

Preface

When a mathematician or biologist hears the term “mathematical biology,” the mental picture that comes to mind for many may be that of calculus-based techniques such as differential equations. There is, of course, much more of a diversity than this, though other types of mathematical biology often live under an umbrella with a different name. For example, many problems and techniques involving discrete mathematics have been relegated to the world of bioinformatics. Another large area of mathematical work in the life sciences is biostatistics, and yet another one emerging more recently is data science. Indeed, the lines between these fields are blurred and subjective. An area that involves mathematics and biology may be considered mathematical biology to some but not to others. Some research projects blend so many different fields that it is unnatural to separate into distinct silos such as “mathematics,” “genomics,” “computational biology,” etc. Rather, they are true transdisciplinary *science* problems: a project on epidemiology might draw from applied mathematics, biology, public health, statistics and data science, computer science, network science, and economics; a project in phylogenetics might involve researchers from mathematics, computer science, a number of fields in biology, statistics, data science, and genomics; and a research group working on protein folding might consist of biologists, biochemists, biophysicists, mathematicians, statisticians, and computer scientists.

Early work involving discrete and algebraic methods to model biological systems can be traced back to (at least) the 1960s. In 1969, theoretical biologist Stuart Kauffman proposed modeling gene regulatory network with Boolean functions. Around the same time, biologist René Thomas pursued a similar modeling framework that he called “logical models.” These types of models have been studied since under different names, such as Boolean networks, automata networks, generalized cellular automata, and others. In some cases, the models are not Boolean, but ternary, or feature a larger state space. If the state space is a finite field (if not, one can just expand it until it is), then the individual functions describing the model are polynomials. This opens a door to using the rich toolbox of computational algebra for analyzing such network models, leading to the province of *Algebraic Biology*. Among the many other examples where discrete mathematics and algebra facilitate progress in modern biology are the field of *Algebraic Statistics* that has proved instrumental for a number of problems in genomics and phylogenetics.

One hallmark of transdisciplinary research is that its results and subsequent publications could not have been produced only by expertise from a subset of the participating disciplines. This is a far cry from some multidisciplinary work where researchers from each discipline may work somewhat independently on individual “modules,” then write separate sections for the project report and subsequent publication. Transdisciplinary research is also a powerful catalyst for accelerating advancement for each of the individual disciplines. In biology, the advent of high-throughput technology in the late 20th and early 21st century such as gene sequencers, RNA-Seq, and CRISPR, along with the rise of high-performance computing, has put this discipline firmly in the spotlight as a prime field to be transformed by mathematics and technology. In 2004, biologist Joel Cohen famously predicted that this is a two-way process when he published the paper titled “Mathematics is biology’s next microscope, only better. Biology is mathematics’ next physics, only better.” The following year, mathematician Bernd Sturmfels asked in the title of a paper he wrote “Can biology lead to new theorems?,” and then proceeded in the body of the paper to answer and support this claim in the affirmative.

The purpose of this book is to highlight some of the new areas of mathematical biology with combinatorial and algebraic flavors and a distinct computational/statistical component. It is in no way meant to be comprehensive, and reflects the personal preferences of the editors to highlight current trends in the discipline. Most importantly, the book reflects our efforts to address the urgent need to connect ongoing advances in discrete and algebraic mathematical biology with the academic curriculum where calculus-based methods still dominate the landscape. While the use of modern algebraic methods is now in the mainstream of mathematical biology research, this trend has been slow to influence the traditional mathematics and biology curricula. Students interested in mathematical biology have relatively easy access to courses that utilize classical analytic methods based on difference and differential equations. By contrast, students interested in algebraic and discrete computational approaches have fewer doors visibly open to them, and indeed may not even know that they exist. Several high-profile national reports have urged the mathematical biology community to enact steps to bridge this gap,¹ and since 2013, the editors have collaborated with groups of like-minded faculty to make headways in addressing this problem. Together, we have led several professional faculty development workshops—at the Mathematical Biosciences Institute at the Ohio State University (2013) and the National Institute for Mathematical and Biological Synthesis (NIMBioS) at the University of Tennessee (2014 and 2016)—focused on developing, disseminating, and classroom-testing novel educational materials based on cutting-edge research in discrete and combinatorial mathematical

1. The report *Vision and change in undergraduate biology education: a call to action* of American Association for the Advancement of Science (2011) and the National Research Council’s report *The Mathematical Sciences in 2025* (2013) are just two examples.

biology. In fact, this book could be viewed as the third publication in a series that has been linked with those workshops.

