A First Course in Network Science

Networks are present in all aspects of our lives: networks of friends, communication and transportation networks, and the Web are all examples that we experience outwardly, while the neurons in our brain and the proteins within our body form networks that determine our intelligence and survival. This modern and accessible textbook introduces the basics of network science required across a wide range of job sectors from management to marketing, from biology to engineering, and from neuroscience to the social sciences. Students will develop important, practical skills and the ability to write code for using networks in their areas of interest — even as they are just learning to program with Python. Extensive sets of tutorials and homework problems provide plenty of hands-on practice and longer programming tutorials online further enhance students' programming skills. This intuitive and direct approach makes the book ideal for a first course, aimed at a wide audience without a strong background in mathematics or computing but with a desire to learn the fundamentals and applications of network science.

Filippo Menczer is Professor of Informatics and Computing at Indiana University, Bloomington. He is an ACM Distinguished Scientist and board member of the Indiana University Network Science Institute (IUNI). He serves in editorial roles for the leading journals *Network Science*, *EPJ Data Science*, and *PeerJ Computer Science*. His research focuses on network science, computational social science, and Web science, with a focus on countering social media manipulation. His work on the spread of misinformation has received worldwide news coverage.

Santo Fortunato is Director of the Network Science Institute (IUNI) and Professor of Informatics at Indiana University, Bloomington. His current research is focused on network science, specifically network community detection, computational social science, and the 'science of science'. He received the German Physical Society's Young Scientist Award for Sociophysics and Econophysics in 2011 for his important contributions to the physics of social systems. He is Founding Chair of the International Conference on Computational Social Science (IC2S2).

Clayton A. Davis holds a Ph.D. in Informatics and BS and MA degrees in Mathematics from Indiana University, Bloomington. His research is concerned with the development of big-data platforms for social media analytics, machine learning algorithms for combating online abuse, design of crowdsourcing platforms, and the role of social media in social movements. His work on social bot detection was featured in major news outlets worldwide. His Web tools, including Botometer, Kinsey Reporter, and the Observatory on Social Media, answer millions of queries from thousands of users weekly. He won the 2017 Informatics Associate Instructor Award for his role in the development of high-quality teaching material for network science courses.

A First Course in Network Science

FILIPPO MENCZER
Indiana University, Bloomington

SANTO FORTUNATO
Indiana University, Bloomington

CLAYTON A. DAVIS
Indiana University, Bloomington

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

79 Anson Road, #06–04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108471138
DOI: 10.1017/9781108653947

First published 2020

Printed in Singapore by Markono Print Media Pte Ltd

A catalogue record for this publication is available from the British Library.

ISBN 978-1-108-47113-8 Hardback

Additional resources for this publication at www.cambridge.org/menczer.

Colleen, Massimiliano, Iris: thank you.
—Filippo Menczer

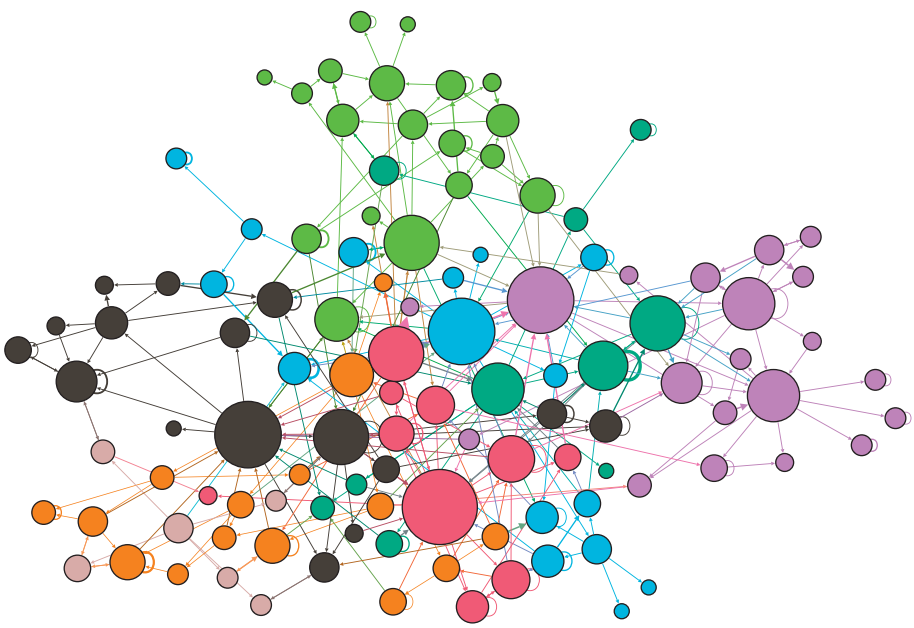
To my parents and my brother.
—Santo Fortunato

