Social Sequence Analysis

Social sequence analysis includes a diverse and rapidly growing body of methods that social scientists have developed to help study complex ordered social processes, including chains of transitions, trajectories, and other ordered phenomena. Social sequence analysis is not limited by content or time scale and can be used in many different fields, including sociology, communication, information science, and psychology. Social Sequence Analysis aims to bring together both foundational and recent theoretical and methodological work on social sequences from the last thirty years. A unique reference book for a new generation of social scientists, this book will aid demographers who study life-course trajectories and family histories, sociologists who study career paths or work/family schedules, communication scholars and micro-sociologists who study conversation, interaction structures, and small-group dynamics, as well as social epidemiologists.

Benjamin Cornwell is an associate professor of sociology at Cornell University. He received his PhD in sociology at the University of Chicago. His research has been published in journals such as American Journal of Sociology, American Sociological Review, and Social Forces. His work has been covered in many media outlets, including CNN, the New York Times, MSNBC, and the Los Angeles Times. In 2012, he taught one of the first graduate courses on social sequence analysis in the United States at Cornell.
Structural Analysis in the Social Sciences
Mark Granovetter, editor

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

1. Mark S. Mizruchi and Michael Schwartz, eds., Intercorporate Relations: The Structural Analysis of Business
3. Ronald L. Brieger, ed., Social Mobility and Social Structure
4. David Knoke, Political Networks: The Structural Perspective
6. Kyriakos M. Kontopoulos, The Logic of Social Structure
7. Philippa Pattison, Algebraic Models for Social Structure
8. Stanley Wasserman, Social Network Analysis: Methods and Applications
9. Gary Herrigel, Industrial Constructions: The Sources of German Industrial Power
10. Philippe Bourgois, In Search of Respect: Selling Crack in El Barrio
11. Per Hage and Frank Harary, Island Networks: Communication, Kinship, and Classification Structures in Oceana
12. Thomas Schweitzer and Douglas R. White, eds., Kinship, Networks, and Exchange
15. Rebecca Adams and Graham Allan, Placing Friendship in Context

(continued after index)
Social Sequence Analysis

Methods and Applications

BENJAMIN CORNWELL

Cornell University
Contents

List of Figures  page xii
List of Tables  xv
Preface  xvii
Acknowledgments  xix

PART I.  INTRODUCTION

1  Sequence Analysis in the Social Sciences  3
   1.1  Timing and Context of the Book  4
   1.2  Contributions  7
   1.3  Audience and Scope  9
   1.4  Related References  14
   1.5  Plan of the Book  16

PART II.  THEORETICAL BACKGROUND

2  Theoretical Foundations of Social Sequence Analysis  21
   2.1  What Are Social Sequences?  21
   2.2  The Structural Origins of Social Sequences  22
   2.3  Why Social Sequences Matter  25
   2.4  Origins of Social Sequence Analysis Techniques  32
      2.4.1  Narrative Positivism  33
      2.4.2  The Relational Nature of Social Phenomena  36
      2.4.3  The Development of Whole-Sequence Analysis Methods  38
      2.4.4  Pivotal Criticisms  41
   2.5  The Second Wave of Social Sequence Analysis  45
      2.5.1  Sequences as Networks  46
      2.5.2  Microsequence Analysis  52
   2.6  Looking Ahead  55
### Contents

**PART III. SOCIAL SEQUENCE ANALYSIS CONCEPTS AND TECHNIQUES**

3 Sequence Analysis Concepts and Data 59
   3.1 Sequence Properties 59
      3.1.1 Positions 60
      3.1.2 Elements 60
      3.1.3 Substructures 61
      3.1.4 Subjects 62
   3.2 Key Issues in Sequence Data 62
      3.2.1 Recurrence 63
      3.2.2 Ties and Multidimensionality 65
      3.2.3 Gaps 66
   3.3 The Sequence Universe 67
      3.3.1 Size 67
      3.3.2 Boundary Specification 69
   3.4 Sequence Data 70
      3.4.1 Data Collection 70
      3.4.2 Element-Position Sampling 73
      3.4.3 Units of Measure 74
   3.5 Data Used in This Book 75
      3.5.1 The Survey of Health, Ageing, and Retirement in Europe (SHARE) 75
      3.5.2 The Correlates of War Global Trade Network 76
      3.5.3 The Davis, Gardner, and Gardner Deep South Study 78
      3.5.4 The Multinational Time Use Study (MTUS) 79
      3.5.5 The American Time Use Survey (ATUS) 80

