

ALGEBRAIC AND DISCRETE MATHEMATICAL METHODS FOR MODERN BIOLOGY

RAINA ROBEVA, EDITOR

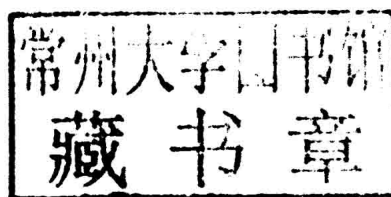


Algebraic and Discrete Mathematical Methods for Modern Biology

Edited by

Raina S. Robeva

Department of Mathematical Sciences, Sweet Briar College, Sweet Briar, VA, USA



AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Academic Press is an imprint of Elsevier



Academic Press is an imprint of Elsevier
32 Jamestown Road, London NW1 7BY, UK
525 B Street, Suite 1800, San Diego, CA 92101-4495, USA
225 Wyman Street, Waltham, MA 02451, USA
The Boulevard, Langford Lane, Kidlington, Oxford OX5 1GB, UK

First edition 2015

Copyright © 2015 Elsevier Inc. All rights reserved.

No part of this publication may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or any information storage and retrieval system, without permission in writing from the publisher. Details on how to seek permission, further information about the Publisher's permissions policies and our arrangements with organizations such as the Copyright Clearance Center and the Copyright Licensing Agency, can be found at our website: www.elsevier.com/permissions.

This book and the individual contributions contained in it are protected under copyright by the Publisher (other than as may be noted herein).

Notices

Knowledge and best practice in this field are constantly changing. As new research and experience broaden our understanding, changes in research methods, professional practices, or medical treatment may become necessary.

Practitioners and researchers must always rely on their own experience and knowledge in evaluating and using any information, methods, compounds, or experiments described herein. In using such information or methods they should be mindful of their own safety and the safety of others, including parties for whom they have a professional responsibility.

To the fullest extent of the law, neither the Publisher nor the authors, contributors, or editors, assume any liability for any injury and/or damage to persons or property as a matter of products liability, negligence or otherwise, or from any use or operation of any methods, products, instructions, or ideas contained in the material herein.

Library of Congress Cataloging-in-Publication Data

A catalog record for this book is available from the Library of Congress

British Library Cataloguing in Publication Data

A catalogue record for this book is available from the British Library

For information on all Academic Press publications
visit our website at <http://store.elsevier.com/>

Printed and bound in the USA

ISBN: 978-0-12-801213-0



Algebraic and Discrete Mathematical Methods for Modern Biology

Contributors

Numbers in parentheses indicate the pages on which the authors' contributions begin.

- Réka Albert** (65), Pennsylvania State University, University Park, PA, USA
- Todd J. Barkman** (261), Department of Biological Sciences, Western Michigan University, Kalamazoo, MI, USA
- Mary Ann Blätke** (141), Otto-von-Guericke University, Magdeburg, Germany
- Hannah Callender** (193,217), University of Portland, Portland, OR, USA
- Margaret (Midge) Cozzens** (29), Rutgers University, Piscataway, NJ, USA
- Kristina Crona** (51), Department of Mathematics and Statistics, American University, 4400 Massachusetts Ave NW, Washington, DC 20016
- Robin Davies** (93,321), Department of Biology, Sweet Briar College, Sweet Briar, VA, USA
- Monika Heiner** (141), Brandenburg University of Technology, Cottbus, Germany
- Terrell L. Hodge** (261), Department of Mathematics, Western Michigan University, Kalamazoo, MI, USA
- Qijun He** (93,321), Department of Mathematical Sciences, Clemson University, Clemson, SC, USA
- John R. Jungeck** (1), Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, USA
- Winfried Just** (193,217), Department of Mathematics, Ohio University, Athens, OH, USA
- Bessie Kirkwood** (237), Sweet Briar College, Sweet Briar, VA, USA
- M. Drew LaMar** (193,217), The College of William and Mary, Williamsburg, VA, USA
- Matthew Macauley** (93,321), Department of Mathematical Sciences, Clemson University, Clemson, SC, USA
- Wolfgang Marwan** (141), Otto-von-Guericke University, Magdeburg, Germany
- David Murrugarra** (121), Department of Mathematics, University of Kentucky, Lexington, KY, USA
- Christian M. Reidys** (347), Department of Mathematics and Computer Science, University of Southern Denmark, Odense M, Denmark
- Raina Robeva** (65), Sweet Briar College, Sweet Briar, VA, USA
- Janet Steven** (237), Christopher Newport University, Newport News, VA, USA
- Blair R. Szymczyna** (261), Department of Chemistry, Western Michigan University, Kalamazoo, MI, USA
- Natalia Toporikova** (193), Washington and Lee University, Lexington, VA, USA
- Alan Veliz-Cuba** (121), Department of Mathematics, University of Houston, and Department of BioSciences, Rice University, Houston, TX, USA
- Rama Viswanathan** (1), Beloit College, Beloit, WI, USA
- Grady Weyenberg** (293), Department of Statistics, University of Kentucky, Lexington, KY
- Emilie Wiesner** (51), Department of Mathematics, Ithaca College, 953 Danby Rd, Ithaca, NY 14850
- Ruriko Yoshida** (293), Department of Statistics, University of Kentucky, Lexington, KY

