

(and, as a consequence, is not 1-factorable). In Chapter 8, the reader encounters a host of interesting examples where the edge set of a graph G is partitioned strategically, including Steiner triple systems, cycle decompositions, Veblen's theorem, and graceful graphs. The chapter closes with a detailed graph-theoretical examination of the puzzle Instant Insanity.

No overview of graph theory would be complete without discussing directed graphs (where edges are each provided a specific direction), the topic of Chapter 9, "Orienting Graphs." Curiously, the authors use the term "oriented graph" rather than the more commonly used "directed graph" for this notion. Here, it appears that they desire to consider all possible orientations of edges, rather than specifying a particular orientation. Nonetheless, their discussions of tournaments and voting procedures illustrate the underlying concepts nicely. Directed graphs (or "digraphs") are considered again in Chapter 12.

Chapter 10, "Drawing Graphs," introduces planar graphs by revisiting the problems of drawing K_5 and $K_{3,3}$ without crossing edges. Euler's famous formula $V - E + F = 2$ for planar graphs is developed along with consequences that show K_5 and $K_{3,3}$ are not planar. The discussion leads to Kuratowski's theorem, crossing numbers, a brief mention of the art gallery problem, and then a nice introduction to graph embeddings on surfaces. This entertaining chapter closes by discussing graph minors, including Wagner's theorem and the graph minor theorem.

The book finishes with two chapters on colorings and graphs. Chapter 11, "Coloring Graphs," focuses on vertex colorings, while Chapter 12, "Synchronizing Graphs," concentrates on edge colorings. The discussion in Chapter 11 opens with the famous Four Color Problem of whether the regions in any map can be colored using four colors in such a way so that no two regions sharing a border are assigned the same color. The authors successfully provide an extensive history of this problem within only twelve pages. The chapter closes by considering vertex coloring in general, including problems in scheduling and traffic control. The final chapter discusses edge coloring, begin-

ning with the chromatic index and Vizing's theorem. Next, several pages are devoted to Ramsey's theorem and Ramsey numbers; the problem of whether K_6 has a monochromatic triangle if the edges are 2-colored is resolved here. The chapter closes by discussing graph synchronization and the road coloring theorem, a result concerning digraphs that has applications in finite automata.

Ultimately, *The Fascinating World of Graph Theory* is an aptly named book, able to present a wide variety of central topics in graph theory, including the history behind them, to a general audience in a lively and entertaining manner. This book would probably not stand alone as a textbook for an introductory course in graph theory; many results in graph theory have proofs that rely on mathematical induction or equivalent proof techniques that are not pursued here. Nonetheless, this book would make for an excellent companion reference in such a course and it provides a superb example of approachable mathematical writing aimed at a broad audience.

MATTHEW G. HUDELSON
Washington State University

ReCombinatorics: The Algorithmics of Ancestral Recombination Graphs and Explicit Phylogenetic Networks. By Dan Gusfield. MIT Press, Cambridge, MA, 2014. \$60.00. xx+580 pp., hardcover. ISBN 978-0-262-02752-6.

Trees—connected graphs that contain no cycles—have been used by biologists for over a century to model evolutionary relationships. In this context, a *phylogenetic tree* models the evolution of a set of taxa that bijectively label the leaves of the tree. The fundamental assumption that underlies the inference and application of trees in evolutionary biology is that the taxa under consideration have evolved from a most recent common ancestor (MRCA) by means of strictly dividing and differentiating. The division results in the bifurcating, and sometimes multifurcating, patterns in the tree shape, and the differentiation results in the varying characteristics (e.g.,

DNA sequences or morphological traits) of the taxa at the leaves of the tree. As the author states [4], the field of phylogenetic trees is extremely well developed—a Google Scholar search of the phrase “phylogenetic tree” returns about 1,160,000 results, and several excellent books have been written on this subject (the author lists four prominent ones in section 2.5).

However, despite the ubiquity of trees, one or more of several processes could take place during an evolutionary history that would violate the aforementioned fundamental assumption. Horizontal gene transfer, which is the movement of segments of genetic material between the genomes of different organisms by means other than ancestor-descendant inheritance, is one such process, and occurs frequently among bacteria, for example. Hybridization, which is the mating of individuals from different species, is another process which has long been acknowledged to occur in certain eukaryotic groups, and its occurrence more broadly than previously thought is increasingly documented in the literature. Meiotic recombination is yet another process which consists of the exchange of pieces of genetic material among chromosome pairs (sister chromatids) during meiosis and results in mosaic sequences in each generation. It is this last process and the graphical model that incorporates it that are the focus of *ReCombinatorics* by Dan Gusfield [4].

