref. [39]), and the thresholds used to identify highly predictable and poorly predictable networks [see Fig. 3(a)].

To summarize the above discussion, we expect results of our study to remain valid at least for the following parameter ranges: weak connectivity in [0.15, 0.6] (for small networks lower bound is greater since it is difficult to observe a meaningful amount of cycles at low connectivity there),

proportion of positive links in [0.3, 0.65], network size between 4 and 40 (but likely larger networks as well, since random sampling strategy does not change), predictability threshold between 1 and 3 standard deviations away from randomness, and a threshold between 0.01 and 0.1 of total dataset size for composing sets of highly and poorly predictable networks. A meaningful lower bound on the number of observed distinct attractors can be chosen by analyzing the distribution of those numbers for fixed connectivity, positive/negative link ratio and network size. For $n = 16$ with $m_+ = m_- = 32$, which has been dominantly considered in this study, we require at least 50 distinct attractors.

In Ref. [42] the parallel of the Boolean model to an asymmetric spin glass model is used to define frustration (in the sense of Ref. [46]). It is argued that the level of frustration may indeed be a nonrandom feature of biological regulatory networks. As in Ref. [42], our research is motivated by two questions, which to date are not fully answered: (1) Which constraints does the gene regulatory network impose on attractors? and (2) Which nonrandom features do gene regulatory networks have and how are they of relevance for their biological function?

We point out that there is no one-to-one relationship between eigenvectors (whose number is given by network size) and attractors (which for the networks discussed here are typically a much larger set). We observe that often several attractors have their maximal correlation (see Methods) with the same eigenvector. However, further evaluating these attractor groups (and these "eigenvector multiplicities," i.e., how often an eigenvector is providing maximal correlation to some attractor) is not related to the nature of network-level predictability.

V. CONCLUSION

We provide a way of quantifying predictability of collective dynamical states by eigenvectors and we offer heuristics when networks have a high predictability (a strong relationship between eigenvectors and attractors) and when they have a low predictability. We also show that the link between eigenvectors and attractors are not universal or omnipresent, but the details of the underlying network matters, in a systematic and predictable way.

In particular, our findings install some confidence that genetic programs and patterns of gene activity are tightly constrained by the architecture of the underlying regulatory network. Our findings also highlight the importance of accumulating enough knowledge about regulatory machineries.

We have not been able to achieve a more mathematical understanding of this relationship between network-level predictability and 3-cycle asymmetry. The fact that this relationship is seen even in very small graphs (see Fig. 5) suggests that indeed 3-cycles might mediate the relationship between attractors and eigenvectors. However, a deeper understanding will need to be developed in future work.

It is an obvious next step to extend these findings to real gene expression patterns. Such an application is limited at the moment by two major factors: (1) the incompleteness of our knowledge of gene regulatory networks, even for the simplest organisms. Today, the most comprehensive transcriptional regulatory network is the one stored in RegulonDB [47]. However, our knowledge about most interactions in this network is partial. In particular, the regulation of many promoters is presently unknown [48]. Transition to multicellular organisms further complicates research, because the data become less and less complete [47]. (2) The fact that real gene expression patterns do not exclusively arise from the action of the transcriptional regulatory network but are subject to a multitude of other biological mechanisms; in the case of bacterial gene regulation, for example, the influence of chromosomal spatial organization has been explored in much detail over the past few years [49].

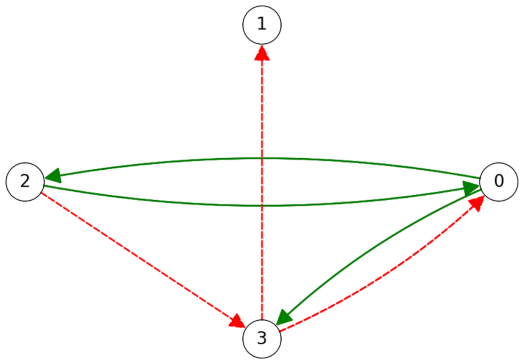


FIG. 7. The network of four nodes with six links used throughout this section. Red dashed and green solid arrows denote negative and positive links, respectively.

