Using Sequence Analysis to Quantify How Strongly Life Courses Are Linked

Tim F. Liao

University of Illinois

Abstract: Dyadic or, more generally, polyadic life course sequences can be more associated within dyads or polyads than between randomly assigned dyadic/polyadic member sequences, a phenomenon reflecting the life course principle of linked lives. In this article, I propose a method of U and V measures for quantifying and assessing linked life course trajectories in sequence data. Specifically, I compare the sequence distance between members of an observed dyad/polyad against a set of randomly generated dyads/polyads. The U measure quantifies how much greater, in terms of a given distance measure, the members in a dyad/polyad resemble one another than do members of randomly generated dyads/polyads, and the V measure quantifies the degree of linked lives in terms of how much observed dyads/polyads outperform randomized dyads/polyads. I present a simulation study, an empirical study analyzing dyadic family formation sequence data from the Longitudinal Study of Generations, and a random seed sensitivity analysis in the online supplement. Through these analyses, I demonstrate the versatility and usefulness of the proposed method for quantifying linked lives analysis with sequence data. The method has broad applicability to sequence data in life course, business and organizational, and social network research.

Keywords: sequence analysis; life course; linked lives; dyads; polyads

A MONG the five general life course principles of life span development—agency, time, space, timing, and linked lives (Elder, Johnson, and Crosnoe 2003)—the principle of linked lives is the only one that directly connects life course trajectories of people in salient relationships. As such, the concept of linked lives emphasizes the generational dimension of time in that one individual's life can be and most often is embedded within the lives of their family members, including those from other generations (Elder 1995; Macmillan and Copher 2005).

This article deals with the analysis of linked lives by proposing a method for quantifying the degrees of linked lives for polyads, which has been illustrated with an example of dyadic life course sequences that are associated with two generations of family members. The objective is to provide a proper assessment of linked lives in the form of two measures—one of life course distances and the other of the degree of life course linkage—between members of dyads or more generally polyads.

Since its introduction from biology by Abbott and Forrest (1986) more than three decades ago, sequence analysis has been widely applied in the social sciences by second-wave sequence analysis researchers (see Aisenbrey and Fasang 2017; Fasang and Raab 2014). There have been rapid and continuing methodological advances in social sequence analysis (Barban et al. 2020; Blanchard, Bühlmann, and Gauthier 2014; Cornwell 2015; Fasang and Liao 2014; Piccarretta 2017; Raab et al. 2014; Studer et al. 2011; Studer 2013; Studer, Struffolino, and Fasang 2018). In this article, I follow this exciting research tradition of developing sequence analysis.

There are three main approaches to analyzing dyadic sequence data, a special case of polyadic sequence data. First, dyadic sequence data can be analyzed with multichannel sequence analysis (Gauthier et al. 2010; Pollock 2007). Fasang and Raab (2014), among others, provide a good example of applying multichannel sequence analysis to parent-child family formation sequence data. Second, dyadic sequence data can be formed into grid-sequences, based on the state space grid method; this approach is called grid-sequence analysis (Brinberg et al. 2018). Typically, cluster analysis follows multichannel sequence analysis and grid-sequence analysis, and the generated clusters often become the dependent variable in a subsequent substantive analysis. Third, based on earlier formal work by Elzinga, Rahmann, and Wang (2008), Liefbroer and Elzinga (2012) proposed a subsequence-based approach to analyzing dyadic sequence data by focusing on the similarities in subsequences between dyadic members compared with unrelated persons, and they compared the subsequence-based approach to similarities based on optimal matching. This approach is rather different in purpose from the first two because multichannel sequence analysis and grid-sequence analysis provide a way to analyze sequence data by summarizing them in different domains and channels, but neither method gives an individual measure of such relatedness. The third approach is the only one in the literature that offers a method for measuring individual dyadic members' similarities. A recent study by Karhula et al. (2019) followed the principle of the third approach and found strong similarity in siblings' early socioeconomic trajectories as compared with unrelated persons. Yet another fourth potential candidate for analyzing dyadic and triadic sequence data is joint sequence analysis proposed by Piccarreta (2017) for analyzing multiple domain sequence data with a purpose similar to that of multichannel sequence analysis, although a full exploration is beyond the scope of this article.

