Random Graphs and Complex Networks

Volume 1

This rigorous introduction to network science presents random graphs as models for real-world networks. Such networks have distinctive empirical properties, and a wealth of new models have emerged to capture them. Classroom tested for over ten years, this text places recent advances in a unified framework to enable systematic study.

Designed for a master's-level course, where students may only have a basic background in probability, the text covers such important preliminaries as convergence of random variables, probabilistic bounds, coupling, martingales, and branching processes. Building on this base – and motivated by many examples of real-world networks, including the Internet, collaboration networks, and the World-Wide Web – it focuses on several important models for complex networks and investigates key properties, such as the connectivity of nodes. Numerous exercises allow students to develop intuition and experience in working with the models.

REMCO VAN DER HOFSTAD is full professor of probability at Eindhoven University of Technology and acting scientific director of the European Institute in Stochastics (Eurandom). He has authored over 100 research articles and has taught courses on random graphs at over 10 institutions. He received the Prix Henri Poincaré 2003 jointly with Gordon Slade, the Rollo Davidson Prize 2007, and is a laureate of the 'Innovative Research Vidi Scheme' 2003 and the 'Innovative Research Vici Scheme' 2008.

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:

17. *Elements of Distribution Theory*, by Thomas A. Severini
18. *Statistical Mechanics of Disordered Systems*, by Anton Bovier
19. *The Coordinate-Free Approach to Linear Models*, by Michael J. Wichura
20. *Random Graph Dynamics*, by Rick Durrett
21. *Networks*, by Peter Whittle
22. *Saddlepoint Approximations with Applications*, by Ronald W. Butler
23. *Applied Asymptotics*, by A. R. Brazzale, A. C. Davison and N. Reid
24. *Random Networks for Communication*, by Massimo Franceschetti and Ronald Meester
25. *Design of Comparative Experiments*, by R. A. Bailey
26. *Symmetry Studies*, by Marlos A. G. Viana
27. *Model Selection and Model Averaging*, by Gerda Claeskens and Nils Lid Hjort
28. *Bayesian Nonparametrics*, edited by Nils Lid Hjort *et al.*
29. *From Finite Sample to Asymptotic Methods in Statistics*, by Pranab K. Sen, Julio M. Singer and Antonio C. Pedrosa de Lima
30. *Brownian Motion*, by Peter Mörters and Yuval Peres
31. *Probability (Fourth Edition)*, by Rick Durrett
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*, by Remco van der Hofstad

Random Graphs and Complex Networks

Volume 1

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/9781107172876

First published 2017

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

ISBN 978-1-107-17287-6 Hardback

Aan Mad, Max en Lars
het licht in mijn leven

Ter nagedachtenis aan mijn ouders
die me altijd aangemoedigd hebben

Contents

Preface

In this book, we study *random graphs* as models for *real-world networks*. Since 1999, many real-world networks have been investigated. These networks turned out to have rather different properties than classical random graph models, for example in the number of connections that the elements in the network make. As a result, a wealth of new models were invented to capture these properties. This book summarizes the insights developed in this exciting period.

This book is intended to be used for a master-level course where students have a limited prior knowledge of special topics in probability. We have included many of the preliminaries, such as convergence of random variables, probabilistic bounds, coupling, martingales and branching processes. This book aims to be self-contained. When we do not give proofs of the preliminary results, we provide pointers to the literature.

The field of random graphs was initiated in 1959–1960 by Erdős and Rényi (1959; 1960; 1961a; 1961b). At first, the theory of random graphs was used to prove deterministic properties of graphs. For example, if we can show that a random graph with a positive probability has a certain property, then a graph must exist with this property. The method of proving deterministic statements using probabilistic arguments is called the *probabilistic method*, and goes back a long way. See among others the preface of a standard work in random graphs by Bollobás (2001), or the classic book devoted to *The Probabilistic Method* by Alon and Spencer (2000). Erdős was one of the pioneers of this method; see, e.g., Erdős (1947), where he proved that the Ramsey number $R(k)$ is at least $2^{k/2}$. The Ramsey number $R(k)$ is the value n for which any graph of size at least n or its complement contains a complete graph of size at least k. Erdős (1947) shows that for $n \leq 2^{k/2}$ the fraction of graphs for which the graph or its complement contains a complete graph of size k is bounded by $1/2$, so that there must be graphs of size $n \leq 2^{k/2}$ for which neither the graph nor its complement contains a complete graph of size k.

The initial work by Erdős and Rényi on random graphs has incited a great amount of work 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) give a rather complete picture of the various phase transitions that occur in the Erdős-Rényi random graph. An interesting quote (Erdős and Rényi, 1960, pages 2–3) is the following:

> It seems to us worthwhile to consider besides graphs also more complex structures from the same point of view, i.e. to investigate the laws governing their evolution in a similar spirit. This may be interesting not only from a purely mathematical point of view. In fact,

> the evolution of graphs can be seen as a rather simplified model of the evolution of certain communication nets...

