

Emergence of diversity in homogeneous coupled Boolean networks

Chris Kang,¹ Boris Aguilar,² and Ilya Shmulevich²

¹*Department of Applied Mathematics, University of Washington, Seattle, Washington*

²*Institute for Systems Biology, Seattle, Washington*

(Dated: May 12, 2018)

The origin of multicellularity in Metazoa is one of the fundamental questions of evolutionary biology. We have modeled the generic behaviors of gene regulatory networks in isogenic cells as stochastic, nonlinear dynamical systems - coupled Boolean networks with perturbation. Model simulations under a variety of dynamical regimes suggest that the central characteristic of multicellularity, permanent spatial differentiation (diversification), indeed can arise. Additionally, we observe that diversification is more likely to occur near the critical regime of Lyapunov stability.

Usage:

PACS numbers:

Structure:

INTRODUCTION

Ernst Haeckel's Gastraea theory has been a widely accepted hypothesis in understanding the origin of multicellularity in Metazoa. Haeckel proposed that spatial differentiation in cells was the result of a morphological organization of unicellular flagellates into a hollow sphere, Blastaea, which upon invagination would form a new internal layer of cells of a different type and which is the evolutionary precursor of the endoderm [1]. A fundamental assumption of the colonial theory of Haeckel is that multicellularity preceded spatial cell differentiation. However, the inability of the Gastraea Theory to predict certain traits in a large monophyletic clade named Opisthokonta, such as the presence of multiple Metazoan genes that exist across animal phyla and that are involved in processes such as cellular adhesion, has given rise to alternate hypotheses of the origin of Metazoa [2]. Among the most prominent of these is the Synzoospore theory proposed by Alexey Zakhvatkin in 1949 [3]. A key assumption of this hypothesis is that cell *types* predated multicellularity.

Zakhvatkin suggested that cell types are formed during different phases of the life cycle of protists, the unicellular ancestor of Metazoa. During ontogenesis, genetically identical protists differentiate and are marked by distinct cellular behaviors, such as monotony, hypertrophic growth, and palintomy. The *temporal differentiation* of these cell types transitioned to permanent *spatial differentiation* when the various cell types were integrated into a colonial body. Thus, according to the Synzoospore theory, the origin of multicellularity lies in the aggregation of different cell types of isogenic organisms. The cell type, in the sense of temporal differentiation, refers to one of multiple functional cellular states that the cell is capable of exhibiting. Upon a particular stimulus, the cell may transition from one functional state, or phenotype, to another. However, when cells are aggregated in physical space, they acquire the ability to communicate by

means of diffusion-based or receptor-ligand interactions. Consequently, the signal that may be required for the cell to transition to a particular functional state or cell *type* may well originate from one or more of its neighboring cells. In this manner, a diversification of cellular phenotypes can emerge and become "locked into" a particular spatial pattern - an attractor of the cellular aggregate or *tissue*. This mutual reinforcement of cellular phenotypes is a manifestation of spatial differentiation. A fundamental question is what generic conditions can give rise to such spatial differentiation.

Motivated by this question, we explore the possibility of diversification of phenotypes in an aggregate (*tissue*) of cells modeled with stochastic, nonlinear dynamical systems, specifically Boolean networks. Since their introduction in 1968, Boolean networks have been widely used to model gene regulatory networks [4, 5]. Random Boolean networks, in particular, which are ensemble models having randomly chosen Boolean functions, have been used as a means to study general behaviors of networks without assumptions about the specific structure of their genetic regulatory mechanisms.

Kauffman initially characterized cell types as attractor states of the Boolean network representing the genetic network of the cell [4, 5]. This is a fitting interpretation for the temporal differentiation in the Synzoospore theory as phases in the life cycle are metastable, meaning that the cell can switch from one attractor to another upon an appropriate stimulus or perturbation. However, when noise is introduced to the states (i.e., Boolean network with random perturbation) there are, strictly speaking, no attractors in the network, since the presence of the random perturbation entails ergodicity of the underlying Markov chain describing the dynamics and therefore, a non-zero steady-state probability for each state of the system (by virtue of rendering the Markov chain irreducible and aperiodic). This fact called into question the interpretation of attractors as cell types [6], although subsequent work by Ribeiro *et al.* [7]