4 Detecting Sequence Structure 83
   4.1 Descriptive Analysis 83
      4.1.1 Sequence Length 84
      4.1.2 Element Frequencies 84
      4.1.3 n-grams 85
      4.1.4 Position Reports 85
   4.2 Describing Stochastic Patterns 86
      4.2.1 Transition Matrices 86
      4.2.2 Markov Chains 91
   4.3 Sequential Connection 91
      4.3.1 Empirical Illustration: Gender Differences in the Parenthood-Stress Link 93
   4.4 Stationarity 94
      4.4.1 Empirical Illustration: Stationarity in the World System 96
   4.5 Spells 97
Contents ix

4.6 Homogeneity 98
4.7 On Using Summary Statistics and Tests 99
4.8 Visual Aids 100
   4.8.1 Transition Bubble Graphs 100
   4.8.2 State Transition Diagrams 101
   4.8.3 Sequence Index Plots 102
   4.8.4 State Distribution Graphs 104
   4.8.5 Tempograms 104
   4.8.6 Sequence-Network Diagrams 105
   4.8.7 On the Use of Color and Grayscale 105
4.9 Looking Ahead 108

5 Whole-Sequence Comparison Methods 109
   5.1 Sequence Alignment 110
      5.1.1 Sequence Alignment Operations 110
      5.1.2 Operation Costs and Distance 111
   5.2 Classical Optimal Matching (OM) 111
      5.2.1 Finding the Optimal Solution 113
      5.2.2 The Needleman–Wunsch Algorithm 114
   5.3 Basic Operation Cost Regimes 115
      5.3.1 Levenshtein Distance 115
      5.3.2 Levenshtein II Distance 115
      5.3.3 Hamming Distance 116
   5.4 Considerations in Setting Operation Costs 116
      5.4.1 Order versus Timing 116
      5.4.2 Time Warping 118
   5.5 Criticisms of Classical OM 119
      5.5.1 Theory/Method Fit 120
      5.5.2 Arbitrary Operation Costs 120
   5.6 Improvements on Classical OM 122
      5.6.1 Variable Substitution Costs 122
      5.6.2 Distance Normalization 124
      5.6.3 Reference Sequence Comparison 125
      5.6.4 Spell-Adjusted Distances 126
      5.6.5 The Dynamic Hamming Method 127
   5.7 Detecting Whole Sequence Patterns 129
      5.7.1 Hierarchical Cluster Analysis 130
      5.7.2 Identifying Sequence Classes 133
      5.7.3 Assessing Validity and Reliability 136
      5.7.4 Describing Sequence Classes 139
   5.8 Recent Advances in Sequence Comparison 143
      5.8.1 Multidimensional Sequence Analysis 144
      5.8.2 Two-Stage Optimal Matching (2SOM) Analysis 146
      5.8.3 Transition Sequence Analysis 147
Contents

