For decades, molecular geneticists have been intensively studying the individual genes of various organisms and how these genes influence their phenotypic behavior. Unfortunately, it is usually very difficult, if not impossible, to isolate specific genetic signals for any arbitrary behavioral aspect or trait. The problem is analogous to that of finding a grass skirt in a very large haystack. Even if one locates a plausible-looking bit of grass, until its connections are laboriously traced out, one cannot know if it is part of the skirt or, as is much more likely, just an unrelated piece of straw. As an example, there are over 100 genes that are known to affect flowering time in the model plant Arabidopsis thaliana. Together, the interactions of these genes comprise a complex signal processing network that integrates multiple internal and external cues to make one of the most critical decisions in a plant’s life cycle—when to reproduce. Yet, all together, these genes comprise only 0.4% of the species’ complete gene network.

Recent advances in molecular genetic technologies are beginning to shed light on the complex interplay between genes that elicit phenotypic behavior characteristic of any given organism. Even so, unraveling the specific details about how these genetic pathways interact to regulate development, shape life histories, and respond to environmental cues remains a very daunting task.

A wide variety of models depicting gene-gene interactions, which are commonly referred to as gene regulatory networks (GRNs), have been proposed in recent literature. While a GRN must be able to mimic experimentally observed behavior, reproducing complex behaviors accurately may entail computationally prohibitive costs. Under these circumstances, model simplicity is an important trade-off for functional fidelity. Consequently, modeling approaches taken are wide and disparate. Machine learning based GRN models are specifically meant for simplicity and/or algorithmic tractability. They rely heavily on computational learning theory, and usually are used to simulate qualitatively, phenotypic behavior of GRNs. We refer to these as high level models. At the other end are more detailed models that take into account the underlying biochemical processes. These models are capable of reproducing realistic gene expressions with great fidelity.

This book is a collection of articles on the various computational tools that are available to decode, model, and analyze GRNs. It is conveniently organized into separate sections, beginning with an introductory section. Each section contains a handful of chapters written by researchers in the field that focus on a specific computational approach.

SECTION 1: INTRODUCTION

The first section contains two introductory chapters on GRNs. Chapter 1 ("What are Gene Regulatory Networks") provides a conceptual framework for GRNs. It shows how the complex nonlinear biochemical processes can be linearized and portrayed as simple graphical models. The nodes of such a network
are either individual genes or groups of functionally related ones. The network can have both directed as well as undirected edges. The chapter also highlights the differences between such networks and two other similar structures, transcriptional regulatory networks and co-expression networks.

The next chapter in this section (Chapter 2) is entitled “Introduction to Gene Regulatory Networks” and has a slightly different focus. While introducing the GRN as a graph, it also details further biological insights into the various underlying biochemical processes within GRNs. The chapter also surveys recent advances in array-based technologies that are available to study such processes. Only minimum background in advanced mathematics is assumed here, making the chapter very useful to biologists interested in this subject.

SECTION 2: NETWORK INFERENCE

While the previous section introduces GRNs as graphical structures, the chapters in this section focus on systems identification; they shed light on how GRNs can be reverse engineered from experimental data. While simply arranging genes into various functional units may be accomplished easily through simple statistical means, depicting causality between these units is more challenging.

To varying degrees, all four chapters in this section deal with Bayesian network approach. Bayesian networks, a marriage between graph theory and probability theory, are a high level abstraction of GRNs. An introductory, yet thorough mathematical description of Bayesian networks in provided in Chapter 3 (“Bayesian Networks for Modeling and Inferring Gene Regulatory Networks”). This chapter considers both discrete probabilities as well as continuous probability distributions. Dynamic Bayesian networks are taken up briefly to show how cyclic graphs can be modeled. The latter half of the chapter casts the tasks of discovering the structure of the Bayesian network and estimating the parameters of its probability distribution(s) as two aspects of learning. Lastly, it addresses issues relating to assessing the performance of inferred networks.

Chapter 4 (“Inferring Gene Regulatory Networks from Genetical Genomics Data”) addresses techniques that can be applied to establish causality between the various nodes in a GRN. These techniques are based on the joint analysis of DNA marker and expression as well as DNA sequence information. In addition to Bayesian networks, another modeling approach, statistical equation modeling, is discussed.

Boolean networks are a GRN modeling approach where each gene is associated with a simple logical function. Chapter 5 (“Inferring Genetic Regulatory Interactions with Bayesian Logic-Based Model”) combines this modeling approach with Bayesian networks. Using simple Boolean semantics to depict underlying interactions among gene products allows for the analysis of larger networks, while the Bayesian framework helps penalize overly complex models. As examples, results of applying this method to data from *S. cerevisiae* and to *Plasmodium falciparum* are provided.