In our investigation we focus on the predictability of attractors from eigenvectors of the underlying regulatory network. Noise, a topic of high relevance in studying gene expression patterns [41,50,51], therefore is less important in our investigation, as it can be thought of as an influencing factor of the transient *towards* an attractor and consequently as a selective mechanism affecting only the statistics of the attractors. However, the robustness of an attractor with respect to (e.g., update timing) noise [36,52,53] is an important characteristic that deserves further investigation. In future work we intend to explore whether highly predictable attractors are more likely

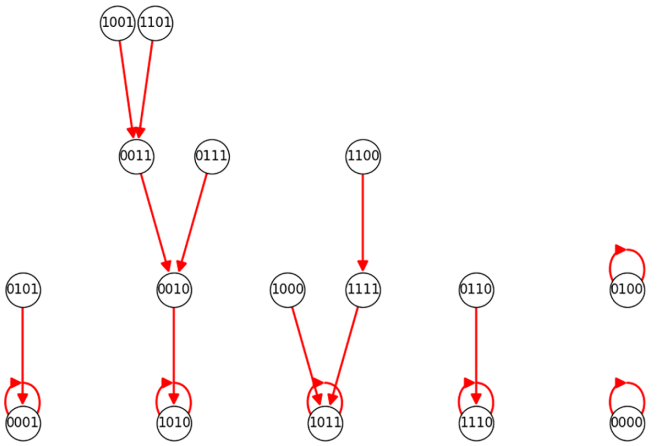
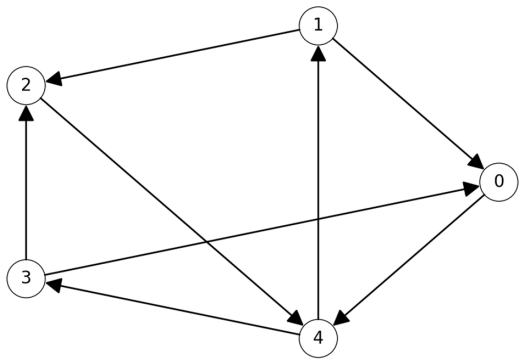
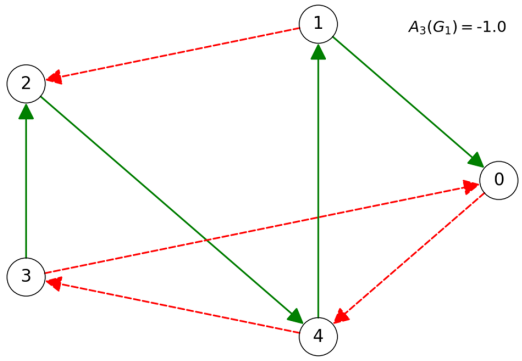


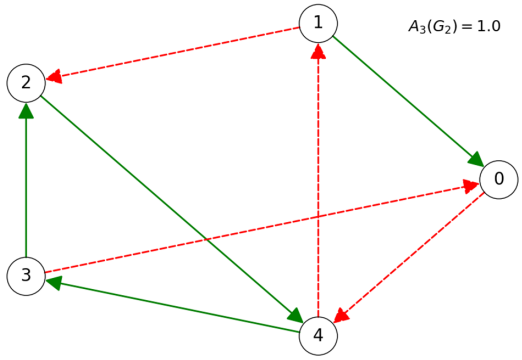
FIG. 8. Structure of the state space, with nodes representing states and arrows representing transitions under the Boolean dynamical rule.



(a)



(b)



(c)

FIG. 9. (a) Directed topology of a network on five nodes with eight links used in this section. (b) Arrangement of link signs G_1 with 3-cycle asymmetry $A_3(G_1) = -1$. Dashed red and solid green arrows denote negative and positive links, respectively. (c) Arrangement of link signs G_2 with 3-cycle asymmetry $A_3(G_2) = 1$. Dashed red and solid green arrows denote negative and positive links, respectively.

robust than poorly predictable attractors, which would suggest that the robustness of a dynamical state would be structurally determined. Some evidence of such structural effects of robustness is provided in Ref. [52].

More generally, our work suggests to resort to eigenvectors to predict collective dynamical states, substantially beyond the well-known example of Turing patterns. So far, our results are a set of statistical observations. However, this statistical approach appears very powerful to unravel unexplored relationships between network architecture and collective states. The challenge is now open to elaborate definite theoretical statements in specific situations from these broad statistical relationships.

ACKNOWLEDGMENTS

M.T.H. acknowledges support by Volkswagen Stiftung (Grant No. 9A174). M.T.H. thanks LPTMC (Paris) for hospitality and the Physics Institute of CNRS (French National Center for Scientific Research) for funding his stays, during which part of this work has been performed.