5.8.4 Nonalignment Techniques 148
5.8.5 Beyond Cluster Analysis 150

PART IV. NEW DIRECTIONS IN SOCIAL SEQUENCE ANALYSIS

6 Network Methods for Sequence Analysis 155
  6.1 Theoretical Rationale 156
    6.1.1 Networks of Sequences 156
    6.1.2 Sequences as Bases of Affiliations 160
  6.2 Network Concepts and Terms 163
    6.2.1 Network Matrices and Components 163
    6.2.2 Directed and Undirected Networks 164
    6.2.3 One- and Two-Mode Networks 165
    6.2.4 Subject Comembership and Element Overlap 165
  6.3 Sequence-Network Construction 167
    6.3.1 Nonrecurrent Sequence Networks 167
    6.3.2 Recurrent Sequence Affiliation Networks 170
    6.3.3 Directed Sequence Networks 171
    6.3.4 Multidimensional Sequence Networks 172
  6.4 Visualization 172
    6.4.1 Relationships among Subjects 173
    6.4.2 Two-Mode Ordered Event Networks 174
    6.4.3 Two-Mode Recurrent Sequences 176
    6.4.4 Directed Sequence Networks 179
  6.5 Structural Measures for Sequence Networks 184
  6.6 Identification of Subject/Element Subsets 189
    6.6.1 Subsets in Sequence Affiliation Networks 190
    6.6.2 Element Subsets in Directed Sequence Networks 193
    6.6.3 Sequence Motifs 195
  6.7 Event Structure Analysis 196
  6.8 Statistical Network Models for Sequential Events 200
    6.8.1 Exponential Random Graph (p*) Models 200
    6.8.2 ERGMs for Ordered Nonrecurrent Sequence Networks 201
  6.9 Dynamic Sequence Networks 202
    6.9.1 Subject- and Element-Level Change 203
    6.9.2 Sequence Subset Evolution 206
    6.9.3 Whole Sequence-Network Evolution 207
  6.10 Conclusion 209

7 Social Microsequence Analysis 210
  7.1 An Exemplary Sequence Context 211
  7.2 The Elements of Social Microsequences 214
    7.2.1 Relationship to Microinteraction Sequences 215
Contents

7.2.2 Elements and Positions 215
7.2.3 Data Collection and Availability 216

7.3 Nonnetwork Sequence Approaches 217
7.3.1 OM Sequence Classification 220
7.3.2 Transition and Switching Analysis 222

7.4 Sequence-Network Approaches 226
7.4.1 Synchrony 227
7.4.2 Measurement of Synchrony 230
7.4.3 Identifying Sources of Synchrony 238
7.4.4 Routine 241
7.4.5 Measurement and Visualization of Routine 244

7.5 Next Steps 251

PART V. CONCLUSIONS

8 The Promise of Social Sequence Analysis 255
8.1 Limitations 256
8.2 Future Research 257
8.2.1 Routine and Routinization 257
8.2.2 Sequence Networks, Network Sequences 258
8.2.3 Sequential Statistical Inference 259
8.2.4 Data Collection 261

Appendix A Recent Whole-Sequence Pattern Analyses 263
Appendix B Linkage Criteria for Agglomerative Hierarchical Clustering 270
References 275
Index 303
Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>How to read this book</td>
<td>12</td>
</tr>
<tr>
<td>4.1</td>
<td>Frequency state transition matrix (N)</td>
<td>87</td>
</tr>
<tr>
<td>4.2</td>
<td>Probability state transition matrix (P)</td>
<td>88</td>
</tr>
<tr>
<td>4.3</td>
<td>Frequency state transition matrix for SHARELIFE life events data</td>
<td>90</td>
</tr>
<tr>
<td>4.4</td>
<td>Probability state transition matrix for SHARELIFE life events data</td>
<td>90</td>
</tr>
<tr>
<td>4.5</td>
<td>Transition bubble graph showing probabilities of first-time life-course transitions among SHARELIFE respondents (N = 2,191)</td>
<td>101</td>
</tr>
<tr>
<td>4.6</td>
<td>State transition diagram graph showing sequential relationships among first-time life-course events among SHARELIFE respondents (N = 2,191)</td>
<td>102</td>
</tr>
<tr>
<td>4.7</td>
<td>Sequence index plot showing the number of SHARELIFE respondents (N = 2,191) who reported experiencing first-time life-course events in different sequence orders</td>
<td>103</td>
</tr>
<tr>
<td>4.8</td>
<td>State distribution graph showing the percent of SHARELIFE respondents (N = 2,191) reporting a given element at each sequence position</td>
<td>105</td>
</tr>
<tr>
<td>4.9</td>
<td>Sequence index plot of life event sequences from Figure 4.7 in grayscale</td>
<td>106</td>
</tr>
<tr>
<td>4.10</td>
<td>Grayscale sequence index plot of life event sequences from Figure 4.9 stretched vertically</td>
<td>107</td>
</tr>
<tr>
<td>5.1</td>
<td>Frequency state transition matrix for world-system positions</td>
<td>123</td>
</tr>
<tr>
<td>5.2</td>
<td>Dendrogram for hierarchical clustering of 42 unique world-system-position sequences from 129 countries between 1967 and 2006</td>
<td>133</td>
</tr>
</tbody>
</table>
Figures