dispensed with this criticism by defining so-called ergodic sets, which are equivalent to irreducible Markov chains (equivalently, connected components in the state transition graph), and showing that while multiple such sets are highly unlikely when all genes are subject to noise perturbation, multiple ergodic sets do emerge when some genes are ‘protected’ from noise. It was later shown, by construction, that there exist noisy Boolean networks with a large number of ergodic sets and that ergodic sets can be arbitrarily large [8]. In a different strategy, Serra *et al.* generalized ergodic sets by defining so-called threshold ergodic sets, each of which consists of multiple attractors among which the system can jump due to noise, but the set of which the system never leaves [9]. In this paper, since our focus is diversification of long-term behaviors, we allow random gene perturbations to all genes and thus, associate the long-term behavior of the noisy system, defined by its steady-state distribution, with its entire collection of physiological states, such as motility, apoptosis, or various metabolic states [10, 11]. That is to say, functional cellular states correspond to the states of the underlying genetic network that carry the highest steady-state probability mass.

When Boolean networks with perturbation are constructed in such a way that an aggregate of identical cells communicates intercellularly by taking on inputs from neighboring cells, the synchronous updating of the cell states models the temporal progression of the tissue toward a potentially spatially differentiated state. The individual cells possess *distinct* steady-state distributions, which correspond to the sets of phenotypes of the cells in the tissue. If a diverse set of phenotypes were to emerge in an organized manner from the aggregate Boolean network model, it would recapitulate the manifestation of spatial differentiation as postulated by the Synzoospore theory. We then ask under what regime of dynamical complexity (stability) of cells can such spatial differentiation of phenotypes occur?

The dynamical behaviors of Boolean networks can be classified into three regimes: ordered, critical, and chaotic. In the ordered regime, the underlying network is robust and resilient to change, thus maintaining relatively simple dynamics of state trajectories and dampening small perturbations. In contrast, in the chaotic regime, a small perturbation can cause a large propagation of changes throughout the network, leading to divergence of state trajectories. Multiple studies show that the most information-rich and complex dynamics arise in systems operating in the critical regime, conferring adaptive and evolutionary functional advantages [12–17]. There is evidence that living systems operate at or near the critical regime, affording them the ability to optimally balance responsiveness to external signals with maintenance of stability and information propagation with minimal loss [18–27].

By means of numerical simulations, we show that vary-

ing degrees of cellular diversity arise in the phenotypes of homogeneous tissues, depending on the network regime, which is characterized by the Lyapunov exponent [28]. Specifically, we show that the maximal diversity of the phenotypes occurs near the critical phase transition. Furthermore, we find the emergence of spatial pattern formation of cellular phenotypes near the critical regime.

Most research on dynamics of Boolean networks has been carried out on a closed, single-cell network. However, in recent years, there have been several models and analyses for intercellularly interacting cells with Boolean networks on a two-dimensional grid [29–34]. A generic set of rules for modeling a homogeneous tissue of interacting Boolean networks as cellular automata was first developed by Serra *et al.* and Villani *et al.* [29, 30]. In cells near the critical regime, they observed a growth of dynamical disorder with a higher intensity level of interaction, measured by fractions of simulation runs required to reach attractors. Damiani *et al.* explored whether interactions among cells may contribute to expanding (or reducing) the variety of behaviors that are possible for a cell with a given genetic content and showed that short-distance interactions among Boolean models of genetic regulatory networks exhibit generic properties that are robust for different interaction mechanisms and concentration thresholds [31]. Flann *et al.* considered Kolmogorov complexity-based measures of set complexity of coupled Boolean networks on a two-dimensional grid [32]. The most information-rich patterns were found to emerge in ordered and critical networks. Damiani *et al.* and Serra *et al.* also analyzed the change of networks caused by a single gene perturbation on coupled Boolean Networks and proposed possible classifications of the cells according to the response to intercellular interactions [33, 34]. Taking these works as motivation, we explore the possibility of emergence of phenotypic diversity in a system of coupled noisy Boolean networks.

BOOLEAN NETWORKS

A Boolean network is defined on a set of n binary-valued variables, often termed ‘genes’, $x = \{x_1, \dots, x_n\}$, where $x_i \in \{0, 1\}$, with corresponding Boolean functions $\{f_1, f_2, \dots, f_n\}$ for each of the genes. Each gene value at $t + 1$ is determined by the values of the genes, $x_{j_1}, x_{j_2}, \dots, x_{j_{k_i}}$ at time t by means of the Boolean function, $f_i : \{0, 1\}^{k_i} \rightarrow \{0, 1\}$. There are k_i regulatory genes assigned to gene x_i that determine the wiring of the gene. Thus, the Boolean value of gene x_i is given by

$$x_i(t + 1) = f_i(x_{j_1}(t), x_{j_2}(t), \dots, x_{j_{k_i}}(t)).$$

In a random Boolean network (RBN), the regulatory genes for any given gene (i.e., its inputs) are chosen randomly. Thus, RBNs are ensemble models. The connectivity k_i can be selected from a given distribution, such

as Poisson. In the original formulation of RBNs, $k_i = k$ for all i .

