Exploratory Social Network Analysis with Pajek

This is an extensively revised and expanded third edition of the successful textbook on analysis and visualization of social networks integrating theory, applications, and professional software for performing network analysis. The main structural concepts and their applications in social research are introduced with exercises. Pajek software and data sets are available, so readers can learn network analysis through application and case studies. In the end, readers will have the knowledge, skills, and tools to apply social network analysis across different disciplines. A fundamental redesign of the menu structure and the capability to analyze much larger networks required a new edition. This edition presents several new operations, e.g., community detection, generalized main paths searches, new network indices, advanced visualization approaches, and instructions for installing Pajek under MacOSX. This third edition is up-to-date with Pajek version 5 and it introduces PajekXXL for very large networks and Pajek3XL for huge networks.

Wouter de Nooy is Associate Professor in the Department of Communication Science at the University of Amsterdam, the Netherlands. He is a member of the Amsterdam School of Communication Research. His research interests include dynamic models for social networks and their application in culture and politics.

Andrej Mrvar is Professor of Social Science Informatics at the Faculty of Social Sciences, University of Ljubljana, Slovenia. He started developing the programs *Pajek*, *PajekXXL*, and *Pajek3XL* as part of his PhD dissertation at the Faculty of Computer and Information Science in 1996. Since then Pajek has been in constant development.

Vladimir Batagelj is Professor Emeritus of the University of Ljubljana, Slovenia. He is also a member of the Institute of Mathematics, Physics, and Mechanics (IMFM), Ljubljana, and of AMI, UP, Koper. His coauthored book *Generalized Blockmodeling* was awarded the 2007 Harrison White Outstanding Book Award by the Mathematical Sociology Section of the American Sociological Association. From the International Network for Social Network Analysis he was awarded the Georg Simmel Award (2007) and the Richards Software Award for the program *Pajek* (2013). In 2014 he coauthored a book entitled *Understanding Large Temporal Networks and Spatial Networks*.

Structural Analysis in the Social Sciences

Series Editor: 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

Wouter De Nooy, Andrej Mrvar and Vladimir Batagelj, *Exploratory Social Network Analysis with Pajek: Revised and Expanded Edition for Updated Software. Third Edition*
Darius Mehri, *Iran Auto*
Navid Hassanpour, *Leading from the Periphery and Network Collective Action*
Cheol-Sung Lee, *When Solidarity Works* (Second Edition)
Benjamin Cornwell, *Social Sequence Analysis* (Second Edition)
Mariela Szwarcberg, *Mobilizing Poor Voters* (Second Edition)
Luke M. Gerdes, ed., *Illuminating Dark Networks* (Second Edition)
Silvia Domínguez and Betina Hollstein, eds., *Mixed Methods in Studying Social Networks* (Second Edition)
Dean Lusher, Johan Koskinen, and Garry Robins, eds., *Exponential Random Graph Models for Social Networks: Theory, Methods, and Applications* (Second Edition)
Sean F. Everton, *Disrupting Dark Networks* (Second Edition)
Wouter de Nooy, Andrej Mrvar, and Vladimir Batagelj, *Exploratory Social Network Analysis with Pajek* (First Edition)
Noah E. Friedkin and Eugene C. Johnsen, *Social Influence Network Theory* (Second Edition)
Zeev Maoz, *The Networks of Nations: The Evolution and Structure of International Networks, 1815–2002* (Second Edition)

(continued after the index)

Exploratory Social Network Analysis with Pajek

Revised and Expanded Edition for Updated Software. Third Edition

WOUTER DE NOOY
University of Amsterdam

ANDREJ MRVAR
University of Ljubljana

VLADIMIR BATAGELJ
University of Ljubljana

CAMBRIDGE UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, 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

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781108474146
DOI: 10.1017/9781108565691

First published 2018

Printed in the United States of America by Sheridan Books, Inc.

