

Network Analysis of Sequence Structures



Benjamin Cornwell

1 From Sequence Pathways to Sequence-Networks

Many social sequences can be analyzed as independent linear chains of events. For example, a given firm's trajectory of growth can be analyzed as a set of developmental stages that may be similar to, but not connected to, other firms' experiences. A common goal in recent analyses of such phenomena is to describe multiple chains and compare them to each other, then classify them or describe their degree of dissimilarity (e.g., see Aisenbrey and Fasang 2010; Piccarreta 2017; Studer et al. 2011). Expanding on the interest of scholars in the relational nature of social phenomena, some social scientists (e.g. Bearman et al. 1999; Bearman and Stovel 2000; Bison 2014; Cornwell 2015) have begun to explore sequences in terms of sets of intersecting events that constitute a larger network of pathways, in which multiple sequence chains are inextricably entwined and not considered as separate. These larger structures of intersecting sequences can be called “sequence-networks.” See also Hamberger (2018) in this bundle.

Researchers who study sequence-networks are often interested in characterizing the overall landscape of interconnectedness and intersection—considering these structures as whole standalone entities, opportunity structures (Merton 1996), or social systems (Parsons 1951). Visually speaking, this involves seeing sequences as a vast, integrated web which, in two-dimensional space, resembles a roadmap. The ordered nature of the phenomena that make up these structures naturally invites the use of existing sequence-analytic tools such as discrepancy analysis (Studer

B. Cornwell (✉)

Department of Sociology, Cornell University, Ithaca, NY, USA

e-mail: btc49@cornell.edu

© The Author(s) 2018

G. Ritschard, M. Studer (eds.), *Sequence Analysis and Related Approaches*,

Life Course Research and Social Policies 10,

https://doi.org/10.1007/978-3-319-95420-2_7

103

et al. 2011) and parallel coordinates plot (Bürgin and Ritschard 2014) to help describe them. This has also prompted more focused investigation of the structure of whole sequence-networks, as well as the role that particular elements within those structures play in tying those networks together.

Toward that end, this chapter presents a set of explicitly network-analytic methods for understanding the larger structure that reflects the connections that exist between the individual sequences that one observes within a given setting. This approach expands on other useful sequence-oriented tools. The notion that social life is a set of sequenced phenomena still holds. But the question now shifts to collectivist questions such as “Is the set of action pathways forged by group A more complex than that forged by group B?” I argue that network methods can offer unique insight into these kinds of issues and therefore provide a useful supplement to existing sequence methods. The goal of this chapter, then, is to outline some affinities between network analysis and sequence analysis, and how the former can be used to enhance the latter. To illustrate, this chapter maps out the complex network of sequence pathways by which individuals get through everyday life. To highlight the capacity for network methods to help shed light on the overall structure of these larger maps, it then compares the complex landscape of interconnected pathways that characterize the everyday lives of younger and older adults, respectively.

2 Sequence Pathways in Everyday Life

What does a sequence-network look like? A context that invites this approach is that of individual-level behavior that unfolds on the hourly or daily time scale. How individuals order everyday action has long been an issue in the social sciences (e.g. Bales 1951). Microsequential activity (Gershuny 2000) is composed of series of successive acts (e.g., talking, eating) that form temporally ordered chains. These sequences contain activities that span a specific period of time, usually a 24-h day. Social actions that occur in a given individual’s schedule on a given day—e.g., “I took a shower, got dressed, and then went to work”—are temporally connected and therefore provide the basis for larger chains of social action. These are not trivial sequences, as the way people sequence their activities during the course of the day provides a key basis for social organization, and reflects their positions within a larger system of social roles and social norms (e.g. Giddens 1984; Merton 1968; Parsons 1951). Thus, understanding how sets of individuals order their everyday activities—and how those activities are related to each other—is vital to understanding larger social structure.

2.1 Activity Sequences in Networks

Assume that we are interested in the activity sequences reported by g individuals. Assume also that we have activity sequence data from each of these individuals that describe which of the k elements (e.g., number of different types of activities) each individual was doing at each of t time points (e.g., 1,440 min in a given day). Here, the activity data for a given set of individuals are observed in one-minute intervals. An element in a given person's activity sequence may be: "Working between 12:00 p.m. and 12:01 p.m." That particular time-activity will then be followed by another time-activity, such as: "Eating between 12:01 p.m. and 12:02 p.m."

These sequence data can be treated as network data. Every network has two aspects: Nodes, or vertices, and the links, or ties, that associate specific pairs of nodes with each other (Wasserman and Faust 1994). Here, the activities that occur at specific times—in other words, the activity-time elements are treated as the nodes in the network. The "links" between them are temporal connections, which are inferred from their adjacency in the sequence of events that is observed in one or more individuals' activity sequences.

2.2 Organizing the Data as a Sequence-Network

The ordered activity sequence data that are contained in a typical time-diary dataset can be recorded in a rectangular matrix, which we will call A . In A , the g individuals are arranged down the rows of the matrix and the $k \times t$ activity-time combinations described above are arranged along the matrix's columns. In network terms, this is known as an affiliation or "2-mode" network matrix (Borgatti and Everett 1997; Wasserman and Faust 1994).¹ The left side of Fig. 1 shows how this matrix would be constructed using a simple hypothetical sample of 10 individuals (1 thru 10), who are observed for three time periods (1, 2, and 3) for evidence regarding which of four activities (A, B, C, and D) they were doing at each time point. Each cell contains information about whether the individual in the corresponding row reported doing the specific activity at the specific time that is designated in the corresponding column. For example, a "1" will appear in the upper-left-most cell if individual 1 reported doing activity A in time period 1. That cell would contain a "0" otherwise.