There are at least three differences between the proposed method in this article, which provides an individual measure of linked lives, and the third approach. The proposed method also examines similarities of dyadic (or polyadic) members, but it is a large sample–based statistical similarity in that the method bases the intradyadic or intrapolyadic distance or (dis)similarity comparison on *a large number of randomly assigned intradyadic or intrapolyadic distances*. In contrast, Liefbroer and Elzinga's (2012) approach relies on an *array of unrelated parent-child pairing possibilities existing in the data matrix*, and Karhula et al.'s (2019) study assigns *for each focal person one sibling and one randomly selected unrelated person to create one sibling dyad and one unrelated dyad*. The proposed method here has three major advantages: First, it provides a flexible application for any distance measures (beyond just optimal matching [OM] typically applied in earlier studies). The flexibility of using different distance measures affords us the functionality of separately analyzing timing, duration, and order of life course trajectories. Second, every dyad/polyad in the sample receives not only their own estimate of life course resemblance but also a statistical confidence of that estimate. Third, the proposed method has flexibility for applications in the analysis of tetrads, pentads, hexads, and, more generally, polyads, as presented in the section on the analytic procedure.

Researchers can apply the proposed method to at least three types of data. First, the concept of linked lives features importantly in intergenerational research, as

evidenced by my reanalysis of Fasang and Raab's (2014) data. Second, analysis of sibling dyads has been a focal issue in sibling studies, such as Karhula et al.'s (2019) recent study. Finally, life course linkage can take place between different sources of data, such as self-reported employment histories and the administrative employment histories analyzed by Wahrendorf et al. (2019).

Furthermore, researchers can use the proposed method for analyzing sequence data in a range of (sub)disciplines. The current article focuses on linked life courses in life course studies, especially family formation. Other types of life courses can also be linked and analyzed as such, such as linked employment and health trajectories. In recent years, sequence analysis has begun to see applications in business and organizational research (Dinovitzer and Garth 2020; Heimann-Roppelt and Tegtmeier 2018; Ho et al. 2020; Nee et al. 2017) and social network research (Cornwell 2015; Nee et al. 2017). Take, for example, Nee et al.'s (2017) study of Chinese entrepreneurs' egocentric network tie trajectories. If data on their six alters' network tie trajectories were also collected, we would be able to apply the proposed method to quantify similarity of the entrepreneurs' network tie trajectories within each of the heptadic networks versus across unrelated networks.

The article proceeds as follows. I first introduce the proposed method for measuring and analyzing life course linkages of polyadic members in two ways. The method allows us to capture the similarity in distance and the degree of linked lives between members of polyads, compared with randomly constructed polyads. It also allows us to define the similarity and the degree of linked lives in terms of how much observed dyadic/polyadic sequences resemble one another more than randomly generated dyadic/polyadic sequences. I then present a simulation study of the two proposed measures (U and V, to be defined later) and an application using an empirical data set, the Longitudinal Study of Generations (LSOG) survey, involving dyadic family data. The application demonstrates usage of the two measures—a measure of dyadic/polyadic distance and a measure of the degree of linked lives indicating statistical confidence—by focusing on three different aspects of the life course: timing, duration, and order of events. Because the approach relies on randomization, I also analyze the effect of random seed selections and further demonstrate such selections in a sensitivity analysis of the LSOG data (reported in the online supplement). Finally, I draw some conclusions for the measures of linked lives proposed here.

Measuring Linked Lives in Life Course Research

The principle of linked lives is influential in life course research. However, to this day, scholars have not had a formal, general way to assess or measure the concept at the individual level (i.e., for every single dyad or polyad) other than the third approach for dyads only. The proposed measure here is based on the relative principle of comparing the observed and a large set of randomized (i.e., unrelated) dyadic life course sequences. Comparing observed and simulated data provides an effective statistical analysis for researchers in many disciplines (for application examples, see Amory et al. [2015] and Furman et al. [2018]). The proposed method

also allows one to use any of the distance measures available in the R package TraMineR.

