

Use R!

Douglas A. Luke

# A User's Guide to Network Analysis in R

 Springer

---

# Use R!

## Series Editors

Robert Gentleman

Div. Public Health Sciences, San Mateo, California, USA

Kurt Hornik

Accounting & Statistics, WU Wien, Dept of Finance, Wien, Austria

Giovanni Parmigiani

Dana Farber Cancer Institute, Boston, USA

More information about this series at <http://www.springer.com/series/6991>

*Albert*: Bayesian Computation with R (2nd ed. 2009)

*Bivand/Pebesma/Gómez-Rubio*: Applied Spatial Data Analysis with R (2nd ed. 2013)

*Cook/Swayne*: Interactive and Dynamic Graphics for Data Analysis: With R and GGobi

*Hahne/Huber/Gentleman/Falcon*: Bioconductor Case Studies

*Paradis*: Analysis of Phylogenetics and Evolution with R (2nd ed. 2012)

*Pfaff*: Analysis of Integrated and Cointegrated Time Series with R (2nd ed. 2008)

*Sarkar*: Lattice: Multivariate Data Visualization with R

*Spector*: Data Manipulation with R

---

*Douglas A. Luke*

---

# **A User's Guide to Network Analysis in R**

**1st ed. 2015**



Douglas A. Luke

---

Center for Public Health Systems Science George Warren Brown School of Social Work, Washington University, St. Louis, MO, USA

ISSN 2197-5736 e-ISSN 2197-5744

ISBN 978-3-319-23882-1 e-ISBN 978-3-319-23883-8

DOI 10.1007/978-3-319-23883-8

Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2015955739

© Springer International Publishing Switzerland 2015

Use R!

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made.

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

---

*To my most important social network—Sue, Alina, and Andrew*

---

# Preface

---

In early 2000, Stephen Hawking said that “... the next century will be the century of complexity.” If his prediction is true, the implication is that we will need new scientific theories, data collection methods, and analytic techniques that are appropriate for the study of complex systems and behavior. Network science is one such approach that views the world through a network lens, where physical and social systems are made up of heterogeneous actors who are connected to one another through different types of relational ties. Network analysis is the set of analytic tools used to study these types of systems. Over the past several decades network analysis has become an increasingly important part of the analytic toolbox for social, health, and physical scientists.

Until recently, network analysis required specialized software, both for network data management and analyses. However, starting around 2000, network analytic tools became available in the R statistical programming environment. This not only made network analytic techniques more visible to the broader statistical community but also provided the breadth and power of R’s data management, graphic visualization, and general statistical modeling capabilities to the network analyst community.

As the title suggests, this book is a user’s guide to network analysis in R. It provides a practical hands-on tour of the major network analytic tasks that can currently be done in R. The book concentrates on four primary tasks that a network analyst typically concerns herself with: network data management, network visualization, network description, and network modeling. The book includes all the R code that is used in the network analysis examples. It also comes with a set of network datasets that are used throughout the book. (See Chap. 1 for more details on the structure of the book, as well as instructions on how to obtain the network data.) The book is written for anybody who has an interest in doing network analysis in R. It can be used as a secondary text in a network science or analysis class or can simply serve as a reference for network techniques in R.

This book would not exist without the help, support, guidance, and mentoring I have received over the last 30 years from my own personal and professional social networks. In the mid-1980s I took a graduate network analysis class from Stan Wasserman at the University of Illinois in Champaign. I remember being excited about this new way to analyze data, but thought that I was not likely to ever use it in my career. However, my colleagues in psychology and public health encouraged me in my early work exploring how network analysis could answer important research and evaluation questions. These include Julian Rappaport, Ed Seidman, Bruce Rapkin, Kurt Ribisl, Sharon Homan, Ross Brownson, and Matt Kreuter. Whether they know it or not, I have been inspired and encouraged by an amazing group of network and systems scientists, including Tom Valente, Steve Borgatti, Martina Morris, Tom Snijders, Scott Leischow, Patty Mabry, Stephen Marcus, and Ross Hammond. My best network ideas have come from my friends and colleagues at the Center for Public Health Systems Science, particularly Bobbi Carothers, Amar Dhand, Chris Robichaux, and Nancy Mueller. I am especially grateful to the students in my network analysis classes and workshops over the years; they have not only improved this book, but they have improved my thinking about network analysis. A very special thank you to Jenine Harris. Jenine was my first doctoral student, now I am inspired by the rigor and elegance of her own work in network science. I would also like to thank the Centers for Disease Control and Prevention, the National Institutes of Health, and the Missouri Foundation for Health for providing research and evaluation support that allowed me to develop and refine my approach to network analysis. Finally, my deepest thanks go to my family. They gave me specific suggestions about the content, provided me space and time to work hard on this book (including a crucial Father’s Day gift), and cheered me on when I most needed it. Thank you, Sue, Ali, and Andrew.

