Inferring Spatial Dynamics through Interpretable Neural Cellular Automata

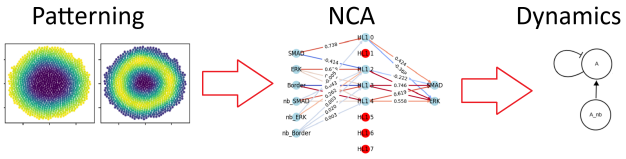
Jakob H. Schauser[1], Ala Trusina[1]
[1]Niels Bohr Institute, Denmark jakob.schauser@nbi.ku.dk

Abstract

Morphogenetic patterning can be reliably modeled using the self-organizing capabilities of Neural Cellular Automata (NCA). But given that we can emulate a dynamical system, what can we do to interpret it? Taking our onset in a specific biological model-system, we propose a method where regulatory elements driving the dynamics can be inferred by probing the internal interactions trough estimation of the Jacobian. Learning the emergent rules behind the NCA's allows us to better understand and engineer biological patterning.

Submission type: **Late Breaking Abstracts**

Graphical Abstract



Introduction

Patterning is ubiquitous in nature. Everyone knows the characteristic Turing patterns visible on a zebra, but inarguably more important is the invisible patterning that defines and stabilizes cell fate decisions, driving embryonic development. This Morphogenetic patterning is determined by the interaction between a number of chemicals called morphogens, whose spatial organization is fundamental for morphogenesis, and, by extension, multicellular life itself.

There are infinitely many interactions that can lead to the same outcome, making solving the inverse problem both computationally and conceptually challenging. The purely local interactions of the NCA allow for modeling the dynamics while respecting the inherent spatial symmetries. But how can we pry open the black box?

In this work we propose a novel Jacobian-based interpretability method that allows for estimation of the inter- and intra-cellular dynamics that can lead to certain spatial distributions of cell types. Specifically, we are interested looking at morphogens in two-dimensional mouse gastruloid - a model system for morphogenetic patterning. Our method allows for finding intuitive and transparent ways to model the system. This can help both design experimental setups, run simulations, and possibly aid our understanding of certain biochemical processes.

From Neural Self-Organization to interpretable dynamics

We present a brief outline of the proposed methodology:

1. Patterning is observed in the system
2. NCAs reproduce the pattern and give us the Jacobian
3. The Jacobian helps us infer the interactions in terms of a number of chemicals and their interactions
4. This understanding guides new simulation or experiments

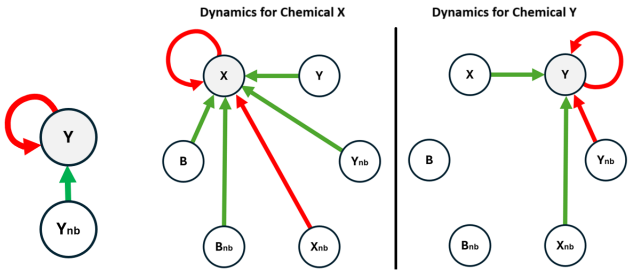
Now, the specifics:
For two chemicals X and Y, to first order, their internal dynamics can be described by the Jacobian:

$$\mathbf{J} = \begin{pmatrix} J_X^X & J_Y^X \\ J_X^Y & J_Y^Y \end{pmatrix} = \begin{pmatrix} \partial_X \dot{X} & \partial_Y \dot{X} \\ \partial_X \dot{Y} & \partial_Y \dot{Y} \end{pmatrix} \tag{1}$$

Where $\dot{X}$ is $\frac{dX(X,Y,y)}{dt}$ and $\partial_X Y$ is shorthand for $\frac{\partial X}{\partial Y}$

The sign, magnitude, and dependencies of the entries in the Jacobian reveal the temporal and spatial interplay of the system variables, forming an *interaction network* similar to a gene regulatory network. Up- and down-regulation between neighboring cells can of course be direct reactions, but can also arise from diffusion processes, as shown in Equation 2 for a chemical Y at point i and neighboring points j:

$$\nabla^2 Y_i \approx \frac{\sum_j Y_j - Y_i}{\Delta x} \propto J^Y_{Y_{nb}} - J^Y_Y \tag{2}$$



(a) (b)

Figure 1: **(a)** An example of an interaction network that corresponds to diffusion **(b)** An example of interaction networks for two cross-inhibiting, diffusing chemicals as inferred by the proposed method.
`X` and `Y` are the active variables, `B` is the system boundary, and the `nb`-subscript corresponds neighbors. Red and green arrows are inhibition and promotion, respectively.