Note that matrix A does not yet show how the time-activity elements are linked to each other in a chain. In network-analytic terms, sequential order can be captured in what is referred to as a "directed" network. One can transform matrix A (which

¹The network framework is very flexible and can be used in different ways. One might fashion a sequence-network that is composed of multiple sequences from a single individual. For example, individuals' activity sequences from different days might interest, thus providing insight into the emergence of routine in everyday life. Alternatively, one can use networks to show how different activities that occur in potentially non-adjacent time periods link together for a given individual.

Matrix **A**: Which individuals did which activities at which time points

		Time-Activities											
		Time											
		1				2				3			
		Activity				Activity				Activity			
		A	B	C	D	A	B	C	D	A	B	C	D
Subjects	1	1	0	0	0	1	0	0	0	1	0	0	0
	2	1	0	0	0	1	0	0	0	0	1	0	0
	3	1	0	0	0	0	1	0	0	0	0	0	1
	4	0	1	0	0	0	0	1	0	1	0	0	0
	5	0	1	0	0	0	0	1	0	0	0	1	0
	6	0	0	1	0	0	0	1	0	0	0	1	0
	7	0	0	0	1	1	0	0	0	1	0	0	0
	8	0	0	0	1	1	0	0	0	0	1	0	0
	9	0	0	0	1	0	1	0	0	0	0	0	1
	10	0	0	0	1	0	0	0	1	0	0	0	1

Matrix **B**: Frequencies of transitions between successive time-activities in the activity sequences shown in matrix **A**

		Time-Activities											
		1A	1B	1C	1D	2A	2B	2C	2D	3A	3B	3C	3D
Time-Activities	1A	-	-	-	-	2	1	0	0	-	-	-	-
	1B	-	-	-	-	0	0	2	0	-	-	-	-
	1C	-	-	-	-	0	0	1	0	-	-	-	-
	1D	-	-	-	-	2	1	0	1	-	-	-	-
	2A	-	-	-	-	-	-	-	-	2	2	0	0
	2B	-	-	-	-	-	-	-	-	0	0	0	2
	2C	-	-	-	-	-	-	-	-	1	0	2	0
	2D	-	-	-	-	-	-	-	-	0	0	0	1
	3A	-	-	-	-	-	-	-	-	-	-	-	-
	3B	-	-	-	-	-	-	-	-	-	-	-	-
	3C	-	-	-	-	-	-	-	-	-	-	-	-
	3D	-	-	-	-	-	-	-	-	-	-	-	-

Fig. 1 Sequence-network matrices **A** and **B**, describing a hypothetical set of ten individuals who engaged in three potential activities (A, B, and C) at four specific time points (1, 2, 3, and 4)

is an undirected affiliation matrix) into a directed matrix that reflects the ordered relationships among the activities. The cells in the resulting matrix reflect how many times a given pair of successive time-activities—for example, 1A and 2B—occur in order across the different sequences. The cells of this new matrix are valued, so that relationships between pairs of successive activities—for example, “Working between 12:00 p.m. and 12:01 p.m.” and “Eating between 12:01 p.m. and 12:02 p.m.”—are expressed in terms of the number of individuals who reported this transition.

The example from the left side of Fig. 1 is carried over into the right side, which shows the contents of matrix **B** for this group. For the purposes of constructing the directed sequence-network, we null all of the blocks along the diagonal (which would represent connections between elements in the same time period, or simultaneous activities), the blocks below the diagonal (which would represent time in reverse), and most of the blocks above the diagonal (which represent links between elements in non-adjacent time periods). This leaves us with the partial, asymmetric matrix that is shown, which contains only the shaded regions. The valid cells in matrix **B** tell us in how many individuals’ sequences a given activity at a given time is followed by another specific activity at the subsequent time. This is therefore a “valued” network matrix. For example, this matrix would suggest that of these 10 people, two transitioned from doing activity D at time 1 to doing activity A at time 2.

Note that matrix B essentially consists of a set of nested first-order transition matrices, but one could expand to include higher-order transition matrices in the nulled regions above the diagonal. There are also numerous ways one could manipulate the values in matrix B , such as by dichotomizing at a given level or by normalizing by row and/or column values. The methods presented here are intentionally left simple and accessible. For an overview of the analysis of temporal networks, see Batagelj et al. (2014).

The following sections describe how the network-analytic framework can be utilized to understand this complex system and what it can reveal about the sequential structure of social phenomena that conventional sequence methods cannot reveal.