In an RBN, the tables of Boolean update functions (truth table) are assigned values of 0 or 1 with a bias $p = P\{f_i = 1\}$, which is the probability that the function takes on the value of 1 for a combination of input values. During simulation of an instance of an RBN, the initial state is selected randomly out of all possible 2^n states. Additionally, to model stochastic dynamics, following the convention used in probabilistic Boolean networks [35, 36], we introduce a perturbation probability q for each time step and a random perturbation vector, $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_n)$, where $\gamma_i \in \{0, 1\}$ and $P\{\gamma_i = 1\} = q$, such that

$$\begin{cases} x(t+1) = x(t) \oplus \gamma, & \text{with probability } (1 - (1 - q)^n) \\ x(t+1) = \{f_1(x(t)), f_2(x(t)), \dots, f_n(x(t))\}, & \text{otherwise,} \end{cases}$$

where \oplus indicates the modulo-2 sum. The fact that any state transition has a nonzero probability under this perturbation model implies that the dynamics of the network is described by an ergodic Markov chain with a unique steady-state distribution. It is easy to see that without perturbation (when $q = 0$), the coupled Boolean networks is not an ergodic Markov chain, and the coupled Boolean networks could be considered as a large, deterministic Boolean network, which is the model described by Serra *et al.* and Villani *et al.*. Without perturbation, one would be unable to characterize the full dynamics of a network when multiple attractors are present.

MODEL TISSUE SETUP

We have constructed a model tissue as a two-dimensional square lattice (grid) of $m \times m$ cells, each represented by a Boolean network containing n genes. At the intracellular level, the Boolean network is selected randomly with bias p and a fixed connectivity k . Each cell has an identical Boolean network with a perturbation probability q . For the simulation, initial states of all the cells are randomly assigned.

At the intercellular level, all Boolean networks of the cells on the square grid are connected by the Von Neumann neighborhood (i.e., in the North, South, West, and East directions). Additionally, periodic boundary conditions are imposed on the tissue, so all cells are communicating identically. For all cells, a randomly designated gene i takes on inputs from another randomly selected gene j of its neighboring cells such that the Boolean value of this gene is determined by a fixed *linear threshold function* and its *linear threshold value*, θ . Formally, the linear threshold function is given by,

$$\begin{cases} x_i = 1, & \text{if } x_{i_{N_j}} + x_{i_{S_j}} + x_{i_{W_j}} + x_{i_{E_j}} \geq \theta, \\ x_i = 0, & \text{otherwise,} \end{cases}$$

where θ is either 1, 2, or 3, and $x_{i_{N_j}}$, $x_{i_{S_j}}$, $x_{i_{W_j}}$, $x_{i_{E_j}}$ are the j th genes, $i \neq j$, of the adjoining cells in the North, South, West, and East directions, respectively. The linear threshold function and θ are fixed for all cells in the tissue.

METHODS FOR QUANTIFYING PHENOTYPIC DIVERSITY

We measured the phenotypic diversity as a function of the Lyapunov exponent of the tissue. In dynamical systems, the Lyapunov exponent, λ , measures the rate of divergence in trajectories between two nearby initial conditions and defines the stability of a system. The Lyapunov exponent for Boolean networks uses the Hamming distance as the natural distance between states such that the smallest perturbation (to one bit) results in a Hamming distance of 1 [28]. We define a Boolean network as being ordered when $\lambda < 0$, critical when $\lambda = 0$, and chaotic when $\lambda > 0$. It has been shown that the Lyapunov exponent of a finite size Boolean network is the logarithm of the mean of the average sensitivities of the functions in the network [37]

$$\lambda = \log \left[\frac{1}{n} \sum_{i=1}^n s^{f_i} \right].$$

The average sensitivity, s^f , of a Boolean function f is the sum of the activities $\{\alpha_1^f, \alpha_2^f, \dots, \alpha_k^f\}$ of its inputs,

