Network Analysis

The size and availability of network information has exploded over the last decade. Social scientists now share the stage of network analysis with computer scientists, physicists, and statisticians. While a number of introductions to network analysis are now available, most focus on theory, methods, or application alone. This book integrates all three. *Network Analysis* is an introduction to both the why and how of social network analysis (SNA). It presents a broad theoretical overview rooted in social scientific approaches and guides users in how network analysis can answer core theoretical questions. It provides a comprehensive overview of descriptive and analytical approaches, including practical tutorials in R with sample data sets. Using an integrated approach, this book aims to quickly bring novice network researchers up to speed while avoiding common programming and analysis mistakes so that they might gain insight into the fundamental theories, key concepts, and methodological application of SNA.

CRAIG M. RAWLINGS is Associate Professor of Sociology at Duke University, where he is affiliated with the Duke Network Analysis Center. His work focuses on the connections between social structures and culture, including belief systems, knowledge, meaning-making processes, and attitude change. His publications have appeared in the *American Journal of Sociology, American Sociological Review, Social Forces, Sociological Science*, and *Poetics*.

JEFFREY A. SMITH is Senior Policy Analyst in Mental Health and Addictions at the Nova Scotia Health Authority. He has done methodological work on network sampling and missing data, as well as more substantive work on network processes, drug use, and health outcomes. His work has been published in the *American Sociological Review*, *Sociological Methodology*, *Social Networks*, and other venues.

JAMES MOODY is Professor of Sociology at Duke University and focuses on the network foundations of social cohesion and diffusion, using network analysis to help understand topics including racial segregation, disease spread, and the development of scientific disciplines. He has won the Freeman Award for contributions to network analysis and a James S. McDonnel Foundation Complexity Scholars award.

> DANIEL A. MCFARLAND is Professor of Education and (by courtesy) Sociology and Organizational Behavior at Stanford University, where he founded Stanford's Center for Computational Social Science. His past work studied social network dynamics of communication, relationships, affiliations, and knowledge structures in educational contexts. His current work integrates social network analysis and natural language processing to study the development of scientific knowledge.

STRUCTURAL ANALYSIS IN THE SOCIAL SCIENCES

Edited by Mark Granovetter

The series Structural Analysis in the Social Sciences presents studies that analyze social behavior and institutions by reference to relations among such concrete social entities as persons, organizations, and nations. Relational analysis contrasts on the one hand with reductionist methodological individualism and on the other with macro-level determinism, whether based on technology, material conditions, economic conflict, adaptive evolution, or functional imperatives. In this more intellectually flexible, structural middle ground, analysts situate actors and their relations in a variety of contexts. Since the series began in 1987, its authors have variously focused on small groups, history, culture, politics, kinship, aesthetics, economics, and complex organizations, creatively theorizing how these shape and in turn are shaped by social relations. Their style and methods have ranged widely, from intense, long-term ethnographic observation to highly abstract mathematical models. Their disciplinary affiliations have included history, anthropology, sociology, political science, business, economics, mathematics, and computer science. Some have made explicit use of social network analysis, including many of the cutting-edge and standard works of that approach, whereas others have kept formal analysis in the background and used "networks" as a fruitful orienting metaphor. All have in common a sophisticated and revealing approach that forcefully illuminates our complex social world.

Recent Books in the Series

- Mario L. Small, Brea L. Perry, Bernice A. Pescosolido, and Edward B. Smith, Personal Networks: Classic Readings and New Directions in Egocentric Analysis
- David Knoke, Mario Diani, and Dimitris Christopolous, and James Holloway, Multimodal Political Networks
- Claire BIdart, Alain Degenne, and Michel Grossetti, Living in Networks: The Dynamics of Social Relations
- William Sims Bainbridge, The Social Structure of Online Communities
- Michael Kenney, The Islamic State in Britain: Radicalization and Resilience in an Activist Network
- Wouter De Nooy, Andrej Mrvar, and Vladimir Batagelj, *Exploratory Social Network Analysis with Pajek: Revised and Expanded Edition for Updated Software*

