

Cambridge University Press
978-1-107-10250-7 - Social Sequence Analysis: Methods and Applications
Benjamin Cornwell
Excerpt
[More information](#)

Part I

Introduction

Cambridge University Press

978-1-107-10250-7 - Social Sequence Analysis: Methods and Applications

Benjamin Cornwell

Excerpt

[More information](#)

1

Sequence Analysis in the Social Sciences

Social actors of all kinds – individuals, families, organizations – are connected to each other and to larger society through ordered elements, states, and events. To be socially connected is to experience fluctuating environments, evolving relationships, shifting preferences and strategies, changing statuses, and constant flows of information and resources. As several important social science theories suggest, examining the timing and order of these elements of social life is the only way to understand social structure (e.g., Bales 1951; Blumer 1969; Bourdieu 1984; Gershuny 2000; Giddens 1984; Parsons 1951; Sacks 1995; Schutz [1932] 1967; Strauss 1993; White 2008; Zerubavel 1981).

And yet, in contemporary social science, the effort to describe and analyze the sequential nature of social reality is often regarded as somewhat of an “offshore operation” (Abbott 2000:66) – something that is done by a few enthusiasts who are engaged in peripheral or exotic research. The overarching goal of this book is to change that perception. It does this both by pointing out that many scholars are already engaged in sequence analysis (often without realizing it) and by promulgating conventional and new sequence approaches.

Researchers are becoming both increasingly interested in and capable of analyzing sequential social phenomena. This book describes some of the most widely used methods that have been developed to study these dynamics and introduces some new techniques and applications as well. Sequence analysis takes a variety of forms in the social sciences, ranging from statistical Markovian and survival analysis approaches to conversation analysis techniques. These are all extraordinarily valuable approaches. This book will focus primarily on another class of sequence methods, which treat whole sequences, as well as larger networks of sequences, as units of study. These approaches allow researchers to study, measure, classify, and visualize sequences of social phenomena. I view these methods – which include sequence description techniques, optimal

4 *Sequence Analysis in the Social Sciences*

matching analysis, cluster analysis, sequence-network analysis, event structure analysis, and social microsequence analyses – as one part of the larger body of *social sequence analysis* methods.

Unfortunately, there are few reference texts available to social scientists who wish to learn (or teach) these particular aspects of social sequence analysis. Furthermore, few texts catalog the range of approaches that are used across the social sciences. This book serves these purposes. Given the recent growth of interest in sequence-oriented questions, data, and methods throughout the social sciences (which are discussed in Chapter 2), there is a growing need to document in one place both the theoretical perspectives that motivate sequence analyses of social phenomena and the methods and applications that have been developed in that spirit. This book will provide the theoretical foundations for these methods throughout, but it will also function as a primer and reference for social sequence analysts.

1.1 Timing and Context of the Book

Scholars in psychology, economics, anthropology, demography, communication, sociology, computational social science, and political science have been using various kinds of sequence methods for decades. Whole-sequence comparison methods were first imported into the social sciences from the biological sciences (inspired largely by efforts to analyze DNA sequences) in the 1980s, and they have since developed in ways that are unique to the social sciences. Sociologists, for example, have used sequence methods to understand the interaction process, life-course and career trajectories, hierarchy formation, and a variety of personal and societal historical processes (Abbott 1995; Aisenbrey and Fasang 2010; Bearman, Faris, and Moody 1999; Blair-Loy 1999; Blanchard, Bühlmann, and Gauthier 2014; Brzinsky-Fay 2007; Chase 1980; Gauthier et al. 2010; Griffin 1993; Han and Moen 1999b; Joseph et al. 2012; Pollock, Antcliff, and Ralphs 2002; Stovel 2001, 2010; Stovel, Savage, and Bearman 1996). As another example, psychologists have used other sequence methods to study how the order of information affects learning and to identify structure in interactions between individuals (Bakeman and Gottman 1997; Clegg, DiGirolamo, and Keele 1998; Cohen, Ivry, and Keele 1990; Gottman and Roy 1990; Ritter et al. 2007). There are many other recent applications in other social science disciplines as well, which will be covered in greater depth in Chapter 2.