**Douglas A. Lul**  
**St. Louis, MO, US**  
**July, 201**

---

---

# Contents

---

## **1 Introducing Network Analysis in R**

### **1.1 What Are Networks?**

### **1.2 What Is Network Analysis?**

### **1.3 Five Good Reasons to Do Network Analysis in R**

#### **1.3.1 Scope of R**

#### **1.3.2 Free and Open Nature of R**

#### **1.3.3 Data and Project Management Capabilities of R**

#### **1.3.4 Breadth of Network Packages in R**

#### **1.3.5 Strength of Network Modeling in R**

### **1.4 Scope of Book and Resources**

#### **1.4.1 Scope**

#### **1.4.2 Book Roadmap**

#### **1.4.3 Resources**

## **Part I Network Analysis Fundamentals**

## **2 The Network Analysis ‘Five-Number Summary’**

### **2.1 Network Analysis in R: Where to Start**

### **2.2 Preparation**

### **2.3 Simple Visualization**

### **2.4 Basic Description**

#### **2.4.1 Size**

#### **2.4.2 Density**

#### **2.4.3 Components**

#### **2.4.4 Diameter**

### **2.5 Clustering Coefficient**



## **3 Network Data Management in R**

---

### **3.1 Network Data Concepts**

#### **3.1.1 Network Data Structures**

#### **3.1.2 Information Stored in Network Objects**

### **3.2 Creating and Managing Network Objects in R**

#### **3.2.1 Creating a Network Object in `statnet`**

#### **3.2.2 Managing Node and Tie Attributes**

#### **3.2.3 Creating a Network Object in `igraph`**

#### **3.2.4 Going Back and Forth Between `statnet` and `igraph`**

### **3.3 Importing Network Data**

### **3.4 Common Network Data Tasks**

#### **3.4.1 Filtering Networks Based on Vertex or Edge Attribute Values**

#### **3.4.2 Transforming a Directed Network to a Non-directed Network**

## **Part II Visualization**

## **4 Basic Network Plotting and Layout**

### **4.1 The Challenge of Network Visualization**

### **4.2 The Aesthetics of Network Layouts**

### **4.3 Basic Plotting Algorithms and Methods**

#### **4.3.1 Finer Control Over Network Layout**

#### **4.3.2 Network Graph Layouts Using `igraph`**

## **5 Effective Network Graphic Design**

### **5.1 Basic Principles**

### **5.2 Design Elements**

#### **5.2.1 Node Color**

#### **5.2.2 Node Shape**

### **5.2.3 Node Size**

---

### **5.2.4 Node Label**

### **5.2.5 Edge Width**

### **5.2.6 Edge Color**

### **5.2.7 Edge Type**

### **5.2.8 Legends**

## **6 Advanced Network Graphics**

### **6.1 Interactive Network Graphics**

#### **6.1.1 Simple Interactive Networks in igraph**

#### **6.1.2 Publishing Web-Based Interactive Network Diagrams**

#### **6.1.3 Statnet Web: Interactive statnet with shiny**

### **6.2 Specialized Network Diagrams**

#### **6.2.1 Arc Diagrams**

#### **6.2.2 Chord Diagrams**

#### **6.2.3 Heatmaps for Network Data**

### **6.3 Creating Network Diagrams with Other R Packages**

#### **6.3.1 Network Diagrams with ggp1ot2**

## **Part III Description and Analysis**

## **7 Actor Prominence**

### **7.1 Introduction**

### **7.2 Centrality: Prominence for Undirected Networks**

#### **7.2.1 Three Common Measures of Centrality**

#### **7.2.2 Centrality Measures in R**

#### **7.2.3 Centralization: Network Level Indices of Centrality**

#### **7.2.4 Reporting Centrality**

## 7.3 Cutpoints and Bridges

---