To date, the only serious analytic attempts related to the concept of linked lives have been the subsequence-based approach (Elzinga et al. 2008; Liefbroer and Elzinga 2012) and, more recently, dyadic sibling comparison with unrelated nonsibling dyads (Karhula et al. 2019). Liefbroer and Elzinga's (2012) method computes the number of shared subsequences between dyadic members, and the resulting value is normed to fall in the range of [0, 1]. The method is conceptually straightforward. However, if a dyad has a resemblance score of 0.35, is it high or low? The question is, in Liefbroer and Elzinga's (2012:4) words, whether "actual pc-dyads do not resemble each other more than nr-dyads" (here the dyadic data consist of any person from the parental generation and any person from the children's generation) where "pc-dyads" stand for "parent-child dyads" and "nr-dyads" is shorthand for "nonrelated dyads." Therefore, we need a formal assessment of such resemblance. The proposed randomization method in the current article makes possible a formal assessment. In fact, by choosing random sequence generation mechanism 1 (to be described later in the section), we can assess the similarity of linked lives by comparing related dyads or polyads with randomly selected unrelated dyads or polyads. The method proposed here allows a variety of dissimilarity or distance measures for analyzing life course (state or event) sequences as discussed by Studer and Ritschard (2016), and it differs from earlier randomization attempts at distinguishing number of transitions only.[1]

The Procedure

The method described below provides an individual measure of linked lives computable using any available distance measures. It evaluates life course sequences in every linked polyad against randomly generated polyads based on a random generation assumption or mechanism (to be discussed later). The evaluation uses a specific distance measure that can be sensitive to timing, duration, or order (Studer and Ritschard 2016). The method follows the principle of randomization tests as a nonparametric statistical method, as described in Liao (2002) and Onghena (2018). Under the null hypothesis of no (treatment) effect, the randomization test applies a random assignment procedure that produces a random shuffle of responses (Onghena 2018). For sequence data, the null hypothesis is no different in terms of sequence distance between members of an observed polyad and randomly assigned members. The random shuffle allows reassignment of any unrelated polyadic member sequences into "related" polyads.

Dyadic/polyadic sequences are three-dimensional objects. There are in total N number of dyads/polyads for $i = 1, 2, \ldots, N$; each dyad/polyad i has a total number of J members for $j = 1, 2, \ldots, J$; and each member j of dyad/polyad i has a sequence of length l. Because the treatment of sequence length is specific to distance measures (e.g., Hamming distance requires equal sequence lengths but OM does not) and because randomization does not involve the length dimension, for the sake of simplicity in presentation without losing generality, we disregard the third dimension and consider polyadic sequences as two-dimensional objects. More
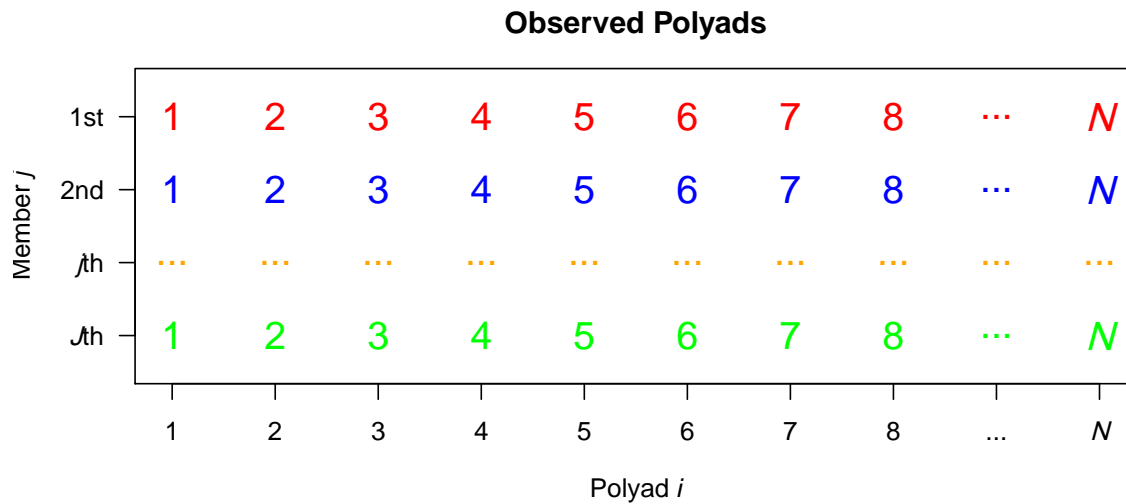
Observed Polyads