This was an excellent prediction indeed! Later, interest in different random graphs arose due to the analysis of real-world networks. Many of these real-world networks turned out to share similar properties, such as the fact that they are *small worlds*, and are *scale free*, which means that they have degrees obeying *power laws*. The Erdős-Rényi random graph does not obey these properties, and therefore new random graph models needed to be invented. In fact, Erdős and Rényi (1960) already remark that

> Of course, if one aims at describing such a real situation, one should replace the hypothesis of equiprobability of all connections by some more realistic hypothesis.

See Newman (2003) and Albert and Barabási (2002) for two reviews of real-world networks and their properties to get an impression of what 'more realistic' could mean, and see the recent book by Newman (2010) for detailed accounts of properties of real-world networks and models for them. These other models are also partly covered in the classical works by Bollobás (2001) and Janson, Łuczak and Rucínsky (2000), but up until today there is no comprehensive text treating random graph models for complex networks. See Durrett (2007) for a recent book on random graphs, and, particularly, dynamical processes living on them. Durrett covers part of the material in the present book, and more, but the intended audience is different. Our goal is to provide a source for a 'Random Graphs' course at the master level.

We describe results for the Erdős-Rényi random graph as well as for random graph models for complex networks. Our aim is to give the simplest possible proofs for classical results, such as the phase transition for the largest connected component in the Erdős-Rényi random graph. Some proofs are more technical and difficult. The sections containing these proofs are indicated with a star * and can be omitted without losing the logic behind the results. We also give many exercises that help the reader to obtain a deeper understanding of the material by working on their 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.

I have tried to give as many references to the literature as possible. However, the number of papers on random graphs is currently exploding. In MathSciNet (see `http://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 increased to 1,614, on April 9, 2013, to 2,346 and, on April 21, 2016, to 2,986. These are merely the papers on the topic in the math 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 `http://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 increased to 7,928; on April 9, 2013, to 13,987 and, on April 21, 2016, to 19,841. It can be expected that these numbers will continue to increase, rendering it impossible to review all the literature.

In June 2014, I decided to split the preliminary version of this book up into two books. This has several reasons and advantages, particularly since the later part of the work is more tuned towards a research audience, while the first part is more tuned towards an audience of master students with varying backgrounds. For the latest version of Volume II, which focusses on connectivity properties of random graphs and their small-world behavior, we refer to

`http://www.win.tue.nl/~rhofstad/NotesRGCN.html`

For further results on random graphs, or for solutions to some of the exercises in this book, readers are encouraged to look there. 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 Network Pages at `http://www.networkspages.nl`. The Network Pages are an interactive website developed by and for all those who are interested in networks. One can find demos for some of the models discussed here, as well as of network algorithms and processes on networks.

This book would not have been possible without the help and encouragement of many people. I thank Gerard Hooghiemstra for the encouragement to write it, and for using it at Delft University of Technology almost simultaneously while I used it at Eindhoven University of Technology in the Spring of 2006 and again in the Fall of 2008. I particularly 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 you, as well as with Henri van den Esker, Dmitri Znamenski, Mia Deijfen and 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. This work has deepened my understanding of the basic properties of random graphs, and many of the proofs presented here have been inspired by our work in Borgs et al. (2005a, b, 2006). Special thanks go to Gordon Slade, who has introduced me to the world of percolation, which is closely linked to the world of random graphs (see also the classic on percolation by Grimmett (1999)). It is peculiar to see that two communities work on two so closely related topics with different methods and even different terminology, and that it has taken such a long time to build bridges between the 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!