$$s^f = \sum_{i=1}^k \alpha_i^f,$$

where the activity, α_i^f , is the probability that toggling the i th input bit changes the function value:

$$\alpha_i^f = P\{f(x^{(i,0)}) \neq f(x^{(i,1)})\},$$

where $x^{(i,j)} = (x_1, \dots, x_{i-1}, j, x_{i+1}, x_k)$, $j \in \{0, 1\}$ and which, under the uniform distribution over the hypercube, is equal to

$$\alpha_i^f = \frac{1}{2^k} \sum_{x \in \{0,1\}^k} f(x^{(i,0)}) \oplus f(x^{(i,1)}).$$

Then, the Lyapunov exponent of the model tissue is the logarithm of the mean of the average sensitivities of all cells' $n \times m \times m$ Boolean functions $\{f_1, f_2, \dots, f_{n \times m \times m}\}$ (including the linear threshold functions that are used to connect the neighboring cells).

Phenotypes, which can be described as functional cellular states, have long been associated with dynamic attractors in networks [4, 5]. Attractors were initially interpreted as cell types by Kauffman. However, it has

also been shown that (high-dimensional) attractors correspond to phenotypic states of cells, such as proliferation, differentiation, or apoptosis [10, 11]. Much work has been done in determining the number of attractors in Boolean networks [38–43], showing that their number in critical networks scales super-polynomially. However, as already mentioned, under the noise model with a nonzero perturbation probability q , there are no attractors and, therefore, we describe the long-run behavior of the network by the steady-state distribution of the ergodic Markov chain. Empirically, the steady-state distribution is determined by counting the number of times each cellular state is visited during the simulation. The attractor states of the noiseless network (corresponding to $q = 0$) typically carry the largest probability mass (i.e., peaks) in the steady-state distribution.

In order to quantify the diversity of the cellular phenotypes in the tissue, we utilized the symmetric form of the Kullback-Leibler divergence to compare the converged steady-state distributions between pairs of cells in the entire model tissue, namely,

$$D_{KLS} = \frac{1}{2}(D_{KL}(P_i||P_j) + D_{KL}(P_j||P_i)),$$

where

$$D_{KL}(P_i||P_j) = \sum_{x \in \{0,1\}^n} P_i(x) \log \frac{P_i(x)}{P_j(x)},$$

and $P_i(x)$ and $P_j(x)$ are the steady-state distributions of cells i and j . Two similar cellular steady-state distributions in a tissue would result in a low D_{KLS} value, whereas a highly contrasting pair of cellular steady-state distributions would result in a high D_{KLS} value. We collected statistics on average D_{KLS} values under multiple simulations with different network parameters of connectivity, k , and bias, p .

We also employed a random restart strategy whereby with a probability, g , the ergodic Markov chain is restarted from a random state for each cell in order to improve the state space sampling and speed up convergence to the steady-state distribution. That is,

$$\begin{cases} x(t+1) = \sigma, \text{ with probability } g \\ x(t+1) = \{f_1(x(t)), f_2(x(t)), \dots, f_n(x(t))\}, \text{ otherwise,} \end{cases}$$

where $\sigma = (\sigma_1, \sigma_2, \dots, \sigma_n)$, is a random binary vector.

We used an empirical approach to set a *convergence threshold*, ϵ , for all model experiments. To determine the convergence of steady-state distributions of cells in the experiments, thus halting the simulation, we implemented an exponential checkpoint strategy. We compared the steady-state distributions at times $t_n = 2^n$ and $t_{n+1} = 2^{n+1}$ for each cell, once again utilizing D_{KLS} . When the D_{KLS} of $(P_i(t_n), P_i(t_{n+1}))$ for all cells, $i = 1, \dots, m \times m$ reached some small ϵ , we deemed the distribution to have converged and stopped the simulation.

To set ϵ , we used the maximum of the convergence time in simulations, which occurred in a 10×10 tissue with $k = 9$ and $p = 0.5$, with $\lambda \approx 1.4420$ (when the network is most chaotic). Figure 1 shows the comparison of convergence between maximum chaotic, maximum ordered, and critical network regimes with the model setup showing that convergence time generally increases with λ , and D_{KLS} decreases uniformly over time. From the empirical comparison, we have determined the convergence threshold value of $\epsilon < .02$ to be an appropriate stopping point for simulation.