Figure 1: A diagram of observed polyadic sequences $S_{ij}(i = 1, 2, \ldots N$ and $j = 1, 2, \ldots J)$.

specifically, the method for analyzing dyadic (e.g., parent-child or sibling-sibling) or polyadic sequence data takes the following steps:

1. (a) Let us use S_{ij} to indicate a polyadic sequence S for the jth member of an observed polyad i for $i = 1$ to N and $j = 1$ to J, where N is the total number of polyads and J is the total number of the polyadic members in each polyad under study. For example, a three-generation triadic sequence data set has three members for each triad, with $J = 3$ and $N = 300$ for a total of 300 triads. Figure 1 shows the general polyadic data setup.

(b) We compute a distance vector D_i of the ith polyad between each member pair of the N polyads, using a user-defined dissimilarity measure, such as the following:

$$D_i = \{d(S_{ij}, S_{ik})\} \quad \text{for } i = 1 \text{ to } N \text{ and } j, \ k = 1 \text{ to } J \text{ but } j \neq k, \qquad (1)$$

where $d(\cdot)$ is a user-chosen distance function. For dyadic data, Equation (1) reduces to $D_i = d(S_{i1}, S_{i2})$; for triadic data, Equation (1) becomes $D_i = \{d(S_{i1}, S_{i2}), d(S_{i1}, S_{i3}), d(S_{i2}, S_{i3})\}$. Thus, for each D_i where $i = 1$ to N, there are Q entries in Equation (1). When the total number of polyadic members is J (i.e., J number of members in a single polyad), there exists $Q = \binom{J}{2} = \frac{J!}{2!(J-2)!}$ number of unique pairwise distances between the members (pairwise because distances can only be computed pairwise).[2]

2. Compute distances between reassigned polyadic sequences randomly drawn from observed polyadic member sets:
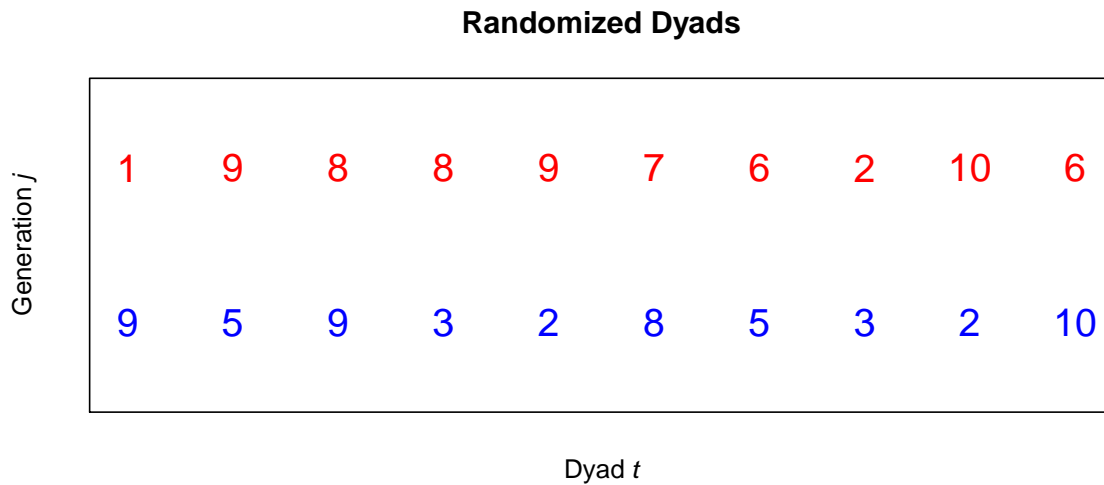
Randomized Dyads

1	9	8	8	9	7	6	2	10	6
9	5	9	3	2	8	5	3	2	10

Generation j

Dyad *t*

Figure 2: A diagram of 10 randomized dyads of two generations for computing R_t ($T = 10$; values in the box indicate the *t*th dyadic sequence of member *j*).