## 8 Subgroups

### 8.1 Introduction

### 8.2 Social Cohesion

#### 8.2.1 Cliques

#### 8.2.2 k-Cores

### 8.3 Community Detection

#### 8.3.1 Modularity

#### 8.3.2 Community Detection Algorithms

## 9 Affiliation Networks

### 9.1 Defining Affiliation Networks

#### 9.1.1 Affiliations as 2-Mode Networks

#### 9.1.2 Bipartite Graphs

### 9.2 Affiliation Network Basics

#### 9.2.1 Creating Affiliation Networks from Incidence Matrices

#### 9.2.2 Creating Affiliation Networks from Edge Lists

#### 9.2.3 Plotting Affiliation Networks

#### 9.2.4 Projections

### 9.3 Example: Hollywood Actors as an Affiliation Network

#### 9.3.1 Analysis of Entire Hollywood Affiliation Network

#### 9.3.2 Analysis of the Actor and Movie Projections

## Part IV Modeling

## 10 Random Network Models

### 10.1 The Role of Network Models

### 10.2 Models of Network Structure and Formation

## **10.2.1 Erdős-Rényi Random Graph Model**

---

## **10.2.2 Small-World Model**

## **10.2.3 Scale-Free Models**

## **10.3 Comparing Random Models to Empirical Networks**

# **11 Statistical Network Models**

## **11.1 Introduction**

## **11.2 Building Exponential Random Graph Models**

### **11.2.1 Building a Null Model**

### **11.2.2 Including Node Attributes**

### **11.2.3 Including Dyadic Predictors**

### **11.2.4 Including Relational Terms (Network Predictors)**

### **11.2.5 Including Local Structural Predictors (Dyad Dependency)**

## **11.3 Examining Exponential Random Graph Models**

### **11.3.1 Model Interpretation**

### **11.3.2 Model Fit**

### **11.3.3 Model Diagnostics**

### **11.3.4 Simulating Networks Based on Fit Model**

# **12 Dynamic Network Models**

## **12.1 Introduction**

### **12.1.1 Dynamic Networks**

### **12.1.2 RSiena**

## **12.2 Data Preparation**

## **12.3 Model Specification and Estimation**

### **12.3.1 Specification of Model Effects**

### **12.3.2 Model Estimation**

## **12.4 Model Exploration**

---

### **12.4.1 Model Interpretation**

### **12.4.2 Goodness-of-Fit**

### **12.4.3 Model Simulations**

## **13 Simulations**

### **13.1 Simulations of Network Dynamics**

#### **13.1.1 Simulating Social Selection**

#### **13.1.2 Simulating Social Influence**

## **References**

---

# 1. Introducing Network Analysis in R

Douglas A. Luke<sup>1</sup> 

(1) Center for Public Health Systems Science George Warren Brown School of Social Work,  
Washington University, St. Louis, MO, USA

*Begin at the beginning, the King said, very gravely, and go on till you come to the end: then stop*  
Lewis Carroll, *Alice in Wonderland*

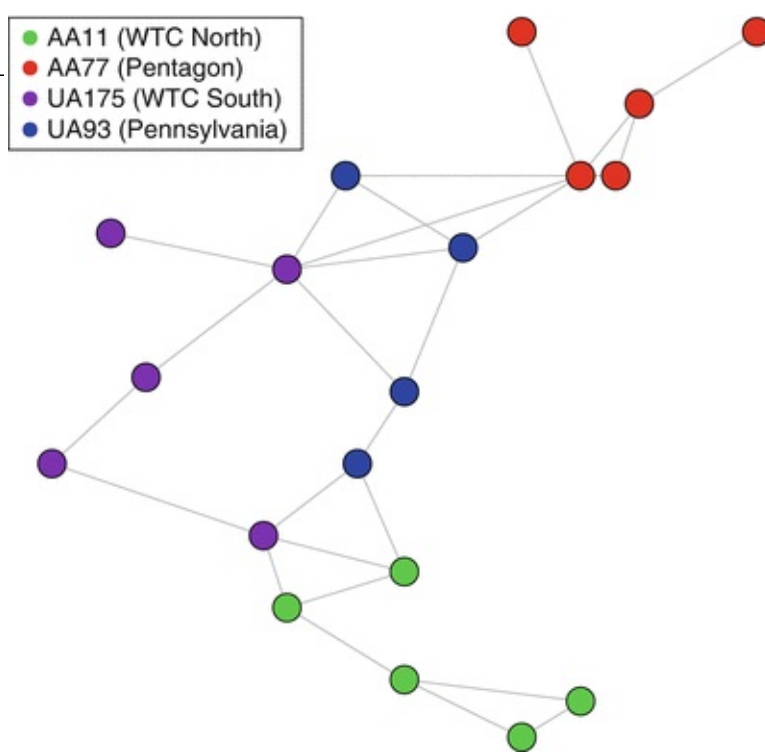
---

## 1.1 What Are Networks?

This book is a user's guide for conducting network analysis in the R statistical programming language. Networks are all around us. Humans naturally organize themselves in networked systems. Our families and friends form personal social networks around each of us. Neighborhoods and communities organize themselves in networked coalitions to advocate for change. Businesses work with (and against) each other in complex, interlocking networks of trade and financial partnerships. Public health is advanced through partnerships and coalitions of governmental and NGO organizations (Luke and Harris 2007). Nations are connected to one another through systems of migration, trade, and treaty obligations.

