Random Graphs and Complex Networks

Volume 2

Complex networks are key to describing the connected nature of the society that we live in. This book, the second of two volumes, describes the local structure of random graph models for real-world networks and determines when these models have a giant component and when they are small-, and ultra-small, worlds.

This is the first book to cover the theory and implications of local convergence, a crucial technique in the analysis of sparse random graphs. Suitable as a resource for researchers and PhD-level courses, it uses examples of real-world networks, such as the Internet and citation networks, as motivation for the models that are discussed, and includes exercises at the end of each chapter to develop intuition. The book closes with an extensive discussion of related models and problems that demonstrate modern approaches to network theory, such as community structure and directed models.

REMCO VAN DER HOFSTAD is Full Professor of Probability at Eindhoven University of Technology. He received the 2003 Prix Henri Poincaré (jointly with Gordon Slade) and the 2007 Rollo Davidson Prize, and he is a laureate of the 2003 Innovative Research VIDI Scheme and the 2008 Innovative Research VICI Scheme. He served as scientific director at Eurandom from 2011 to 2019, and is a member of the Dutch Royal Academy of Science. Van der Hofstad is one of the principal investigators of the prestigious NETWORKS Gravitation Program and is the creator of the interactive website networkpages.nl, aimed at a broad audience.

CAMBRIDGE SERIES IN STATISTICAL AND PROBABILISTIC MATHEMATICS

This series of high-quality upper-division textbooks and expository monographs covers all aspects of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, optimization, and mathematical programming. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasizing rigorous treatment of theoretical methods, the books also contain applications and discussions of new techniques made possible by advances in computational practice.

A complete list of books in the series can be found at www.cambridge.org/statistics. Recent titles include the following:

31. *Probability: Theory and Examples* (*Fourth Edition*), by Rick Durrett
32. *Analysis of Multivariate and High-Dimensional Data*, by Inge Koch
33. *Stochastic Processes*, by Richard F. Bass
34. *Regression for Categorical Data*, by Gerhard Tutz
35. *Exercises in Probability* (*Second Edition*), by Loïc Chaumont and Marc Yor
36. *Statistical Principles for the Design of Experiments*, by R. Mead, S. G. Gilmour and A. Mead
37. *Quantum Stochastics*, by Mou-Hsiung Chang
38. *Nonparametric Estimation under Shape Constraints*, by Piet Groeneboom and Geurt Jongbloed
39. *Large Sample Covariance Matrices and High-Dimensional Data Analysis*, by Jianfeng Yao, Shurong Zheng and Zhidong Bai
40. *Mathematical Foundations of Infinite-Dimensional Statistical Models*, by Evarist Giné and Richard Nickl
41. *Confidence, Likelihood, Probability*, by Tore Schweder and Nils Lid Hjort
42. *Probability on Trees and Networks*, by Russell Lyons and Yuval Peres
43. *Random Graphs and Complex Networks* (*Volume 1*), by Remco van der Hofstad
44. *Fundamentals of Nonparametric Bayesian Inference*, by Subhashis Ghosal and Aad van der Vaart
45. *Long-Range Dependence and Self-Similarity*, by Vladas Pipiras and Murad S. Taqqu
46. *Predictive Statistics*, by Bertrand S. Clarke and Jennifer L. Clarke
47. *High-Dimensional Probability*, by Roman Vershynin
48. *High-Dimensional Statistics*, by Martin J. Wainwright
49. *Probability: Theory and Examples* (*Fifth Edition*), by Rick Durrett
50. *Model-based Clustering and Classification for Data Science*, by Charles Bouveyron *et al.*
51. *Spectral Analysis for Univariate Time Series*, by Donald B. Percival and Andrew T. Walden
52. *Statistical Hypothesis Testing in Context*, by Michael P. Fay and Erica H. Brittain
53. *The Fundamentals of Heavy Tails*, by Jayakrishnan Nair, Adam Wierman and Bert Zwart
54. *Random Graphs and Complex Networks* (*Volume 2*), by Remco van der Hofstad
55. *Modern Discrete Probability*, by Sébastien Roch
56. *Generalized Additive Models for Location, Scale and Shape*, by Mikis D. Stasinopoulos *et al.*

Random Graphs and Complex Networks

Volume 2

Remco van der Hofstad
Technische Universiteit Eindhoven

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, 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

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment, a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107174009

DOI: 10.1017/9781316795552

First published 2024

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

A Cataloging-in-Publication data record for this book is available from the Library of Congress