APPENDIX A: ILLUSTRATION OF THE NUMERICAL PROCEDURE FOR A SMALL GRAPH

In this Appendix, we provide a step by step computation of network-level predictability for a small network. The network G of four nodes with six links is given in Fig. 7. Its adjacency matrix A is given by:

$$a = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & -1 \\ -1 & -1 & 0 & 0 \end{pmatrix}.$$

Normalized left eigenvectors of A are as follows:

$$\vec{e}_1 = \begin{bmatrix} -0.356i \\ 0.3062 - 0.5303i \\ 0.3062 + 0.1768i \\ 0.6124 \end{bmatrix}, \quad \vec{e}_2 = \begin{bmatrix} 0.356i \\ 0.3062 + 0.5303i \\ 0.3062 - 0.1768i \\ 0.6124 \end{bmatrix},$$

$$\vec{e}_3 = \begin{bmatrix} -0.7071 \\ 0 \\ -0.7071 \\ 0 \end{bmatrix}, \quad \vec{e}_4 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

Since network size allows, Boolean dynamics is run from all possible initial conditions. State space structure is summarized in Fig. 8:

After discarding $\vec{0}$ (for which computation of correlations is impossible), we are left with the following set of fixed-point attractors:

$$\vec{a}_1 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}, \ \vec{a}_2 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 0 \end{bmatrix}, \ \vec{a}_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \ \vec{a}_4 = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \ \vec{a}_5 = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix}.$$

To calculate predicability of every attractor, we compute absolute values of correlation coefficients for every pair $(\vec{e}_k, \vec{a}_j)$:

$\lvert\mathrm{corr}(\vec{e}_k, \vec{a}_j)\rvert$	$\vec{e}_1$	$\vec{e}_2$	$\vec{e}_3$	$\vec{e}_4$
$\vec{a}_1$	0.5	0.5	1	0.5773
$\vec{a}_2$	0.5773	0.5773	0.5773	0.3333
$\vec{a}_3$	0.5773	0.5773	0.5773	0.3333
$\vec{a}_4$	0.5773	0.5773	0.5773	1
$\vec{a}_5$	0.5773	0.5773	0.5773	1

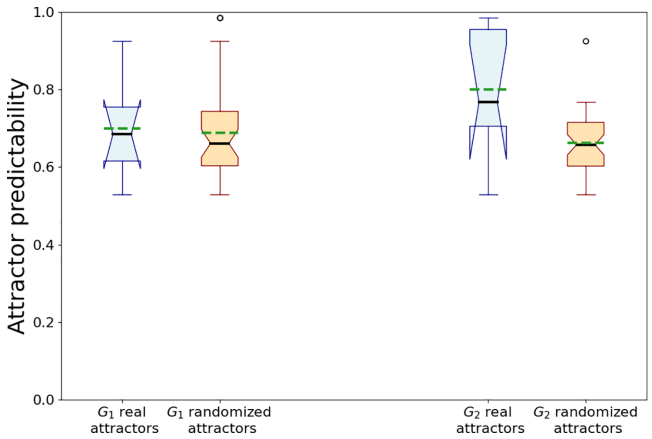


FIG. 10. Predictability of actual and randomized attractors for G_1 and G_2, computed similarly to Appendix A. Dashed green and solid black lines represent means and medians, respectively.

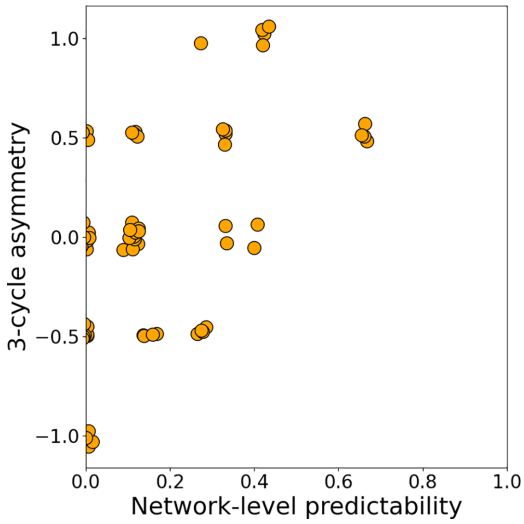


FIG. 11. Scatterplot of 3-cycle asymmetry and network-level predictability of all 70 labeled balanced sign arrangements of topology in Fig. 9(a). To clearly show the number of networks with coinciding pairs of values, small random shifts have been applied.

Based on the above table, $\pi(\vec{a}_1) = \pi(\vec{a}_4) = \pi(\vec{a}_5) = 1$, $\pi(\vec{a}_2) = \pi(\vec{a}_3) = 0.5773$.