In the presence of any of the three processes mentioned above, utilizing the richer model of phylogenetic networks provides a more powerful framework for evolutionary thinking and analysis [1]. A phylogenetic network that is used to model evolutionary histories of sequences in the presence of recombination is often referred to as an *ancestral recombination graph*, or ARG. The underlying topology of an ARG as defined in [4] is a rooted, directed, acyclic graph (rDAG) $g = (V, E)$, where $E \subseteq V \times V$ and V can be partitioned into four sets (we use $\delta^-(u)$ and $\delta^+(u)$ to denote the in- and out-degree of node u , respectively):

- $\{r\}$, where $\delta^-(r) = 0$ and $\delta^+(r) = 2$: the (unique) root of the rDAG;
- $L = \{u : \delta^-(u) = 1 \wedge \delta^+(u) = 0\}$: the set of the rDAG’s leaves;

- $C = \{u : \delta^-(u) = 1 \wedge \delta^+(u) = 2\}$: the set of the coalescence nodes; and,
- $R = \{u : \delta^-(u) = 2 \wedge \delta^+(u) = 1\}$: the set of recombination nodes.

Further, for every node $u \in R$, one of its two incoming edges is labeled “P” (for *prefix*) and the other is labeled “S” (for *suffix*). To complete the description of an ARG (as applicable to [4]), the rDAG is coupled with two functions that are parameterized by a positive integer m (we drop the superscript later since this value is always the length of the sequences and is clear from the context):

- $\mu^m : \{1, 2, \dots, m\} \rightarrow (E \cup \{\emptyset\} \setminus (V \times R))$ such that $\mu^m(i) = \emptyset$ denotes that no edge is labeled with i .
- $\rho^m : R \rightarrow \{1, 2, \dots, m - 1\}$.

Let s_u denote the binary sequence associated with node u , and let s_u^i denote the state of site i in that sequence. *Flipping* s_u^i means replacing it by its complement, that is, setting $s_u^i \leftarrow (1 \oplus s_u^i)$. An ARG $\mathcal{N} = (g, m, \mu, \rho)$ with binary sequence s_r associated with its root can be viewed as a *deterministic* generative process of sequences of length m at the leaves as follows:

- For each node $u \in (V - \{r\})$ in a top-down fashion,
 1. if $u \in (L \cup C)$ is the head of edge $e = (v, u)$, then s_u is assigned s_v and then s_u^i is flipped for every $i \in \mu^{-1}(e)$ where $i \neq \emptyset$;
 2. if $u \in R$ and v is the parent node such that the edge (v, u) is labeled “P” and w is the parent node such that the edge (w, u) is labeled “S”, then s_u is formed by concatenating the first $\rho(u)$ sites from s_v with the last $m - \rho(u)$ sites from s_w .

This formalism reflects the *infinite sites model* for mutations, whereby a site in a molecular sequence mutates at most once during the evolution of the sequences. (Since each site mutates at most once under the infinite sites model, we observe at most two different states at each site across the set of sequences, and hence the focus on binary sequences.) This model is assumed for most of the results in the book. While the outcome of this process is a set of binary sequences of length m , one per node

in the ARG, in biological applications only the sequences at the leaves constitute the observed data.

Minichiello and Durbin [7] wrote that

if the true ARG were known, it would provide the optimal amount of information for mapping—no extra information would be available from the genotypes. Not only would disease-associated regions be identified, but the ARG would give the ages of the causative mutations, would specify the haplotypic background of those mutations, and so forth.

Indeed, the author dedicates Chapters 12 and 13 to applications of ARGs in haplotype inference and association mapping, respectively.

For all practical purposes, though, the premise in the aforementioned quote from Minichiello and Durbin is false; that is, the true ARG is almost never known. Therefore, the major task in this area is to infer the ARG—the rDAG, as well as the two functions μ and ρ —and the binary sequence at the root from the observed data, which consist of a set of binary sequences, all of equal length. However, as Minichiello and Durbin [7] and Siepel [14] pointed out, in practical settings inferring the *true* ARG is impossible, since too much time has elapsed for most mutations and recombinations to be “visible” and, in turn, reconstructible. This is why, for example, Minichiello and Durbin [7] sought to reconstruct “plausible” ARGs, ambiguous as it may sound (the authors discuss plausibility in terms of the accuracy of mapping trait loci using the inferred ARG).

A more systematic treatment of this inference task under a set of assumptions traces its roots, per Gusfield’s account in the preface, to [15] and [10]. The precise mathematical definitions and elegant combinatorial questions in these two papers gave rise to two related lines of investigation: estimating the minimum number of recombination nodes needed in an ARG to generate the observed sequences, and inferring an ARG with the minimum number of recombination nodes along with the functions μ and ρ and the sequence at the root that could generate the observed sequences.