ISBN 978-1-107-17400-9 Hardback
ISBN 978-1-316-62632-0 Paperback

Aan Mad, Max en Lars
het licht in mijn leven

Ter nagedachtenis aan mijn ouders
die me altijd aangemoedigd hebben

Contents

Preface

Targets. In this book, which is Volume 2 of a sequence of two books, we study *local limits*, *connected components*, and *small-world properties* of random graph models for complex networks. Volume 1 describes the preliminaries of *random graphs* as models for *real-world networks*, as investigated since 1999. These networks turned out to be rather different from classical random graph models, for example in the number of connections that the elements make. As a result, a wealth of new models was invented to capture these properties. Volume 1 studies these models as well as their *degree structure*. Volume 2 summarizes the insights developed in this exciting period related to the *local, connectivity*, and *small-world structure* of the proposed random graph models. While Volume 1 is intended to be used for a master level course, where students have a limited prior knowledge of special topics in probability, Volume 2 describes the more involved notions that have been the focus of attention of the research community in the past two decades.

Volume 2 is intended to be used for a PhD level course, a reading seminar, or for researchers wishing to obtain a consistent and extended overview of the results and methodologies developed in this scientific area. Volume 1 includes many of the preliminaries, such as the convergence of random variables, probabilistic bounds, coupling, martingales, and branching processes, and we frequently rely on these results.

The sequence of Volumes 1 and 2 aims to be self-contained. In Volume 2, we briefly repeat some of the preliminaries on random graphs, including an introduction to the key models and their degree distributions, as discussed in detail in Volume 1. In Volume 2, we aim to give detailed and complete proofs. When we do not give proofs, we provide heuristics, as well as extensive pointers to the literature. We further discuss several more recent random graph models that aim to more realistically model real-world networks, as they incorporate their *directed* nature, their *community structure*, and/or their *spatial embedding*.

Developments. The field of random graphs was pioneered in 1959–1960 by Erdős and Rényi (1959; 1960; 1961a; 1961b), in the context of the *probabilistic method.* The initial work by Erdős and Rényi incited a great amount of follow-up in the field, initially mainly in the combinatorics community. See the standard references on the subject by Bollobás (2001) and Janson, Łuczak, and Ruciński (2000) for the state of the art. Erdős and Rényi (1960) gives a rather complete picture of the various phase transitions that occur in the Erdős–Rényi random graph. This initial work did not aim to model real-world networks realistically.

In the period after 1999, owing to the fact that data sets of large real-world networks became abundantly available, their structure has attracted enormous attention in mathematics as well as in various applied domains. This is exemplified by the fact that one of the first articles in the field, by Barabási and Albert (1999), has attracted over 40,000 citations. One of the main conclusions from this overwhelming body of work is that many real-world networks share two fundamental properties. The first is that they are highly *inhomogeneous*, in the sense that different vertices play rather different roles in the networks. This property is exemplified by the degree structure of the real-world networks obeying power laws: these networks are *scale-free*. This scale-free nature of real-world networks has prompted the community to come up with many novel random graph models that, unlike the Erdős–Rényi random graph, do have power-law degree sequences. This was the key focus in Volume 1.

Content. In this book, we pick up on the trail left in Volume 1, where we now focus on the *connectivity structure* between vertices. Connectivity can be summarized in two key aspects of real-world networks: the facts that they are *highly connected*, as exemplified by the fact that they tend to have one giant component containing a large proportion of the vertices (if not all of them), and that they are *small world*, in that most pairs of vertices are separated by short paths. We discuss the available methods for these proofs, including path-counting techniques, branching-process approximations, exchangeable random variables, and de Finetti's theorems. We pay particular attention to a recent technique, called *local convergence*, that makes the statement that random graphs "locally look like trees" precise.

This book consists of four parts. In Part I, consisting of Chapters 1 and 2, we start in Chapter 1 by repeating some definitions from Volume 1, including the random graph models studied in the present book, which are inhomogeneous random graphs, configuration models, and preferential attachment models. We also discuss general topics that are important in random graph theory, such as power-law distributions and their properties. In Chapter 2, we continue by discussing *local convergence*, an extremely powerful technique that plays a central role in the theory of random graphs and in this book. In Part II, consisting of Chapters 3–5, we discuss local limits and large connected components in random graph models. In Chapter 3, we further extend the definition of the generalized random graph to general inhomogeneous random graphs. In Chapter 4, we discuss the local limit and large connected components in the configuration model, and in Chapter 5, we discuss the local structure in, and connectivity of, preferential attachment models. In Part III, consisting of Chapters 6–8, we study the small-world nature of random graphs, starting with inhomogeneous random graphs, continuing with the configuration model, and ending with the preferential attachment model. In Part IV, consisting of Chapter 9, we study related random graph models and their structure.