Depicting the dynamic interactions of genes within a network as a set of ordinary differential equations helps preserve biochemical fidelity. Unfortunately, this detailed approach is too complex to be extended beyond a few genes. Chapter 6 (“A Bayes Regularized Ordinary Differential Equation Model for the Inference of Gene Regulatory Networks”), makes use of the stochastic nature of GRNs to integrate the differential equation models within a probabilistic network. Bayesian learning is applied to determine the parameters of the differential equation model. The effectiveness of this overall approach is demonstrated by applying it to the yeast cell.
SECTION 3: MODELING METHODS

As noise and delays are intrinsic to biochemical processes, they must be accounted for when dealing with the most detailed differential equation models of GRNs. This issue is addressed in Chapter 7 (“Computational Approaches for Modeling Intrinsic Noise and Delays in Genetic Regulatory Networks”) and in the following one, Chapter 8 (“Modeling Gene Regulatory Networks with Delayed Stochastic Dynamics”).

A basic Monte Carlo simulation technique to simulate noisy biochemical reactions, as well as a generalization to include delays, are described in both chapters, although to different ends. Chapter 7 follows this with a study into ‘coarse grain’ approaches, which reduce computational costs when dealing with larger biochemical systems. The methodology is demonstrated with a few case studies. In contrast, Chapter 8 discusses simulation studies with single genes as well as simple networks of genes. It concludes with a genetic algorithm1 based simulation to investigate how simple GRNs evolve.

Chapter 9 (“Nonlinear Stochastic Differential Equations Method for Reverse Engineering of Gene Regulatory Networks”) is a study on how structures of GRNs can be obtained from expression data. It uses stochastic differential equation models, where noise is depicted as a Brownian process. The authors show how regulators for genes are selected using heuristics based on statistical and information theoretic principles, and demonstrate this concept with a few case studies.

The last chapter in this section, Chapter 10 (“Modelling Gene Regulatory Networks with Computational Intelligence Techniques”) introduces computational intelligence techniques in GRNs with a focus on genetic algorithms. The authors propose the guided genetic algorithm as an optimization method for causal modeling of GRNs. Case studies involving both simulated data as well as real yeast data are described to show how their approach works.

SECTION 4: STRUCTURE AND PARAMETER LEARNING

This section contains a set of chapters that are most directly related to algorithmic approaches for learning structures and parameters of GRNs. It begins with Chapter 11 (“A Synthesis Method of Gene Regulatory Networks based on Gene Expression by Networking Learning”), which addresses how GRNs can be modeled to produce oscillatory behavior. This is an important problem as oscillations such as circadian rhythm are routinely observed in gene expression patterns. The chapter proposes a recurrent neural network modeling approach to derive networks of low complexity that can produce desired oscillatory sequences.

Chapter 12 (“Structural Learning of Genetic Regulatory Networks Based on Prior Biological Knowledge and Microarray Gene Expression Measurements”) is a survey of current methods on Bayesian network models of GRNs. It focuses on structure priors derived from experimental results such as protein-protein interactions, transcription factor binding locations, evolutionary relationships as well as existing literature.

The following chapter, Chapter 13 (“Problems for Structure Learning: Aggregation and Computational Complexity”) offers a critique on current approaches to inferring model structure using standard machine learning techniques. The authors identify three specific factors in support of their argument: that the methods reported in the literature make use of synthetic as opposed to real data, that they claim success when the actual gene network structure is not known, and that only isolated successes are published.
SECTION 5: ANALYSIS AND COMPLEXITY

Large, heterogeneous datasets arising from a variety of experiments, intricacies involved at various stages of the modeling process, as well as the intrinsically complex nature of the genetic interactions within the organisms themselves—shaped through millennia of evolution—all contribute to models that are often difficult to analyze and comprehend. A collection of articles that address this issue is included in this section.

Chapter 14 (“Complexity of the BN and the PBN Models of GRNs and Mappings for Complexity Reduction”) is intended to provide a generic framework for model complexity reduction in Boolean and probabilistic Boolean networks. Statistical and information theoretic views of complexity are described. Approaches to map larger GRNs into smaller, more tractable ones, while preserving the overall dynamical behavior, are considered within this scheme.

Chapter 15 (“Abstraction Methods for Analysis of Gene Regulatory Networks”) also addresses the issue of reducing the complexity in GRNs. It details steps that can be taken to merge similar reactions and eliminate insignificant ones from large-scale models of biochemical reactions. Using these simplifications, models based on chemical kinetics can be abstracted into higher level ones called finite state systems.

Chapter 16 (“Improved Model Checking Techniques for State Space Analysis of Gene Regulatory Networks”) describes a software tool that applies model checking—a technique used to analyze computer programs—to discrete GRN models. Using this technique, steady state characteristics of the models can be examined. Two case studies, the gene network for cell cycle of yeast, as well as that for wing formation in D. melanogaster, illustrate the effectiveness of this technique.

Chapter 17 (“Determining the Properties of Gene Regulatory Networks from Expression Data”) shows how topological properties of GRNs can be applied to the practical analysis of experimental gene expression data. Using examples that apply this approach, the authors argue that there is much more to regulation between genes than just transcription factors.