Instead of computing $d(S_{ij}, S_{ik})$ where $i = i$ and $j \neq k$, here we compute

$$R_t = \{d(S_{aj}, S_{bk})\}, \tag{2}$$

where $j \neq k$ still, and subscripts *a* and *b* are instances of *i* and represent a randomly selected "*a*th" (instead of *i*th) polyad (e.g., for parent) and a randomly selected "*b*th" (instead of *i*th) polyad (e.g., for child). Unlike in Equation (1) where $i = i$, here $a \neq b$ is true albeit not necessarily a required condition because each *t* for $t = 1$ to *T* (when it is large) represents a randomly drawn member sequence for either the *j*th or *k*th member of an *i*th polyad (where $i = a$ or *b*). In other words, each *t* represents a new cross-polyadic matching of a randomly drawn *j* member (e.g., father) sequence with a randomly drawn counterpart *k* member (e.g., offspring) sequence.

3. Repeat step 2 *T* number of times, with *T* being a large number preferably $\geq 1{,}000$. For dyadic data, R_t is a single vector with *T* entries when $Q = 1$. R_t remains a vector of $1 \times T$ entries for triadic data when $Q = 3$ and more generally for polyadic data when $Q > 3$.[3] For a simple example of how the random assignment works, let us use a dyadic example where $J = 2$ and $N = 10$ for $t = 1$ to *T* number of randomized matchings of dyadic sequences (Figure 2).

Once again, we ignore the length dimension in sequences. In the figure, the *x* axis gives the value of a randomized *t*th dyad. In Figure 2, when $t = 1$, or the first randomization, the first sequence in the first generation (member) in red is randomly matched with the ninth member sequence of the second

generation in blue, or *a* and *b*, as in Equation (2). When $t = 6$ for the sixth randomization, the seventh sequence in the first generation in red (or *a*) is randomly matched with the eighth member of the second generation in blue (or *b*). Randomization continues until the last position, when $t = T$.

4. Using Equations (1) and (2), we obtain two statistics:

First, we define dyadic/polyadic distance U_i as follows:

$$U_i = \frac{\sum_t R_t}{T} - D_i. \tag{3}$$

In other words, we subtract the observed D_i from the mean of R_t for each of the *i*th polyad; however, we could calculate a "unique" mean of R_t for each *i*th polyad by computing T number of R_t for each polyad (resulting in $T \times N$ R_t in total), it would be computationally intensive and unnecessary so long as we keep $T \geq 1,000$, because when T is very large, the individually computed R_t mean would not be distinguishable from one another. Besides, using the same global mean in Equation (3) to compute U_i provides the same benchmark for all observed polyadic distances. A greater U_i value suggests a greater linkedness between the members of a polyad, although its actual value depends on the chosen distance measure.

Second, record in a new variable the degree of dyadic/polyadic linkage V_i (for $i = 1$ to N) of the proportion out of T times when $D_i < R_t$, with the new variable value falling in the [0, 1] interval (e.g., $V_1 = 990$ of $1,000 = 0.990$, $V_2 = 891$ of $1,000 = 0.891, \ldots, V_N = 995$ of $1,000 = 0.995$ for $T = 1,000$). This computation essentially performs a randomization test and forms a test statistic. It differs from the first statistic U_i that is based on the contrast between D_i and the mean of R_t; the second statistic, V_i, is based on the contrast between D_i and each of the T number of R_t. By using individual R_t instead of its mean, one can expect V_i to have a greater variability than U_i. V_i has the range of [0, 1], and a greater V_i value means a greater linkedness between the members of a polyad. Additionally and optionally, for each V_i in step 4, if it is >0.95, a value of 1 can be recorded in a new vector of length N, otherwise, 0. This provides a confidence indicator of $p \geq 0.95$ for each of the *i*th polyads.

The computation of steps 1 to 4 involves two separate loops: a loop of T times and another loop of N times (as steps 1 and 4 can be computed in the same loop), separate rather than embedded. As stated earlier, when T is large enough, there is no need to repeat the randomization for each of the *i*th observed sequence. An author-written R program `seqpolyads`[4] that depends on `TraMineR` (Gabadinho et al. 2011) performs these computations and is available in the `TraMineRextras` package at CRAN. To illustrate how the proposed procedure works, let us take a simple example where $J = 2$ and $N = 10$, a two-generation dyadic set of family formation sequences for ages 21 to 30. The trajectories have three states: s = single, m = married, and c = having a child (Figure 3).