Along the way, we give many exercises that should help the reader to obtain a deeper understanding of the material by working on the solutions. These exercises appear in the last section of each of the chapters, and, when applicable, we refer to them at the appropriate place in the text. We also provide extensive notes in the penultimate section of each chapter, where we discuss the links to the literature and some extensions.

Literature. We have tried to give as many references to the literature as possible. However, the number of papers on random graphs has exploded. In MathSciNet (see `www.ams.org/mathscinet`), there were, on December 21, 2006, a total of 1,428 papers that contain the phrase "random graphs" in the review text; on September 29, 2008, this number had increased to 1,614, to 2,346 on April 9, 2013; to 2,986 on April 21, 2016; and to 12,038 on October 5, 2020. These are merely the papers on the topic in the mathematics community. What is special about random graph theory is that it is extremely multidisciplinary, and many papers using random graphs are currently written in economics, biology, theoretical physics, and computer science. For example, in Scopus (see `www.scopus.com/scopus/home.url`), again on December 21, 2006, there were 5,403 papers that contain the phrase "random graph" in the title, abstract or keywords; on September 29, 2008, this had increased to 7,928; to 13,987 on April 9, 2013; to 19,841 on April 21, 2016; and to 30,251 on October

5, 2020. It can be expected that these numbers will continue to increase, rendering it utterly impossible to review all the literature.

In June 2014, we decided to split the preliminary version of this book up into two books. This has several reasons and advantages, particularly since Volume 2 is more tuned towards a research audience, while Volume 1 is aimed at an audience of master students of varying backgrounds. The pdf-versions of both Volumes 1 and 2 can be obtained from

`www.win.tue.nl/∼ rhofstad/NotesRGCN.html`.

For errata for this book and Volume 1, or possible outlines for courses based on them, readers are encouraged to look at this website or e-mail me. Also, for a more playful approach to networks for a broad audience, including articles, videos, and demos of many of the models treated in this book, we refer all readers to the NetworkPages at `www.networkspages.nl`. The NetworkPages provide an interactive website developed by and for all those who are interested in networks. Finally, we have relied on various real-world networks data sets provided by the KONECT project; see `http://konect.cc` as well as Kunegis (2013) for more details.

Thanks. This book, as well as Volume 1, would not have been possible without the help and encouragement of many people. I particularly thank Gerard Hooghiemstra for encouraging me to write it, and for using it at Delft University of Technology almost simultaneously while I was using it at Eindhoven University of Technology in the spring of 2006 and again in the fall of 2008. I thank Gerard for many useful comments, solutions to exercises, and suggestions for improvements of the presentation throughout the book. Together with Piet Van Mieghem, we entered the world of random graphs in 2001, and I have tremendously enjoyed exploring this field together with them, as well as with Henri van den Esker, Dmitri Znamenski, Mia Deijfen, Shankar Bhamidi, Johan van Leeuwaarden, Júlia Komjáthy, Nelly Litvak and many others.

I thank Christian Borgs, Jennifer Chayes, Gordon Slade, and Joel Spencer for joint work on random graphs that are like the Erdős–Rényi random graph but do have geometry. Special thanks go to Gordon Slade, who introduced me to the exciting world of percolation, which is closely linked to the world of random graphs (see the classic text on percolation by Grimmett (1999)). It is striking to see two communities working on two such closely related topics with different methods and even different terminology, and it has taken a long time to build bridges between the two subjects. I am very happy that these bridges are now rapidly appearing, and the level of communication between different communities has increased significantly. I hope that this book helps to further enhance this communication. Frank den Hollander deserves a special mention. Frank, you have been important as a driving force throughout my career, and I am very happy now to be working with you on fascinating random graph problems!