CAMBRIDGE

Tear Down

Cambridge University Press & Assessment 978-1-107-03778-6 — Network Analysis
Craig M. Rawlings , Jeffrey A. Smith , James Moody , Daniel A. McFarland
Frontmatter
More Information

Darius Mehri, Iran Auto Navid Hassanpour, Leading from the Periphery and Network Collective Action Cheol-Sung Lee, When Solidarity Works Benjamin Cornwell, Social Sequence Analysis Mariela Szwarcberg, Mobilizing Poor Voters Luke M. Gerdes, ed., Illuminating Dark Networks Silvia Domínguez and Betina Hollstein, eds., Mixed Methods in Studying Social Networks Dean Lusher, Johan Koskinen, and Garry Robins, eds., Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications Sean F. Everton, Disrupting Dark Networks Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj, Exploratory Social Network Analysis with Pajek Noah E. Friedkin and Eugene C. Johnsen, Social Influence Network Theory Zeev Maoz, The Networks of Nations: The Evolution and Structure of International Networks, 1816-2001 Martin Kilduff and David Krackhardt, Interpersonal Networks in Organizations Ari Adut, On Scandal: Moral Disturbances in Society, Politics, and Art Robert C. Feenstra and Gary G. Hamilton, Emergent Economies, Divergent Paths Eiko Ikegami, Bonds of Civility: Aesthetic Networks and the Political Origins of Japanese Culture Peter Carrington, John Scott, and Stanley Wasserman, Models and Methods in Social Network Analysis Patrick Doreian, Vladimir Batagelj, and Anujka Ferligoj, Generalized Blockmodeling James Lincoln and Michael Gerlach, Japan's Network Economy Robert Franzosi, From Words to Numbers Sean O'Riain, The Politics of High-Tech Growth Philippe Bourgois, In Search of Respect: Selling Crack in El Barrio (Second Edition) Isabella Alcañiz, Environmental and Nuclear Networks in the Global South

Sean F. Everton, Networks and Religion: Ties that Bind, Loose, Build-up and

Network Analysis

Integrating Social Network Theory, Method, and Application with R

CRAIG M. RAWLINGS *Duke University*

JEFFREY A. SMITH Nova Scotia Health Authority

JAMES MOODY Duke University

DANIEL A. MCFARLAND *Stanford University*





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/9781107037786

DOI: 10.1017/9781139794985

© Craig M. Rawlings, Jeffrey A. Smith, James Moody, and Daniel A. McFarland 2023

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press & Assessment.

First published 2023

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

Library of Congress Cataloging-in-Publication Data

Names: McFarland, Daniel A., 1971- author. | Smith, Jeffrey A., 1983- author.

Title: Network analysis : integrating social network theory, method, and application with R / Daniel A. McFarland, Stanford University, Jeffrey A. Smith, University of Nebraska, Lincoln, James Moody, Duke University, North Carolina, Craig M. Rawlings, Duke University, North Carolina.

Description: 1 Edition. | New York : Cambridge University Press, [2023] | Series: Sass structural analysis in the social sciences

Identifiers: LCCN 2023011794 (print) | LCCN 2023011795 (ebook) | ISBN 9781107037786 (hardback) | ISBN 9781107611900 (paperback) | ISBN 9781139794985 (epub)

Subjects: LCSH: Social sciences–Network analysis. | Social networks–Research–Methodology. | Application software. | Web applications.