There has been a surge of interest in sequence methods in the social sciences during the past decade, especially in my home discipline of sociology. This has led to some major advances in the sequence methods that are covered in this book. Many of the methodological developments

1.1 Timing and Context of the Book

5

in this branch of social sequence analysis came on the heels of a 2000 special issue devoted to the topic in *Sociological Methods & Research*, which saw an unusually spirited debate (in which the contributors used terms such as “trivial,” “silly,” and “vehement” to characterize each other’s positions) over methods that are designed to compare the structures of whole sequences. That debate has inspired a number of methodological innovations in an effort to address some shortcomings of earlier sequence comparison methods. Enough improvements were made over the ensuing decade that in 2010 *Sociological Methods & Research* organized another special issue on social sequence analysis. These papers report on what some scholars call the “second wave” of sequence analysis (Aisenbrey and Fasang 2010), which involves new methods and applications to a greater range of social phenomena. Following this, papers presented at the 2012 Lausanne Conference on Sequence Analysis produced additional important updates and suggestions for new approaches. These were compiled and published in Springer’s Life Course Research series as *Advances in Sequence Analysis: Methods, Theories and Applications* (Blanchard, Bühlmann, and Gauthier 2014). The methods that are addressed in these works will be covered in this book.

This book is an effort to collect and present these recent developments in one place, and to contextualize these methods relative to other sequence methods that are used in the social sciences. Social scientists in different disciplines have used different terms to refer to this set of methods, including “sequential analysis” and simply “sequence analysis.” I use the phrase *social sequence analysis* in the interest of developing an interdisciplinary term that nonetheless distinguishes this body of methods from those used in the biological sciences and other fields (Stark and Vedres 2006; Vedres and Stark 2010).

One reason for the growing interest in sequence methods is that there have been drastic improvements over the past decade in the technologies for capturing and analyzing sequenced social data. For example, an emerging alternative to survey and direct observational techniques is real-time data capture (Croushore 2011; Stone et al. 2007), which includes methods such as *ecological momentary assessment*, or EMA (e.g., see Shiffman, Stone, and Hufford 2008), and *experience sampling* (e.g., see Hektner, Schmidt, and Csikszentmihalyi 2007). Technology has made the collection of continuous real-time sequence data easier and less burdensome for respondents. Mobile phone applications that passively record respondents’ physical activity and location, social contacts, electronic communications, and other real-time phenomena are increasingly available (see Kim et al. 2012; Lester, Choudhury, and Borriellol 2006). Likewise, “scraping” technology now automates the collection of sequenced data online, including real-time streams of communications such as emails and tweets (e.g., see Marres and Weltevrede 2013).

6 *Sequence Analysis in the Social Sciences*

For many years, scholars felt that the kinds of data that are developed through these methods are too detailed and too unwieldy to properly analyze or describe. But recent improvements in computer processing and software have produced a number of serviceable sequence analysis packages for popular programs such as Stata and R that make it relatively easy for scholars to perform sequence analyses. It is now possible to handle comparisons of large numbers of sequences, in much the same way that it is now possible to analyze the mammoth social network structures that were difficult to study before the 1990s.

The recent explosion of longitudinal and time-sensitive datasets and the rapid development of software that facilitates their analysis reflect (and hopefully inspire) social scientists' growing fascination with the temporal contexts in which social phenomena occur. It is no longer enough to examine the social statuses actors possess, the connections they maintain, or even the causes and consequences of these social facts. The new approach in social science is to view these as elements of larger transitions, trajectories, and sequences that unfold over time. This has shifted the focus to the social implications of *processes*. Lately, then, the goal of much recent research has been to understand the order in which social phenomena unfold, the common patterns of sequenced phenomena, and their origins and significance.

Social scientists have, for the first time, both a wealth of available sequence data and the tools that are needed to analyze them. Numerous datasets – such as the British Household Panel Survey (BHPS) and the Panel Study of Income Dynamics (PSID) – provide detailed individual-level data on lengthy career and marital histories. There has also been an explosion of interest in time-use datasets. Researchers throughout the world have been collecting detailed information about how people spend their time – what they do, with whom, where, and when – throughout the day. In the United States, for example, the Bureau of Labor Statistics (BLS) has conducted the American Time Use Survey (ATUS) every year since 2003. This undertaking has resulted in a dataset that contains retrospective twenty-four-hour time diary records for more than 148,000 Americans from 2003 to 2013.