By such mapping we can begin to catalog the effective building blocks of the emergent patterning, including diffusion, cross-inhibition, self-activation, or long-range feedback. The ability to read these directly from an NCA-trained Jacobian means that otherwise opaque self-organizing systems can be translated into interpretable causal diagrams. The previously mentioned diffusion (Eq. 2) corresponds to the interaction network as seen in Figure 1a

Two-dimensional patterning

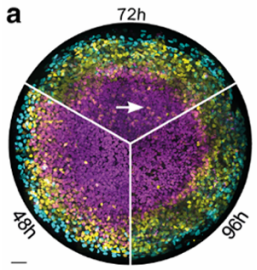


Figure 2: The 2D gastruloid model system **(adapted from Chen et al. (2024))**

For method development, we took our onset in gastruloids as described in Siggia and Warmflash (2018) for example. As a model system, the two-dimensional gastruloid provides a minimal and controllable setup for studying gastrulation patterning *in vitro*, were especially a number of stable, concentric rings are of interest (as seen in Figure 2).

Our lab has recently developed an *in silico* agent-based model, explaining some of the diverse dynamics observed in the system. We utilized this data to develop the approach presented here.

Preliminary Results / Discussion

For simple patterning, NCAs quickly learn stable, reproducible update rules. The inferred interaction networks imply biologically plausible causal relationships, with realistic diffusion, inhibition, and promotion.

As an example, inferring the dynamics for two self-inhibiting chemicals with identical boundary conditions result in the interaction network as seen in Figure 1b. This corresponds to the following short-hand equations:

$$\dot{X}_i \propto B_i + Y_i - X_i + \sum_j^{\{nbs\}} (B_j + Y_j - X_j)$$

$$\dot{Y}_i \propto X_i - Y_i + \sum_j^{\{nbs\}} (X_j - Y_j)$$

When simulating the above equations (after tuning relevant parameters), the resulting patterning correctly resemble the two observed chemicals.

More complex dynamics also show promising results, though further validation is ongoing. For most simple observed patterns, sensible interaction networks are generated, but interpretability depends on the linearity and separability of the active variables; nonlinearities (*and*-logic) remain difficult to disentangle.

The methodology is flexible enough for either training the NCA to reproduce the end-result or the full time-series, depending on the experimental constraints. When observing fewer time steps, the possible solution space increases, leading to a broader variety of discovered solutions. We are currently investigating how to best/most exhaustively search this space, a as well as exploring the generalization of our approach to other systems.

Final remarks

In this work we have presented a framework for interpreting self-organizing solutions through NCAs. The method was applied to a well known model system and while in a very early stage, shows promising result. Our approach, focusing on interpretability, generates hypotheses, which can ultimately be tested experimentally.

By bridging black-box learning and interpretable modeling, Neural Cellular Automata can provide a powerful tool for exploring emergent patterning, allowing us to further our understanding of the chemical basis of morphogenesis.

Acknowledgements

This work was supported by the Novo Nordisk Foundation, grant No. NNF23OC0086722

References

Chen, B., Khan, H., Yu, Z., Yao, L., Freeburne, E., Jo, K., Johnson, C., and Heemskerk, I. (2024). Extended culture of 2d gastruloids to model human mesoderm development. *undefined*.

Siggia, E. D. and Warmflash, A. (2018). Modeling mammalian gastrulation with embryonic stem cells. *Current Topics in Developmental Biology*, pages 1–23.