Bipartite Networks to Study the Genotype-to-Phenotype Relationship in Cellular Automata Models

Christian Darabos
Dartmouth College
Hanover, NH, USA
Christian.Darabos@dartmouth.edu

Britney E. Graham
Dartmouth College
Hanover, NH, USA
Britney.E.Graham.GR@dartmouth.edu

Ting Hu
Dartmouth College
Hanover, NH, USA
Ting.Hu@dartmouth.edu

Jason H. Moore
Dartmouth College
Hanover, NH, USA
Jason.H.Moore@dartmouth.edu

ABSTRACT

In biological organisms, a single genotype may map to several phenotypes and vice-versa. This many-to-many relationship is believed to be a major drive of the phenotypic robustness and genotypic evolvability found in all life forms. Given the inherent complexity of the genotype-to-phenotype (G2P) mappings, we use cellular automata (CAs) as rudimentary proxies for biological organisms. CA models have the same many-to-many G2P mappings, and their sensitivity to initial conditions allows the same genotype to differentiate into different phenotypes. We use a bipartite network to study the G2P landscape, and its projections in either space. The network and its projections all have a Òheavy-tailedÓ degree distribution, hinting at an increased robustness supported by the network structure. We also show a strong correlation between the phenotypeÕs complexity and its robustness. We are currently working on analyzing the relationships between the robustness and the evolvability both at the genotypic and phenotypic level. Preliminary results agree with those of previous similar studies, using different computational models.

Categories and Subject Descriptors

F.1.1 [**Models of Computation**]: Unbounded-action devices—*cellular automata*; I.2.8 [**Artificial Intelligence**]: Problem Solving, Control Methods, and Search—*Heuristic methods*; E.1 [**Data Structures**]: Graphs and networks—*Bipartite networks*

Keywords

cellular automata, genotype, phenotype, bipartite network

1. INTRODUCTION

For the past two decades, geneticists have been studying the intricate genotype-to-phenotype (G2P) relationship in biological organisms. Genome-wide association studies (GWAS), and the recent advances in modern high throughput sequencing technologies, have made understanding how metabolic reactions, cell signaling, and developmental pathways translate the genome of a living organism into its phenotype an achievable goal [4]. However, GWAS have also unveiled unprecedented degrees of complexity, making clinical progress much slower than anticipated. As geneticists learn more about G2P mappings, it becomes more apparent that there is a many-to-many relationship. Indeed, several different genotypes, usually resulting from small perturbations or *neutral mutations*, result in the exact same phenotype. This feature is responsible for the phenotypic robustness of biological organisms, and their relative insensitivity to small genetic perturbations. On the other hand, identical genotypes may develop into dramatically different phenotypes, depending on a set of internal and external signals and factors. The embryonic stem cell, which may potentially develop into any cell type, is a prime example of a single genotype yielding several phenotypes. The ability to adapt to internal and external factors is believed to be at major factor of the evolvability of all life forms. Given the inherent complexity of the G2P mappings, we recognize the need for smart, adaptive mathematical, statistical or computational models to study this relationship.

Cellular automata (CAs) are dynamical, usually deterministic, discrete, abstract models primarily used to simulate and study distributed computation. CAs have also been used for years as a rudimentary proxy for biological organisms. One prevalent example is Kauffmann's Random Boolean Network (RBN) model for genetic regulatory networks [3]. In all CAs, update functions are generally represented as a Boolean lookup table of all possible binary permutations of the cell's neighborhood. Starting with an initial configuration (IC) of cells (i.e. set of states of all cells), the system will possibly pass throughout transient until it reaches a configuration previously visited. Because of its deterministic nature, the CA will get caught in a *attractor* of one or more configurations. Due to their simplicity and flexibility, CAs and their subsequent extensions are very attractive as models to study the robustness and evolvability in biological systems. Indeed, CAs have a genotype, a phenotype, and mimic the many-to-many G2P mappings. The update function of CAs is a direct equivalent of a genotype, which can be mutated at will, and is a set of rules followed by the system to achieve a steady state. The attractor reached by the CAs is the phenotype resulting from

GECCO'13 Companion, July 6–10, 2013, Amsterdam, The Netherlands.
ACM 978-1-4503-1964-5/13/07.