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

ISBN 978-1-108-47414-6 Hardback
ISBN 978-1-108-46227-3 Paperback

Contents

Figures

Tables

Preface to the Third Edition

Two major developments in program Pajek required a new edition of this book: a fundamental redesign of the menu structure and the capability to analyze much larger networks, containing billions of vertices. A proliferation of methods for analyzing a single network necessitated a reorganization of Pajek's menu structure, in particular the former *Net* menu, now called the *Network* menu. The new *Network* menu contains submenus for commands that apply to a particular type of network: a two-mode network, multiple relations network, acyclic network, temporal network, and signed network. It is much easier now to find analyses for special networks. Because the *Network* menu is used most intensively, we had to adjust most of the commands in the Application sections of this book.

Pajek's capability to analyze much larger networks is the second major development. As a result of changes in the Windows® operating system, Pajek can now handle networks with nearly one billion vertices. For even larger networks, PajekXXL and Pajek3XL have been developed, which can handle up to two and ten billion vertices respectively. PajekXXL and Pajek3XL have the same user interface as Pajek; if you can work with Pajek, you can also work with PajekXXL and Pajek3XL. What you need to know is that the latter two programs offer a very limited set of analyses to describe and partition huge networks. Subnetworks extracted from a huge network can be sent directly to (regular) Pajek for further analysis. Appendix A1.5 explains all of this.

Questions on how to install and use Pajek on Mac OS X have been asked repeatedly. This new edition brings an additional appendix (Appendix 3) containing detailed instructions on installing Pajek under Mac OS X. We hope that Mac users will find out that installing and running Pajek on Mac OS X is not a difficult technical problem: Installation takes only few additional minutes compared to installation under native Windows. When Pajek is installed, running Pajek under Mac OS X is the same as running it under Windows.

Finally, we took the opportunity to include some analyses requested by Pajek users. Chapter 1 now includes Pivot MDS and VOS mapping for network layout as well as the correlation between layout coordinates and network geodesics as a measure of layout performance. Chapter 2 introduces partitions on vertex labels using regular expressions and marking partition clusters with Unicode symbols in the Draw screen. It also discusses interactive FishEye magnification and the Adjusted Rand Index. Relaxed balance is explained in Chapter 4; community detection (Louvain Method and VOS Clustering) and the E-I Index appear in Chapter 5. Chapter 6 includes (degree) assortativity and the assortativity coefficient. A collection of main path methods (including key-route searches) and preprint transformation are explained and applied in Chapter 11. Appendices A1 and A2 have been updated, now containing goodies such as dragging and dropping data to a Pajek window, sending Pajek objects to Excel®, defining colors and transparency of vertices and lines, using Unicode symbols and additional vertex shapes (e.g., man, woman, and house), tooltips for vertex labels, drawing curved lines, and so on. We hope that you will continue enjoying social network analysis with Pajek.

The webpage to the third edition of this book (http://mrvar.fdv.uni-lj.si/pajek/be3.htm, mirror http://mrvar2.fdv.uni-lj.si/pajek/be3.htm) contains the example data sets, helper programs, and other online documents referenced in this book.

Preface to the Second Edition

> I go with him out in a shed in back and see he is selling a whole Harley machine in used parts, except for the frame, which the customer already has. He is selling them all for $125. Not a bad price at all.
>
> Coming back I comment, "He'll know something about motorcycles before he gets *those* together."
>
> Bill laughs. "And that's the best way to learn, too."
>
> Robert M. Pirsig, *Zen and the Art of Motorcycle Maintenance*

To some of its readers, this book is an introduction to social network analysis; to other readers, it is a manual to Pajek software (http://mrvar.fdv.uni-lj.si/pajek). To us, it is both. As Patrick Doreian argued in his review of our book (In: *Social Networks* 28 [2006] 269–274), an understanding of social network analysis is required for proper use of Pajek and, vice versa, understanding the concepts and logic of Pajek fosters comprehension of network concepts. In this second edition, we have aimed to strengthen both aspects, updating the discussion of the Pajek interface and commands to include several capabilities that have been implemented since we submitted the text of the first edition, such as multiplex networks (Section 1.3.1), eigenvector centrality (Section 6.5), matrix multiplication (Section 11.3), and using Pajek output in R (Chapters 5 and 13). The new capabilities cover some important advances in social network analysis, including random graph models to which we have dedicated a new chapter.