5.3 Line graph showing the relationship between the number of clusters at given dissimilarity thresholds in the hierarchical clustering of world-system-position sequences 135

5.4 Tempograms showing the proportions of workers who were working at specific time points throughout the day in five weekday clusters, based on analysis of data on 8,997 individuals from the 2008–2010 American Time Use Surveys 141

5.5 Sequence index plots showing sequences of the 129 countries in each of the five world-system-position clusters 143

6.1 Narrative network showing the structure of the identity history of a Nazi 159

6.2 Affiliation matrix (A) showing women’s involvement in different events, as reported in the Deep South study 168

6.3 Comembership matrix (N) showing overlap between women with respect to involvement in different events in the Deep South study 169

6.4 Network showing the strengths of relationships among the women in the Davis, Gardner, and Gardner Deep South study 174

6.5 Bipartite network showing which events were attended by which of the women in the Deep South study 175

6.6 Hypothetical two-mode sequence network 177

6.7 Two-mode sequence network depicting activity sequences of two parents and their two children between 5 P.M. and midnight on a Wednesday evening in October 2000 178

6.8 Annotated directed activity sequence network for two parents and their two children between 5 P.M. and midnight on a Wednesday evening in October 2000 180

6.9 Node-free, variable-path-width depiction of the activity sequences of a U.K. family 182

6.10 Colorized variable-path-width depiction of the activity sequences of a U.K. family 183

6.11 Directed activity sequence network for two parents and their two children between 5 P.M. and midnight on a Wednesday evening in October 2000, with time-equivalent position-elements tiled along the x-axis 184

6.12 Overview of useful structural measures for two-mode sequence networks 186

6.13 Overview of useful structural measures for one-mode sequence networks 187
Figures

6.14 Affiliation matrix (A) from the Deep South study, rearranged to indicate block membership 192
6.15 Event structure analysis (ESA) diagram showing connections among elements of the labor struggle in the meatpacking industry during the Reagan era 199
6.16 Evolution of a firm’s investment network 204
6.17 Three-hour-long snapshots of a U.K. family’s two-mode sequence network, 5:00–8:00 P.M. 206
7.1 Annotated microsequence diagram showing how one woman spent the day on a Monday in May 2011 219
7.2 Tempograms showing the percent of respondents in the 2003–2011 ATUS who reported different types of contact at each 5-minute interval between 6 A.M. and midnight, shown separately for each of five clusters (N = 2,853) 221
7.3 Transition probability matrix showing switches between different types of social contacts in a single workday among working mothers in the 2003–2011 ATUS (N = 4,658) 226
7.4 Transition probability matrix showing switches between different types of social contacts in a single workday among working fathers in the 2003–2011 ATUS (N = 6,752) 226
7.5 The position of work activity throughout the day in the Spanish Daily Sequence Network 237
7.6 Annotated sequence motif illustrating points of synchrony in Spain in 2003 (uncommon transitions not shown) 240
7.7 Histogram showing distribution of average degree of routine (in minutes) between pairs of weekdays among respondents in the 2005 Time Use Survey of the Netherlands (N = 1,807) 250
B.1 The types of relationships between cases that are considered given different linkage criteria for merging separate clusters in hierarchical clustering 271
Tables