3 Analyzing Sequence-Network Structure

As described above, the network framework treats sequences as integrated sets of intersecting pathways. The adoption of this framework does not imply an abandonment of sequence concepts. Rather, it shifts the focus from the individual chains and transitions to the larger structure that results. This system has properties that, once analyzed, reveal additional information that should be of interest to sequence analysts. This section describes some of the properties one might be interested in along these lines (along with some discussion of sequence approaches that share a similar focus), and how to compare those properties across sets of sequences. Note that all of the analyses that were conducted here were executed using Stata 14.2 (StataCorp 2015), but most of these analyses can be conducted using publicly available network software, such as Ucinet (Borgatti et al. 2002) or in R.

3.1 Describing Sequence-Network Structure

I begin by highlighting some affinities that exist between some network concepts and sequence ideas. Table 1 provides a non-exhaustive list of sequential properties (in the middle column) that can be analyzed using an integrated network framework. To give a simple example and to help fix ideas, the concept of *network diameter* reflects the distance between the most distant pair of nodes in the network (that is, how many “steps” they are from each other in terms of the presence of intervening nodes in the observed sequence). In a sequence, this corresponds to the length of the longest observed sequence. Thus, assuming that all of the observed sequences include the same number of time periods, the network diameter is simply the number of time periods being observed.

Network size is the most fundamental network-level property, as it reflects how many nodes are involved. In the sequence context, network size is equivalent to

Table 1 Affinities between key network concepts and sequence concepts, and a comparison of their measurement in the observed sequence-networks of 219 younger and 688 older adults in the ATUS, respectively

Network concept	Corresponding sequence phenomenon	Observed levels in sequence-networks ^a	
		Younger adults	Older adults
<i>Diameter</i> —The largest geodesic distance between any two nodes in the network	The length of the longest observed sequence (number of time periods observed)	1,440	1,440 ^b
<i>Size</i> —The total number of nodes in the network	The boundary of the observed element universe (number of unique activity-times observed)	15,489	14,181** (13,442, 15,107)
<i>Density</i> —The proportion of connections that are observed among all pairs of nodes that could be connected ^c	The extent to which alternative possible sequence pathways emerge at a given time (as opposed to a circumstance in which activities at a given time tend to converge on just one activity at the subsequent time)	0.045	.040*** (.037, .042)
<i>Centralization</i> —The extent to which certain ties are stronger than others	The extent to which particular sequence pathways are more common (some transitions between pairs of specific activity-times are experienced by many, others by few)	35.535	40.250*** (38.782, 41.887)
<i>Homophily</i> —The extent to which nodes are connected to similar types of nodes as opposed to different types of nodes	The extent to which particular sequence pathways are composed of like elements (activities tend to remain the same over successive time periods)	0.986	.989*** (.988, .990)

* $p < .05$, ** $p < .01$, *** $p < .001$

^aValues for younger adults are based on their 219-person network. Values for older adults are derived from the bootstrapping procedure involving 1,000 samples, each with 219 randomly selected older adults. Corresponding 99% confidence intervals for older adults appear in parentheses below the mean.

^bThis value does not differ between groups because the same number of time periods was observed for both groups.

^cNetwork density is the total number of different types of transitions that were observed divided by the total number that could have occurred. Between any two successive time points, there are $17 \times 17 = 289$ possible transitions (e.g., eating \rightarrow working). With 1,440 time points, that means $1,439 \times 289 = 415,871$ possible transitions per group. This is the denominator for calculating network density.

the number of unique activity-time elements that are observed. This can also be expressed as a proportion of the possible element universe. With respect to activity sequence-networks, a greater number of nodes means that there are more possible pathways through the day, and thus more sequence complexity.

Network density is a property that refers to the extent to which all of the nodes which could be connected within a network are indeed connected. In a sequence context, greater density means that more of the available pathways which could be followed are indeed being followed in reality (to some extent). With regard to activity sequences, greater density means that alternative pathways between time periods are being followed by different people—as opposed to a circumstance in which the sequence are characterized by dominant pathways that many people follow. All else equal, lower levels of density mean that there are numerous pathways that are not being followed.

A related concept is *centralization*. Centrality can refer to the number of connections a given node has or how dominant a given pathway is. Those that are more central tend to be more influential or prominent in the context in question (Wasserman and Faust 1994). For example, a more central node corresponds to an activity that serves as a junction between many other preceding and subsequent activities. This corresponds to what (White 1995) refers to as “publics.” Likewise, central ties are those “wide” pathways on which numerous actors are observed at a given time (i.e., common transitions). “Centralization” is the network-level extension of this idea, as it refers to a network in which pathways tend to converge on central events or where certain pathways are more dominant than others. Here, I focus on path centralization, which can be measured simply as the standard deviation around the average width of paths that are observed in the network. (Note that there are numerous other measures of centrality—such as betweenness centrality—that may be employed to understand sequence structure and how particular elements fit into it. But due to space constraints I focus here only on path centralization.)

A final concept is *network homophily*, which refers to the extent to which like elements (e.g., people of the same race) form connections with each other as opposed to with unlike others (McPherson et al. 2001). Sequence-networks are “partitioned” networks, meaning that the nodes have different properties—where temporally adjacent elements may or may not be similar. In this context, homophily can refer to the extent to which sequence elements tend to be followed by like elements in subsequent time period (i.e., a “run” or “spell”). There are numerous ways to measure this, but one straightforward approach is to simply calculate the proportion of times a given sequence element (e.g., sleeping) at one time point is followed by the same element at the next time point, as opposed to transitioning to a different element (e.g., eating).