Preface

In the last 15 years, the field of modern biology has been transformed by the use of new mathematical methods, complementing and driving biological discoveries. Problems from gene regulatory networks and genomics, RNA folding, infectious disease and drug resistance modeling, phylogenetics, and ecological networks and food webs have increasingly benefited from the application of discrete mathematics and computational algebra. Modern algebra approaches have proved to be a natural fit for many problems where the use of traditional dynamical models built with differential equations is not appropriate or optimal.

While the use of modern algebra methods is now in the mainstream of mathematical biology research, this trend has been slow to influence the undergraduate mathematics and biology curricula, where difference and differential equation models still dominate. Several high-profile reports have been released in the past 5 years, including Refs. [1–3], calling urgently for broadening the undergraduate exposure at the interface of mathematics and biology, and including methods from modern discrete mathematics and their biological applications. However, those reports have been slow to elicit the transformative change in the undergraduate curriculum that many of us had hoped for. The anemic response may be attributed to a relative lack of educational undergraduate resources that highlight the critical impact of algebraic and discrete mathematical methods on contemporary biology. It is this niche that our book seeks to fill.

The format of this volume follows that of our earlier book, *Mathematical Concepts and Methods in Modern Biology: Using Modern Discrete Methods*, Robeva and Hodge (Editors), published in 2013 by Academic Press. At the time of its planning, we considered the modular format of that text (with chapters largely independent from one another) experimental, but we felt reassured when the book was selected as 1 of 12 contenders for the 2013 Society of Biology Awards in its category. We have adopted the same format here, as we believe that it provides readers and instructors with the independence to choose biological topics and mathematical methods that are of greatest interest to them.

Due to the modular format, the order of the chapters in the volume does not necessarily imply an increased level of difficulty or the need for more prerequisites for the later chapters. When chapters are connected by a common biological thread, they are grouped together, but they can still be used independently. Each chapter begins with a question or a number of related questions from modern biology, followed by the description of certain mathematical methods and theory appropriate in the search of answers. As in our earlier book, chapters can be viewed as fast-track pathways through the problem, which start by presenting the biological foundation, proceed by covering the relevant mathematical theory and presenting numerous examples, and end by highlighting connections with ongoing research and current publications. The level of presentation varies among chapters—some may be appropriate for introductory courses, while others may require more mathematical or biological background. Exercises are embedded within the text of each chapter, and their execution requires only material discussed up to that point. In addition, many chapters feature challenging open-ended questions (designated as projects) that provide starting points for explorations appropriate for undergraduate research, and supply references to relevant publications from the recent literature. In their most general form, some of the projects feature truly open questions in mathematical biology.

The book's companion website (<http://booksite.elsevier.com/9780128012130>) contains solutions to the exercises, as well all figures and relevant data files for the examples and exercises in the chapters. In addition, the site hosts software code, project guidelines, online supplements, appendices, and tutorials for selected chapters. The specialized software utilized throughout the book highlights the critical importance of computing

applications for visualization, simulation, and analysis in modern biology. We have been careful to feature software that is in the mainstream of current mathematical biology research, while also being mindful of giving preference to freely available software.

