Preface

Why should we care about networks and network data? For one, they are *everywhere*. We are living in networks, we are using networks, and we *are* networks.

Too philosophical? Do you use Google? Did you browse some online news or social media today? They are all built on the web, a giant network of webpages. Whenever we click a link on the web, we are navigating the web. Google was one of the first companies that recognized the value of understanding the web's network structure, which led to the development of "PageRank," the famous algorithm behind their success. In fact, look at any large technological company. You will see that most of them are now built on some type of network and are extracting massive value from those networks.

Whenever we see *relationships*, there are networks. Do you want to build a massive online marketplace? Your chance of success may increase if you understand and leverage the network between customers and products. Do you want to understand how our cells and brains work? It boils down to understanding the cellular network (interactions between biological components) or the brain network.

If your work calls for understanding a complex system, chances are, you will be examining some kind of network. You will be working with network data!

Why this book?

There are great textbooks on network science. We complement these with a focus on the practical side of network science—working with network data. The purpose of this book is to provide a more practical guide for data scientists to use network science.

For instance, think about the process of *defining* a network. Although it is usually not emphasized, the process of defining nodes and edges from data is often critical but non-trivial! What about visual, exploratory analyses of networks? While an integral part of network science research—how you visualize networks can make or break your project and determine whether you discover salient patterns—it is often not discussed or taught in enough detail.

So, here we are. First, we hope that this book can help researchers in day-to-day tasks, starting from the very act of conceptualizing networks through to sophisticated network analysis, from exploratory analysis to statistical modeling and machine learning. We believe that this book can be a useful resource for simple yet critical questions that researchers and practitioners face everyday, such as: How should I define a network from this data? Should I conceptualize this network as a weighted, directed network, or

should we ignore those properties? How can I reduce or simplify this network? How can I visualize this network? How can I interpret the results of community analysis? And so on and so forth.

Second, we aim to give data scientists a foundational understanding of the tools, both mathematical and computational, at their disposal. The breadth and depth of statistical methods we can now use on network data is dizzying. We wish to take the prepared data scientist from their base knowledge of mathematics and statistics forward on a journey through the fundamentals of network data.

We hope to help students and researchers navigate both common yet critical questions and empower them with cutting-edge tools and the understanding to bend them to address their work.

How to use this book

This book is organized into two major parts—*practice* and *fundamentals*—connected by an interlude about good computing practices. As we emphasize the *practice* of network science, we present it first. But that does not mean that a course, or a self-study, should follow the same structure. Depending on the needs, goals, and level of preparation, one may pick and choose either the practical or the theoretical part. For instance, a network science course may use chapters from both parts in parallel to cover theoretical and practical aspects of each module's topic (e.g., see Chs. 11 and 12 for a practical guides to data and Ch. 22 for theoretical models of that data).

These two parts are preceded by an introduction and background that explains the basic concepts of network science as well as mathematical primers for the theory and data ethics. They may be used as a quick review of the basics or an introductory material.

Throughout the text we use *boxes* to emphasize certain points:

> **ⓘ** "Info boxes" provide definitions and contexts.

> **✅** "Good boxes" point out good practices and steps to take.

> **⚠** "Warning boxes" highlight dangers and concerns to be mindful of.

Prerequisites Although not required, you probably want to understand the basics of network science as may be conveyed by a one-semester introductory undergraduate course. Familiarity with probability, statistics, and linear algebra will be necessary, with stronger familiarity needed for Part III. Some experience with basic programming as well as basic knowledge of algorithms and data structures is also assumed and we will on occasion present code snippets, either in pseudocode or in Python. We review many of these topics in Ch. 4.

A website accompanying this book is available online at cambridge.org/network-data.

Acknowledgements

We are grateful to the many students, collaborators, and mentors who inspired us to write this book, enriched our understanding of this material, and helped us improve our text. We are indebted to David Hemsley and Nicholas Gibbons, Stephanie Windows, and Jane Chan at Cambridge University Press for shepherding our book from creation to production. Above all, we thank our loved ones for their support and encouragement during the writing of this book.

<div align="right">

James Bagrow
Williston, Vermont

Yong-Yeol Ahn
Bloomington, Indiana

February, 2024

</div>