4.1 Transition matrices showing frequency of transition from having first child to experiencing first period of major stress among parents in the SHARELIFE data  
4.2 Transition matrices showing movement within the world system between 1967–1986 and 1987–2006  
5.1 The relationship between the operation cost regime and the emphasis on features being used as a basis for sequence comparison  
5.2 Country membership in the five world-system clusters  
7.1 Extent of interpersonal and generalized synchrony in individuals’ activity sequences as reported in the 2002–2003 Spanish Time Use Survey (in minutes and percent), by socio demographic characteristics (N = 20,136)  
7.2 Extent of routine in individuals’ activity sequences as reported in the 2005 Time Use Survey of the Netherlands (in minutes and percent), by socio demographic characteristics (N = 1,807)  
A.1 Studies that have analyzed whole-sequence patterns using OM and other methods since 2000
Preface

During a routine literature search a few years ago, I stumbled onto a prickly set of articles in a 2000 special issue of the journal *Sociological Methods & Research*. I was seeking methodological guidance for an analysis of the association between individuals’ stress levels and their frequency of switching between social roles and contexts. Several of the articles seemed generally relevant to what I was trying to do, so I looked at the entire issue. As I read on, I noticed that there was a measure of antagonism among some of the issue’s authors. This in itself is not unusual, as methodological debates are common in the social sciences and can lead to conflict (escalating, in some cases, to near vehicular assault in campus parking lots). Nothing so serious was going on in the case of this special issue. But the contributors were using markedly spirited terms – such as “trivial” and “silly” – to characterize each other’s contributions. The subject of this particular debate was sequence analysis.

In the research that I had done to that point, I had never used the kinds of sequence analysis methods that were being discussed in that special issue – in particular, optimal matching. But I have long been fascinated by complex dynamic social processes, so the idea of learning more about how to detect general patterns in such processes appealed to me. My methodological training in graduate school focused primarily on multivariate analysis and social network techniques, and it seemed that these would not take me where I wanted to go with my new research. (I turned out to be only half wrong about that.) I broadened my literature search, and soon discovered numerous alternative approaches that are concerned with assessing the timing and order of social phenomena. They all shared a concern with sequencing. And yet, much of that work avoided the language of sequence analysis. As the pile of relevant references on the desk in my study grew taller, I became increasingly annoyed by the fact that I could not find a single source that tied all of this work together. So,
I decided to write this book, both as a methodological reference and as a unifying conceptual framework.

This is a timely book. The growth of sequence-oriented approaches within the social sciences over the past few decades has been steady but slow. But from where I stand, it is apparent that things are about to change. The social sciences have entered a period that will likely be known in retrospect both for the sudden availability of massive streams of complex, real-time social data and for the challenge of making sense of them. The rise of computational social science, the accessibility of dynamic data, an increasing focus on real-time events and time use, and major improvements in analytic technologies call for methods that can make intuitive sense of detailed sequential data.

Perhaps even more importantly, these developments highlight the urgent need for a coherent conceptual framework that can serve larger theories about ordered social processes. The social sciences are full of well-theorized but seldom-tested ideas about the structural causes and consequences of the ordering of social events. Talcott Parsons's grand theory of a social system in which action is predictably coordinated via an interlocking set of social roles is one example. Another is Anthony Giddens's theory that everyday routine is crucial for maintaining individuals' sense of continuity and ontological security in an otherwise fast-paced world. There are many other prominent examples, some of which are addressed in the second chapter of this book. I believe that the best conceptual framework for moving these ideas forward can be found at the intersection of social network analysis and sequence analysis. This book provides the foundation for such a framework.
Acknowledgments

This book has benefited inestimably from the encouragement and insights of many generous friends and colleagues. I will begin with those who had the most direct role in bringing this book to the light of publication. First, I owe a great debt to Robert Dreesen, senior commissioning editor at Cambridge University Press, for seeing the potential of this idea and encouraging me to pursue it as a book project. Mark Granovetter, editor of the Structural Analysis in the Social Science series, shepherded the book through the drafting and revision processes and provided indispensable feedback that shaped the book along the way. This book was only possible due to their highly responsive support, expertise, and good cheer. I also thank Brianda Reyes for providing able editorial assistance throughout the production process.