There are numerous other potentially relevant network concepts one could consider, but measuring the concepts that are described in Table 1 will help to demonstrate how a network approach can help to describe overall sequential structure. To be sure, there are more conventional sequence-analytic approaches one could take to measure related properties of sequence-network behavior. Dissimilarity-based

discrepancy approaches can be used, for example, to assess the diversity of pathways followed by different groups (e.g., see Studer et al. 2011). Likewise, something akin to the homophily measure that was just described can be derived using a measure of the mean of the sequence entropies for subjects belonging to different social groups. Indeed, just about any measure that is derived from matrix \mathbf{B} can be derived from a first-order Markov model.²

Another benefit of treating sequences as networks is that it facilitates a different type of visualization of sequence pathways. Visualization aids in the comparison of the larger systems of action that characterize different groups. One conventional approach is to present a sequence index plot (e.g., Brzinsky-Fay et al. 2006; Gabadinho et al. 2011) or, more germane to the notion of linking elements across time points, parallel coordinate plots (Bürgin and Ritschard 2014). For recent developments in the visualization of sequences, see e.g. Fasang and Liao (2014) and Gabadinho and Ritschard (2013). Network diagrams that use spring-embedding algorithms (Mrvar and Batagelj 1996, e.g., see) can be useful as well by allowing the positioning of particular activity-times within the plot to vary on the vertical axis, thus highlighting how their relationships to other particular types of elements (e.g., another type of activity) depends on the element position/time. This reflects a core network-analytic argument, which is that the significance and meaning of a given node or element depends on the nodes or elements to which it is connected and its position within the larger structure (Wellman 1983). This will also be illustrated momentarily, in a representation of the daily action sequences of younger and older individuals, respectively.

3.2 *Comparing Sequence-Networks*

A key concern for many sequence analysts is whether these properties of sequential action differ between groups. For example, do older adults tend to enact more homogenous activity sequences that are characterized by dominant pathways than do younger adults? Conceptually, this involves testing whether older adults' sequence-networks are smaller, evince less variation with respect to tie strength, and have lower density.

But the network framework poses an interesting challenge in this respect, as theoretically a given group (e.g., older adults) is characterized by only one sequence-network. If comparing the structure of sequence pathways for younger individuals to that of older individuals, for example (as will be done later in this chapter), there are only two structures to compare. This is akin to attempting to compare two roadmaps. *Prima facie*, it seems that this makes any statistically definitive comparison of sequential patterns between the groups difficult. Yet, as will be shown below, with a nonparametric approach, these comparisons are relatively easy to conduct.

²I am indebted to an anonymous reviewer for pointing out these very useful affinities between network analysis and sequence analysis concepts and methods.

The first step is to identify some feature of the sequence-networks that one wants to compare. For example, as just discussed, we may be interested in the extent to which the two groups differ with respect to the total number of elements that are involved in their sequence-networks (i.e., sequence-network size).

A challenge emerges when the groups being compared differ in size. This makes the estimates of their key network properties incomparable, as network properties automatically vary by group size. For example, the greater the number of individuals in a group, the larger and less dense the network will be (Borgatti and Everett 1997). Because each group is a different size, any apparent differences in network properties often merely reflect group size differences. Therefore, the procedure will involve comparing the value of a given measure that is calculated for one group to the value of the same measure for the second group that would be expected if that second group were the same size of the first group.

It is relatively easy to use a bootstrap approach to obtain these comparison values (Efron and Tibshirani 1993). Begin by treating the smaller of the two groups being compared as the baseline group. For example, if the first group includes 500 individuals and the second group includes 300, then begin by recording the properties of the network of the second group of 300, and treat these as baseline estimates.

The next task is to obtain comparable estimates for the first group, which contains 500 people. To do this, we draw a random subset of the same size (300 people) from the sample of 500 people in the second group, without replacement. We then calculate the network measure of interest for that random subset, and record it as a first comparable observation. For precision, we do this 1,000 times using different combinations of 300 randomly chosen respondents, thus generating 1,000 comparison estimates. We then average these 1,000 estimates together to assess the “typical” value of the measure in question for 300-person sets from within the larger second group. We use the percentile method (Mooney and Duval 1993) to obtain a 95% confidence interval around this estimate. We can thereby assess whether the observed values of various network measures for the smaller baseline 300-person group fall within the confidence interval that we see for the random subsets that are drawn from the 500-person group. This makes it possible to determine whether there are statistically significant differences in the structure of the sequence-networks of people in the groups being compared.

4 Illustrative Analysis: Activity Sequencing by Age

To demonstrate the potential analytic value of networks for studying sequential social phenomena, a network analysis of the activity sequences that emerge in different age groups is presented below. There are good reasons to analyze how age shapes sequential daily behavior. Most importantly, the notion of “successful aging” is an influential concept in efforts to conceptualize the lives of adults who remain physically, cognitively, and psychologically healthy, as well as physically

active and socially engaged throughout the life course (Rowe and Kahn 1987). A substantial body of research (Adams et al. 2011, e.g.,) shows that older adults who remain physically and socially active lead healthier and happier lives in general.