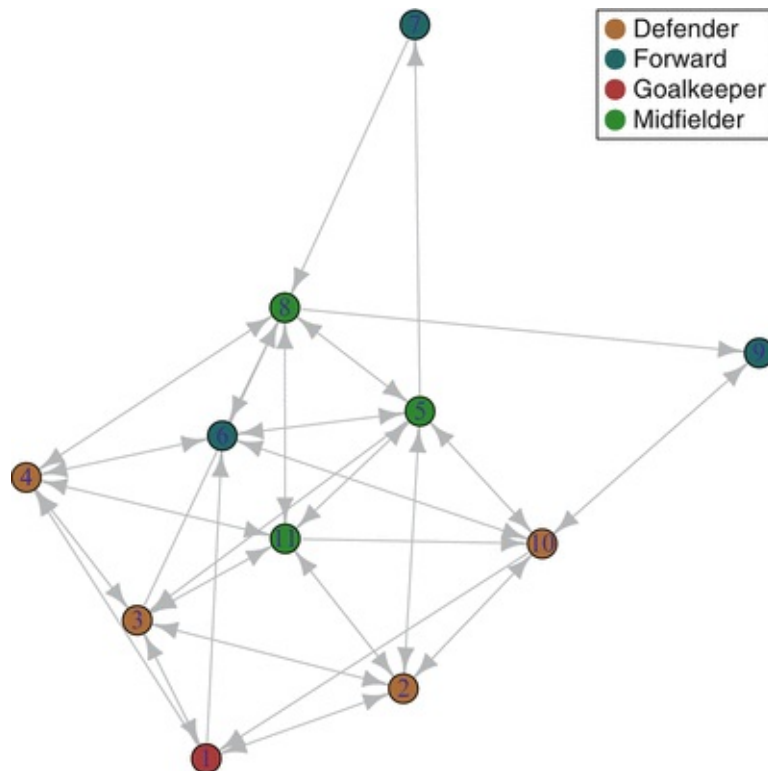
Moreover, non-human networks exist almost anywhere you look. Our genes and proteins interact with one another through complex biological networks. The human brain is now viewed as a complex network, or 'connectome' (Sporns 2012). Similarly, human diseases and their underlying genetic roots are connected as a 'diseasome' (Barabási 2007). Animal species interact in many complex ways, one of which is a networked food-web that describes interactions in 'who-eats-whom' relationships. Information itself is networked. Our legal system is built on an interconnecting network of prior legal decisions and precedents. Social and scientific progress is driven by a diffusion of innovation processes by which information is disseminated across connected social systems, whether they are Iowa corn farmers (Rogers 2003) or public health scientists (Harris and Luke 2009). It appears that one of the ways the universe is organized is with networks.

So what is a network? Figures 1.1 and 1.2 present two examples of important and interesting social networks. Figure 1.1 presents the contact network of the 19 9–11 hijackers, based on the work of Valdis Krebs (2002). Every social network is made up of a set of actors (also called nodes) that are connected to one another via some type of social relationship (also called a tie). In the figure, nodes are the circles and the ties are the lines connecting some of the nodes. The network shows us that the hijackers had some contact with one another before September 11th, but the network is not very densely connected and there appears to be no prominent network member who is connected to all or even most of the other hijackers.



**Fig. 1.1** Network of 9–11 hijackers

The second example in Fig. 1.2 is from a very different sort of social network. Here the nodes are members of the 2010 Netherlands FIFA World Cup team, who went on to lose in the final to Spain. The ties represent passes between the different players during the World Cup matches. The arrows show the directional pattern of the passes. We can see that the goalkeeper passed primarily to the defenders, and the forwards received passes primarily from the midfielders (except for #6, who appears to have a different passing pattern than the other two forwards).



**Fig. 1.2** Network of Netherlands 2010 World Cup soccer team

These two examples may appear to have little in common. However, they both share a

fundamental characteristic common to all social networks. The social patterns that are displayed in the network figures are not random. They reflect underlying social processes that can be explored using network science theories and methods. The terrorist network has no prominent leader and is not tightly interconnected because it makes the network harder to detect or disrupt. The pattern of passing ties in the soccer network reflects the assigned positions of the players, the rules of the game, and the strategies of the coach. The network analysis does not ‘know’ about any of those rules or strategies. Yet, network analysis can be used to reveal these patterns that reflect the underlying rules and regularities.

---

## 1.2 What Is Network Analysis?

Network science is a broad approach to research and scholarship that uses a relational lens to study and understand biological, physical, social, and informational systems. The primary tool for network scientists is network analysis, which is a set of methods that are used to (1) visualize networks, (2) describe specific characteristics of overall network structure as well as details about the individual nodes, ties, and subgroups within the networks, and (3) build mathematical and statistical models of network structures and dynamics. Because the core question of network science is about relationships, most of the methods used in network analysis are quite distinct from the more traditional statistical tools used by social and health scientists.

Network analysis as a distinct scientific enterprise with its own theories and methods grew out of developments in many other disciplines, particularly graph theory and topology in mathematics, the study of kinship systems in anthropology, and social groups and process from sociology and psychology. Although network analysis was not invented by one person at a specific place and time, the initial development of what we now recognize as modern network analysis can be traced back to the work of Jacob Moreno in the 1930s. He defined the study of social relations as sociometry, and founded the journal *Sociometry* that would publish the early studies in this area. He also invented the sociogram, which was a visual way to display network structures. The first published sociogram appeared in the *New York Times* in 1933, and it was a network diagram of the friendship ties among a 4th grade class. (These data are available as part of the network dataset package that accompanies this book, see Sect. 1.4.3 below.)

The theories and methods of network analysis were developed throughout the rest of the twentieth century, with important contributions from sociology, psychology, political science, business, public health, and computer science. Network science as an empirical practice was propelled by the development of a number of network specific software tools and packages, including UCINET, STRUCTURE, Negopy, and Pajek. The interest in network science has exploded in the last 20–30 years, driven by at least three different factors. First, mathematicians, physicists, and other researchers developed a number of influential theories of network structure and formation that brought attention and energy to network science (see Chap. 10 for some discussion of these theories). Second, advances in computational power and speed allowed network methods to be applied to large and very large networks, such as the internet, the population of the planet, or the human brain. Finally, advances in statistical network theory allowed analysts for the first time to move beyond simple network description to be able to build and test statistical models of network structures and processes (see Chaps. 11 and 12).