Classification: LCC HM741 .M3864 2023 (print) | LCC HM741 (ebook) | DDC 302.3072-dc23/ eng/20230410

LC record available at https://lccn.loc.gov/2023011794

LC ebook record available at https://lccn.loc.gov/2023011795

ISBN 978-1-107-03778-6 Hardback ISBN 978-1-107-61190-0 Paperback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Contents

<i>page</i> ix
XV
xvii
1
19
45
67
88
117
143
161
190
216
246
269
301
340

vii

viii	Contents
15 Models for Social Influence	364
16 Conclusion: Network Analysis Tomorrow	390
References	421
Index	447

Figures

1.1	Types of systems	page 4
1.2	High school sexual relations network (Bearman, Moody, & Stovel 2004)	10
1.3	Structural forces in international relations	10
2.1	Schematic of rendering reality into knowledge	20
2.2	Networks from slices of interactions using classroom	
	observation data (Bender-deMoll & McFarland 2006)	22
2.3	Schematic of social structure (adapted from Hinde 1976)	25
2.4	Social structure constructed in notes passed between two his	gh
	school students (McFarland & Wolff 2022)	30
2.5	Hypothetical romantic network	36
2.6	Role structure of a Western family unit	37
2.7	Role structure of the Trobriand Island Kula Ring	
	exchange network	39
3.1	The Königsberg bridge problem and its graph representation	46
3.2	A basic network structure	48
3.3	Four types of networks	52
3.4	Kinship/business relations among Florentine families in	
	the Renaissance	53
3.5	Southern women and their social event attendance	55
3.6	Example graph and basic network definitions	57
3.7	Walks of length 3 from Tim	57
4.1	The difference between complete and partial network data	68
4.2	The consequences of missing network data	80
4.3	Simple imputation options for missing network data	82
5.1	Scatterplots of three bivariate distributions shown in	
	Table 5.1	89

ix

List of Figures х 5.2 "Circus Sideshow" by Georges Seurat, with detail 90 Girls' cabin signed network from Moreno (1934) 5.3 91 5.4 Three visualizations of Zachary's (1977) karate club data 94 (a-c)5.5 Visualization of a dyadic distance matrix of select US cities 96 5.6 98 Spatial representation of Moreno's (1934) data 5.7Hospital exchange network using ZIP codes for coordinates 99 5.8 Tree diagram 101 5.9 Family tree diagram (Kaplanis et al. 2018) 102 5.10 Girls' cabin network as positive, negative, and joint ties, 104 using positive ties for layout 5.11 Using colors and weights to modify the default image for publication display 106 Modified image of Moreno data 5.12107 5.13 Removing nodes to clarify relations 108 5.14 Suppressing nodes on a large, dense graph clarifies relations 109 (a and b) between authors and topics in large-scale science network 5.15 Contour sociogram of natural science disciplines 109 5.16 Contour sociogram of Scholars@Duke data 110 5.17 Alluvial flow diagram of small STD simulation data (Bender-deMoll 2016) 112 5.18 Dynamic senate co-voting blocks (Moody & Mucha 2013) 112 6.1 Ideal-typical self in Western society 120 6.2 Typified egocentric networks 128 (a and b) 6.3 A typical Facebook friend network 130 6.4 Racial homophily as a function of high school racial heterogeneity (Moody 2001) 134 6.5 Racial mixing matrix in one high school 135 6.6 The density of an egocentric network 136 6.7 Egocentric networks and structural holes 136 6.8 Redundancy and constraint in an ego's network structure 137 6.9 Five types of brokers (Gould & Fernandez 1989) 141 7.1 Three types of dyads in directed networks: mutual, asymmetric, and null 146 Example calculation of reciprocity 7.2 147 7.3 The four types of triads for undirected networks 149 7.4 The sixteen types of triads for directed networks 150 7.5 Triad census 151 The bank wiring room friendship network (Roethlisberger, 7.6 Dickson, & Wright 1947 [1939]) 151 7.7 Forty triadic motifs in networks with two types of symmetric 153 ties (shown in red and blue) 7.8 156 Fritz Heider's POX System