Attempts to study this aspect of successful aging have typically involved examining age differences in generic estimates or summaries of activity tendencies or levels. Empirical measures capture overall levels of activity—for example, frequencies of social group participation. One problem with such measures is that they lack temporal context. Rates of various types of activity may be higher for members of a given group on the whole, and yet be lower than another group's at certain times of day. The more fundamental problem is that the conventional approach treats activities as independent events, as if they do not occur within a larger chain of activities. As discussed above, activities are elements in a larger chain or path. As numerous scholars have argued, how people experience activity sequences can be just as consequential as their levels of certain activities (Bales 1951; White 1995). Some sequences, for example, include more transitions between activities, which in turn implies greater movement between social roles (Cornwell and Watkins 2015). The higher rates of switching that result require that people move from automatic to deliberative modes of cognition, reduce their dependence on pre-established routines, and heighten their awareness of environmental stimuli and cues (Hitlin and Elder Jr. 2007). The higher rates of cognitive decline among older adults thus provides some motivation for examining the extent to which older adults experience more or less complex activity sequences.

Unfortunately, few studies have analyzed activities in the context of their sequencing—and fewer still that consider this in the context of aging. Research that uses the conventional approach to assessing older adults' activities has generally supported the general assumption that later life is a period of relative simplicity. Do the activity pathways people tend to take through the day in later life indeed wind through a less complex network of action possibilities?

4.1 The Activity Sequence Data

I study this using time diary data that are collected in the annual American Time Use Surveys (ATUS). The ATUS is a nationally representative telephone survey that has been conducted every year since 2003 to assess individuals' work schedules and community involvement. For the sake of illustration, the following analysis relies on data from 2015 only.

To obtain a sample, the ATUS begins by drawing a random sample of households from those leaving the Current Population Survey (CPS) rotation each month. An eligible person from the household (who is at least 15 years old) is randomly selected from the household to be interviewed by phone. In 2015, the response rate was 48.5%, yielding a total of 10,905 respondents (Bureau of Labor Statistics and U.S. Census Bureau 2016).

The main goal of the ATUS is to collect 24-h recall diaries from these respondents, yielding a full account of their activities on the day immediately preceding the interview. ATUS interviewers start by asking respondents about the beginning of the previous day: “So, let’s begin. Yesterday [e.g., Thursday], at 4:00 a.m. What were you doing?” They then work forward through the day, collecting information about: (1) what the respondent was doing; (2) the times each activity began and ended; (3) where each activity occurred; and (4) whom the respondent was with. This analysis focuses on the activity data.

In the interview, respondents provide finely grained accounts of how they spent the day in question, in time intervals as small as one minute. All activities except those that occur during periods of paid work are coded using a detailed list of hundreds of possible activities. For the purposes of this analysis, these are collapsed to the ATUS’s first-tier coding scheme, which includes 17 possible activities.³ Because weekday and weekend activities are so different, and because weekdays are bound to provide more structure for most people, this analysis ignores individuals whose diaries were collected on weekends (roughly half).

The goal is to examine any differences in the activity sequences of younger and older adults. This can be done by defining two groups of people who fall into different age ranges and constructing their sequence-networks separately.⁴ The two age groups are constructed using twenty-year ranges. For the younger age group, the time diaries of those ATUS respondents who were between 25–44 years of age are included (25 being a typical starting point in studies of time use among working adults.) For the older group, those between 65–84 are included (65 being a typical starting point for studying retirees). It is likely that any differences that are observed between the younger and older age groups will reflect difference in employment status and thus paid work activity. In addition, the data on work activity in the ATUS is fairly poor and non-specific, which will artificially delimit the activity codes for workers. To ensure that this is not the case, this study only includes those who are not in the labor force. Taking all of these restrictions into account, 1,106 respondents remain in scope. Finally, 199 of these individuals are eliminated because they have at least one missing activity code during the 24-h period in question, resulting in a final valid sample of 907 individuals.

The sample includes 219 non-working people between 25–44 years old and 688 non-working people between 65–84 years old. Because the younger age group is smaller (a common sampling problem), it is treated as the baseline group against which the bootstrap samples for the larger, older group are compared.

³These activities are: personal care, household (HH), caring for/helping HH members, caring for/helping non-HH members, work/work-related, education, consumer purchases, professional/personal care services, HH services, government services/civic obligations, eating/drinking, socializing/relaxing/leisure, sports/exercise/recreation, religious/spiritual activities, volunteering, telephone calls, and travel.

⁴One can develop more comparable group by implementing a rigorous matching process, whereby members of the two groups are not only different by age, but also similar with respect to other attributes (apart from employment status, which is already controlled here) that could shape activity sequences, such as socioeconomic status and income.