We hope that the book will be a valuable resource to mathematics and biology programs, as it describes methods from discrete mathematics and modern algebra that can be presented, for the most part, at a level completely accessible to undergraduates. Yet the book provides extensions and connections with research that would also be helpful to graduate students and researchers in the field. Some of the material would be appropriate for mathematics courses such as finite mathematics, discrete structures, linear algebra, abstract/modern algebra, graph theory, probability, bioinformatics, statistics, biostatistics, and modeling, as well as for biology courses such as genetics, cell and molecular biology, biochemistry, ecology, and evolution.

The selection of topics for the volume and the choice of contributors grew out of the workshop “Teaching Discrete and Algebraic Mathematical Biology to Undergraduates” organized by Raina Robeva, Matthew Macauley, and Terrell Hodge and funded and hosted by the Mathematical Biosciences Institute (MBI) on July 29–August 2, 2013 at The Ohio State University. The editor and contributors of this volume greatly appreciate the encouragement and assistance received from the MBI’s leadership and staff. Without their support, this volume would not have been possible. We also acknowledge with gratitude the support of the National Institute for Mathematical and Biological Synthesis (NIMBioS) in providing an opportunity to further test selected materials as part of the tutorial “Algebraic and Discrete Biological Models for Undergraduate Courses” offered on June 18–20, 2014 at NIMBioS.

I would like to express my personal thanks to all contributors who embraced the project early on and committed time and energy into producing the chapter modules for this unconventional textbook. Your enthusiasm for the project was remarkable, and you have my deep gratitude for the dedication and focus with which you carried it out. My special thanks also go to Daniel Hrozencik and Timothy Comar for providing feedback on a few of the chapter drafts. I am indebted to the editorial and production teams at Elsevier and particularly to the book’s editors, Paula Callaghan and Katey Birtcher, our editorial project managers, Sarah Watson and Amy Clark, and our production manager, Vijayaraj Purushothaman. It has been a pleasure and a privilege to work with all of you. Finally, I would like to thank my husband, Boris Kovatchev, for his patience and support throughout.

Raina S. Robeva
October 20, 2014

REFERENCES

- [1] Committee on a New Biology for the 21st Century: Ensuring the United States Leads the Coming Biology Revolution, Board on Life Sciences, Division on Earth and Life Studies, National Research Council. A new biology for the 21st century. Washington, DC: The National Academies Press; 2009.
- [2] Brewer CA, Anderson CW (eds). Vision and change in undergraduate biology education: a call to action. Final report of a National Conference organized by the American Association for the Advancement of Science with support from the National Science Foundation, July 15–17, 2009, Washington, DC. The American Association for the Advancement of Science; 2011. <http://visionandchange.org/files/2013/11/aaas-VISchange-web1113.pdf> (accessed March 1, 2015).
- [3] Committee on the Mathematical Sciences in 2025, Board on Mathematical Sciences and Their Applications, Division on Engineering and Physical Sciences, National Research Council. The mathematical sciences in 2025. Washington, DC: The National Academies Press; 2013.

Companion website: <http://booksite.elsevier.com/9780128012130/>

Supplementary Resources for Instructors

The website features the following additional resources available for download:

- All figures from the book
- Solutions to all exercises
- Computer code, data files, and links to software and materials carefully chosen to supplement the content of the textbook
- Appendices, tutorials, and additional projects for selected chapters

