Preface

Networks and network analysis are arguably one of the largest recent growth areas in the quantitative sciences. Despite roots in social network analysis going back to the 1930s and roots in graph theory going back centuries, the phenomenal rise and popularity of the modern field of 'network science', as it is sometimes called, is something that likely could not have been predicted 10–15 years ago. Networks have permeated everyday life, far beyond the realm of research and methodology, through now-familiar realities like the Internet, social networks, viral marketing, and more.

Measurement and data analysis are integral components of network research. As a result, there is a critical need for all sorts of statistics for network analysis, both common and sophisticated, ranging from applications, to methodology, to theory. As with other areas of statistics, there are both descriptive and inferential statistical techniques available, aimed at addressing a host of network-related tasks, including basic visualization and characterization of network structure; sampling, modeling, and inference of network topology; and modeling and prediction of network-indexed processes, both static and dynamic.

Software for performing many such network-related analyses is now available in various languages and environments, across different platforms. Not surprisingly, the R community has been particularly active in the development of software for doing statistical analysis of network data. As of this writing there are already dozens of contributed R packages devoted to some aspect of network analysis. Together, these packages address tasks ranging from standard manipulation, visualization, and characterization of network data (e.g., **igraph**, **network**, and **sna**), to modeling of networks (e.g., **igraph**, **eigenmodel**, **ergm**, and **mixer**), to network topology inference (e.g., **glasso** and **huge**). In addition, there is a great deal of analysis that can be done using tools and functions from the R base package.

In this book we aim to provide an easily accessible introduction to the statistical analysis of network data, by way of the R programming language. As a result, this book is not, on the one hand, a detailed manual for using the various R packages encountered herein, nor, on the other hand, does it provide exhaustive coverage of the conceptual and technical foundations of the topic area. Rather, we have attempted

to strike a balance between the two and, in addition, to do so using a (hopefully!) optimal level of brevity. Accordingly, we envision the book being used, for example, by (i) statisticians looking to begin engaging in the statistical analysis of network data, whether at a research level or in conjunction with a new collaboration, and hoping to use R as a natural segue, (ii) researchers from other similarly quantitative fields (e.g., computer science, statistical physics, economics, etc.) working in the area of complex networks, who seek to get up to speed relatively quickly on how to do statistical analyses (both familiar and unfamiliar) of network data in R, and (iii) practitioners in applied areas wishing to get a foothold into how to do a specific type of analysis relevant to a particular application of interest.

More generally, the book has been written at a level aimed at graduate students and researchers in quantitative disciplines engaged in the statistical analysis of network data, although advanced undergraduates already comfortable with R should find much of the book fairly accessible as well. At present, therefore, we anticipate the book being of interest to readers not only in statistics, of course, but also in areas like computational biology, computer science and machine learning, economics, neuroscience, quantitative finance, signal processing, statistical physics, and the quantitative social sciences.

There are a number of people we wish to thank, whose help at various stages of development and writing is greatly appreciated. Thanks go to the editorial team at Springer for their enthusiasm in encouraging us to take on this project and for their feedback along the way and to the students in the course Statistical Analysis of Network Data (MA703) at Boston University in Fall 2013, for their comments on many of the earlier chapters. Special thanks are also due to Xinyu Kang, Heather Shappell, and Yaonan Zhang, who went through the entire first complete draft, carefully reading every chapter and testing the code blocks throughout the book. We are grateful as well to Christophe Ambroise, Alain Barrat, Mark Coates, Suchi Gopal, Emmanuel Lazega, and Petra Staufer for kindly making available their data. More broadly, we would like to express our appreciation in general for the countless hours of effort invested by the developers of the many R packages that we have made use of throughout the pages of this book. Without their work, the breadth and scope of our own here would be significantly reduced. Finally, although surely still inadequate, we wish to express our deepest gratitude to our respective families for their love, patience, and support throughout the writing of this book.

All code and data used in this book have been made available in the R package **sand**, distributed through the CRAN archive.

Boston, MA, USA — Eric D. Kolaczyk
Cambridge, MA, USA — Gábor Csárdi
March, 2014

http://www.springer.com/978-1-4939-0982-7

Statistical Analysis of Network Data with R

Kolaczyk, E.D.; Csárdi, G.

2014, XIII, 207 p. 55 illus., 53 illus. in color., Softcover
ISBN: 978-1-4939-0982-7