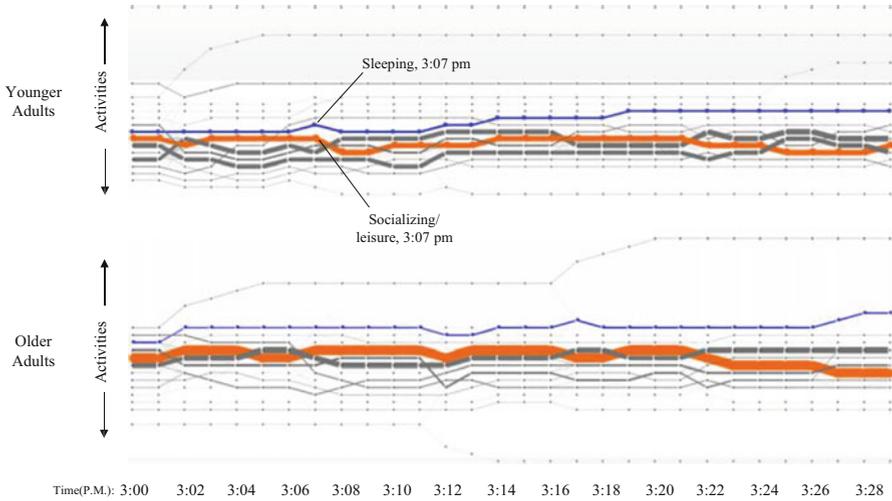


Fig. 2 Sequence-networks representing one half hour in the afternoon among younger and older adults, respectively. Both networks are drawn from equal size groups of 219 individuals. The tiny circles (nodes) represent activities that were reported as occurring during specific one-minute intervals, and the lines represent transitions between those activities from one time point to the next. Blue lines represent pathways characterized by sleep. Orange lines are pathways characterized by socializing/leisure. The thickness of lines is proportionate to the number of respondents who reported those exact activity transitions at that time

4.2 Sequence-Network Analysis Findings

This kind of analysis does not focus on the types of activities people engage in, how that varies over time, or which types of sequences people follow. Rather, it addresses *how sequence* pathways intersect in larger networks. This section presents the unique kinds of findings that such an analysis adds to sequence studies.

A network diagram for the younger group is shown in the top half of Fig. 2. This diagram was created using Pajek (Mrvar and Batagelj 1996), which places points adjacent to each other to the extent that they involve similar actors. For example, when many of the people who are engaged in a given activity at 3:00 p.m. are engaged in another activity at 3:01 p.m., those two activities will appear close together one step to the right. If, on the other hand, very few of the people who are engaged in the 3:00 p.m. activity are engaged in a given activity at 3:01 p.m., those two activities will be separated by substantial space on the vertical axis. This, in turn, means that the diagram adjusts the placement of activities at each time point such that those which fall on pathways that tend to intersect at that time point or adjacent time points are placed more closely together in the diagram.

For illustrative purposes, this diagram shows only a half-hour portion of the 24-h period, but it holds the same structural pattern as other time periods.⁵ This network is composed of nodes (small dots)—which represent activities (e.g., leisure) that occurred during specific minute intervals (e.g., 3:00 p.m. to 3:01 p.m.)—and lines, which represent transitions between those activities. Time flows from the left side of the diagram (i.e., the left-most nodes represent activities that took place during the 3:00 p.m. to 3:01 p.m. interval) to the right side of the diagram (3:29 p.m. to 3:30 p.m.). Pathways and activity time-points that involve spells of socializing/leisure (a common activity during this time) are marked in orange, while pathways that involve spells of sleep are marked with blue.

Between the time intervals that are represented here, there are several thick horizontal lines, which represent popular types of activities that many people reported throughout the half hour that is presented here. For example, this figure shows that there were numerous people in both groups who exhibited spells of socializing and leisure during this time (indicated by the thick orange lines in both panels of the diagram). In general, these more common paths contain people who were engaged in one of three common activity types: (1) leisure/socializing/recreation; (2) household activity; and (3) taking care of household members (e.g., children). Thinner strands that occur throughout the period represent people who were doing less common activities or who were switching between these and other types of activities during this time period. For example, spells of sleep were relatively uncommon during this time period.

Note that there is also considerable switching between activity types during this period. An example of this is marked in the younger adults' (top) panel. The fact that the thickest horizontal lines (which represent the most common types of activities) lay relatively close together in the network diagram attests to the fact that there was a lot of switching between these activities throughout the time period. In other words, the network diagramming software modulates the vertical location of the activities vis-à-vis each other such that those activity pathways which are more interconnected via switches appear closer together. This visual connection is one of the advantages that come with using network-diagrammatic algorithms to place particular sequence elements relative to particular sequence positions. A good example is the link between socializing/leisure (orange) and sleeping (blue). As the afternoon wears on, sleeping becomes more decoupled (i.e., it becomes less compatible) with the more common activities of this time period, including socializing/leisure.

⁵A half hour is used as the analytic time frame because it provides enough granularity to allow a close visualization of activity pathways but without the indecipherable jumble of lines and nodes that would appear if a longer period of time were used. This particular time period (mid-afternoon) is chosen because so few people were asleep.

The measures describing the structure of the younger adults' complete network for the full 24-h period are provided in Table 1. This network is composed of the 15,489 unique activity-times that these respondents reported from among the $1,440 \times 17 = 24,480$ possibilities (63.3% of all possible combinations) available. There were 18,683 different types of transitions observed between these 15,489 activity-times, out of $1,439 \times 17 \times 17 = 415,817$ possible types of transitions between all pairs of successive time points. Accordingly, the density of this network is .045, or 4.5% of the possible level of interconnectedness among pairs of elements between successive time points.