Contents

Contributors	ix	2.6 Conclusions	48
Preface	xi	References	49
1. Graph Theory for Systems Biology: Interval Graphs, Motifs, and Pattern Recognition		3. Adaptation and Fitness Graphs	
<i>John R. Jungck and Rama Viswanathan</i>		<i>Kristina Crona and Emilie Wiesner</i>	
1.1 Introduction	1	3.1 Introduction	51
1.2 Revisualizing, Recognizing, and Reasoning About Relationships	3	3.2 Fitness Landscapes and Fitness Graphs	52
1.2.1 Basic Concepts from Graph Theory	3	3.2.1 Basic Terminology and Notation	52
1.2.2 Interval Graphs in Biology	6	3.2.2 Fitness, Fitness Landscapes, and Fitness Graphs	53
1.3 Example I—Differentiation: Gene Expression	15	3.2.3 Epistasis	55
1.4 Example II—Disease Etiology	20	3.3 Fitness Graphs and Recombination	58
1.5 Conclusion	25	3.4 Fitness Graphs and Drug Cycling	60
Acknowledgments	26	References	63
References	26	 	
 		4. Signaling Networks: Asynchronous Boolean Models	
2. Food Webs and Graphs		<i>Réka Albert and Raina Robeva</i>	
<i>Margaret (Midge) Cozzens</i>		4.1 Introduction to Signaling Networks	65
2.1 Introduction	29	4.2 A Brief Summary of Graph-Theoretic Analysis of Signaling Networks	66
2.2 Modeling Predator-Prey Relationships with Food Webs	29	4.3 Dynamic Modeling of Signaling Networks	69
2.3 Trophic Levels and Trophic Status	30	4.4 The Representation of Node Regulation in Boolean Models	70
2.3.1 Background and Definitions	31	4.5 The Dynamics of Boolean Models	72
2.3.2 Adding Complexity: Weighted Food Webs and Flow-Based Trophic Levels	35	4.6 Attractor Analysis for Stochastic Asynchronous Update	75
2.3.3 Flow-Based Trophic Level	36	4.7 Boolean Models Capture Characteristic Dynamic Behavior	77
2.4 Competition Graphs and Habitat Dimension	37	4.8 How to Deal with Incomplete Information when Constructing the Model	80
2.4.1 Competition Graphs (also Called Niche Overlap Graphs and Predator Graphs)	37	4.8.1 Dealing with Gaps in Network Construction	81
2.4.2 Interval Graphs and Boxicity	37	4.8.2 Dealing with Gaps in Transition Functions	82
2.4.3 Habitat Dimension	40	4.8.3 Dealing with Gaps in Initial Condition	84
2.5 Connectance, Competition Number, and Projection Graphs	41	4.8.4 Dealing with Gaps in Timing Information	85
2.5.1 Connectance	42	4.9 Generate Novel Predictions with the Model	85
2.5.2 Competition Number	43		
2.5.3 Projection Graphs	44		

4.10 Boolean Rule-Based Structural Analysis of Cellular Networks	86	6.8 Conclusion	137
4.11 Conclusions	90	References	138
References	90		
5. Dynamics of Complex Boolean Networks: Canalization, Stability, and Criticality		7. BioModel Engineering with Petri Nets	
<i>Qijun He, Matthew Macauley and Robin Davies</i>		<i>Mary Ann Blätke, Monika Heiner and Wolfgang Marwan</i>	
5.1 Introduction	93	7.1 Introduction	141
5.2 Boolean Network Models	95	7.2 Running Case Study	144
5.2.1 Gene Regulatory Networks	95	7.3 Petri Nets (\mathcal{PN})	146
5.2.2 Network Topology	96	7.3.1 Modeling	146
5.2.3 Network Topology and Random Networks	99	7.3.2 Analysis	153
5.2.4 Boolean Functions	100	7.3.3 Further Reading	159
5.2.5 Boolean Networks	102	7.3.4 Exercises	160
5.3 Canalization	104	7.4 Stochastic Petri Nets (\mathcal{SPN})	162
5.3.1 Canalizing Boolean Functions	104	7.4.1 Modeling	162
5.3.2 Nested Canalizing Functions	105	7.4.2 Analysis	165
5.3.3 Canalizing Depth	109	7.4.3 Further Reading	169
5.3.4 Dominant Variables of NCFs	110	7.4.4 Exercises	170
5.4 Dynamics Over Complex Networks	112	7.5 Continuous Petri Nets (\mathcal{CPN})	172
5.4.1 Boolean Calculus	113	7.5.1 Modeling	172
5.4.2 Derrida Plots and the Three Dynamical Regimes	115	7.5.2 Analysis	173
5.4.3 Ensembles of RBNs	116	7.5.3 Further Reading	175
Acknowledgments	118	7.5.4 Exercises	176
References	118	7.6 Hybrid Petri Nets (\mathcal{HPN})	177
		7.6.1 Modeling	178
6. Steady State Analysis of Boolean Models: A Dimension Reduction Approach		7.6.2 Analysis	180
<i>Alan Veliz-Cuba and David Murrugarra</i>		7.6.3 Further Reading	181
6.1 Introduction	121	7.6.4 Exercises	182
6.2 An Example: Toy Model of the <i>lac</i> Operon	122	7.7 Colored Petri Nets	183
6.3 General Reduction	125	7.7.1 Further Reading	186
6.3.1 Definition	125	7.7.2 Exercises	186
6.3.2 Examples	125	7.8 Conclusions	187
6.4 Implementing the Reduction Algorithm Using Boolean Algebra	128	Acknowledgments	189
6.5 Implementing the Reduction Algorithm Using Polynomial Algebra	129	7.9 Supplementary Materials	189
6.5.1 Background	129	References	189
6.5.2 Using Polynomial Algebra Software to Reduce Boolean Networks	130		
6.6 Applications	131	8. Transmission of Infectious Diseases: Data, Models, and Simulations	
6.6.1 The <i>lac</i> Operon	131	<i>Winfried Just, Hannah Callender, M. Drew LaMar and Natalia Toporikova</i>	
6.6.2 Th-Cell Differentiation	133	8.1 Introduction: Why Do We Want to Model Infectious Diseases?	193
6.7 AND Boolean Models	134	8.2 Mathematical Models of Disease Transmission	198
6.7.1 Background	135	8.2.1 Transmission Probabilities	199
		8.2.2 The Time Line of Within-Host Dynamics	201
		8.2.3 Movement Between Compartments	203
		8.2.4 Basic Model Types: <i>SEIR</i> , <i>SIR</i> , <i>SI</i> , and <i>SIS</i>	206