In order to compute network-level predictability of G, predictability distribution of shuffled attractors is required. For every a_i, we consider all of its permutations that do not equal another a_i:

$$\vec{a}_1 : \begin{bmatrix}1\\1\\0\\0\end{bmatrix}, \begin{bmatrix}1\\0\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\\0\end{bmatrix}, \begin{bmatrix}0\\1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\0\\1\\1\end{bmatrix},$$

$$\vec{a}_2 : \begin{bmatrix}1\\1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \quad \vec{a}_3 : \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix},$$

$$\vec{a}_4 : \begin{bmatrix}1\\1\\0\\1\end{bmatrix}, \begin{bmatrix}0\\1\\1\\1\end{bmatrix}, \quad \vec{a}_5 : \begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \begin{bmatrix}0\\0\\1\\0\end{bmatrix}.$$

Distribution of $\pi(\vec{a})$ for those shuffled attractors has $\langle C\rangle = 0.6543$ and $\sigma = 0.1437$. Using the 2σ threshold yields network-level predictability $\pi(G) = 0.6$.

APPENDIX B: EXAMPLE OF INDIVIDUAL POSITIVE AND NEGATIVE CYCLES

In this Appendix, we illustrate for small graphs how alteration of 3-cycle asymmetry changes network-level predictability of a small network. To that end, we fix the directed unsigned topology depicted in the Fig. 9(a) and endow it with different arrangements of links signs.

For the sake of consistency with the main text, we only consider sign arrangements which have equal number of positive and negative links and refer to those as *balanced* sign arrangement.

Figure 9(b) features an example of a balanced sign arrangement G_1 where $A_3(G_1) = -1$, i.e., every 3-cycle is negative. Predictability of observed and shuffled attractors is shown in Fig. 10. Resulting network-level predictability $\pi(G_1) = 0$.

Conversely, if we consider a sign arrangement G_2 with $A_3(G_2) = 1$ [see Figs. 9(c) and 10], then we see a significant increase in network-level predictability ($\pi(G_2) = 0.42$).

To verify existence of relationship between $A_3(G)$ and $\pi(G)$, we construct all labeled balanced link sign arrangements on the directed topology shown in Fig. 9(a). The scatterplot of 3-cycle asymmetry and network-level predictability for those arrangements is featured in Fig. 11. Based on cycle asymmetry, two distinct groups of networks can be naturally defined: one with $A_3(G) \leqslant -0.5$ and another with $A_3(G) \geqslant 0.5$. Between those two groups, distributions of network-level predictability are markedly different: networks from the first group dominantly exhibit low values of $\pi(G)$, while networks of the second tend to have high network-level predictability. Pearson correlation of $A_3(G)$ and $\pi(G)$ is equal to 0.504 if all balanced networks are considered and becomes 0.4618 without duplicate value pairs.

[1] H. Nakao and A. S. Mikhailov, Turing patterns in network-organized activator–inhibitor systems, Nat. Phys. **6**, 544 (2010).

[2] M.-T. Hütt, D. Armbruster, and A. Lesne, Predictable topological sensitivity of Turing patterns on graphs, Phys. Rev. E **105**, 014304 (2022).

[3] E. Laurence, N. Doyon, L. J. Dubé, and P. Desrosiers, Spectral Dimension Reduction of Complex Dynamical Networks, Phys. Rev. X **9**, 011042 (2019).

[4] T. Nishikawa, J. Sun, and A. E. Motter, Sensitive Dependence of Optimal Network Dynamics on Network Structure, Phys. Rev. X **7**, 041044 (2017).

[5] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, and P. Hagmann, Predicting human resting-state functional connectivity from structural connectivity, Proc. Natl. Acad. Sci. USA **106**, 2035 (2009).

[6] H.-J. Park and K. Friston, Structural and functional brain networks: From connections to cognition, Science **342**, 1238411 (2013).

[7] T. Aavik, R. Holderegger, and J. Bolliger, The structural and functional connectivity of the grassland plant *Lychnis flos-cuculi*, Heredity **112**, 471 (2014).

[8] A. Messé, M.-T. Hütt, P. König, and C. C. Hilgetag, A closer look at the apparent correlation of structural and functional connectivity in excitable neural networks, Sci. Rep. **5**, 7870 (2015).