---

## 1.3 Five Good Reasons to Do Network Analysis in R

As the title suggests, this book is designed as a general guide for how to do network analysis in the R



statistical language and environment. Why is R an ideal platform for developing and conducting network analyses? There are at least five good reasons.

---

### 1.3.1 Scope of R

The R statistical programming language and environment comprise a vast integrated system of thousands of packages and functions that allow it to handle innumerable data management, analysis, or visualization tasks. The R system includes a number of packages that are designed to accomplish specific network analytic tasks. However, by performing these network tasks within the R environment, the analyst can take advantage of any of the other capabilities of R. Most other network analysis programs (e.g., Pajek, UCInet, Gephi) are stand-alone packages, and thus do not have the advantages of working within an integrated statistical programming environment.

### 1.3.2 Free and Open Nature of R

One of the important reasons for R's popularity and success is its free and open nature. This is formally ensured via the GNU General Public License (GPL) that R-code is released under. More informally, there is a vast R user and developer community which is continually working to enhance and improve R base code and the thousands of R packages that can be freely accessed. The social network capabilities of R described in this book have, in fact, been developed by the R user community. This open nature of R facilitates faster (and arguably, cleaner and more powerful) development and dissemination of new statistical and data analytic techniques, such as these network analytic tools.

### 1.3.3 Data and Project Management Capabilities of R

Although there are many good network analysis programs available which can handle a wide variety of network descriptive statistics and visualization tasks, no other network package has the same power to handle often complex data and project management tasks for larger-scale network analyses compared to R. First, as suggested above, network analysis in R can take advantage of the powerful data management, cleaning, import and export capabilities of base R. As described in Chap. 3, network analysis often starts by importing and transforming data from other sources into a form that can be analyzed by network tools. All network packages have some data management capabilities, but no other program can match R's breadth and depth.

Second, when conducting sophisticated scientific or commercial network analyses, it is important to have the right project management tools to facilitate code storage and retrieval, managing analysis outputs such as statistical results and information graphics, and producing reports for internal and external audiences. Traditional statistical analysis platforms such as SAS and SPSS have these sorts of tools, but most network programs do not. By pairing R up with an integrated development environment (IDE) such as RStudio (<http://rstudio.org/>) and taking advantage of packages such as `knitr` and `shiny`, the user has the ability to manage any type of complex network project. In fact, the development and availability of these tools has been one of the driving forces of the *reproducible research* movement (Gentleman and Lang 2007), which emphasizes the importance of combining data, code, results, and documentation in permanent and shareable forms. As one example of the power of the reproducible research tools accessible in R is this book, which was created entirely in RStudio.

### 1.3.4 Breadth of Network Packages in R

The primary reason R is ideal for network analysis is the breadth of packages that are currently available to manage network data and conduct network visualization, network description, and network modeling. There are dozens of network-related packages, and more are being created all the time. R network data can be managed and stored in R native objects by the `network` and `igraph` packages, and the data can be exchanged between formats with the `intergraph` package. Basic network analysis and visualization can be handled with the `sna` package contained within the much broader `statnet` suite of network packages, as well as within `igraph`. More sophisticated network modeling can be handled by `ergm` and its associated libraries, and dynamic actor-based network models are produced by `RSiena`. Freestanding network analysis programs have many strengths (e.g., the visualization capabilities of Gephi), but no single program matches the combined power of the social network analysis packages contained in R.

### 1.3.5 Strength of Network Modeling in R

Finally, the particular network modeling strengths of R should be mentioned. R is the only generally available software package that includes comprehensive facilities to do stochastic network modeling (e.g., exponential random graph models), dynamic actor-based network models that allow study of how networks change over time, and other network simulation procedures.

---

## 1.4 Scope of Book and Resources

---

### 1.4.1 Scope

As the title suggests, the goal of this book is to provide a hands-on, practical guide to doing network analysis in the R statistical programming environment. It is hands-on in the sense that the book provides guidance primarily in the form of short network analysis code snippets applied to realistic network data. The results of the analyses follow immediately. All the code and data are available to the reader, so that it is easy to replicate what is shown in the book, experiment with your own data or code extensions, and thus facilitate learning.

The practical goal of the book is to demonstrate network analytic techniques in R that will be useful for a wide variety of data analysis and research goals. This includes data management, network visualization, computation of relevant network descriptive statistics, and performing mathematical, statistical, and dynamic modeling of networks. The intended audiences include students, analysts and researchers across a wide variety of disciplines, particularly the social, health, business, and engineering domains.

It is also useful to state what this book is not designed to do. First, it does not provide an in-depth treatment of network science theories or history. There are many good books, papers, training courses and online resources available that cover this material. For good general overviews, the classic text by Wasserman and Faust (1994) is still relevant, and John Scott provides a good, more current treatment (2012). For more in-depth treatment of network science and statistical theory, see Newman (2010) or Kolaczyk (2009). Finally, two edited volumes that have good coverage of the recent history of network science as well as well-executed examples of empirical network research are Newman et al. (2006) and Scott and Carrington (2011).