8.2.5 How to Model Time and Run Simulations	208	10.3.1 Nonindependence of Multiple Traits	250
8.3 How Does the Computer Run Simulations?	210	10.3.2 The Genetic Variance-Covariance Matrix	252
8.3.1 Meet the Simulator	210	10.3.3 Simultaneous Selection on Multiple Traits	253
8.3.2 How to Load the Die	212	10.3.4 Predicting the Outcome of Selection on Covarying Traits	255
References	214	10.3.5 Evolution of the G Matrix Itself	257
9. Disease Transmission Dynamics on Networks: Network Structure Versus Disease Dynamics		References	258
<i>Winfried Just, Hannah Callender and M. Drew Lamar</i>		11. Metabolic Analysis: Algebraic and Geometric Methods	
9.1 Introduction	217	<i>Terrell L. Hodge, Blair R. Szymczyna and Todd J. Barkman</i>	
9.2 Models Based on the Uniform Mixing Assumption	218	11.1 Introduction	261
9.2.1 Compartment-Based Models	218	11.2 Encoding the Reactions: Linear Algebraic Modeling	262
9.2.2 The Basic Reproductive Ratio R_0	220	11.3 Adding Reaction Kinetics: Algebraic Formulation of Mass-Action Kinetics	271
9.3 Network-Based Models	224	11.4 Directions for Further Reading and Research: Metabolic Pathways	273
9.3.1 Networks and Graphs	225	11.5 NMR and Linear Algebraic Methods	274
9.3.2 Disease Transmission on Networks	229	11.6 NMR Spectroscopy and Applications to the Study of Metabolism	274
9.3.3 Examples of Contact Networks	230	11.6.1 Principles of NMR Spectroscopy	275
9.3.4 Additional Graph-Theoretic Notions	231	11.6.2 The NMR Spectrum	277
9.3.5 Erdős-Rényi Random Graphs	233	11.6.3 NMR Investigations of Metabolism	281
9.4 Suggestions for Further Study	234	11.7 NMR for Metabolic Analysis and Mathematical Methods: Directions of Further Research	289
Acknowledgments	235	11.8 Supplementary Materials	290
References	235	References	290
10. Predicting Correlated Responses in Quantitative Traits Under Selection: A Linear Algebra Approach		12. Reconstructing the Phylogeny: Computational Methods	
<i>Janet Steven and Bessie Kirkwood</i>		<i>Grady Weyenberg and Ruriko Yoshida</i>	
10.1 Introduction	237	12.1 Introduction	293
10.2 Quantifying Selection on Quantitative Traits	238	12.1.1 Sequences and Alignments	297
10.2.1 Describing Traits Mathematically	238	12.2 Quantifying Evolutionary Change	299
10.2.2 Quantifying Reproduction and Survival	241	12.2.1 Probabilistic Models of Molecular Evolution	299
10.2.3 Describing the Relationship Between Fitness and a Trait	242	12.2.2 Common Model Extensions	306
10.2.4 Determining the Genetic Component of Quantitative Traits	245	12.3 Reconstructing the Tree	306
10.2.5 Estimating Heritability in a Trait	246	12.3.1 Distance-Based Methods	306
10.2.6 The Breeder's Equation	247	12.3.2 Maximum Parsimony	309
10.2.7 The Price Equation	249	12.3.3 Methods Based on Probability Models	310
10.3 Covariance Among Traits Under Selection	249		