The first three dyads have just the two states of "s" and "m" without having had a child. The other seven dyads all have a child. To keep it simple, we delay
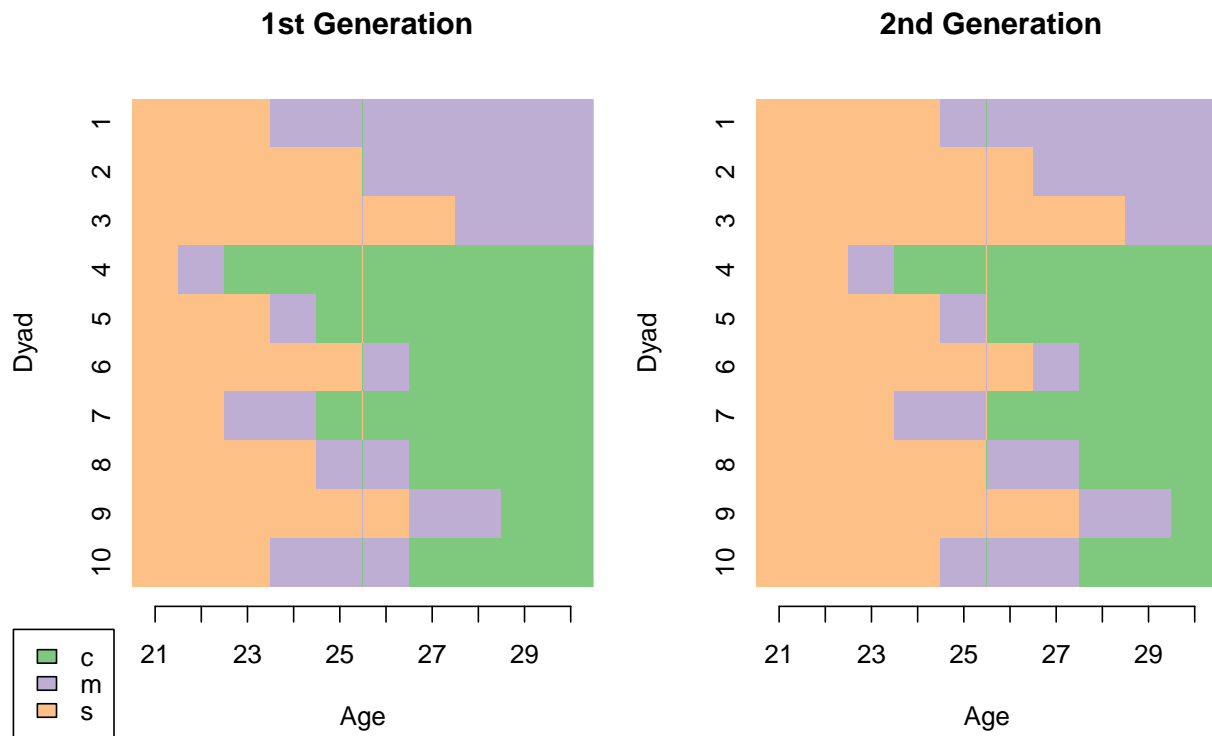
1st Generation   2nd Generation

Figure 3: Sequence index plots of the set of hypothetical two-generation dyadic family formation sequences $S_{ij}(J = 2, N = 10)$.

Table 1: U_i and V_i statistics for the 10 hypothetical dyads

	$i = 1$	$i = 2$	$i = 3$	$i = 4$	$i = 5$	$i = 6$	$i = 7$	$i = 8$	$i = 9$	$i = 10$	Mean
					U_i						
Timing	3.258	3.258	3.258	2.258	2.258	2.258	2.258	2.258	2.258	2.258	2.558
Order	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445
Duration	1.924	1.924	1.924	1.095	1.095	1.095	1.095	1.095	1.095	1.095	1.344
					V_i						
Timing	0.866	0.866	0.866	0.720	0.720	0.720	0.720	0.720	0.720	0.720	0.764
Order	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445	0.445
Duration	0.850	0.850	0.850	0.696	0.696	0.696	0.696	0.696	0.696	0.696	0.742

marriage (and first birth when relevant) of the second generation by one year invariably without varying the order or duration of events (other than that due to changed timing). I applied the proposed procedure to this illustrative dyadic data set, and I present the results in Table 1.