The network for younger adults is also highly centralized. The average path width for this age group is 16.9, and the standard deviation around this mean is 35.5, which is the centralization estimate. This means that path widths tended to vary quite a bit from each other, with some paths being very dominant (witness the long tails at the beginning and end of the sequence-network depicted in Fig. 2) and others being singular and poorly populated. It is evident from the diagram that much of the path centralization reflects the predominant norm of sleeping through the late night and early morning hours.

Finally, the homophily estimate is .986. This means that if we observe one of these younger adults at a given time, s/he will be doing the same activity during the next minute 98.6% of the time. In short, their sequence chains contain long spells of the same activity. This high estimate is due largely to the high granularity of the ATUS data, which contains one observation every minute. (Time diary data that come in longer time period increments will automatically have lower homophily estimates.)

A key question is how these aspects of the structure of the older adults' sequence-network compare to these estimates for younger adults. Results of the bootstrapping procedure are presented in the right-most column of Table 1, along with 99% confidence intervals around the mean estimates for older adults. The main findings are that older adults' sequence-networks have an average of 14,181 unique activity-times, which is significantly smaller than the observed size of the younger adults' network ($p < .01$). The older adults' network also tends to be less dense, with an average of 4.0% of all possible activity pathways being realized ($p < .001$).

In addition, centralization remains high in the older adults' sequence-network. But the average level of centralization here (40.250) is significantly greater than we see in the younger adults' network ($p < .001$) indicating that there is less variation around the average path width for this group. Finally, the average homophily estimate for older adults is .989, which is not substantially different but certainly significantly greater ($p < .001$) than the level of homophily in the younger adults' network.

The differences in structure that were just described can also be seen by visually comparing the network diagram for the older sample, which is shown

in the bottom half of Fig. 2.⁶ The diagrams reveal the same patterns that are evident in the statistical comparison. The younger adults' network includes more activity times overall and, thus, more possible pathways. The network begins and ends with more activities (15 and 14, respectively) than does the older adults' network, in which only 13 different types of activities are represented at both the beginning and end. Furthermore, there are a greater number of moderately trafficked pathways in the younger adults' networks. In contrast, the older adults' network contains a single dominant pathway (leisure/socializing/recreation) and a secondary moderately sized pathway (household activity). This reflects the greater centralization around dominant pathways within the older sample that is revealed by the bootstrap procedure.⁷ In sum, the older adults' network is composed of fewer, wider pathways, which means that they report doing fewer activities and are more likely to follow a few more common activity sequences.

5 Discussion and Conclusion

Increasingly, sequence-oriented scholars are interested in larger structures that consist of multiple intersecting chains of events, or sequence-networks. This chapter has argued that one can gain additional insight into these structures by combining network-analytic with sequence-analytic methods. Scholars have identified several affinities between the network framework and the sequence framework (Bison 2014; Cornwell 2015). This chapter expanded on this work by showing how several key network concepts—including network diameter, size, density, centralization, and homophily—correspond to key sequence concepts—such as sequence length and the extent of variation in the popularity of transitions between different sequence elements. In addition, some network methods, such as visualization of network diagrams, can reveal new things about sequential social phenomena.

This chapter demonstrated the sequence-network approach via an analysis of the structure of the network of activity pathways that are followed by individuals in different age groups. Older adults' pathways are less complex, as they involve fewer

⁶For older adults, a representative network is derived from the reports of 219 in-scope older respondents. The first network that fell within one standard deviation of the means with respect to network size, density, centralization, and homophily was chosen. The observed levels for these measures for this network were 13,929, .039, 40.741, and .989, respectively.

⁷Supplemental analysis (not shown) shows that the structure of this network evolves somewhat over time. At some time points (e.g., nighttime) there is more centralization in both groups. But the degree of centralization tends to be greater among younger adults than it is among older adults in the morning hours (partly because older adults tend to wake up earlier), and this pattern shifts at midday. Thus, a disaggregation of the network estimates by time would reveal significant heterogeneity in levels of observed difference by age.

activities at different times, fewer transitions between different activity-times, are more centralized around fewer and more dominant sequence pathways, and involve less switching between different activities. In light of what we know about the health benefits of complex activity (e.g., Adams et al. 2011), these findings suggest a new area of concern for social gerontology and research on well-being. These findings suggest that older adults inhabit less complex action systems, which may have negative consequences for their social connectedness as well as their cognitive functioning. On the other hand, it is suggestive of greater levels of organization and predictability in everyday life (Zerubavel 1981), which may provide psychological comfort (Giddens 1984) in an otherwise difficult period of the life course. These substantive findings call for further study. But more work is needed to identify which network measures and methods are most useful for examining sequential social action.

I close by reiterating that the network methods that are discussed here are not substitutes for sequence analysis—they are supplements. These methods are a first step in an emerging effort to develop new approaches for understanding how sets of sequences intersect in larger, complex systems. There are several conventional sequence methods already in use which can be used to analyze some of the sequence-related attributes that I have analyzed in this paper. But this paper proposes a new language and analytic framework for the study of sequential phenomena which has as-yet-unknown levels of correspondence with existing sequence methods. I argue that the most insightful analyses of sequential phenomena will come when researchers combine network and sequence methods together.