The first book, titled *Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Models* and published in 2013, was edited by Raina Robeva and Terrell Hodge. Topics include Boolean networks, agent-based, and neuronal models, linear algebra models of populations and metabolic pathways, hidden Markov models in genetics, and geometric approaches in phylogenetics. The second publication, *Algebraic and Discrete Mathematical Methods for Modern Biology*, edited by Raina Robeva and published in 2015, covers topics from graph theory in systems biology, ecology, and evolution, more topics on Boolean networks, Petri nets, epidemiology on networks, linear algebraic approaches in genetics and metabolic analysis, computational phylogenetics, and RNA folding. Most of the material in these books is accessible to undergraduates who have not necessarily taken calculus. In addition to being ideal for undergraduates, these books can provide detailed introductions to the topics for biologists who have limited or even no calculus background. The current "Volume 3" explores a new set of topics with a distinct computational flavor, either not covered in the previous two, or topics that have emerged as fundamental to the field in the last few years. Although our target audience this time is primarily graduate students, we have made every effort to keep most of the topics accessible to advanced undergraduates as well. All three books are filled with examples and exercises to promote their use in the classroom, and feature notes on the use of specialized software for computation, analysis, and simulation. The chapters are designed to be largely independent from one another and can be viewed as starting points for undergraduate research projects or as entryways for graduate students and researchers new to the field of algebraic mathematical biology. They can also be used as "modules" for classroom use and independent studies. Solution guides containing the solutions to most exercises are also available.

The chapters of this volume are organized to highlight several common themes. We begin with a chapter on multiscale modeling, with a focus on the molecular level, followed by two chapters on the assembly of DNA. Chapters 4–6 involve topics on discrete models of the dynamics of molecular networks. More specifically, Chapter 4 introduces the local modeling framework, which attempts to clarify and unify a number of modeling paradigms, including Boolean networks, logical models, and automata networks. Chapter 5 considers these systems with stochastic features, which are sometimes called Stochastic Discrete Dynamical Systems.

Chapter 6 looks at the question of reverse engineering the wiring diagram, using techniques from combinatorial commutative algebra and algebraic geometry—namely Stanley-Reisner theory and the primary decomposition of square-free monomial and pseudomonomial ideals. Though Chapter 7 is on a problem from neuroscience, it also involves the same underlying algebraic framework as Chapter 6. The concept of a pseudomonomial ideal, as far as

we can tell, had not been studied until it arose recently in several diverse areas in mathematical biology, from reverse engineering—molecular networks to encoding the structure of place fields in neuroscience. Researchers are now studying and publishing on these objects and on so-called “neural ideals.” This is a prime example of how biology is leading to new theorems, as predicted by Sturmfels. The neuroscience topic continues into Chapter 8 on threshold linear ODE models over graphs—a framework now used as a simple model of firing patterns in neurons. A central theme in this chapter is how to deduce the dynamics of the system from the structure of the underlying graph.

The focus of Chapter 9 is on multistationarity in biochemical reaction networks. Although this topic may appear unrelated, the aforementioned theme of connecting local network structure to global system dynamics emerges once again, after being introduced in Chapter 4, and being an underlying theme of Chapter 8. This question has appeared throughout the decades in different forms. Back in the 1980s, René Thomas posed these questions both in the context of logical models (recall, a variant of Boolean networks), which were popular models of gene networks, and in continuous differential equation frameworks. He observed that as a rule of thumb, positive feedback is a necessary condition for having multiple steady states (multistationarity), but negative feedback loops are necessary for cyclic attractors, and hence homeostasis. These conjectures have since been formalized and proven in a number of settings, from discrete models to differential equations.

Chapter 10 is on optimization and linear programming in phylogenetics, where the problem to infer and interpret a phylogenetic tree is useful in multiple contexts in biology and medicine. Finally, Chapters 11 and 12 examine classification in biology through clustering and machine learning, with examples ranging from protein families to environmental systems.

This book would not have been possible without the dedicated team of authors who felt passionately about the value of presenting their research results in a way that provides hands-on practical knowledge for readers ranging from advanced undergraduate students to researchers entering the field of algebraic and computational biology. We are grateful for their patience during the editing process and for their willingness to go through multiple revisions with us. We warmly appreciate the support of NIMBioS for the 2016 workshop *Discrete and Algebraic Mathematical Biology: Research and Education*. Work on many of the chapters in this volume started during this workshop and may not have materialized otherwise. Our personal thanks go to Katerina Zaliva, our Editorial Project Manager, who was gracious with her time, prompt to answer questions, and ready to adopt a cheerful attitude during some of the unavoidable challenges in the process. Finally, we thank our spouses, Catherine Gurri and Boris Kovatchev, for their patience and support throughout.

Matthew Macauley
Raina Robeva
 August 27, 2018