[9] V. Voutsa, D. Battaglia, L. J. Bracken, A. Brovelli, J. Costescu, M. Diaz Munoz, B. D. Fath, A. Funk, M. Guirro, T. Hein, C. Kerschner, C. Kimmich, V. Lima, A. Messé, A. J. Parsons, J. Perez, R. Pöppl, C. Prell, S. Recinos, Y. Shi *et al.*, Two classes of functional connectivity in dynamical processes in networks, J. R. Soc. Interface **18**, 20210486 (2021).

[10] S. Kauffman, Homeostasis and differentiation in random genetic control networks, Nature (London) **224**, 177 (1969).

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

[12] T. Mihaljev and B. Drossel, Scaling in a general class of critical random Boolean networks, Phys. Rev. E **74**, 046101 (2006).

[13] F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, The yeast cell-cycle network is robustly designed, Proc. Natl. Acad. Sci. USA **101**, 4781 (2004).

[14] M. I. Davidich and S. Bornholdt, Boolean network model predicts cell cycle sequence of fission yeast, PLoS ONE **3**, e1672 (2008).

[15] S. Bornholdt, Less is more in modeling large genetic networks, Science **310**, 449 (2005).

[16] M. Bartolozzi, T. Surungan, D. B. Leinweber, and A. G. Williams, Spin-glass behavior of the antiferromagnetic Ising model on a scale-free network, Phys. Rev. B **73**, 224419 (2006).

[17] C. Marr and M.-T. Hütt, Topology regulates pattern formation capacity of binary cellular automata on graphs, Physica A **354**, 641 (2005).

[18] P. Moretti and M.-T. Hütt, Link-usage asymmetry and collective patterns emerging from rich-club organization of complex networks, Proc. Natl. Acad. Sci. USA **117**, 18332 (2020).

[19] A. Szejka, T. Mihaljev, and B. Drossel, The phase diagram of random threshold networks, New J. Phys. **10**, 063009 (2008).

[20] C. Marr and M.-T. Hütt, Outer-totalistic cellular automata on graphs, Phys. Lett. A **373**, 546 (2009).

[21] C. Marr and M.-T. Hütt, Cellular automata on graphs: Topological properties of ER graphs evolved towards low-entropy dynamics, Entropy **14**, 993 (2012).

[22] J. Krumsiek, C. Marr, T. Schroeder, and F. J. Theis, Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network, PLoS ONE **6**, e22649 (2011).

[23] M. Choi, J. Shi, S. H. Jung, X. Chen, and K.-H. Cho, Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage, Science Signal. **5**, ra83 (2012).

[24] A. Saadatpour and R. Albert, Boolean modeling of biological regulatory networks: A methodology tutorial, Methods **62**, 3 (2013).

[25] B. C. Daniels, H. Kim, D. Moore, S. Zhou, H. B. Smith, B. Karas, S. A. Kauffman, and S. I. Walker, Criticality Distinguishes the Ensemble of Biological Regulatory Networks, Phys. Rev. Lett. **121**, 138102 (2018).

[26] T. Helikar, B. Kowal, S. McClenathan, M. Bruckner, T. Rowley, A. Madrahimov, B. Wicks, M. Shrestha, K. Limbu, and J. A. Rogers, The cell collective: Toward an open and collaborative approach to systems biology, BMC Systems Biology **6**, 96 (2012).

[27] R. S. Malik-Sheriff, M. Glont, T. V. Nguyen, K. Tiwari, M. G. Roberts, A. Xavier, M. T. Vu, J. Men, M. Maire, S. Kananathan *et al.*, BioModels—15 years of sharing computational models in life science, Nucleic Acids Res. **48**, D407 (2020).

[28] J. G. Zanudo, M. Aldana, and G. Martínez-Mekler, Boolean threshold networks: Virtues and limitations for biological modeling, in *Information Processing and Biological Systems* (Springer, Berlin, 2011), pp. 113–151.

[29] J. Mozziconacci, M. Merle, and A. Lesne, The 3D genome shapes the regulatory code of developmental genes, J. Mol. Biol. **432**, 712 (2020).

[30] R. Albert and H. G. Othmer, The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster, J. Theor. Biol. **223**, 1 (2003).

[31] M. Chaves, R. Albert, and E. D. Sontag, Robustness and fragility of Boolean models for genetic regulatory networks, J. Theor. Biol. **235**, 431 (2005).

[32] J. X. Zhou, A. Samal, A. Fouquier d'Hérouël, N. D. Price, and S. Huang, Relative stability of network states in boolean network models of gene regulation in development, Biosystems **142-143**, 15 (2016).