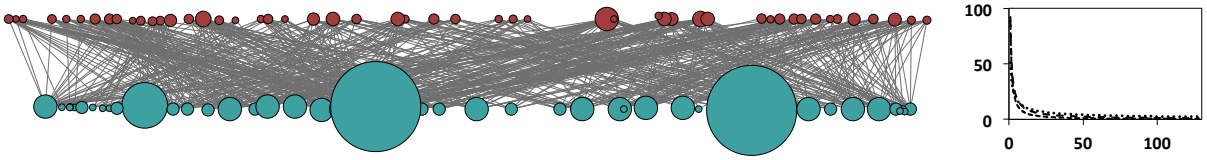

Figure 1: Filtered Bipartite G2P Network and Degree Distributions.

a genotype and an initial configuration. The same attractor can be reached by different BUFs, and a single BUF can result in different attractors depending on the IC. In this work, we explore the evolvability, robustness, and accessibility of pseudo-biological organisms modeled by a small CA. We exhaustively explore all G2P mappings of a CA by representing it in a bipartite network, and also projecting it onto the phenotype and genotype landscape respectively. Additionally, we study the distribution of robustness (also called neutrality) in the genotypic landscape of our model, and its effect on the phenotypic landscape. Similarly, we look at the genotypic and phenotypic evolvability, and the correlations between robustness and evolvability. Indeed, the seemingly contradictory effect of robustness and evolvability has been studied and disproved in many systems, where they in fact facilitate each other [1, 2].

2. METHODS & RESULTS

In order to fully explore the G2P relationship in our CA model, we exhaustively explore all possible genotype mappings for all possible ICs. Unfortunately, the (super-)exponential nature of the genotype and phenotype spaces, we are limited to a small number of cells, $N = 5$, and a *radius* $r = 1$, where the radius defines the number of neighbors each cell arranged on a ring can reach on either side. Therefore, a radius $r = 1$ results in neighborhood sizes of $n = 3$. In CAs, there are $2^{2^n} = 2^{2^3} = 256$ possible genotypes. CAs have $2^N = 2^5 = 32$ possible ICs, and the same number of possible point (i.e. single configuration) attractors, and at most $2^{2^n} \times 2^N = 8192$ possible attractors of any length, as every combination of genotype and IC can potentially result in a different phenotype. Figure 1 represents the filtered bipartite network, genotypes on the top row, phenotypes on the bottom. The vertex size is proportional its degree (i.e. to the number of mapped phenotypes, or mapping genotypes respectively. For readability reasons, we have filtered out vertices of a degree below 5.) We also show the trends of degree distributions for the bipartite network, as well as both projections of the genotypes only and of the phenotypes only. All degree distributions are right skewed, with a heavy tail, denoting the presence of highly connected "hub" vertices and a "scale-free" like topology. The literature reports several measurements of genotypic and phenotypic robustness and evolvability, we use the definitions detailed in [5]. We study the statistical characteristics of the genotype and phenotype space, assigning robustness, evolvability, and accessibility "scores" to each genotype and phenotype. For instance, we report in Figure 2 the strong, quasi linear positive correlation between the number of phenotypes mapping to a genotype, and the evolvability of those genotypes. This correlation is to be intuitively expected from biological organisms in which genotypes responsible for more phenotypes are also considered the most evolvable.

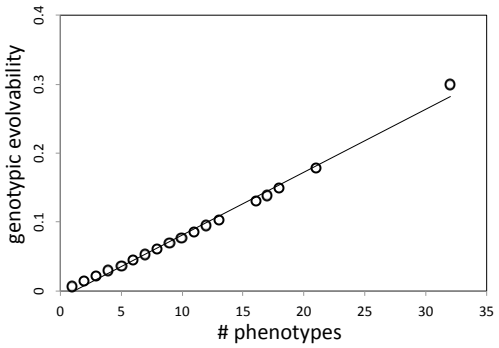

Figure 2: Strong correlation between the number of phenotype mapped by a genotype and the genotypic evolvability.

We also witness a strong negative correlation between the length of the phenotype's attractor and its robustness. This result agrees with Kauffman's work on RBNs. The same negative correlation appears with the phenotypic evolvability. The "bell shaped" distributions of genotypic and phenotypic robustness, or neutrality, are also aligned with results in similar studies. We are currently working on analyzing the complex relationships between robustness and evolvability, both genotypic and phenotypic. These links are, we believe, the most biologically relevant and could confer the most relevance to our model.

Acknowledgments

The authors gratefully acknowledge the financial support of the NIH (grants AI59694, LM010098 and LM009012).

3. REFERENCES

[1] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold. Protein stability promotes evolvability. *Proc Natl Acad Sci U S A*, 103(15):5869–5874, Apr 2006.

[2] E. Ferrada and A. Wagner. Protein robustness promotes evolutionary innovations on large evolutionary time-scales. *Proceedings of the Royal Society London B*, 275:1595–1602, 2008.

[3] S. A. Kauffman. Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol.*, 22:437–467, 1969.

[4] S. V. Nuzhdin, M. L. Friesen, and L. M. McIntyre. Genotype-phenotype mapping in a post-gwas world. *Trends Genet*, 28(9):421–426, Sep 2012.

[5] A. Wagner. Robustness and evolvability: a paradox resolved. *Proc Biol Sci*, 275(1630):91–100, Jan 2008.