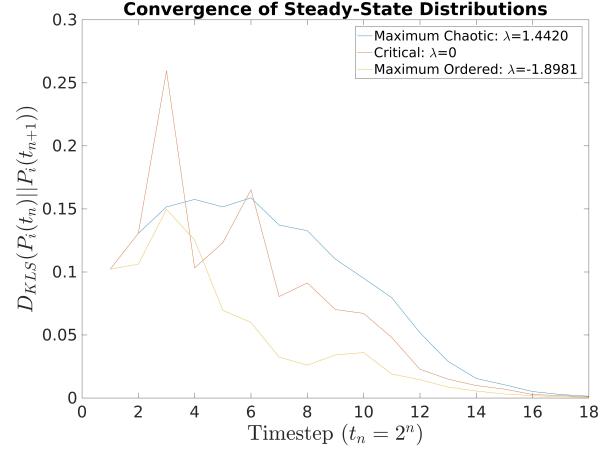


FIG. 1. Convergence of maximum chaotic, maximum ordered, and critical regimes for 10×10 tissues.

RESULTS

We performed simulations on 2×2 , 3×3 , 4×4 , and 10×10 grid tissues with 10 genes in each cell, with connectivity of k ranging from 1 to 9 (one gene designated for intercellular communication), and with the bias p ranging from 0.05 to 0.95, with increments of 0.025. Three different linear thresholds, $\theta = 1, 2, 3$, were considered for all sets of experiments. Perturbation probability of $q = 0.01$ was used for the noise model. The restart probability was set to $g = 0.05$. Each experiment was repeated 50 times. The first and second moments of the distribution for the pairwise D_{KLS} data of cells were collected for each experiment.

All 10×10 grid simulation results are shown in Figure 2. In all the plots, the x-axis indicates the Lyapunov exponent. In plots Figure 2a, Figure 2c, and Figure 2e the y-axis indicates the average value of all cellular pairwise D_{KLS} 's in an experiment. In plots Figure 2b, Figure 2d, and Figure 2f, the y-axis indicates the variance of all cellular pairwise D_{KLS} 's.

These data suggest a lack of diversity when $\theta = 1$ for 10×10 tissues (Figure 2). With such a low bar in the Boolean function for the intercellular gene to turn “on,”

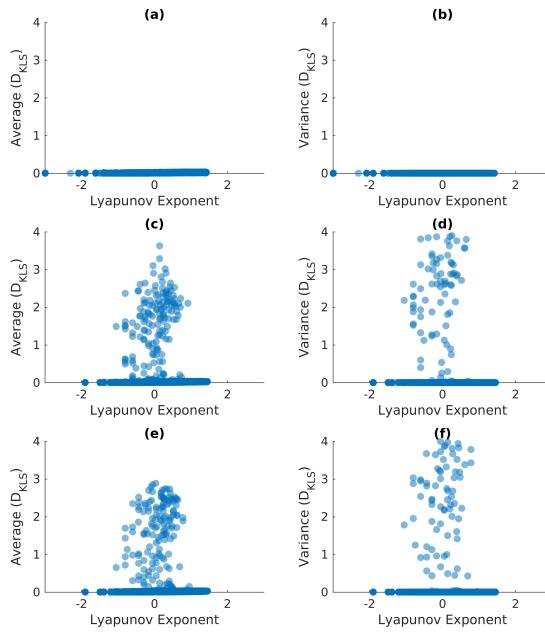


FIG. 2. Diversity vs. Lyapunov exponent in a 10×10 tissue, all experiments. Panels (a) and (b) are simulation results for $\theta = 1$; panels (c) and (d) are simulation results for $\theta = 2$; and panels (e) and (f) are simulation results for $\theta = 3$.

the tissues behave uniformly across the cells and always form similar steady-state distributions. In the plots with $\theta = 2$ and $\theta = 3$, there is a clear concentration of high-diversity (high average and variance of D_{KLS} values) in model tissues that are close to the critical regime (near $\lambda = 0$). 95% of the simulations that demonstrated high diversity fell within the range $-0.5 \leq \lambda \leq 0.5$, near the critical regime. Similar, but reduced effects were found on grids of even-sized lower dimension: 2×2 and 4×4 . A noticeable high diversity was observed starting with grid dimension of 6×6 . On grids of odd-sized dimension, 3×3 and 9×9 , diversity failed to arise (see Supplemental Fig. S1 [44]).