List of Figures		xi
8.1	Conceptualizing the social cohesiveness of networks	163
8.2	Network density versus connectivity	165
8.3	A simple ridge structure	166
8.4	Connectivity as robustness to node removal	169
8.5	Ideal-typical small-world network structure	171
8.6	Small network with many overlapping cliques	176
8.7	Network with two 3-cores	177
8.8	Comparing four approaches to deriving	
	cohesive subgroups	181
8.9	Resolution parameter sweep	182
8.10	Connectivity sets for Figure 8.3	185
9.1	The conundrum of network degree centrality	197
9.2	Poisson distribution of centrality in a random network	204
9.3	Degree centrality distribution, 10,000 node network with a	
	nearly scale-free degree distribution	206
9.4	Varieties of hierarchy	207
9.5	Two macrostructures with no violations of balance rules	209
9.6	Relaxing the A4 balance rule allows for multiple clusters	210
9.7	Relaxing the A4 and A3 rules (a) and relaxing the A4, A3,	
(a and b)		210
9.8	Idealized image matrices of hierarchical structures depicted	
	in Figure 9.4	213
10.1	Stereotypical school role relations	219
10.2	Roles derived from compound relations	220
10.3	Structural equivalence in a formal hierarchy	222
10.4	Automorphic equivalence	223
10.5	Regular equivalence	224
10.6	A typical hierarchical structure	225
10.7	Reduced macrostructure of hierarchy in Figure 10.6	227
10.8	Reduced macrostructure of hierarchy in Figure 10.6 based	227
10.0	on regular equivalence	227
10.9	Illustrating stacking multiple relations within a family	220
10.10	exchange network	228
10.10	Illustration of the CONCOR algorithm	230
10.11	Illustrating hierarchical cluster analysis	232
10.12	Triad-position census for deriving role equivalence	236
10.13	Role positions in a classroom friendship network	238
(a–d)	Commenting role structures in two high schools	240
10.14	Comparing role structures in two high schools	240
10.15	Generalized blockmodels with four or more positions	241
(a and b)	Two structures of overlapping social similar of	
11.1	Two structures of overlapping social circles of	247
(a and b)	four individuals	247

xii	List of Fig	gures
11.2	Two-mode network of faculty and departments via	
	joint appointments	251
11.3	One-mode projections of southern women's affiliations	252
(a and b)		253
11.4	Academic careers in high school math course-taking (McFarland 2006)	260
11.5	Correspondence analysis of southern women and club events	263
11.6	Five organizational forms in a two-dimensional Blau	
(a and b)		266
12.1	Journal co-citation network in the social sciences (Moody &	
	Light 2006)	273
12.2	Correspondence analyses of tastes in music and	
(a and b)		279
12.3	Three-dimensional factor analysis of music tastes with	
	undergraduate major centroids	282
12.4	Two-dimensional MDS of music tastes (1993 GSS)	284
12.5	Four correlational classes of music tastes (1993 GSS)	288
(a–d)		• • •
12.6	Attitude network of music tastes (1993 GSS)	292
12.7	Topic contour plots of core sociology journals, 1990–92 and 2009–11	294
12.8	A priori decision tree for choosing clustering methods	
	(Pimentel 2014)	296
13.1	QAP on "same race" in one Add Health school	315
13.2	Edgewise expected values from "p1" simple random	
	graph models	317
13.3	Exemplar data matrix for simple ERGM change statistics	320
13.4	Three latent space models of Sampson's monastery data	328
13.5	3D plot of Sampson's monastery data	329
13.6	Role structure mobility matrix	331
13.7	The STERGM framework jointly models tie formation and tie	
	dissolution (Statnet Development Team; see Morris et al.	225
1 4 1	2014)	335
14.1	Common SIR model representation	343
14.2	Example of SIR compartmental model dynamics	344
14.3	US COVID-19 infections over time	345
14.4	School 2 from the Add Health network, with edges weighted by the number of activities students reported doing with	
	one another	350
14.5	Simulated SIR on network from Figure 14.4, with constant transmission probability (0.1), proportional to	
	edge weights	352
14.6	Temporal constraints on diffusion exposure	354