However, sequence-oriented data collection and analysis projects have grown most rapidly elsewhere in the world. Indeed, the existence of time-use data projects in other countries has led to the creation of a huge international repository for detailed social microsequence data. The organization that has played the biggest role in organizing this research is the Centre for Time Use Research (CTUR) at the University of Oxford. The CTUR initiated and manages the Multinational Time Use Study (MTUS), which is a conglomeration of more than sixty harmonized time-use datasets that have been collected since the 1970s and numerous other historical and special time-use data sources (see Fisher et al. 2012). These

1.2 Contributions

7

datasets yield volumes of extraordinarily detailed data on the everyday activities individuals do throughout the day. As of this writing, the CTUR is a gateway to more than half a million individuals' time diaries from scores of countries. This provides social scientists with ample opportunities for exploratory and comparative research on time use and recurrent activity sequences throughout the world. These data will be discussed at several points in this book.

Researchers are also increasingly interested in over-time processes among larger social units, including organizations (e.g., firms), neighborhoods, cities, and nations – as well as changes in the networks that connect these units to each other (e.g., transportation or exchange networks). Time-series data that are organized around these units of analysis provide valuable insights into large-scale social-ecological and developmental processes. For example, one of the foundational research traditions in sociology concerned the evolution of cities. Theories in this tradition argued that metropolitan growth must be understood in terms of patterns of residential mobility and the sequence of changes in neighborhood characteristics such as ethnic composition, property values, and the presence of social institutions. To provide another example, researchers in political science are interested in national trends in attitudes, values, and behaviors. Many data on these topics can be accessed directly through online repositories such as the Inter-University Consortium for Political and Social Research (ICPSR). Data on trends concerning neighborhoods, metropolitan areas, and nations can also be built relatively easily using aggregate data from available sources such as census records, the World Values Survey (WVS), and the International Social Survey Programme (ISSP). Thus, the data that are needed to study foundational questions concerning the sequential order of social phenomena are increasingly available throughout the social sciences.

1.2 Contributions

This book describes a set of flexible analytic techniques that can be employed to understand a variety of social processes. Like social network analysis, social sequence analysis is largely agnostic with respect to content and time scale, and it can be applied to questions that concern social scientists in different fields. As such, this book is intended for a diverse social science audience. The methods described here can be employed by scholars who study life-course transitions, career paths, work/family schedules, micro (e.g., individual) or macro (e.g., economic) stages of development, conversation and interaction structure and process, contact networks, and network change, as well as processes such as diffusion, evolution, and hierarchy formation. To underscore their

8 *Sequence Analysis in the Social Sciences*

broad applicability, I illustrate sequence analysis techniques using data from a variety of sources and with different foci – micro and macro, individual-level and national, short-term and long-term processes alike.

This book makes several contributions to social sequence analysis. First, it packages sequence analysis methods from different disciplines together as a family of methods. Because few works in the social sciences (apart from theoretical reviews) have addressed or used more than one or two forms of sequence analysis at a time, there are no comprehensive reference texts or primers on this topic specifically for social scientists. This book thus provides an overview of fundamental forms of sequence analysis, including basic definitions and notation, various descriptive and visualization techniques, more advanced forms of sequence analysis – such as whole sequence pattern detection, optimal matching, and cluster analysis techniques – as well as theoretical rationales for using these approaches. This book discusses the advantages, disadvantages, and limitations of different techniques and provides solutions to common problems, applications, and a variety of colorful empirical illustrations. It also provides an overview of data sources and software packages. Where possible, I identify the common threads, similarities, and complementary features that link together sequence methods that are typically used for different applications.

Second, this book expands on existing social sequence analysis methods by highlighting some new techniques. One involves treating sequence elements as nodes in a network, in which social actors and dynamic elements of the social world are tied together. This approach opens up an entire set of methods for examining linkages between actors, social elements, and the larger sequences they compose. This “sequence-network” framework, as it might be called, provides a different approach to exploring the connections among subjects (e.g., people) and elements of sequential social processes (e.g., everyday activities). Many sequence techniques work to identify common sequence patterns by examining differences and dissimilarities among multiple actors’ sequences. Social network approaches bring to sequence analysis a greater emphasis on *connections* that exist between these actors. Network methods provide new ways to examine how sequence elements relate to each other as well. Sequences are usually treated as linear constructs that are composed of sets of adjacent elements that are stochastically related. A network approach affords more flexibility in tracing connections that exist between both adjacent and nonadjacent elements. Reciprocally, social sequence approaches bring to network analysis a greater concern about time and order, and new ways to measure it. The interplay between these methods is given considerable attention later.

