

Designing Empirical Social Networks Research

A user-friendly introductory guide to the empirical study of social networks. Jennifer M. Larson presents the fundamentals of social networks in an intuition-forward way that guides theory-driven research design. Substantial attention is devoted to a framework for developing a network theory that will steer data collection to be maximally informative and minimally frustrating.

Other features include: coverage of a range of practical topics including selecting operationalizations, cutting survey costs, and cleaning data; a tutorial for getting started in analyzing networks in R; and technical sections full of examples, points to hone intuition, and practice problems with solutions.

Designing Empirical Social Networks Research will be a valuable tool for advanced undergraduates, PhD students in the social sciences, especially political science, and researchers across the social sciences who are new to the study of networks.

Jennifer M. Larson is Associate Professor in the Department of Political Science at Vanderbilt University.

Cambridge University Press & Assessment
978-1-009-48418-3 — Designing Empirical Social Networks Research
Jennifer M. Larson
Frontmatter
[More Information](#)

“A terrific explainer of how to study social networks by one of the stars of the field.”

David Lazer

University Distinguished Professor, Northeastern University

“This book delivers exactly what the title promises by providing clear, concise, step-by-step guidance for designing empirical social networks research. Larson introduces readers to the conceptual language necessary to theorize the role and nature of networks, and explains how to analyze networks empirically. The book will be an invaluable resource for social scientists interested in adding social networks research to their toolkit.”

Imke Harbers

Associate Professor of Political Science, the University of Amsterdam

“Much has been written about the analysis of network data, but the field has lacked a comprehensive guide to the design of network research until now. Particularly important in Larson’s book is the focus on developing a sound network theory and using that to guide the rest of the process, from data gathering and processing to analysis. With practical examples from an experienced field researcher throughout, this book is a definitive reference for those embarking on network research.”

Skyler J. Cranmer

Carter Phillips and Sue Henry Professor and Director,
Machine Intelligence and Data Science (MIDS) Laboratory,
The Ohio State University

Methods for Social Inquiry

Editors

Colin Elman, Syracuse University
Diana Kapiszewski, Georgetown University
James Mahoney, Northwestern University

The *Methods for Social Inquiry* series comprises compact texts offering practical instructions for qualitative and multi-method research. Each book is accompanied by pedagogical data and exercises.

The books in the series offer clear, straightforward, and concrete guidance for teaching and using methods. While grounded in their relevant prescriptive logics, the books focus on the “how-to” of the methods they discuss – the practical tasks that must be undertaken to effectively employ them. The books should be useful for instruction at both the advanced undergraduate and graduate levels.

The books are tightly integrated with digital content and online enhancements through the Qualitative Data Repository (QDR). QDR is a new NSF-funded repository housing digital data used in qualitative and multi-method social inquiry. The pedagogical data (and related documentation) that accompany the books in the series will be housed in QDR.

Books in the series

Larson, Jennifer M., *Designing Empirical Social Networks Research*
Schneider, Carsten Q., *Set-Theoretic Multi-Method Research: A Guide to Combining QCA and Case Studies*
Kreuzer, Markus, *The Grammar of Time: A Toolbox for Comparative Historical Analysis*
Oana, Ioana-Elena, Schneider, Carsten Q., and Thomann, Eva, *Qualitative Comparative Analysis Using R: A Beginner's Guide*
Cyr, Jennifer, *Focus Groups for the Social Science Researcher*

Designing Empirical Social Networks Research

Jennifer M. Larson

Vanderbilt University



CAMBRIDGE
UNIVERSITY PRESS

Cambridge University Press & Assessment
978-1-009-48418-3 — Designing Empirical Social Networks Research
Jennifer M. Larson
Frontmatter
[More Information](#)



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/9781009484183

DOI: 10.1017/9781009484206

© Jennifer M. Larson 2024

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.