Second, this book is not in any way an adequate introduction to R programming and statistical analysis. Although every attempt is made to make each code example clear and succinct, a novice R user will find some of the techniques and code syntax hard to follow. In particular, understanding R's capabilities for data management, graphics, and the object-oriented approach to statistical modeling will be very helpful for getting the most out of this user-guide.

Thus, the book is designed for the interested student, analyst, or researcher who is familiar with R and has some understanding of network science theories and methods. It could serve as a secondary text for a graduate level class in network analysis. It also could be useful as a primer for an experienced R analyst who wants to incorporate network analysis into her programming and analytic toolbox.

### 1.4.2 Book Roadmap

The book is organized into four main sections, which correspond to the four fundamental tasks that network analysts will spend most of their time on: data management, network visualization, network description, and network modeling. The first section has two chapters that cover both a simple introduction to basic network techniques, then a more in-depth presentation of data management issues in network analysis. The three chapters in the Visualization section cover basic network graphics layout, network graphic design suggestions, and some discussion of advanced graphics topics and techniques. The Description and Analysis section has three chapters that cover the most widely used techniques for describing important network characteristics, including actor prominence, network subgroups and communities, and handling affiliation networks. The final section, Modeling, includes four chapters that present advanced techniques for mathematical modeling, statistical modeling, modeling of dynamic networks, and network simulations. Table 1.1 presents this roadmap

**Table 1.1** User's Guide roadmap

Chapter	Packages	Datasets
Introduction		FIFA_Nether, Krebs
5 number summary	statnet, sna	Moreno
Network data	statnet, network, igraph	DHHS, ICTS
Basic visualization	statnet, sna	Moreno, Bali
Graphic design	statnet, sna, igraph	Bali
Advanced graphics	arcdiagram, circlize, visNetwork, networkD3	Simpsons, Bali
Prominence	statnet, sna	DHHS, Bali
Subgroups	igraph	DHHS, Moreno, Bali
Affiliation networks	igraph	hwd
Mathematical models	igraph	lhds
Stochastic models	ergm	TCnetworks
Dynamic models	RSiena	Coevolve
Simulations	igraph	

### 1.4.3 Resources

The most important resource for this user guide is a collection of network datasets that have been curated and made available to the readers of this book. Over a dozen network datasets are included in the form of an R package called `UserNetR`. These datasets are used throughout the book to support the coding and analysis examples. The network data included in the `UserNetR` package mostly come from published network studies, while a few are created to help illustrate particular analytic options.

Table 1.1 lists the names of the datasets that are featured in each chapter.

The `UserNetR` package is maintained on GitHub, and must be downloaded and installed to make the network data available. This can be done using the following code. (The `devtools` package must also be installed if it is not on your system.)

```
library(devtools)
install_github("DougLuke/UserNetR")
```

Once this is done, the package must be loaded to make the various datafiles available. This can be done with the `library()` function, just like for any R package. This command will not always be explicitly shown throughout the book, so make sure to load the package prior to executing any of the included R code.

```
library(UserNetR)
```

Finally, the documentation for the `UserNetR` package can be viewed through the R help system.

```
help(package='UserNetR')
```

## References

Barabási AL (2007) Network medicine from obesity to the diseasome. *N Engl J Med*. <http://www.nejm.org/doi/full/10.1056/nejme078114>

Barabási AL, Albert R (1999) Emergence of scaling in random networks. *Science*. <http://www.sciencemag.org/content/286/5439/50short>

Broder A, Kumar R, Maghoul F (2000) Graph structure in the web. *Computer Netw*. <http://www.sciencedirect.com/science/article/pii/S1389128600000839>