Acknowledgements Early versions of this chapter were presented at colloquia hosted by the Departments of Sociology at Duke University, Columbia University, and the University of Oxford in 2015 and 2016. I thank the many members of those audiences who commented and asked questions—and especially Peter Bearman, Erin York Cornwell, Jonathan Gershuny, Sang Kyung Lee, James Moody, Gilbert Ritschard, Matthias Studer, Oriell Sullivan, Giacomo Vagni, and four anonymous reviewers—for suggesting changes that improved this chapter.

References

- Adams, K. B., Leibbrandt, S., & Moon, H. (2011). A critical review of the literature on social and leisure activity and wellbeing in later life. *Ageing and Society*, 31, 683–712.
- Aisenbrey, S., & Fasang, A. E. (2010). New life for old ideas: The “second wave” of sequence analysis bringing the “course” back into the life course. *Sociological Methods & Research*, 38, 420–462.
- Bales, R. F. (1951). *Interaction process analysis: A method for the study of small groups*. Cambridge: Addison-Wesley.
- Batagelj, V., Doriean, P., Ferligoj, A., & Kejzar, N. (2014). *Understanding large temporal networks and spatial networks*. Chichester: Wiley.
- Bearman, P. S., Faris, R., & Moody, J. (1999). Blocking the future: New solutions for old problems in historical social science. *Social Science History*, 23, 501–33.

- Bearman, P. S., & Stovel, K. (2000). Becoming a nazi: A model for narrative networks. *Poetics*, 27, 69–90.
- Bison, I. (2014). Sequence as network: An attempt to apply network analysis to sequence analysis. In P. Blanchard, F. Bühmann, & J.-A. Gauthier (Eds.), *Advances in sequence analysis: Theory, method, applications* (pp. 231–248). New York: Springer.
- Borgatti, S. P., & Everett, M. G. (1997). Network analysis of 2-mode data. *Social Networks*, 19, 243–269.
- Borgatti, S. P., Everett, M. G., & Freeman, L. C. (2002). *UCINET 6 for windows*. Harvard: Analytic Technologies.
- Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal*, 6(4), 435–460.
- Bureau of Labor Statistics and U.S. Census Bureau. (2016). American time use survey user's guide. User's guide, Bureau of Labor Statistics and U.S. Census Bureau, Washington, DC
- Bürgin, R., & Ritschard, G. (2014). A decorated parallel coordinate plot for categorical longitudinal data. *The American Statistician*, 68(2), 98–103.
- Cornwell, B. (2015). *Social sequence analysis*. New York: Cambridge University Press.
- Cornwell, B., & Watkins, K. (2015). Sequence-network analysis: A new framework for studying action in groups. *Advances in Group Processes*, 32, 31–63.
- Efron, B., & Tibshirani, R. (1993). *An introduction to the bootstrap*. London: Chapman and Hall.
- Fasang, A. E., & Liao, T. F. (2014). Visualizing sequences in the social sciences: Relative frequency sequence plots. *Sociological Methods & Research*, 43(4), 643–676.
- Gabadinho, A., & Ritschard, G. (2013). Searching for typical life trajectories applied to childbirth histories. In R. Levy & E. Widmer (Eds.), *Gendered life courses – Between individualization and standardization. A European approach applied to Switzerland* (pp. 287–312). Vienna: LIT-Verlag.
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
- Gershuny, J. (2000). *Changing times: Work and leisure in postindustrial society*. Oxford: Oxford University Press.
- Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press, Berkeley, CA.
- Hamberger, K. (2018). Relational sequence networks as a tool for studying gendered mobility patterns. In G. Ritschard & M. Studer (Eds.), *Sequence analysis and related approaches: Innovative methods and applications*. Cham: Springer (this volume).
- Hitlin, S., & Elder Jr., G. H. (2007). Time, self, and the curiously abstract concept of agency. *Sociological Theory*, 25, 170–91.
- McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27, 415–444.
- Merton, R. K. (1968). *Social theory and social structure*. New York: Free Press.
- Merton, R. K. (1996). Opportunity structure. In *On social structure and science* (pp. 153–161). Chicago: University of Chicago Press.
- Mooney, C. Z., & Duval, R. D. (1993). *Bootstrapping: A non-parametric approach to statistical inference*. Newbury Park: Sage.
- Mrvar, A., & Batagelj, V. (1996). *Pajek 64-XXL (version 5.01). Program for large network analysis*. University of Ljubljana, Ljubljana.
- Parsons, T. (1951). *The social system*. New York: Free Press.
- Piccarreta, R. (2017). Joint sequence analysis: Association and clustering. *Sociological Methods & Research*, 46, 252–287.
- Rowe, J. W., & Kahn, R. L. (1987). Human aging: Usual and successful. *Science*, 237(4811), 143–149.
- StataCorp. (2015). *Stata statistical software: Release 14*. College Station: StataCorp LP.
- Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy analysis of state sequences. *Sociological Methods & Research*, 40, 471–510.

- Wasserman, S., & Faust, K. (1994). *Social network analysis*. New York: Cambridge.
- Wellman, B. (1983). Network analysis: Some basic principles. *Sociological Theory*, 1, 155–200.
- White, H. C. (1995). Network switchings and Bayesian forks. Reconstructing the social and behavioral sciences. *Social Research*, 62, 1035–1063.
- Zerubavel, E. (1981). *Hidden rhythms: Schedules and calendars in social life*. Chicago: University of Chicago Press.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