When citing this work, please include a reference to the DOI 10.1017/9781009484206

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 978-1-009-48418-3 Hardback

ISBN 978-1-009-48422-0 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>List of Figures</i>	<i>page</i> x
<i>List of Tables</i>	xiii
<i>Acknowledgments</i>	xiv
<i>Note on the text</i>	xv
1. Introduction to Social Networks Research	1
Studying Social Networks	1
The Vast Scope of Social Networks Research	3
Types of Empirical Social Networks Research	6
Why Networks Warrant Special Consideration	8
Should I Undertake Networks Research?	10
How to Use This Book	12
2. Describing and Interpreting Social Network Features	14
Defining a Network	14
Representing a Network	15
When the Links Have Direction	18
Structural Network Features	22
Indirect Connections: Walks and Paths	22
Network Shapes	26
A Node’s Importance: Centrality Measures	30
Tight-Knittedness	38
Practice	40
3. Accounting for Substantive Network Features	42
Node Attributes	42
Homophily	43
Assortativity	48
Link Attributes	50
Link Weights	50
Tie Strength	51
Practice	52
4. Crafting a Network Theory	54
Conceptualizing the Network of Interest	54

viii	Contents
	Defining a Set of Nodes 55
	Specifying the Type of Link 56
	Specifying the Function of the Links 60
	Multilayer Networks 63
	Building a Theory 64
	Network Hypotheses 65
	Descriptive Research 65
	Network as a Dependent Variable 66
	Network as an Independent Variable 67
	Groundwork for Theory Building 70
	Practice 71
	5. Moving from Theory to Empirics 73
	Determining Which Nodes to Include 73
	Network Boundary Issues 73
	Network Sampling Issues 75
	Ego Networks vs. Full Networks 81
	Operationalizing Link Type 82
	Consequences of Operationalization for Variation 83
	Biasing Results with Operationalization 86
	Practice 88
	6. Acquiring Network Data 90
	Measuring Networks via Surveys 91
	Lists and Name Generators 91
	Survey Questions to Elicit Links 93
	Collecting Node and Link Attributes 95
	Tradeoffs of Design Decisions 95
	Capping the Number of Links 96
	Measuring Aggregates Instead of Individuals 97
	Using Proxies for Network Position 98
	Applying Principles of Survey Design to Other Methods of Data Collection 99
	7. Preparing Network Data 102
	Anonymity and Ethical Issues 102
	Constructing the Network 103
	Cleaning the Data 104
	Identity Matching 104
	Spelling Discrepancies 105
	Automating the Cleaning 105
	Links to Unsampled Nodes 108

	Contents	ix
Creating Aggregate Nodes	109	
Making Links with “And” and “Or” Rules	109	
Multiple Link Operationalizations	110	
Constructing Measures of Tie Strength	112	
8. Working with Network Data in R	113	
Practice with the Network Functions	113	
Importing and Preparing Network Data	116	
Scenario 1: Data Stored as Edgelist	117	
Scenario 2: Data Stored as Node List	119	
Creating and Inspecting a Network Object	121	
Modifying a Network Object	124	
Describing Structural Features	125	
Adding Substantive Attributes	130	
Working with Node Attributes	131	
Working with Link Attributes	132	
Generating Random Networks	137	
Assessing the Significance of Network Features	139	
Practice	141	
9. Conclusion: Where to Turn Next	143	
Seeking Inspiration	143	
Resources for More Technical Detail	144	
Resources for Network Data Analysis	145	
Resources for Using Social Media Data	146	
Resources for Visualizing Networks	146	
Resources for Integrating Qualitative and Quantitative Information	147	
Resources for Network Experiments	147	
Resources for Networks in Game Theory	148	
<i>Selected Answers</i>	149	
Answers to Chapter 2 Practice	149	
Answers to Chapter 3 Practice	151	
Answers to Chapter 8 Practice	152	
<i>Glossary</i>	155	
<i>Bibliography</i>	159	
<i>Index</i>	168	