- Butts CT (2008) network: a package for managing relational data in R. J Stat Softw. <http://cran.repo.bppt.go.id/web/packages/network/vignettes/networkVignette.pdf>
- Freeman LC (2004) The development of social network analysis: a study in the sociology of science. Empirical Press, p 205. ISBN:1594577145. <https://books.google.com/books?id=VcxqQgAACAAJ&pgis=1>
- Fruchterman TMJ, Reingold EM (1991) Graph drawing by force-directed placement. Softw Pract Exp. [ftp://132.180.22.143/axel/papers/reingold:graph\\_drawing\\_by\\_force\\_directed\\_placement.pdf](ftp://132.180.22.143/axel/papers/reingold:graph_drawing_by_force_directed_placement.pdf)
- Galaskiewicz J (1985) The influence of corporate power, social status, and market position on corporate interlocks in a regional network. Social Forces. <http://sf.oxfordjournals.org/content/64/2/403.abstract>
- Gentleman R, Lang DT (2007) Statistical analyses and reproducible research. J Comput Graph Stat. <http://www.jstor.org/stable/27594227>
- Goodreau SM (2007) Advances in exponential random graph ( $p^*$ ) models applied to a large social network. Soc Netw 29(2):231–248. ISSN:03788733. doi:10.1016/j.socnet.2006.08.001
- Granovetter MS (1973) The strength of weak ties. Am J Sociol. <http://www.jstor.org/stable/2776392>
- Handcock MS et al (2008) statnet: software tools for the representation, visualization, analysis and simulation of network data. J Stat Softw 24(1):1–11. ISSN:1548–7660
- Harris JK (2013) An introduction to exponential random graph modeling. SAGE, p 136. ISBN:148332205X. <https://books.google.com/books?hl=en&lr=&id=lkYXBAAAQBAJ&pgis=1>
- Harris JK, Luke DA (2009) Forty years of secondhand smoke research: the gap between discovery and delivery. Am J Prev Med. <http://www.sciencedirect.com/science/article/pii/S0749379709001548>
- Holland PW, Leinhardt S (1978) An omnibus test for social structure using triads. Sociol Methods Res. <http://smr.sagepub.com/content/7/2/227.short>
- Hunter DR et al (2008) ergm: a package to fit, simulate and diagnose exponential-family models for networks. J Stat Softw 24(3):nihpa54860. ISSN:1548-7660. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2743438&tool=pmcentrez&rendertype=abstract>
- Knoke D, Burt RS (1983) Prominence. Appl Netw Anal. [https://scholar.google.com/scholar?q=knoke+burt+prominence&btnG=&hl=en&as\\_sdt=0%2C26#0](https://scholar.google.com/scholar?q=knoke+burt+prominence&btnG=&hl=en&as_sdt=0%2C26#0)
- Kolaczyk ED (2009) Statistical analysis of network data: methods and models. Springer, p 398. ISBN:0387881468. <https://books.google.com/books?id=Q-GNLsq7QwC&pgis=1>
- Krebs VE (2002) Uncloaking terrorist networks. <http://firstmonday.org/ojs/index.php/fm/article/view/941/863>
- Leischow SJ, Luke DA, et al. (2010) Mapping US government tobacco control leadership: networked for success? Nicotine Tob Res. <http://ntr.oxfordjournals.org/content/12/9/888.short>
- Liljeros F, Edling CR, Amaral LAN (2001) The web of human sexual contacts. Nature. <http://www.nature.com/nature/journal/v411/n6840/full/411907a0.html>
- Luke DA, Harris JK (2007) Network analysis in public health: history, methods, and applications. Annu Rev Public Health 28:69–90. ISSN:0163-7525. doi:10.1146/annurev.publhealth.28.021406.144132
- Luke DA, Stamatakis KA (2012) Systems science methods in public health: dynamics, networks, and agents. Annu Rev Public Health. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3644212/>
- Luke DA, Wald LM (2013) Network influences on dissemination of evidence-based guidelines in state tobacco control programs. Health Educ Behav. [http://heb.sagepub.com/content/40/1\\_suppl/33S.short](http://heb.sagepub.com/content/40/1_suppl/33S.short)
- Luke DA, Harris JK, Shelton S (2010) Systems analysis of collaboration in 5 national tobacco control networks. Am J Public Health. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2882404/>
- McPherson M, Smith-Lovin L, Cook JM (2001) Birds of a feather: homophily in social networks. Annu Rev Sociol. <http://www.jstor.org/stable/265682>