Cambridge University Press & Assessment
978-1-107-03778-6 – Network Analysis
Craig M. Rawlings , Jeffrey A. Smith , James Moody , Daniel A. McFarland
Frontmatter
More Information

List of F	List of Figures	
14.7	Varying thresholds for adoption given a single source set	
	(yellow)	359
14.8	Distribution of the proportion of students joining the moveme	ent
	by average threshold level – ensemble of 500 random	
	threshold distributions	360
15.1	Simulated peer influence on heterodox opinion	368
15.2	The structural alignment of friends and interests	381
16.1	GitHub network	395
16.2	Two details of GitHub network	396
(a and b		
16.3	Follower network of 300,000+ Twitter users	396
16.4	Block-image network for a PROSPER school, with lines	
	shaded by difference from expected value	399
16.5	Second-order clustering of triadic role positions	400
16.6	Example of a multiplex network as a multilayer network	414
16.7	Idealized adjacency matrices from multilayering of networks	
	of southern women's clubs data	415
16.8	Newcomb's fraternity data presented as a multilayer	
	(temporal) network	416

Tables

2.1	Types of causal social-structural questions and social network	
	research agendas	page 33
3.1	Adjacency matrix of Florentine family relations	60
3.2	Affiliation matrix of southern women data	61
3.3	Edgelist of Florentine family relations	63
3.4	Adjacency list of Florentine family relations	64
4.1	Sample network questionnaire	73
4.2	Comparison of realist and nominalist data collection strategies	79
5.1	Three bivariate distributions	89
5.2	Distance matrix of five US cities	95
6.1	Common ego network measures related to	
	structuration features	139
7.1	Uncertainty in relating	154
8.1	Odds ratios as a measure of group segregation	174
8.2	Blocking matrix	186
9.1	Typology of centrality scores (Borgatti & Everett 2006)	196
9.2	Triadic forms forbidden by one or more balance theory rules	211
10.1	Summarized positional relations	219
10.2	Permuted adjacency matrix based on the blockmodel of	
	hierarchy in Figure 10.6	225
10.3	Image matrix reduction of the blockmodel solution in	
	Table 10.2	226
12.1	Factor analysis of music tastes (1993 GSS)	282
13.1	Common structural and actor effects included in ERGMs	322
13.2	Statistical models for longitudinal networks	332
13.3	Inventory of p-shifts with examples (Gibson 2003)	337
15.1	SAOM of peer influence on smoking (see Schaefer, Haas, &	
	Bishop 2012)	384

xv

Acknowledgments

This book is the product of an invisible college that spans hundreds of campuses with thousands of faculty, students, and researchers in academia and industry. Members belong to disciplines across the social and natural sciences, engineering, and the humanities. But this college has no lecture halls or gardens of its own. Some of its members meet regularly at conferences or online, while many know each other only on paper. And yet, it is a college in the sense that it exists as a network of scholars who share the same intellectual goals and jointly add to a shared body of knowledge. We were educated in this college when it was somewhat smaller and mainly consisted of social scientists. We offer this book as an attempt to affirm and strengthen some of those social scientific roots, and in the hope that doing so will help nourish and integrate the many branches of this invisible college today.

Our primary thanks go to those who helped to build this college. And to be clear, there are so many we are certain to have forgotten to cite and thank all those who deserve it. We stand on the shoulders of generations of scholars who pioneered network analysis. We have known many of these scholars personally, and they continue to inspire us. We first acknowledge our direct teachers and formal mentors in network analysis: Peter Bearman, Charles Bidwell, Peter Blau, Noah Friedkin, Roger Gould, Maureen T. Hallinan, Edward Laumann, J. Miller McPherson, John Mohr, John Padgett, and Doug White. We also thank our teachers' teachers: Harrison White, James Coleman, Paul DiMaggio, Walter Powell, and Scott Boorman. We have learned a great deal through the informal networks of friendship and advice that are also vital to the integration of the invisible college, especially the following individuals: jimi adams, Chris Bail, Ronald Breiger, Ronald Burt, Carter Butts, Karen Cook, Linus Dahlander, David Diehl, Paul DiMaggio, Jan Fuhse, Amir Goldberg, Sharique Hasan, Henning Hillman, Lisa A. Keister, John Levi Martin, Paul McLean, Ann