Figures

2.1	Example network with seven nodes.	<i>page</i> 16
2.2	Directed example network with seven nodes.	19
2.3	One walk from node 4 to node 7: 4 to 5 to 6 to 4 to 2 to 6 to 7.	23
2.4	One path from node 4 to node 7: 4 to 2 to 6 to 7.	23
2.5	A directed network where the directed path from 4 to 1 is longer than the undirected path between 4 and 1.	26
2.6	Example network with two components.	27
2.7	Example network with a bridge between nodes 5 and 8.	28
2.8	When the bridge link between 5 and 8 is removed, the network breaks apart into two components.	28
2.9	Examples of canonical network shapes: (a) ring network and (b) star network.	30
2.10	More examples of canonical network shapes: (a) tree network and (b) complete network.	30
2.11	Node 5 has the highest betweenness centrality in this network because it occupies a bottleneck position.	35
2.12	Node 2’s neighborhood. Node 2’s clustering is calculated in reference to these nodes.	39
2.13	Node 5’s neighborhood. Node 5’s clustering is calculated in reference to these nodes.	40
3.1	Example network in which student nodes have an attribute corresponding to their major: math (vertical lines), physics (horizontal lines), and political science (no lines).	44
3.2	Example network in which student nodes have an attribute corresponding to their major: math (vertical lines) or political science (no lines).	53
3.3	Example network in which student nodes have an attribute corresponding to their major: math (vertical lines), physics (horizontal lines), or political science (no lines).	53
4.1	Close friendship network for the ten people in Table 4.1.	58
4.2	Shared alma mater network for the ten people in Table 4.1.	59
4.3	Shared vowel in names network for the ten people in Table 4.1.	60

	List of Figures	xi
5.1	Friendship network with links operationalized as birthday party invitations. Arrows point from inviter to invitee.	74
5.2	The result of randomly sampling students 1, 3, 5, 9, and 11 for inclusion in the friendship network with links operationalized as birthday party invitations.	76
5.3	The more the network features of interest depend on interconnections throughout a network, the more precisely a network needs to be measured. Studies interested in network features such as centrality, path lengths, transitivity, and reciprocated ties suffer most acutely from small samples.	78
5.4	The consequences of random sampling from four real social networks at different sample sizes for six network attributes. The plots display the average Kendall rank correlation between the nodes' true network feature and the observed network feature when a sample of a certain size is drawn. Networks A and B are undirected social networks from Banerjee <i>et al.</i> (2013), and Networks C and D are directed social networks from Larson and Lewis (2017). On average, small samples perform especially badly for more demanding structural features.	80
5.5	Four different options for how to operationalize "social ties" led to four very different views of the network in this Ugandan village. (a) With whom do you regularly spend time? (b) With whom do you regularly discuss politics? (c) To whom would you confide a secret? (d) Whom do you regularly speak to on the phone?	84
5.6	A network where mobilization ties are operationalized as shared membership in an organization. Since everyone in the data is a member of the PTA, everyone is linked to everyone else.	86
5.7	A network where affinity ties are operationalized as shared membership in a committee. Since there are two committees in the data, there are two components, with each person connected to every other within a component.	88
8.1	Example network that appeared as Figure 2.7.	114
8.2	Network visualized with R.	115
8.3	Creating a dataset in edgelist form.	118
8.4	Creating a dataset in node list form.	118
8.5	Visualization of our network in R with the isolate node Carl added.	125
8.6	Degree distribution of our co-classmate network.	126

xii	List of Figures
8.7	Co-classmate network with the node attribute “year” indicated by color. Blue nodes are seniors, yellow are juniors. 131
8.8	Co-classmate network with the node attribute “year” indicated by color and the link weights (number of classes taken together) determines link thickness. 133
8.9	Two Erdős–Rényi random networks with fifty nodes and the probability of link formation set to .2 (a) and .4 (b). 138
8.10	Sampling distribution of the mean clustering value for networks with the same number of nodes and links as our data’s network. Vertical line shows the mean value of our measured network’s clustering – it is much larger than would be expected in networks whose links were added randomly. 141