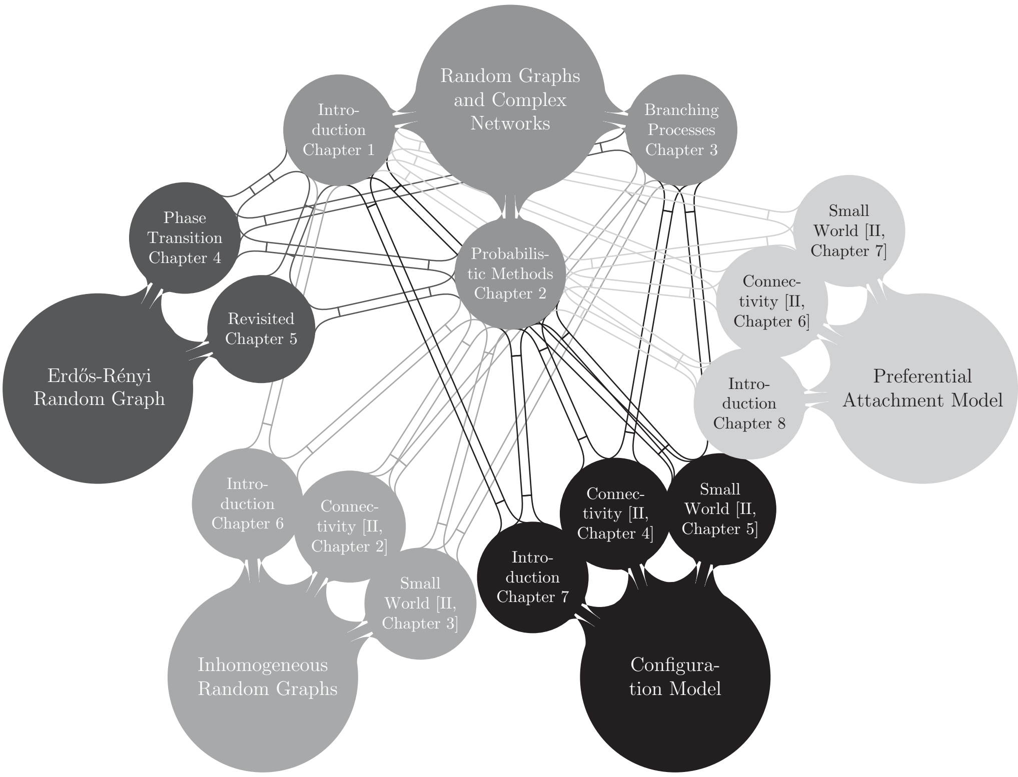
Further I thank Marie Albenque, Gianmarco Bet, Shankar Bhamidi, Finbar Bogerd, Marko Boon, Francesco Caravenna, Rui Castro, Kota Chisaki, Mia Deijfen, Michel Dekking, Henri van den Esker, Lucas Gerin, Jesse Goodman, Rajat Hazra, Markus Heydenreich, Frank den Hollander, Yusuke Ide, Lancelot James, Martin van Jole, Willemien Kets, Júlia Komjáthy, John Lapeyre, Nelly Litvak, Norio Konno, Abbas Mehrabian, Mislav Mišković, Mirko Moscatelli, Jan Nagel, Sidharthan Nair, Alex Olssen, Mariana Olvera-Cravioto, Helena Peña, Nathan Ross, Karoly Simon, Dominik Tomecki, Nicola Turchi, Thomas Vallier and Xiaotin Yu for remarks and ideas that have improved the content and presentation of these notes substantially. Wouter Kager has entirely read the February 2007 version of this book, giving many ideas for improvements of the arguments and of the methodology. Artëm Sapozhnikov, Maren Eckhoff and Gerard Hooghiemstra read and commented the October 2011 version. Sándor Kolumbán read large parts of the October 2015 version, and spotted many errors, typos and inconsistencies.

I especially thank Dennis Timmers, Eefje van den Dungen and Joop van de Pol, who, as, my student assistants, have been a great help in the development of this book, in making figures, providing solutions to the exercises, checking the proofs and keeping the references up to date. Maren Eckhoff and Gerard Hooghiemstra also provided many solutions to the exercises, for which I am grateful! Sándor Kolumbán and Robert Fitzner 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.

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

Course Outline

The relation between the chapters in Volumes I and II of this book is as follows:



Here is some more explanation as well as a possible itinerary of a master course on random graphs. We include Volume II (see van der Hofstad (2015)) in the course outline:

Start with the introduction to real-world networks in Chapter 1, which forms the inspiration for what follows. Continue with Chapter 2, which gives the necessary probabilistic tools used in all later chapters, and pick those topics that your students are not familiar with and that are used in the later chapters that you wish to treat. Chapter 3 introduces branching processes, and is used in Chapters 4 and 5, as well as in most of Volume II.

After these preliminaries, you can start with the classical Erdős-Rényi random graph as covered in Chapters 4 and 5. Here you can choose the level of detail, and decide whether you wish to do the entire phase transition or would rather move on to the random graphs models for complex networks. It is possible to omit Chapter 5 before moving on.

After this, you can make your own choice of topics from the models for real-world networks. There are three classes of models for complex networks that are treated in this book. You can choose how much to treat in each of these models. You can either treat few models and discuss many aspects, or instead discuss many models at a less deep level. The introductory chapters about the three models, Chapter 6 for inhomogeneous random graphs, Chapter 7 for the configuration model, and Chapter 8 for preferential attachment models, provide a basic introduction to them, focussing on their degree structure. These introductory chapters need to be read in order to understand the later chapters about these models (particularly the ones in Volume II). The parts on the different models can be read independently.