12.4 Model Selection	312		
12.5 Statistical Methods to Test Congruency Between Trees	313		
References	316		
13. RNA Secondary Structures: Combinatorial Models and Folding Algorithms			
<i>Qijun He, Matthew Macauley and Robin Davies</i>			
13.1 Introduction	321		
13.2 Combinatorial Models of Noncrossing RNA Structures	324		
13.2.1 Partial Matchings and Physical Constraints	324		
13.2.2 Loop Decomposition	327		
13.3 Energy-Based Folding Algorithms for Secondary Structure Prediction	329		
13.3.1 Maximizing Bond Strengths via Dynamic Programming	329		
13.3.2 Minimum Free Energy Folding	333		
13.4 Stochastic Folding Algorithms via Language Theory	335		
13.4.1 Languages and Grammars	335		
		13.4.2 The Knudsen-Hein Grammar for RNA Secondary Structures	337
		13.4.3 Secondary Structure Prediction Using SCFGs	340
		13.4.4 Summary	341
		13.5 Pseudoknots	341
		Acknowledgments	344
		References	344
14. RNA Secondary Structures: An Approach Through Pseudoknots and Fatgraphs			
<i>Christian M. Reidys</i>			
14.1 Introduction	347		
14.2 Fatgraphs and Shapes	349		
14.3 Genus Recursion	354		
14.4 Shapes of Fixed Topological Genus	357		
Acknowledgments	361		
References	361		
		Index	363

Graph Theory for Systems Biology: Interval Graphs, Motifs, and Pattern Recognition

John R. Jungck¹ and Rama Viswanathan²

¹Center for Bioinformatics and Computational Biology, University of Delaware, Newark, DE, USA, ²Beloit College, Beloit, WI, USA

1.1 INTRODUCTION

Systems thinking is perceived as an important contemporary challenge of education [1]. However, *systems biology* is an old and inclusive term that connotes many different subareas of biology. Historically two important threads were synchronic: (a) the systems ecology of the Odum school [2–4], which was developed in the context of engineering principles applied to ecosystems [5, 6], and (b) systems physiology that used mechanical principles [7] to understand organs as mechanical devices integrated into the circulatory system, digestive system, anatomical system, immune system, nervous system, etc. For example, the heart could be thought of as a pump, the kidney as a filter, the lung as a bellows, the brain as a wiring circuit (or later as a computer), elbow joints as hinges, and so on. It should be noted that both areas extensively employed *ordinary* and *partial differential equations* (ODEs and PDEs). Indeed, some systems physiologists argued that all mathematical biology should be based on the application of PDEs. On the other hand, evolutionary biologists argued that these diachronic systems approaches too often answered only “how” questions that investigated optimal design principles and did not address “why” questions focusing on the constraints of historical contingency.

Not surprisingly, one of the leading journals in the field—*Frontiers in Systems Biology*—announces in its mission statement, [8] “Contrary to the reductionist paradigm commonly used in Molecular Biology, in Systems Biology the understanding of the behavior and evolution of complex biological systems need not necessarily be based on a detailed molecular description of the interactions between the system’s constituent parts.” Therefore, in this chapter we emphasize two major macroscopic and global aspects of contemporary systems biology: (i) the graph-theoretic relationships between components in networks and (ii) the relationship of these patterns to the historical contingencies of evolutionary constraints. Numerous articles and several books [9, 10] exist on graph theory and its application to systems biology, so the reader may ask what are we doing in this chapter that is different. Our main purpose is to help biologists, mathematicians, students, and researchers recognize which graph-theoretic tools are appropriate for different kinds of questions, including quantitative analyses of interactions for mining large data sets, visualizing complex relationships, modeling the consequences of perturbation of networks, and testing hypotheses.