Liz, Gina, Mary Jo and Jay: your love and support mean everything to me.
—Clayton Davis

Contents

Preface

Networks are present in all aspects of our lives: networks of friends, communications, computers, the Web, and transportation are examples that we experience outwardly, while our brain cells and the proteins within our body form networks that determine our survival and intelligence. When people communicate through Facebook or Twitter, buy stuff on Amazon, search on Google, or buy an air ticket to visit family, they use networks without knowing it. Today, a basic understanding of network processes is required in job sectors from technology to marketing, from management to design, and from biology to the arts and humanities. This textbook explores the study of networks and how they help us understand the complex patterns of relationships that shape our lives.



This book is also a network! The relationships between chapters, sections, and subsections are depicted in the above image. Links represent both the hierarchical structure of the book (as seen in the Contents) and cross-references among chapters, sections, figures, tables, equations, and boxes. Node colors represent chapters and node size is proportional to the number of neighbors.

Why a "First Course" in Network Science?

This is not the first book on network science — in fact there are several excellent ones to choose from, and we list a few in Chapter 1. We have been teaching these topics for several years at Indiana University, to a broad audience of undergraduate students in informatics, computer science, data science, information science, business, and the natural and social sciences. This experience has taught us that students are eager to "get their hands dirty" and write code to both understand and use networks in their application domains of interest — even as they are just learning to program and lack math and computing background beyond high school and entry-level college courses. So we developed a wide set of tutorials and problems, both theoretical and computational, providing students with an abundance of hands-on practice in network science. Using such an approach, the book introduces networks to a wide audience of learners with no technical prerequisites other than some introductory programming, and a willingness to learn by doing. This makes the textbook appropriate for a "first course" in network science.

Synopsis

After surveying networks in many areas, we talk about social networks, which are the most familiar to students. This allows the introduction of concepts like the small-world property (short paths) and clustering (triangles and transitivity). These topics are explained through fun learning activities such as the *Six Degrees of Kevin Bacon* game. Then we delve into the role of hubs by using the Friendship Paradox and discuss network robustness. Next, we introduce directed and weighted networks, respectively. The Web, Wikipedia, citations, traffic, and Twitter are used to illustrate the role of direction and weights. The final three chapters cover more advanced topics, namely models for the emergence of networks, community detection methods, and dynamic processes that take place on networks.

Each chapter brings into focus the basic concepts necessary to understand a fundamental aspect of networks; advanced topics and formalism are avoided. When a bit of math is helpful, we include it in boxes. This slightly more technical content can be skipped without loss of basic understanding of a topic. But students who can follow these extra notes will be able to gain a deeper comprehension of the material. Each chapter includes programming tutorials and exercises, allowing readers to apply and test their knowledge through hands-on activities for building and analyzing networks. These tutorials work on examples of real-world networks that are used to illustrate concepts throughout the book. Both tutorial code and network data are available on the book's GitHub repository.[1]

[1] github.com/CambridgeUniversityPress/FirstCourseNetworkScience

Target Audience

With the explosion of popularity and commercial success of online social media, many students are interested in learning a bit about what is "under the hood" of networks. This textbook is aimed at all of those students, mainly at the undergraduate level, although the book may be useful for introductory graduate courses in non-technical fields as well. Students in programs such as data science, informatics, business, computer science, engineering, information science, biology, physics, statistics, and social sciences will benefit from courses based on this textbook. Their interest will be piqued enough to study network science in greater depth, and perhaps they will choose a career that might land them a job at Google, Facebook, Twitter, or a network start-up of their own.

Pedagogy

No technical mathematical or programming background is required, making the book feasible for introductory courses at any level, including network literacy and programming literacy courses. Such courses may skip the math boxes. By working through the programming tutorials in a collaborative computing lab and assigning the coding exercises, instructors will enable students to acquire sufficient technical skills to perform data-analytic tasks involving networks. This is our approach at Indiana University, where we teach the material in the book over two courses: a first introductory class aimed at sophomore/junior students who have taken or are taking concurrent programming courses in Python; and a second class aimed at junior/senior students. The first course roughly covers the material in Chapters 0–4. The second focuses on Chapters 5–7 after an extended review and some more advanced tutorials on the earlier material.