- Morris M, Hancock MS, Hunter DR (2008) Specification of exponential-family random graph models: terms and computational aspects. *J Stat Softw* 24(4):1548–7660. ISSN:1548–7660. <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2481518&tool=pmcentrez&rendertype=abstract>
- Murrell P (2005) *R graphics*. Taylor & Francis, p 328. ISBN:158488486X. <https://books.google.com/books?id=fUUVngEACAAJ&pgis=1>
- Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA*. <http://www.pnas.org/content/103/23/8577.short>
- Newman M (2010) *Networks: an introduction*. Oxford University Press, Oxford, p 784. ISBN:0191500704. <https://books.google.com/books?id=LrFaU4XCsUoC&pgis=1>
- Newman MEJ, Girvan M (2004) Finding and evaluating community structure in networks. *Phys Rev E Stat Nonlinear Soft Matter Phys* 69(2):1–15. ISSN:1063651X. doi:10.1103/PhysRevE.69.026113. arXiv: 0308217 [cond-mat]
- Newman M, Barabási A-L, Watts DJ (2006) *The structure and dynamics of networks*. Princeton University Press, p 582. ISBN:0691113572. <https://books.google.com/books?id=0FNQ1LYKTMwC&pgis=1>
- Rogers EM (2003) *Diffusion of innovations*, 5th edn. Simon and Schuster, p 576. ISBN:0743258231. <https://books.google.com/books?id=9U1K5LjUOwEC&pgis=1>
- Scott J (2012) *Social network analysis (3rd Ed.)* SAGE Publications.
- Scott J, Carrington PJ (2011) *The SAGE handbook of social network analysis*. SAGE Publications.
- Snijders TAB, Pattison PE (2006) New specifications for exponential random graph models. *Sociol Methodol*. url: <http://smx.sagepub.com/content/36/1/99.short>.
- Snijders TAB, Van de Bunt GG, Steglich CEG (2010) Introduction to stochastic actor-based models for network dynamics. *Soc Netw*. <http://www.sciencedirect.com/science/article/pii/S0378873309000069>
- de Solla Price DJ (1976) A general theory of bibliometric and other cumulative advantage process. *J Am Soc Info Sci*. [https://scholar.google.com/scholar?q=price+1976+bibliometric&btnG=&hl=en&as\\_sdt=0%2C26#4](https://scholar.google.com/scholar?q=price+1976+bibliometric&btnG=&hl=en&as_sdt=0%2C26#4)
- Sporns O (2012) *Discovering the human connectome*. MIT, p 232. ISBN:026017903. <https://books.google.com/books?id=uoNf2x0J8LMC&pgis=1>
- Tufte ER (1990) *Envisioning information*, vol 914. Graphics Press, p 126. <https://books.google.com/books?id=1uloAAAIAAJ&pgis=1>
- Tufte ER (2001) *The visual display of quantitative information*. Graphics Press, p 197. ISBN:0961392142. <https://books.google.com/books?id=GTd5oQEACAAJ&pgis=1>
- Tukey JW (1977) *Exploratory data analysis*. Addison-Wesley, p 688. ISBN:0201076160. <https://books.google.com/books?id=UT9dAAAIAAJ&pgis=1>
- Valente TW (2010) *Social networks and health: models, methods, and applications*. Oxford University Press, Oxford, p 296. ISBN:0199719721. <https://books.google.com/books?id=xnMzd1-7iGgC&pgis=1>
- Wasserman S, Faust K (1994) *Social network analysis: methods and applications*, vol 25. Cambridge University Press, p 825. ISBN:0521387078. <https://books.google.com/books?id=CAm2DpIqRUIC&pgis=1>
- Watts DJ, Strogatz SH (1998) Collective dynamics of 'small-world' networks. *Nature*. <http://www.nature.com/nature/journal/v393/n6684/abs/393440a0.html>

Part I

---

# Network Analysis Fundamentals

## 2. The Network Analysis ‘Five-Number Summary’

Douglas A. Luke<sup>1</sup>✉

- (1) Center for Public Health Systems Science George Warren Brown School of Social Work, Washington University, St. Louis, MO, USA

*There is nothing like looking, if you want to find something. You certainly usually find something, you look, but it is not always quite the something you were after. ( J.R.R. Tolkien – The Hobbit )*

---

### 2.1 Network Analysis in R: Where to Start

How should you start when you want to do a network analysis in R? The answer to this question rests of course on the analytic questions you hope to answer, the state of the network data that you have available, and the intended audience(s) for the results of this work. The good news about performing network analysis in R is that, as will be seen in subsequent chapters, R provides a multitude of available network analysis options. However, it can be daunting to know exactly where to start.

In 1977, John Tukey introduced the five-number summary as a simple and quick way to summarize the most important characteristics of a univariate distribution. Networks are more complicated than single variables, but it is also possible to explore a set of important characteristics a social network using a small number of procedures in R.

In this chapter, we will focus on two initial steps that are almost always useful for beginning a network analysis: simple visualization, and basic description using a ‘five-number summary.’ This chapter also serves as a gentle introduction to basic network analysis in R, and demonstrates how quickly this can be done.

---

### 2.2 Preparation

Similar to most types of statistical analysis using R, the first steps are to load appropriate packages (installing them first if necessary), and then making data available for the analyses. The `statnet` suite of network analysis packages will be used here for the analyses. The data used in this chapter (and throughout the rest of the book) are from the `UserNetR` package that accompanies the book. The specific dataset used here is called `Moreno`, and contains a friendship network of fourth grade students first collected by Jacob Moreno in the 1930s.

```
library(statnet)
library(UserNetR)
data(Moreno)
```

---

### 2.3 Simple Visualization



- [click Life Is Short \(No Pun Intended\)](#)
- [click Life in the Balance: A Physician's Memoir of Life, Love, and Loss with Parkinson's Disease and Dementia pdf, azw \(kindle\), epub](#)
- [click Best Little Stories from the Civil War: More Than 100 True Stories](#)
- [read Chez Panisse Cafe Cookbook book](#)
- [My Pink Road to Russia: Tales of Amazons, Peasants, and Queers online](#)
  
- <http://aseasonedman.com/ebooks/Life-Is-Short--No-Pun-Intended-.pdf>
- <http://berttrotman.com/library/Life-in-the-Balance--A-Physician-s-Memoir-of-Life--Love--and-Loss-with-Parkinson-s-Disease-and-Dementia.pdf>
- <http://www.celebritychat.in/?ebooks/Best-Little-Stories-from-the-Civil-War--More-Than-100-True-Stories.pdf>
- <http://flog.co.id/library/Hunt-for-Jade-Dragon--Michael-Vey--Book-4-.pdf>
- <http://dadhoc.com/lib/Thinking--Recording--and-Writing-History-in-the-Ancient-World.pdf>