Third, this book describes new applications of sequence methods to old social science problems. To illustrate, I devote Chapter 7 to the

1.3 Audience and Scope

9

application of social sequence analysis techniques to the analysis of social microsequences, or the chains of contacts, settings, and experiences that social actors forge over the course of minutes, hours, and days (Gershuny 2000). The spirit of narrative explanation for microsocial phenomena has long been alive in methods such as conversation analysis (Sacks 1995; Schegloff 2007), interaction process analysis (Bales 1951), and dramaturgical analysis (Goffman 1959). With the aid of sequence analysis techniques that are typically used to study other phenomena (e.g., life-course and career trajectories), we can gain fresh insight into important but understudied phenomena such as interpersonal synchrony and the routine nature of everyday social interactions.

Throughout this book I remain mindful of the value of visualization. I agree with Edward R. Tufte (1983) that good visual aids are often the most effective and efficient way to tell a complex story and to understand it (see also Healy and Moody 2014; Hofstadter 1979). As with social network data, social sequence data are inherently complex. One quickly finds that visual aids are indispensable tools for conveying social structure. As with social network analysis, “visualization is a central feature of discovery” in the analysis of sequential social phenomena (Healy and Moody 2014:106). Thus, an ancillary contribution of this book is that it provides some new ideas for visualizing sequenced social phenomena. Where warranted, I present figures in color. But in light of the constraints of many print-based outlets, I make several suggestions for maximizing the effectiveness of grayscale visualization.

Ultimately, this book is an attempt to draw social scientists’ attention to a growing and exciting set of research methods – an invitation to social scientists to adopt a sequence framework for studying social phenomena. Of course, this book is not an exhaustive account of all conceivable social sequence methods, some of which are more handily discussed by statisticians, qualitative researchers, and historians. I take up this issue in the next section.

1.3 Audience and Scope

Who should read this book? I initially approached the planning and writing of this book with junior colleagues and graduate students in mind. Thus, I imagined the steps necessary to introduce an intrigued scholar to social sequence analysis – first by explaining why it is useful, then introducing sequence methods gradually via preliminary descriptive measures and tests, then moving on to more sophisticated whole-sequence-pattern search techniques, and finally describing brand new applications. But this book is also potentially valuable to a general audience of senior colleagues who are interested in sequence analysis but do not know much

10 *Sequence Analysis in the Social Sciences*

about it. It should be of use to anyone who is curious about the possibilities of seeing social life through the lens of sequences, and of the prospects of using sequence methods to understand dynamic social phenomena in whatever context.

In short, this book is intended as a primer on sequence analysis for social scientists of various stripes. It is ideal both for graduate-level courses that address the topic and as a bookshelf reference for scholars who are interested in studying sequenced social phenomena. At the same time, this book addresses cutting-edge sequence methods, and thus speaks to veteran sequence analysts (and veteran network analysts) who are interested in new ideas. It is intended as an update on methods that have been presented in recent issues of prominent methods journals, and it suggests some new techniques that have not been covered in the literature.

There are several parts of this book that will be of particular interest to scholars who work in certain areas. Because of the integration of network and sequence approaches in this book, some of the methods that are described here will be of interest to network researchers – especially those whose concern is social network dynamics. Scholars are increasingly cognizant of the fact that social networks are constantly in flux, and that real-time network changes have important implications for social actors (e.g., Doreian et al. 1996; Jackson and Rogers 2007; Snijders, van de Bunt, and Steglich 2010; Spiro, Acton, and Butts 2013; Toivonen et al. 2009). Sequence analysis provides some useful methods for doing this, especially for seeing patterns in abstruse network change data (e.g., see Stark and Vedres 2006). This book may also be of interest to network researchers simply because it shows how sequence analysis can be done using social network techniques (which, for some, are more familiar). As such, the book may provide new avenues of exploration for social network analysts. Chapters 6 and 7 are particularly relevant to network scholars.

Portions of this book will be of particular interest to scholars who work on time use, microinteraction, small-group processes, the sociology of time, and related fields. Chapter 7 deals specifically with the topic of microsequences (Collins 2004; Gershuny 2000), which refer to the chains of interactions, social contacts, and contexts that people experience during the course of a bounded period of time, usually a day or week. This book's treatment of microsequence data is intended as a comment on work that treats time solely as something that is "allocated," "budgeted," or "used," as if it is a divisible substance or commodity. The goal is to encourage scholars to see the value in viewing time as something that flows and links together – or sequences – elements of social experience (Emirbayer 1997). Scholars who work on the issue of time can get more mileage out of time-specific data (as are available in time diaries and