Every network construct in systems biology is a hypothesis. For example, Rios and Vendruscolo [11] describe the network hypothesis as the assumption “according to which it is possible to describe a cell through the set of interconnections between its component molecules.” They then conclude, “it becomes convenient to focus on

these interactions rather than on the molecules themselves to describe the functioning of the cell.” In this chapter, we go a step further. We believe that a mathematical biology perspective also studies such questions as: Which molecules are involved? What do they do functionally? What is their three-dimensional structure? Where are they located in a cell? We stress that every network and pathway that we discuss is a useful construct from a biological perspective. They do not exist *per se* inside of cells. Imagine a series of biological macromolecules (proteins, nucleic acids, polysaccharides) that are crowded and colliding with one another in a suspension. The networks and pathways for the interactions between these molecules constructed by biologists may represent preferred associations defined by tighter bindings of specific macromolecules or the product of a reaction catalyzed by one macromolecule (an enzyme) as the starting material (substrate) of another enzyme. Thus, biologists have already drawn mathematical diagrams and graphs in the sense that they have abstracted, generalized, and symbolized a set or relationships.

Too often biologists produce networks as visualizations without further analysis. In this chapter, using Excel and Java-based software that we have developed, we show readers how to make mathematical measurements (average degree, diameter, clustering coefficient, etc.) and discern holistic properties (small world versus scale-free, see Hayes [12] for a complete overview) of the networks being studied and visualized, and obtain insights that are relevant and meaningful in the context of systems biology. We show how the network hypothesis can be investigated by complementary and supplementary mathematical and biological perspectives to yield key insights and help direct and inform additional research.

Palsson [10] suggests that twenty-first century biology will focus less on the reductionist study of components and more on the integration of systems analysis. He identifies four principles in his “systems biology paradigm”: “First, the list of biological components that participated in the process of interest is enumerated. Second, the interactions between these components are studied and the ‘wiring diagrams’ of genetic circuits are constructed Third, reconstructed network[s] are described mathematically and their properties analyzed.... Fourth, the models are used to analyze, interpret, and predict biological experimental outcomes.” Here, we assume that the first two steps exist in databases or published articles; this allows us to focus on the mathematics of the third step as a way that allows biologists to better direct their work on the fourth step. Thus, the goals for this chapter are as follows.

- Learn how graph theory can be used to help obtain meaningful insights into complex biological data sets.
- Analyze complex biological networks of diverse types (restriction maps, food webs, gene expression, disease etiology) to detect patterns of relationships.
- Visualize ordering of modules/motifs within complex biological networks by first testing the applicability of simple linear approaches (interval graphs).
- Demonstrate that even when strict mathematical assumptions do not apply fully to a given biological data set, there is still benefit in applying an analytical approach because of the power of the human mind to discern prominent patterns in data rearranged through the application of mathematical transformations.
- Show that the visualizations help biologists obtain insights into their data, examine the significance of outliers, mine databases for additional information about observed associations, and plan further experiments.

To accomplish this, we first emphasize *how* graph theory is a natural fit for biological investigations of relationships, patterns, and complexity. Second, graph theory lends itself easily to questions about *what* biologists should be looking for among representations of relationships. We introduce concepts of hubs, maximal cliques, motifs, clusters, interval graphs, complementary graphs, ordering, transitivity, Hamiltonian circuits, and consecutive ones in adjacency matrices. Finally, graph theory helps us interrogate *why* these relationships are occurring. Basically, we examine the triptych of form, function, and phylogeny to differentiate between evolutionary and engineering constraints.

The chapter is structured as follows. We begin by introducing some background concepts from graph theory that will be utilized later in the chapter. We then introduce interval graphs through two biological examples related to chromosome sequencing and food webs. The rest of the chapter is devoted to two extended examples of biological questions related to recently published studies on gene expression and disease etiology. The analyses