Cornell University has been an ideal setting for the development of this project. For one, because Cornell is rife with network researchers, there is a constant demand here for fresh and interesting network-related ideas. The basic notion of studying sequences as networks has met with great enthusiasm. Cornell also provided substantial institutional support. I am grateful to Cornell’s Institute for the Social Sciences, in conjunction with the Department of Sociology, for funding my leave from teaching in the spring semester of 2013, during which the initial draft of this book was written. The opportunity to offer a new graduate course on social sequence analysis at Cornell also benefited this project tremendously by providing me with the impetus to develop and refine many of the ideas that are presented here. The course was first offered in spring 2012 at Cornell University, attended by Rachel Behler, Chris Cameron, Dan DellaPosta, Michael Genkin, Ningzi Li, Noona Oh, Kelly Lee Patterson, Victoria Sosik, and Dana Warmsley. Their critical engagement with the material and our discussions about potential applications had an immeasurable impact on my approach to this subject. Dan DellaPosta also provided valuable research assistance later,
including tracking down references to empirical examples of whole sequence comparison techniques.

In many ways, this book has its roots in the Department of Sociology at the University of Chicago. As my advisor, mentor, and now close friend, Ed Laumann has nurtured in me an appreciation for the dynamic properties of social structure. He has continued to challenge me to think about and model these in terms of networks. Many of the new contributions of this book reflect his influence on me as a scholar. Obviously, the very topic of this book owes much to Andy Abbott and his decades of work in bringing sequence analysis to the social sciences. Throughout the development of this manuscript, Andy has been gracious in discussing sequence analysis methods themselves and in providing advice about some potential references and contacts.

My engagement with the topic of microsequences specifically was partly inspired by the work of Jay Gershuny, Director of the Centre for Time Use Research (CTUR) at the University of Oxford. He and his colleagues – including Kimberly Fisher, Teresa Harms, and Oriel Sullivan – graciously hosted me at the CTUR at St. Hugh’s College in the fall of 2014. That was the beginning of what is sure to be a fruitful collaboration. This part of the book also benefited from the comments and suggestions of Matt Brashears, Tom Buchanan, Jessica Collett, Karen Danna-Lynch, Ed Lawler, Michael Macy, Brian Rubineau, Jeremy Schulz, and participants at the Perspectives on Time Use in the U.S. Conference at the U.S. Bureau of Labor Statistics in Washington, DC, in June 2014; the annual meeting of the American Sociological Association in Atlanta, Georgia in 2010; and the Cornell Population Program seminar series in March 2010.

A variety of colleagues played important roles in the development of this book by talking through sundry sequence-related ideas, helping me to recognize the theoretical and methodological scope of the work, and pointing me to related lines of research. At Cornell, Richard Swedberg humored me by engaging in discussions about the deeper theoretical relevance of sequential social phenomena, such as in the work of Talcott Parsons. Steve Morgan saw promise in this idea early on and arranged my initial introduction to Cambridge University Press. I am also grateful to Jim Moody for first introducing me to network analysis in general and the idea of narrative networks in my first year as a graduate student, more than a decade ago. Other scholars who discussed aspects of this work, responded to queries, supplied data, or otherwise engaged in exchanges about sequence analysis include Jason Beckfield, Cliff Brown, John Brueggemann, Chris Marcum, and Kate Stovel. The anonymous reviewers of the manuscript draft also made many wise suggestions that improved this book.
Acknowledgments

Finally, I could not have done this without the understanding and patience of those closest to me. My extended family unwittingly did a great service simply by forcing me to explain my ideas to them in plain language. I am also particularly grateful to my wife, Erin York Cornwell, who provided infinite support and advice. Her influence on this work has been immense, and is absorbed into nearly every passage of this book. Erin provides the ultimate combination of wisdom, companionship, and inspiration. I cannot imagine getting through such an undertaking as this without knowing that there is someone like her, and our Watson, nearby.