We expanded the Further Reading sections with references to seminal, much cited texts. This should allow the reader to trace the literature on the selected topic in bibliographic and citation databases. For more comprehensive lists of literature, we refer to two other volumes in this series: S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge: Cambridge University Press, 1994) and P. J. Carrington, J. Scott, and S. Wasserman, *Models and Methods in Social Network*

Analysis (Cambridge: Cambridge University Press, 2005). A concise history of social network analysis is published by L. C. Freeman, *The Development of Social Network Analysis. A Study in the Sociology of Science* (Vancouver, Canada: Empirical Press, 2004).

We hope that this second edition will continue to stimulate analysts to sharpen their understanding of social networks and expand their command of network analytic tools.

Preface to the First Edition

In the social sciences, social network analysis has become a powerful methodological tool alongside statistics. Network concepts have been defined, tested, and applied in research traditions throughout the social sciences, ranging from anthropology and sociology to business administration and history.

This book is the first textbook on social network analysis integrating theory, applications, and professional software for performing network analysis. It introduces structural concepts and their applications in social research with exercises to improve skills, questions to test the understanding, and case studies to practice network analysis. In the end, the reader has the knowledge, skills, and tools to apply social network analysis.

We stress learning by doing: Readers acquire a feel for network concepts by applying network analysis. To this end, we make ample use of professional computer software for network analysis and visualization: Pajek. This software, operating under Windows 95 and later, and all example data sets are provided on a Web site (http://vlado.fmf.uni-lj.si/pub/networks/book/) dedicated to this book. All the commands that are needed to produce the graphical and numerical results presented in this book are extensively discussed and illustrated. Step by step, the reader can perform the analyses presented in the book.

Note, however, that the graphical display on a computer screen will never exactly match the printed figures in this book. After all, a book is not a computer screen. Furthermore, newer versions of the software will appear, with features that may differ from the descriptions presented in this book. We strongly advise using the version of Pajek software supplied on the book's Web site (http://vlado.fmf.uni-lj.si/pub/networks/book/) while studying this book and then updating to a newer version of Pajek afterward, which can be downloaded from http://vlado.fmf.uni-lj.si/pub/networks/pajek/default.htm.

Overview

This book contains five sections. The first section (Part I) presents the basic concepts of social network analysis. The next three sections present the three major research topics in social network analysis: cohesion (Part II), brokerage (Part III), and ranking (Part IV). We claim that all major applications of social network analysis in the social sciences relate to one or more of these three topics. The final section discusses an advanced technique (viz., blockmodeling), which integrates the three research topics (Part V).

The first section, titled Fundamentals, introduces the concept of a network, which is obviously the basic object of network analysis, and the concepts of a partition and a vector, which contain additional information on the network or store the results of analyses. In addition, this section helps the reader get started with Pajek software.

Part II on cohesion consists of three chapters, each of which presents measures of cohesion in a particular type of network: ordinary networks (Chapter 3), signed networks (Chapter 4), and valued networks (Chapter 5). Networks may contain different types of relations. The ordinary network just shows whether there is a tie between people, organizations, or countries. In contrast, signed networks are primarily used for storing relations that are either positive or negative such as affective relations: liking and disliking. Valued networks take into account the strength of ties, for example, the total value of the trade from one country to another or the number of directors shared by two companies.

Part III on brokerage focuses on social relations as channels of exchange. Certain positions within the network are heavily involved in the exchange and flow of information, goods, or services; whereas others are not. This is connected to the concepts of centrality and centralization (Chapter 6) or brokers and bridges (Chapter 7). Chapter 8 discusses an important application of these ideas, namely, the analysis of diffusion processes.