In the visual examination of the tissues that exhibited high-diversity, such as the one shown in Figure 3, definite alternating pattern formations of steady-state distributions were found. Figure 4a is a snapshot of Figure 3, at time $t = 1$, showing the states (as color squares) of the cells in the tissue. At $t = 2986$ (Figure 4b) and thereafter, the states of each cell were found generally to be arranged in an alternating fashion, barring random noise propagation. This checkerboard patterning was found in every tissue with diversity $D_{KLS} > 1.0$. This emergence of two different cell behaviors from identical Boolean networks could be a manifestation of a similar mechanism that may be at play in cellular differentiation, whereby genetically identical cells converge to different (sets of)

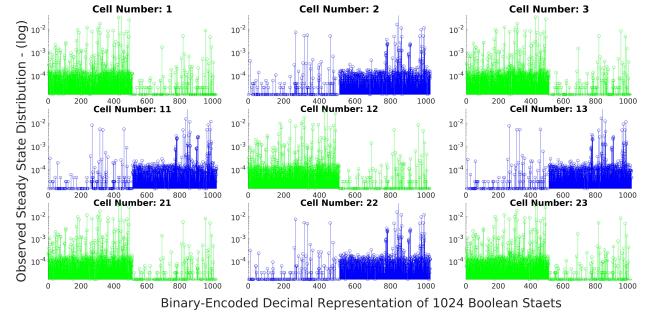


FIG. 3. Steady-state distributions of nine upper-left hand corner cells of a 10×10 “outlier” tissue on logarithmic scale. For each cell, the x-axis is the binary-encoded decimal representation of 1024 Boolean states, and the y-axis is the relative frequency of the states.

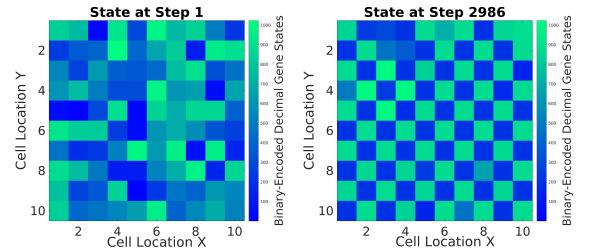


FIG. 4. (a) Randomly initialized states. (b) Formation of checkerboard pattern of cellular states at $t = 2986$. The individual squares are binary-encoded color representations of 1024 Boolean states of the 10×10 cells. See Supplemental Video S2 for the full temporal behaviors of the states [44].

states and, therefore, phenotypes.

See Supplemental Videos S1, S2, and S3 for tissue-level state transitions of coupled Boolean networks in ordered, near critical (Figure 4), and chaotic regimes. See Supplemental Videos S4, S5, and S6, for the dynamics of those tissues, colored by arithmetic mean of cell states in order to observe cumulative, long-term behaviors of cells. See Supplemental Table S1 for model parameters of an example Boolean network that does not diversify; and Supplemental Table S2 for model parameters of the diversifying Boolean network used in Figures 3 and 4 [44].

CONCLUSIONS AND DISCUSSION

Through numerical simulations, we have shown that on a homogeneous model tissue of cells, highest diversification of phenotypes occurs near the critical regime. Additionally, in those tissues that demonstrated such diversification, we observed an emergence of differentiation of phenotypes. The steady-state distributions formed a checkerboard pattern, which suggests the effects of spatial arrangement of cells and intercellular communication. However, not every tissue near the critical regime demonstrated this diversification of cells. Given that not

all model tissues near the critical regime necessarily exhibited high diversity, and the high diversity “outlier” tissues fell within a specific range of λ , we can speculate that there may be an inherent property of the Boolean networks, in addition to being in the critical regime, that affords the ability to diversify. Damiani *et al.* have previously observed distinct classes of behaviors in coupled, isolated critical Boolean networks in the instance of single gene perturbation [28]. Similarly, a certain class or property within our noisy tissue model in the critical regime may be the determinant of diversification. This property of the Boolean network, if it exists, that drives spatial differentiation in cells remains an open question to be further investigated.