Both problems proved to be computationally very hard and, consequently, a large body of work appeared in the literature to address them. And this is what this book is about; it elegantly describes this large body of work in a unified approach that also connects the work on ARGs to that on the broader category of phylogenetic networks. The first four chapters provide an accessible introduction to topics in mathematics and biology that are central to understanding ARGs. Chapters 5–11 describe the work on the aforementioned lines of investigation. As mentioned above, Chapters 12 and 13 discuss mathematical models and computational solutions to ARG-based applications. In Chapter 14, the author discusses extensions and connections to other types of phylogenetic networks.

This book is true to its title and stated scope in that it focuses exclusively on the combinatorial aspects of ARGs, mostly under the infinite sites model and the parsimony criterion of minimizing the number of recombinations to model the evolution of a set of sequences. In this regard, some of the readers might be frustrated at the almost exclusive focus of the book on ARGs under the parsimony criterion. Recent papers have described statistical approaches to ARG inference [6, 8, 18, 13] that go beyond the parsimony model. However, these formulations of ARGs and their inference have received much less attention from the mathematical and computational communities than the parsimony formulations that this book addresses. It is worth mentioning, though, that while the parsimony criterion of minimizing the number of recombinations could result in “good” ARGs, this is not always the case. In fact, one could concoct an evolutionary scenario where the true number of recombinations is arbitrarily much larger or much smaller than the number of recombinations needed under the parsimony criterion and infinite sites model. In practice, this could happen due to a variety of reasons, including violating the infinite sites model, recombination that “swaps” identical genomic regions, etc. Furthermore, the locations of the mutations and recombination nodes in an ARG inferred based on the parsimony criterion might not necessarily correspond

to the true events. In the case of mutations, the location refers to the edge that function μ assigns to a specific site in the sequences. In the case of recombinations, the location refers to the two parents of the recombination node as well as the value that function ρ assigns to the recombination node.

Another major direction that is worth mentioning here is one that links the work described in the book to a large body of work that has emerged in recent years and is focused on inference of species phylogenies. By the nature of the assumed models and stated definitions, the work described in the book is applicable to population-level data (sequences obtained from individuals within a single species). The “phylogenetic ARG” [14], which combines ARGs with species trees by viewing genomic genealogies as ARGs embedded within the branches of the species tree, provides a model for ARGs in a phylogenetic context. A model that has garnered much more attention and seen much more development is the *multispecies coalescent* [3]. Since recombination decouples the evolutionary histories of adjacent loci in the genomes and creates incongruence among them (a phenomenon known as incomplete lineage sorting), the genealogies of individual loci could be estimated and a phylogeny on the set of species could be inferred under this model. Indeed, a wide array of inference methods were developed very recently based on the multispecies coalescent; see [12] for a recent survey of the developments in this area. Further, this gave rise to work where the species phylogeny takes the shape of a network and the genealogies of the various loci can differ due to recombination [11], which has led to parsimony [16] and statistical formulations [2, 17] of species network inference from loci that are independent due to recombination.

In the preface to the book, the author writes that with the maturation of the field of ancestral recombination graphs and, more broadly, phylogenetic networks, “it was time to revisit the whole area, to write a book for a broad audience of computer scientists, mathematicians, and biologists.” I am a big proponent of introducing mathematics and computer science into the biology curricula. Until that happens, I feel that *ReCombinatorics* will be accessible only to

very few biologists. In terms of the audience that it targets for exposure to phylogenetic networks, Gusfield’s new book is more similar to [5] than [9]; the latter is aimed at the practitioner who wishes to learn what tools for phylogenetic networks are out there, what data they use, what assumptions they make, and how to interpret their inferences. The last point cannot be overemphasized in this regard, since the term “phylogenetic network,” as introduced in the growing literature on the subject, means different things, some of which differ significantly in their modeling objectives. In terms of the first two audiences—computer scientists and mathematicians—this is a very well written and much needed book on various combinatorial aspects of ARGs delivered by a major player in the field. The work described in the book is combinatorial in nature, and the results are derived under potentially restrictive assumptions about the evolutionary history and model of sequence evolution. Nevertheless, the results are very fundamental and would prove essential for future developments on ARG methodologies that go beyond the models, criteria, and assumptions of the book.

Acknowledgment. This work was supported by the National Science Foundation under grant CCF-1302179.

REFERENCES

- [1] E. BAPTESTE, L. VAN IERSEL, A. JANKE, S. KELCHNER, S. KELK, J. O. MCINERNEY, D. A. MORRISON, L. NAKHLEH, M. STEEL, L. STOUGIE, AND J. WHITFIELD, *Networks: Expanding evolutionary thinking*, Trends in Genetics, 29 (2013), pp. 439–441.
- [2] E. W. BLOOMQUIST AND M. A. SUCHARD, *Unifying vertical and nonvertical evolution: A stochastic ARG-based framework*, Systematic Biology, 59 (2010), pp. 27–41.
- [3] J. H. DEGNAN AND N. A. ROSENBERG, *Gene tree discordance, phylogenetic inference and the multispecies coalescent*, Trends in Ecology & Evolution, 24 (2009), pp. 332–340.
- [4] D. GUSFIELD, *ReCombinatorics: The Algorithmics of Ancestral Recombination Graphs and Explicit Phylogenetic Networks*, MIT Press, 2014.