The direction of ties (e.g., who initiates the tie) is not very important in the section on brokerage, but it is central to ranking, presented in Part IV. Social ranking, it is assumed, is connected to asymmetric relations. In the case of positive relations, such as friendship nominations or advice seeking, people who receive many choices and reciprocate few choices are deemed as enjoying more prestige (Chapter 9). Patterns of asymmetric choices may reveal the stratification of a group or society into a hierarchy of layers (Chapter 10). Chapter 11 presents a particular type of asymmetry, namely, the asymmetry in social relations caused by time: genealogical descent and citation.

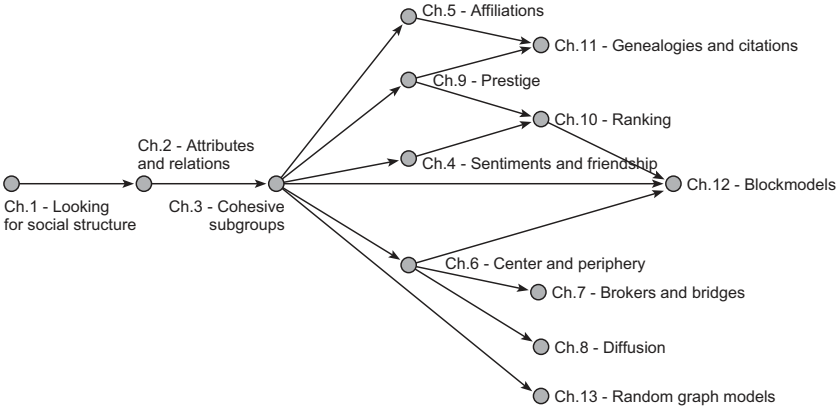

Figure 1. Dependencies between the chapters (for the second and third editions).

The final section, Part V, on roles concentrates on rather dense and small networks. This type of network can be visualized and stored efficiently by means of matrices. Blockmodeling is a suitable technique for analyzing cohesion, brokerage, and ranking in dense, small networks. It focuses on positions and social roles (Chapter 12).

The book is intended for researchers and managers who want to apply social network analysis and for courses on social network analysis in all social sciences as well as other disciplines using social methodology (e.g., history and business administration). Regardless of the context in which the book is used, Chapters 1, 2, and 3 must be studied to understand the topics of subsequent chapters and the logic of Pajek. Chapters 4 and 5 may be skipped if the researcher or student is not interested in networks with signed or valued relations, but we strongly advise including them to be familiar with these types of networks. In Parts III (Brokerage) and IV (Ranking), the first two chapters present basic concepts and the third chapter focuses on particular applications.

Figure 1 shows the dependencies among the chapters of this book. To study a particular chapter, all preceding chapters in this flowchart must have been studied before. Chapter 10, for instance, requires understanding of Chapters 1 through 4 and 9. Within the chapters, there are no sections that can be skipped.

In an undergraduate course, Part I and II should be included. A choice can be made between Part III and Part IV; or, alternatively, just the first chapter from each section may be selected. Part V on social roles and blockmodeling is quite advanced and more appropriate for a

postgraduate course. For managerial purposes, Part III is probably more interesting than Part IV.

Justification

This book offers an introduction to social network analysis, which implies that it covers a limited set of topics and techniques, which we feel a beginner must master to be able to find his or her way in the field of social network analysis. We have made many decisions about what to include and what to exclude, and we want to justify our choices now.

As reflected in the title of this book, we restrict ourselves to exploratory social network analysis. The testing of hypotheses by means of statistical models falls outside the scope of this book. In social network analysis, hypothesis testing is important but complicated; it deserves a book on its own. Aiming our book at people who are new to social network analysis, our first priority is to have them explore the structure of social networks to give them a feel for the concepts and applications of network analysis. Exploration involves visualization and manipulation of concrete networks, whereas hypothesis testing boils down to numbers representing abstract parameters and probabilities. In our view, exploration yields the intuitive understanding of networks and basic network concepts that are a prerequisite for well-considered hypothesis testing.