Chapter 18 (“Generalized Boolean Networks: How Spatial and Temporal Choices Influence Their Dynamics”) relaxes the requirements in random Boolean network models, that genes operate in synchrony and that their connectivity remain fixed. These modifications, it is argued, enable Boolean networks to better capture some characteristics present in gene expression, such as activation sequences in genes and periodic attractors.

SECTION 6: HETEROGENEOUS DATA

Linear programming—a simple technique for the constrained optimization of linear functions—can be used to synthesize GRNs from multiple data sources, as the next two chapters show.

In Chapter 19 (“A Linear Programming Framework for Inferring Gene Regulatory Network by Integrating Heterogeneous Data”), the authors use linear differential equation models of GRNs to which matrix decomposition methods and linear programming are applied. Data from heterogeneous sources, such as documented literature, protein-protein interaction data, and so forth are added as constraints. Using this formulation, the authors attempt to obtain robust GRN models that are consistent with multiple datasets.

Chapter 20 (“Integrating Various Data Sources for Improved Quality in Reverse Engineering of Gene Regulatory Networks”) shows how to reverse engineer large-scale GRNs by integrating various data sources, such as information gleaned by text mining of published research. Using this prior knowledge as
soft evidence, a methodology is proposed to obtain GRN models that can account for large error distributions in microarrays. Simulations with yeast cell data corroborate the effectiveness of this method.

SECTION 7: NETWORK SIMULATION STUDIES

Chapter 21 (“Dynamic Links and Evolutionary History in Simulated Gene Regulatory Networks”) describes computational studies on the evolution of GRNs. Using evolutionary strategies, an algorithmic approach similar to genetic algorithms, the authors are able to simulate the evolution of GRNs that produce stable multicellular growth. They observe that the evolutionary process favors the appearance of negative feedback in the evolved networks. They hypothesize that this is because negative feedback imparts the network with robustness to potentially deleterious mutations.

A new GRN model that incorporates greater biological detail than traditional methods is outlined in the other simulation study in this section (Chapter 22 “A Model for a Heterogeneous Genetic Network”). The authors report computer experiments to generate GRNs using this biologically-motivated approach. They examine the topological features and dynamic behaviors of models obtained in this manner, and provide arguments that such models possess features that correlate well with biological observations.

SECTION 8: OTHER STUDIES

One of the purposes of GRNs is to model cellular dynamics, which are usually characterized by stable attractors. In this context, planned external interventions to redirect these networks from abnormal states (as in with the onset of cancer) to more regular ones is important for many applications, such as prescribing effective drugs. In Chapter 23 (“Planning Interventions for Gene Regulatory Networks as Partially Observable Markov Decision Processes”), this intervention problem is modeled as a Markov decision process. Two well known algorithms borrowed for artificial intelligence are proposed to solve the problem.

There are two modes of propagation of a bacterial virus known as the λ phage: direct replication and integration with the host bacterium. The decision concerning which mode to adopt is controlled by a simple GRN called the λ switch. Chapter 24 “Mathematical Modeling of the λ Switch: A Fuzzy Logic Approach” uses fuzzy logic to model the switch, making it tractable to mathematical treatment. Using this approach, the chapter suggests explanations for certain behavioral aspects of the λ switch, particularly how the bacterium switches to the direct replication mode of transmission when DNA damage occurs in the host.

Chapter 25, “Petri Nets and GRN Models,” introduces Petri nets, a graphical modeling approach for modeling GRNs. An introduction to Petri nets as well as related techniques useful in modeling biochemical processes is provided. The application of this approach for the gene regulation in Duchenne muscular dystrophy (DMD) is taken up. An analysis of the results sheds lights on the advantages and disadvantages of the method.

CONCLUSION

This book provides a bird’s eye view of the vast range of computational methods used to model GRNs. It contains introductory material and surveys, as well as articles describing in-depth research in various
aspects of GRN modeling. The editors expect it to be useful to researchers in a variety of ways. It can provide a comprehensive overview of artificial intelligence approaches for learning and optimization and their use in gene networks to biologists involved in genetic research. It can assist computer science and artificial intelligence theorists in understanding how their methodologies can be applied to GRN modeling. Although not intended to be a textbook, the book can be of immense use as a reference for students and classroom instructors. As the book would bridge the gap between computer science and genomic research communities, it will be very useful to graduate students considering research in this direction. Finally, this book would be useful to industrial researchers involved in gene regulatory modeling.

Sanjoy Das
Doina Caragea
Stephen M. Welch
William H. Hsu

**ADDITIONAL READING**


**ENDNOTE**

1 Genetic algorithms are a class of approaches borrowed from computational intelligence for stochastic optimization. The usage of the word “genetic” does not imply a direct relationship with GRNs, but stems from the fact that these algorithms loosely mimic biological evolution.