- [5] D. H. HUSON, R. RUPP, AND C. SCORNAVACCA, *Phylogenetic Networks: Concepts, Algorithms and Applications*, Cambridge University Press, New York, 2010.
- [6] F. LARRIBE, S. LESSARD, AND N. J. SCHORK, *Gene mapping via the ancestral recombination graph*, *Theoret. Population Biol.*, 62 (2002), pp. 215–229.
- [7] M. J. MINICHELLO AND R. DURBIN, *Mapping trait loci by use of inferred ancestral recombination graphs.*, *Amer. J. Human Genetics*, 79 (2006), pp. 910–922.
- [8] A. P. MORRIS, J. C. WHITTAKER, AND D. J. BALDING, *Fine-scale mapping of disease loci via shattered coalescent modeling of genealogies*, *Amer. J. Human Genetics*, 70 (2002), pp. 686–707.
- [9] D. A. MORRISON, *Introduction to Phylogenetic Networks*, RJR Productions, 2011.
- [10] S. R. MYERS AND R. C. GRIFFITHS, *Bounds on the minimum number of recombination events in a sample history*, *Genetics*, 163 (2003), pp. 375–394.
- [11] L. NAKHLEH, *Evolutionary phylogenetic networks: Models and issues*, in *The Problem Solving Handbook for Computational Biology and Bioinformatics*, L. Heath and N. Ramakrishnan, eds., Springer, Berlin, 2010, pp. 125–158.
- [12] L. NAKHLEH, *Computational approaches to species phylogeny inference and gene tree reconciliation*, *Trends in Ecology & Evolution*, 28 (2013), pp. 719–728.
- [13] M. D. RASMUSSEN, M. J. HUBISZ, I. GRONAU, AND A. SIEPEL, *Genome-wide inference of ancestral recombination graphs*, *PLoS Genetics*, 10 (2014), p. e1004342.
- [14] A. SIEPEL, *Phylogenomics of primates and their ancestral populations*, *Genome Res.*, 19 (2009), pp. 1929–1941.
- [15] L. WANG, K. ZHANG, AND L. ZHANG, *Perfect phylogenetic networks with recombination*, *J. Comput. Biol.*, 8 (2001), pp. 69–78.
- [16] Y. YU, R. M. BARNETT, AND L. NAKHLEH, *Parsimonious inference of hybridization in the presence of incomplete lineage sorting*, *Systematic Biol.*, 62 (2013), pp. 738–751.
- [17] Y. YU, J. DONG, K. J. LIU, AND L. NAKHLEH, *Maximum likelihood inference*

of reticulate evolutionary histories, *Proc. Natl. Acad. Sci. USA*, 111 (2014), pp. 16448–16453.

- [18] S. ZOLLNER AND J. K. PRITCHARD, *Coalescent-based association mapping and fine mapping of complex trait loci*, *Genetics*, 169 (2005), pp. 1071–1092.

LUAY NAKHLEH
Rice University

Finite Element Methods for Computational Fluid Dynamics: A Practical Guide. By Dmitri Kuzmin and Jari Hämäläinen. SIAM, Philadelphia, 2014. \$104.00. viii+313 pp., softcover. ISBN 978-1-611973-60-0.

In great clarity this book gives a thorough introduction to the finite element method and its application to flow problems. The focus is on applicability and efficiency of powerful numerical methods. Outstanding parts of this book are chapters on modern stabilization techniques. To allow for easy reading, the authors first present basic ideas and analysis for one-dimensional problems, followed by an exhaustive discussion of multidimensional systems. A further highlight of this monograph is the practical guide to the finite element software Elmer, allowing for a direct implementation and realization of the different approximation schemes. This book is a substantial help for graduate students and researchers looking for a state-of-the-art description of modern numerical tools in the field of computational fluid dynamics.

THOMAS RICHTER
University of Heidelberg

Programming Projects in C for Students of Engineering, Science, and Mathematics. By Rouben Rostamian. SIAM, Philadelphia, 2014. \$69.00. xvi+393 pp., softcover. ISBN 978-1-611973-49-5.

This is a textbook for a scientific computing course for graduate students and advanced undergraduates. It can also be used for self study. Why C? C is relatively old and well established; it's not this week's fashionable language. Compilers are freely available for