From the vast array of network analytic techniques and indices we discuss only a few. We have no intention of presenting a survey of all structural techniques and indices because we fear that the readers will not be able to see the forest for the trees. We focus on as few techniques and indices as are needed to present and measure the underlying concept. With respect to the concept of cohesion, for instance, many structural indices have been proposed for identifying cohesive groups: *n*-cliques, *n*-clans, *n*-clubs, *m*-cores, *k*-cores, *k*-plexes, lambda sets, and so on. We discuss only components, *k*-cores, 3-cliques, and *m*-slices (*m*-cores) because they suffice to explain the basic parameters involved: density, connectivity, and strength of relations within cohesive subgroups.

Our choice is influenced by the software that we use because we have decided to restrict our discussion to indices and techniques that are incorporated in this software. Pajek software is designed to handle very large networks (up to millions of vertices). Therefore, this software package concentrates on efficient routines, which are capable of dealing with large networks. Some analytical techniques and structural indices are known to be inefficient (e.g., the detection of *n*-cliques), and for others no efficient algorithm has yet been found or implemented. This limits our options; we present only the detection of small cliques (of size 3), and we cannot extensively discuss an important concept such as *k*-connectivity. In

summary, this book is neither a complete catalogue of network analytic concepts and techniques nor an exhaustive manual to all commands of Pajek. It offers just enough concepts, techniques, and skills to understand and perform all major types of social network analysis.

In contrast to some other handbooks on social network analysis, we minimize mathematical notation and present all definitions verbatim. There are no mathematical formulae in the book. We assume that many students and researchers are interested in the application of social network analysis rather than in its mathematical properties. As a consequence, and this may be very surprising to seasoned network analysts, we do not introduce the matrix as a data format and display format for social networks until the end of the book.

Finally, there is a remark on the terminology used in the book. Social network analysis derives its basic concepts from mathematical graph theory. Unfortunately, different "vocabularies" exist within graph theory, using different concepts to refer to the same phenomena. Traditionally, social network analysts have used the terminology employed by Frank Harary, for example, in his book *Graph Theory* (Reading: Addison-Wesley, 1969). We choose, however, to follow the terminology that prevails in current textbooks on graph theory, for example, R. J. Wilson's *Introduction to Graph Theory* (Edinburgh: Oliver and Boyd, 1972; published later by Wiley, New York). Thus, we hope to narrow the terminological gap between social network analysis and graph theory. As a result, we speak of a vertex instead of a node or a point, and some of our definitions and concepts differ from those proposed by Frank Harary.

Acknowledgments

The text of this book has benefited from the comments and suggestions from our students at the University of Ljubljana and the Erasmus University Rotterdam, who were the first to use it. In addition, Michael Frishkopf and his students of musicology at the University of Alberta gave us helpful comments. Mark Granovetter, who welcomed this book to his series, and his colleague Sean Farley Everton have carefully read and commented on the chapters. In many ways, they have helped us make the book more coherent and understandable to the reader. We are also very grateful to an anonymous reviewer, who carefully scrutinized the book and made many valuable suggestions for improvements. Ed Parsons (Cambridge University Press) and Nancy Hulan (TechBooks) helped us through the production process. Finally, we thank the participants of the workshops we conducted at the Sunbelt International Conference on Social Network Analysis in New Orleans (XXII) and Cancun (XXIII) for their encouraging reactions to our manuscript.

Most data sets that are used in this book have been created from sociograms or listings printed in scientific articles and books. Notwithstanding our conviction that reported scientific results should be used and distributed freely, we have tried to trace the authors of these articles and books and ask for their approval. We are grateful to have obtained explicit permission for using and distributing the data sets from them. Authors or their representatives whom we have not reached are invited