We note that all the experiments in this work were carried out on a two-dimensional lattice with symmetric interactions of neighboring cells, resembling cellular automata. The emergence of checkerboard patterns of cellular phenotypes in only even-sized grids implies a role of spatial dependency and interaction rules. The inability for the tissue to diversify on an odd $m \times m$ grid could be explained by the vertex coloring problem in graph theory. On a connected graph where the vertex set is odd, and each of indegrees and outdegrees are 4, which is the model connectivity, it is not possible to label vertices with two colors such that no two vertices sharing the same edge have the same color. A state transition video of a 9×9 tissue with the diversifying network topology utilized in Figure 3 and Figure 4 is available as Supplemental Video S7 [44]. As can be seen in the video, cell states propagate and clearly diversify in small regions, but the tissue-wide checkerboard pattern never fully emerges, as it fails to occupy the sites with alternating colors due to deadlocks in ‘vertex coloring.’

While the checkerboard pattern formation resembles many known physical, dynamical systems such as anti-ferromagnetic coupled spin models, Turing pattern formation by reaction-diffusion system, or a noisy coupled oscillatory system, we note that as we are defining phenotypes of cells by their steady state distributions, the mechanisms underlying the pattern formation is not trivial. The perturbation in the coupled Boolean network model is an internal property, which means it impacts every cell in the tissue, at any time with probability q . While the resulting pattern formation of steady state distributions demonstrates the checkerboard pattern, the states of the cells across the cells almost never transition synchronously because of the perturbation. Because we are dealing with emerging behaviors of steady state distributions, the mechanisms of pattern formation in the tissue is less likely to be directly linked to the aforementioned systems, which are characterized by convergence of states at a specific time point.

An interesting area of further research is examining the effect of such self-organizing phenotypes in different spatial arrangements such as hexagonal grid or

three-dimensional (off-lattice) free space with agent-based modeling. In an off-lattice setup, the rules of interaction would not depend on Von Neumann neighborhoods, but rather on diffusion of molecules, which requires asynchronous updating of intracellular networks. Such interactions may result in Turing-type pattern formation in the context of gene regulatory networks, which is a well-studied natural phenomenon in biology (modeled as reaction-diffusion equations) [45, 46]. We have demonstrated the possibility of diversification of phenotypes in a restricted non-linear stochastic model. Thus, we believe biological realism could be achieved in future research in more complex model setups, with results not restricted to checkerboard patterns.

This study corroborates previous findings that complexity of computation or information processing is maximized in systems operating near a phase transition between an ordered and a chaotic regime [12–14]. Additionally, we see that even in a relatively simple model system of a tissue with isogenic cells, a natural emergence of cellular differentiation can take place.

The authors would like to thank Simon Kahan (Bionillion, SPC), Felix Ye (University of Washington), and Moriah Echlin (ISB) for helpful discussions. The authors would also like to credit Charlie Fieseler and Katherine Lacy of University of Washington for early help in model coding. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system and funded by the STF at the University of Washington.

-
- [1] E. Haeckel, Jena Z. Naturwiss. **8**, 1 (1874).
 - [2] K. V. Mikhailov, A. V. Konstantinova, M. A. Nikitin, P. V. Troshin, L. Y. Rusin, V. A. Lyubetsky, Y. V. Panchin, A. P. Mylnikov, L. L. Moroz, S. Kumar, *et al.*, Bioessays **31**, 758 (2009).
 - [3] A. Zakhvatkin, Sources and method of the origin of metazoan development. Moscow, Russia: Soviet Science (1949).
 - [4] S. A. Kauffman, Journal of theoretical biology **22**, 437 (1969).
 - [5] S. A. Kauffman, in *Spin Glasses and Biology* (World Scientific, 1992) pp. 61–100.
 - [6] L. Kadanoff, S. Coppersmith, and M. Aldana, arXiv preprint nlin/0204062 (2002).
 - [7] A. S. Ribeiro and S. A. Kauffman, Journal of theoretical biology **247**, 743 (2007).
 - [8] C. J. Kuhlman and H. S. Mortveit, Theoretical Computer Science **559**, 20 (2014), non-uniform Cellular Automata.
 - [9] R. Serra, M. Villani, A. Barbieri, and A. Colacci, Journal of theoretical biology **265**, 185 (2010).
 - [10] H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, and S. Huang, Nature **453**, 544 (2008).
 - [11] S. Huang, G. Eichler, Y. Bar-Yam, and D. E. Ingber, Physical review letters **94**, 128701 (2005).
 - [12] C. G. Langton, Physica D: Nonlinear Phenomena **42**, 12