The extensive programming tutorials and exercises will allow instructors to easily lead and assign hands-on activities, and empower students to reinforce and test their understanding of network concepts. The activities include tutorials on *NetworkX*, a widely adopted library for network analytics; and on all topics covered in the book, from basic exercises to advanced techniques. For example, one tutorial guides students through the steps of extracting social network data from the Web. Using the Twitter application programming interface (API), students will be able to analyze popular topics, identify influential users, and reconstruct information diffusion networks showing how hashtags spread online. Students who carry out the programming tutorials and exercises will become proficient in building, importing/exporting, analyzing, manipulating, and visualizing networks of any type.

The tutorials are in Python, which is the most popular scripting/programming language. A primer that reviews the basic concepts of programming in Python is included in Appendix A. All tutorials are available online as Ipython Notebooks. Over time, NetworkX (and even Python) may evolve and some of the code in the book may need to be updated. We will note such updates in the book's GitHub repository.

There are, of course, other libraries for programming networks, for example *igraph*, *SNAP*, and *graph-tool*. Our selection of NetworkX is based on the fact that it is written in pure Python, which makes it easy to debug for students who are familiar with Python. Many alternatives have Python interfaces but are written in C, making them more efficient but also harder to debug.

Finally, some chapters leverage interactive models to demonstrate network phenomena such as giant components, small worlds, PageRank, preferential attachment, and epidemics. These models run in NetLogo, a popular simulation platform. A tutorial on NetLogo and some of the most relevant models is presented in Appendix B.

About the Cover

The network on the cover, generated by Onur Varol (Ferrara *et al.*, 2016), depicts the diffusion of the `#SB277` hashtag on Twitter. This hashtag refers to a 2015 California law on vaccination requirements and exemptions, and the network represents the discussion that was taking place online among supporters and opponents of the bill. Nodes represent Twitter users, and links show information spreading among users via retweets. Node size represents account influence (how many times a user is retweeted) and node colors represent bot scores: red nodes are likely to be bot accounts, blue nodes are likely to be humans.

Acknowledgments

The initial idea for this book originated from conversations with our former colleague Alex Vespignani. Over the years, our colleagues Sandro Flammini, YY Ahn, and Filippo Radicchi have provided precious advice. Several students have assisted in the teaching of our network science courses at Indiana University. Among them, we want to acknowledge Mike Conover, who first conceived some of the exercises found in these pages. We are also grateful to colleagues who provided feedback on early drafts of the book, especially Claudio Castellano, Chato Castillo, and several anonymous reviewers.

We thank our wonderful collaborators, students, postdocs, and visitors: Ana Maguitman, Ben Markines, Bruno Gonçalves, Chengcheng Shao, Diep Thi Hoang, Dimitar Nikolov, Emilio Ferrara, Giovanni Luca Ciampaglia, Jacob Ratkiewicz, Jasleen Kaur, Jose Ramasco, Kai-Cheng Yang, Karissa McKelvey, Kazu Sasahara, Le-Shin Wu, Lilian Weng, Luca Maria Alello, Mark Meiss, Markus Jakobsson, Mihai Avram, Nicola Perra, Onur Varol, Pik-Mai Hui, Prashant Shiralkar, Przemek Grabowicz, Ruj Akavipat, Xiaodan Lou, Xiaoling Sun, Zoher Kachwala, and too many others to name. They are an amazingly bright and fun bunch of people, who have contributed in many ways to the ideas, datasets, and illustrations in the book.

Our work would not be possible without the support of many dedicated staff members in the Center for Complex Networks and Systems Research, the School of Informatics, Computing, and Engineering, and the Indiana University Network Science Institute. Above all, we gratefully acknowledge Tara Holbrook, Michele Dompke, Rob Henderson, Dave Cooley, Patty Mabry, Ann McCranie, Val Pentchev, Matthew Hutchinson, Chathuri Peli Kankanamalage, and Ben Serrette. Thanks also to Nick Gibbons at Cambridge University Press for his encouragement and feedback.

We are grateful to Aric Hagberg, Pieter Swart, and Dan Schult, the authors of NetworkX; and to Uri Wilenski and the Center for Connected Learning and Computer-Based Modeling at Northwestern University for developing and maintaining NetLogo.

Finally, we owe a huge debt of gratitude to our families who love, support, and bear with us even as we work more than we should.