Tables

1.1 Framework for defining a network with sufficient precision and relevance. Topic of Chapter 4.	<i>page 2</i>
2.1 The edgelist representation of the network in Figure 2.1. Each row is a link.	16
2.2 The node list representation of the network in Figure 2.1. Link information is organized by node, one row for each.	17
2.3 The edgelist representation of the directed network in Figure 2.2. Since the network is directed, we need to include each direction of each link separately.	20
2.4 The node list representation of the directed network in Figure 2.2. Now a node's list only includes nodes to which that node has a directed link.	20
2.5 Length of the shortest paths between 1 and all other nodes, used to calculate node 1's closeness centrality.	33
2.6 Information needed to calculate node 5's betweenness centrality. Total number of shortest paths between the row and column nodes (denominator), and the number of these that go through node 5 (numerator). B_5 is calculated by summing these values.	36
2.7 Rank of nodes in Figure 2.11 by four different centrality measures. Centrality score is in parentheses. Nodes at the top are most central by that column's measure.	38
4.1 Friendship and college information for ten people.	57
4.2 Three elements of a well-defined network.	61
5.1 Organization and committee membership information for ten people.	85
8.1 Example data in node list form. Each row contains all links for the node listed first in the row. We will first convert this to an edgelist.	117

Acknowledgments

First and foremost, I am grateful to the many students at Harvard, New York University, and Vanderbilt who have embarked on the study of networks by taking my classes. Their thoughtful engagement helped my instruction evolve over the years, inspiring and ultimately shaping this book. Looking back even further, I owe a debt of gratitude to David Lazer, who generously opened his lab to me when I was an eager (read: clueless) graduate student just getting started with networks research. I can think of no better introduction to the field.

I greatly appreciate the PhD students at Vanderbilt who worked through an early draft of this manuscript and offered feedback. I also received valuable feedback at the Authors' Workshops at the Institute for Qualitative and Multi-Methods Research, especially from my discussants Steven Wilson, Diana Kapiszewski, Alan Jacobs, Jim Mahoney, Matt Ingram, and Hillel Soifer. Cassy Dorff's insights on this book were yet another example of her substantially improving my work – I am lucky to have her as a colleague. I am also grateful to Janet Lewis, with whom I have been fortunate to be able to collaborate repeatedly to empirically study networks. Our collaborations have taught me so much and continue to remind me how enjoyable research can be.

And of course, I owe a tremendous thanks to my parents, Jim and Mary Jo Larson, and my husband, Andrew Coe. Their endless support is the reason why I am in any position to write a book. Having a husband who can and will join me in any weeds I want to get into is pretty wonderful.

Note on the Text

This book is primarily designed for (1) advanced undergraduates; (2) graduate students in the social sciences, especially political science; and (3) scholars who have some background in empirical research methods but who are new to the study of networks. Its goals are to introduce the fundamentals of the study of social networks and guide theory-driven research design.

Instructors of undergraduates may prefer to assign the first three chapters (which cover the basics of descriptive measures of network features) and perhaps also the eighth chapter (which is a tutorial for calculating these network features in R). Those teaching graduate students and planning their own research projects may want to add the middle four chapters as well (which cover how to formulate a network theory and use it to design data collection).

This book does not presume fluency in mathematical notation. That does not mean it omits the details. To the contrary, this book covers the fundamentals of structural and substantive network features that can be found in other textbooks aimed at audiences with more technical proficiency (Chapters 2 and 3). However, it does so by adding a lot of words around the mathematical expressions, referring to an abundance of examples, and presenting opportunities throughout to hone intuition. Readers looking for a terse mathematical presentation will be disappointed, but those looking for a rigorous yet intuition-forward presentation may prefer the level of the book.

This book also devotes substantial attention to the development of network theory (Chapters 4–6). It introduces a framework that should make both informal and formal theorizing about the role of networks easier. The framework's objective is to guide data collection so that an empirical project can be maximally informative (and minimally frustrating).

The section on data cleaning (Chapter 7) is particularly valuable to read before embarking on data collection, because it contains tips for avoiding onerous cleaning needs in the first place. Those who want an introduction to network analysis in R may find the tutorial in Chapter 8 useful, as it covers the initial steps of importing and formatting data as well as standard uses of the `igraph` package. And for those using this book as a jumping off point, Chapter 9 points to a variety of resources to assist in richer analyses, experimental designs, game theoretic approaches, and other next steps.