- (1990).
- [13] D. J. Galas, M. Nykter, G. W. Carter, N. D. Price, and I. Shmulevich, *IEEE Transactions on Information Theory* **56**, 667 (2010).
 - [14] M. Nykter, N. D. Price, A. Larjo, T. Aho, S. A. Kauffman, O. Yli-Harja, and I. Shmulevich, *Physical review letters* **100**, 058702 (2008).
 - [15] P. Krawitz and I. Shmulevich, *Physical review letters* **98**, 158701 (2007).
 - [16] C. Torres-Sosa, S. Huang, and M. Aldana, *PLoS computational biology* **8**, e1002669 (2012).
 - [17] J. Hidalgo, J. Grilli, S. Suweis, M. A. Muñoz, J. R. Banavar, and A. Maritan, *Proceedings of the National Academy of Sciences* **111**, 10095 (2014).
 - [18] E. Balleza, E. R. Alvarez-Buylla, A. Chaos, S. Kauffman, I. Shmulevich, and M. Aldana, *PLoS One* **3**, e2456 (2008).
 - [19] I. Shmulevich, S. A. Kauffman, and M. Aldana, *Proceedings of the National Academy of Sciences of the United States of America* **102**, 13439 (2005).
 - [20] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, *Proceedings of the National Academy of Sciences* **105**, 1897 (2008).
 - [21] P. Rämö, J. Kesseli, and O. Yli-Harja, *Journal of Theoretical Biology* **242**, 164 (2006).
 - [22] D. Krotov, J. O. Dubuis, T. Gregor, and W. Bialek, *Proceedings of the National Academy of Sciences* **111**, 3683 (2014).
 - [23] W. Bialek, A. Cavagna, I. Giardina, T. Mora, O. Pohl, E. Silvestri, M. Viale, and A. M. Walczak, *Proceedings of the National Academy of Sciences* **111**, 7212 (2014).
 - [24] T. Mora and W. Bialek, *Journal of Statistical Physics* **144**, 268 (2011).
 - [25] J. M. Beggs, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* **366**, 329 (2008).
 - [26] W. L. Shew and D. Plenz, *The neuroscientist* **19**, 88 (2013).
 - [27] B. Del Papa, V. Priesemann, and J. Triesch, *PloS one* **12**, e0178683 (2017).
 - [28] B. Luque and R. V. Solé, *Physica A: Statistical Mechanics and its Applications* **284**, 33 (2000).
 - [29] R. Serra, M. Villani, C. Damiani, A. Graudenzi, A. Colacci, and S. Kauffman, in *Proceedings of ECCS07: European Conference on Complex Systems* (2007).
 - [30] M. Villani, R. Serra, P. Ingrami, and S. A. Kauffman, in *ACRI* (Springer, 2006) pp. 548–556.
 - [31] C. Damiani, R. Serra, M. Villani, S. Kauffman, and A. Colacci, *IET systems biology* **5**, 137 (2011).
 - [32] N. S. Flann, H. Mohamadlou, and G. J. Podgorski, *Biosystems* **112**, 131 (2013).
 - [33] C. Damiani, A. Graudenzi, and M. Villani, *Artificial Life and Evolutionary Computation* , 259 (2010).
 - [34] R. Serra, M. Villani, C. Damiani, A. Graudenzi, and A. Colacci, *Cellular automata* , 315 (2008).
 - [35] I. Shmulevich and E. R. Dougherty, *Probabilistic Boolean networks: the modeling and control of gene regulatory networks* (SIAM, 2010).
 - [36] I. Shmulevich, E. R. Dougherty, S. Kim, and W. Zhang, *Bioinformatics* **18**, 261 (2002).
 - [37] I. Shmulevich and S. A. Kauffman, *Physical review letters* **93**, 048701 (2004).
 - [38] B. Drossel, *Physical Review E* **72**, 016110 (2005).
 - [39] B. Drossel, T. Mihaljević, and F. Greil, *Physical Review Letters* **94**, 088701 (2005).
 - [40] V. Kauffman, T. Mihaljević, and B. Drossel, *Physical Review E* **72**, 046124 (2005).
 - [41] K. Klemm and S. Bornholdt, *Physical Review E* **72**, 055101 (2005).
 - [42] T. Mihaljević and B. Drossel, *Physical Review E* **74**, 046101 (2006).
 - [43] B. Samuelsson and C. Troein, *Physical Review Letters* **90**, 098701 (2003).
 - [44] See Supplemental Material at [] for additional information and videos.
 - [45] S. Kondo, *Systems Biology* , 37 (2009).
 - [46] H. Meinhardt and A. Gierer, *Bioessays* **22**, 753 (2000).