xvii

CAMBRIDGE

Cambridge University Press & Assessment 978-1-107-03778-6 — Network Analysis Craig M. Rawlings , Jeffrey A. Smith , James Moody , Daniel A. McFarland Frontmatter <u>More Information</u>

xviii

Acknowledgments

Mische, Jonathan Morgan, Martina Morris, Andrej Mrvar, Peter J. Mucha, Paolo Parigi, Sanne Smith, Lynn Smith-Lovin, Steve Vaisey, Alex Volfovsky, and Stanley Wasserman. We would also never have gotten here without the students we have taught over the years and the need to try to distill so much of this knowledge into a course of learning. Special thanks to graduate students Gabriel Varela, Tom Wolff, and Joe Quinn for sundry reviews of the R labs and specific text details. We have each taught elements of this text in our courses and thank the students who suffered through early and incomplete versions of this manuscript for their feedback on the text and labs, as it was always welcome and wonderful. You know who you are, and we thank you for teaching us how to better teach this material through trial and much error.

Many of the R labs presented in this textbook were built from prior versions developed at Stanford University by Daniel A. McFarland, Solomon Messing, Michael Nowak, Sean J. Westwood, and Sanne Smith. Chapter 5's lab for NDTV drew on Skye Bender-deMoll's materials; Chapter 12 on LDA/CA came from Love Börjeson and Daniel A. McFarland; Chapter 13 concerning "ERGM" and "relevant" drew on Carter Butts' materials; and Chapter 15 on SIENA/SAOM drew on ICS materials. Finally, a great many resources from the Duke Network Analysis Center (DNAC) helped us in formulating elements in many of the labs. For example, Chapter 4 on missing data imputation drew on James Moody and Jeffrey A. Smith's work in the DNAC, as did Chapter 14 on diffusion. We are grateful to these institutions and individuals for sharing code and helping us formulate applications for each chapter's theories.

We have been fortunate to benefit from the time and research opportunities provided by external funding throughout the lifetime of this project. This includes the Social Networks and Health (SN&H) NICHD Workshop grant (NICHD, 2 R25 HD079352–06), which supported the development of many of the pedagogical ideas, and a James S. McDonnell Foundation Complexity Scholars award to James Moody (220020397). Examples in the text draw on data collected or analysis completed in whole or in part from numerous funded projects, including NSF (0624134, 1633036, 1827477, SMA-1829240, 2022435; BCS-2024271-1; SES-2029790), NIH (R01 DA018225–05A1, R01 HD075712–01; R21HD104431), and DARPA (FA8650–18-C-7826).

This manuscript benefited from several individuals who read it in part or whole. We are extremely grateful to Kevin Lewis for reading the entire manuscript and providing detailed comments. We cannot overstate how helpful his insights were in clarifying points large and small. Our deep thanks also go to Tom Snijders for expert advice and detailed comments on Part III of the book. These insights were invaluable. Others have supported this endeavor in important ways. We are very grateful for expert editorial assistance provided by Laura Tesch. We also thank Colin Birkhead for additional editorial help. We are also grateful to Robin Gauthier and Sela Harcey for their helpful comments and suggestions on the R tutorials. Craig Rawlings thanks Moira Killoran and

Acknowledgments

xix

Clayton Childress for their expert advice on many occasions. Finally, we thank Mark Granovetter for believing in this endeavor and supporting it along its long journey. It was a long, well-traveled road, crowded with assistance, and that has made all the difference.

The authors would like to dedicate the book to:

CMR: JMR JAS: GRG & EAAS JWM: LAK DAM: SMS