[33] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, Genetic networks with canalyzing Boolean rules are always stable, Proc. Natl. Acad. Sci. USA **101**, 17102 (2004).

[34] M. Aldana, Boolean dynamics of networks with scale-free topology, Physica D **185**, 45 (2003).

[35] S. Braunewell and S. Bornholdt, Reliability of genetic networks is evolvable, Phys. Rev. E **77**, 060902 (2008).

[36] S. Braunewell and S. Bornholdt, Reliability of regulatory networks and its evolution, J. Theor. Biol. **258**, 502 (2009).

[37] S. Bornholdt and S. Kauffman, Ensembles, dynamics, and cell types: Revisiting the statistical mechanics perspective on cellular regulation, J. Theor. Biol. **467**, 15 (2019).

[38] U. Alon, Network motifs: Theory and experimental approaches, Nat. Rev. Genet. **8**, 450 (2007).

[39] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevE.108.014402 for network used in Fig. 2(b), average numbers of distinct attractors depending on connectivity and fraction of positive links, robustness of our results with respect to various modeling choices, distributions of 3-cycle asymmetry for signed directed regular random networks, stability of $\pi(g)$, descriptions of original and modified ode models and comparison of $\pi(g)$ from Boolean dynamics to that of original ode model, performance of widely used topological quantities in discriminating networks with high and low $\pi(g)$, distribution of $\pi(g)$ for networksused in Fig. 5(a), and extended in Table T1. It includes Refs. [13,43].

[40] K. Kosmidis, K. P. Jablonski, G. Muskhelishvili, and M.-T. Hütt, Chromosomal origin of replication coordinates logically

distinct types of bacterial genetic regulation, npj Syst. Biol. Appl. **6**, 1 (2020).

[41] B. Zhang and P. G. Wolynes, Stem cell differentiation as a many-body problem, Proc. Natl. Acad. Sci. USA **111**, 10185 (2014).

[42] S. Tripathi, D. A. Kessler, and H. Levine, Biological Networks Regulating Cell Fate Choice are Minimally Frustrated, Phys. Rev. Lett. **125**, 088101 (2020).

[43] T. Schaffter, D. Marbach, and D. Floreano, GeneNetWeaver: In silico benchmark generation and performance profiling of network inference methods, Bioinformatics **27**, 2263 (2011).

[44] D. Marbach, R. J. Prill, T. Schaffter, C. Mattiussi, D. Floreano, and G. Stolovitzky, Revealing strengths and weaknesses of methods for gene network inference, Proc. Natl. Acad. Sci. USA **107**, 6286 (2010).

[45] D. Marbach, J. C. Costello, R. Küffner, N. M. Vega, R. J. Prill, D. M. Camacho, K. R. Allison, M. Kellis, J. J. Collins, and G. Stolovitzky, Wisdom of crowds for robust gene network inference, Nat. Methods **9**, 796 (2012).

[46] P. Anderson, The concept of frustration in spin glasses, J. Less-Common Met. **62**, 291 (1978).

[47] M.-T. Hütt and A. Lesne, Gene regulatory networks: Dissecting structure and dynamics, in *Systems Medicine: Integrative, Qualitative and Computational Approaches*, edited by O. Wolkenhauer, Vol. 3 (Academic Press, 2021), pp. 77–85.

[48] S. Gama-Castro *et al.*, RegulonDB version 9.0: High-level integration of gene regulation, coexpression, motif clustering and beyond, Nucleic Acids Res. **44**, D133 (2016).

[49] E. Cakir, A. Lesne, and M.-T. Hütt, The economy of chromosomal distances in bacterial gene regulation, npj Syst. Biol. Appl. **7**, 49 (2021).

[50] M. Thattai and A. Van Oudenaarden, Intrinsic noise in gene regulatory networks, Proc. Natl. Acad. Sci. USA **98**, 8614 (2001).

[51] G. Chalancon, C. N. Ravarani, S. Balaji, A. Martinez-Arias, L. Aravind, R. Jothi, and M. M. Babu, Interplay between gene expression noise and regulatory network architecture, Trends Genet. **28**, 221 (2012).

[52] K. Klemm and S. Bornholdt, Topology of biological networks and reliability of information processing, Proc. Natl. Acad. Sci. USA **102**, 18414 (2005).

[53] S. Bornholdt, Boolean network models of cellular regulation: Prospects and limitations, J. R. Soc. Interface **5**, S85 (2008).