Further, I thank

Marie Albenque, Yeganeh Alimohammadi, Rangel Baldasso, Gianmarco Bet, Shankar Bhamidi, Finbar Bogerd, Marko Boon, Christian Borgs, Hao Can, Francesco Caravenna, Rui Castro, Kota Chisaki, Deen Colenbrander, Nicolas Curien, Umberto De Ambroggio, Mia Deijfen, Michel Dekking, Serte Donderwinkel, Dylan Dronnier, Henri van den Esker, Lorenzo Federico, Federica Finazzi, Allison Fisher, *Lucas Gerin, Cristian Giardinà, Claudia Giberti, Jesse Goodman, Rowel Gündlach, Rajat Hazra, Markus Heydenreich, Frank den Hollander, Yusuke Ide, Simon Irons, Emmanuel Jacob, Svante Janson, Guido Janssen, Lancelot James, Martin van Jole, Joost Jorritsma, Willemien Kets, Heejune Kim, Bas Kleijn, Júlia Komjáthy, Norio Konno, Dima Krioukov, John Lapeyre, Lasse Leskelä, Nelly Litvak, Neeladri Maitra, Abbas Mehrabian, Marta Milewska, Steven Miltenburg, Mislav Mišković, Christian Mönch, Peter Mörters, Mirko Moscatelli, Jan Nagel, Sidharthan Nair, Alex Olssen, Mariana Olvera-Cravioto, Helena Peña, Manish Pandey, Rounak Ray, Nathan Ross, Christoph Schumacher, Matteo Sfragara, Karoly Simon, Lars Smolders, Clara Stegehuis, Dominik Tomecki, Nicola Turchi, Viktória Vadon, Thomas Vallier, Irène Ayuso Ventura, Xiaotin Yu, Haodong Zhu and Bert Zwart*

for remarks and ideas that have improved the content and presentation of these books substantially. Wouter Kager read the February 2007 version of this book in its entirety, giving many ideas for improvements in the arguments and the methodology. Artëm Sapozhnikov, Maren Eckhoff, and Gerard Hooghiemstra read and commented on the October 2011 version. Haodong Zhu read the December 2023 version completely, and corrected several typos.

Particular thanks go to Dennis Timmers, Eefje van den Dungen, Joop van de Pol, Rowel Gündlach and Lourens Touwen, who, as, my student assistants, have been a great help in the development of this pair of books, in making figures, providing solutions to some of the exercises, checking proofs, and keeping the references up to date. Maren Eckhoff also provided many solutions to the exercises, for which I am grateful! Sándor Kolumbán, Robert Fitzner, and Lourens Touwen helped me to turn all pictures of real-world networks as well as simulations of network models into a unified style, a feat that is beyond my LaTeX skills. A big thanks for that! Also my thanks for suggestions and help with figures to Marko Boon, Alessandro Garavaglia, Dimitri Krioukov, Vincent Kusters, Clara Stegehuis, Piet Van Mieghem, and Yana Volkovich. A special thanks to my running mates Jan and Ruud, whose continuing support has been extremely helpful for me.

Support. This work would not have been possible without the generous support of the Netherlands Organization for Scientific Research (NWO) through VIDI grant 639.032.304, VICI grant 639.033.806, and the Gravitation NETWORKS grant 024.002.003.

Possible Course Outlines

The relation between the chapters in Volumes 1 and 2 is as follows:

Here is some more explanation as well as a possible itinerary of a master or PhD course on random graphs, based on Volume 2, in a course outline. For a course outline based on Volume 1, we refer to [V1, Preface] for alternative routes through the material, we refer to the book's website at `www.win.tue.nl/~ rhofstad/NotesRGCN.html`:

▷ Start with the introduction to real-world networks in [V2, Chapter 1], which forms the inspiration for what follows. For readers wishing for a more substantial introduction, do visit Volume 1 for an extensive introduction to the models discussed here.

▷ Continue with [V2, Chapter 2] on the local convergence of (random and non-random) graphs, as this is a crucial tool in the book and has developed into a key methodology in the field.

978-1-107-17400-9 — Random Graphs and Complex Networks
Remco van der Hofstad
Frontmatter
More Information

The material in this book is rather substantial, and probably too much to be treated in one course. Thus, we give two alternative approaches to teaching coherent parts of this book:

▷ You can either take one of the *models* and discuss the different chapters in Volume 2 that focus on them. [V2, Chapters 3 and 6] discuss inhomogeneous random graphs, [V2, Chapters 4 and 7] discuss configuration models, while [V2, Chapters 5 and 8] focus on preferential attachment models.

▷ The alternative is that you take one of the *topics*, and work through them in detail. [V2, Part II] discusses the local limits and largest connected components or phase transition in our random graph models, while [V2, Part III] treats their small-world nature.

If you have further questions and/or suggestions about course outlines, feel free to contact me. Refer to `www.win.tue.nl/∼ rhofstad/NotesRGCN.html` for further suggestions on how to lecture from Volume 2.