The U_i and V_i statistics are reported in Table 1 for each of the ith dyadic sequences, calculated with $T = 1,000$. For timing, I used Hamming distance; for

order, I implemented SVRspell with a spell duration weight exponent of 0 and a subsequence length-weight exponent of 1; for duration, I applied the CHI2 distance with the number of intervals of 1 (or $K = 1$) (for further details of the parameters for the three distances, see Tables 5 and 6 and Figure 1 in Studer and Ritschard [2016]). Because all second-generation sequences are delayed in timing by a year, the observed dyadic distance between members of dyads are 1 for the first three dyads and 2 for the remaining seven dyads (because they involve two transitions). As a result, the first three dyads' U_i are all one unit higher than that of the other dyads because their chance of being randomly matched with a sequence with two delayed transitions is about double. Similarly, the first three dyads' V_i values are higher than those of the others. The identical values of U_i and V_i for order represent a special case. Here, for all dyads, sequence orders remain constant between generations. Thus, the observed order distances are all zeros. However, because a comparison of a sequence with two states with another with three states yields a difference in order (and in this case, a unit difference in distance), the estimated 0.445 represents the proportion of such a difference. Finally, the dyadic duration resemblances are lower for the dyadic sequences 4 to 10 in either U_i or V_i because the durations of their three events of s, m, and c differ from randomly assigned dyads to a smaller degree than do the three dyads with only two events.

Random Sequence Generation Mechanisms

I included two random sequence generation mechanisms for performing step 2:

1. Sequence-conditional random sequence generation: By making this assumption, sequences of length l are randomly drawn from the observed set of polyadic members (for dyads, e.g., set of parents' sequences and set of children's sequences). Using this mechanism preserves the meaningful order of states and is useful when certain states cannot precede certain other states (e.g., divorce cannot precede first marriage for family formation sequences).

2. Sequence-conditional random state generation: By making this assumption, sequences of length l are randomly drawn from the whole set of states from the observed sequences under consideration with state replacement within a selected sequence. Each sequence is randomly selected first before a random reshuffle of the states within the selected sequence. This mechanism can be useful for sequences with no logical orders for the list of states. For example, out of the labor force, employment, and unemployment can occur at any time and can last for any duration.

The choice of a random sequence generation mechanism depends on the nature of sequences and the substantive need of the research.

A Simulation Study

To assess the statistical properties of the proposed method, I conducted a simulation study of randomly generated sequences of 100 positions ($L = 100$) that belong to a

dyadic ($J = 2$) data set with a variable N, contained in subset 1 (for the first member of the dyad) and subset 2 (for the second member of the dyad) of the paired dyadic sequence data. To control how the two subsets are linked, the paired sequences are generated one pair at a time, collected into subset 1 and subset 2 one pair at a time. After a specified proportion of linkedness is reached, the remaining dyads are generated with a different alphabet (see below).

To simulate the two subsets, I randomly generated a varying proportion of dyads with similar distinctive states. Two sequences with no common tokens/states are maximally dissimilar (Dijkstra and Taris 1995; Elzinga 2003:9, Axiom 1) when pairwise substitution costs between states are all equal, an assumption used in the current data setup. To test how two members of a dyadic pair differ, I selected the first three letters of the English alphabet (i.e., A, B, and C) and assigned them randomly with replacement into the entire 100 positions of the first member of a dyad. I also randomly assigned the same three letters with replacement into the 100 positions of the second member of a dyad by keeping a varying degree of similarity (to the first member's 100 positions) at 18 different levels, with percentage of similarities varying from 1 percent to 9 percent by 1 percent increments in order to capture the more sensitive end of shared states, and from 10 percent to 90 percent by 10 percent to cover the remaining range. Note that 0 percent and 100 percent are trivial cases that need no simulation.

I conducted these operations with N = 50, 100, 250, 500, and 1,000 for each of the 18 percentages of linked sequences. The simulation of linked lives computed Hamming distances, using the procedure described in the previous section with $T = 1,000$ for simulating T number of randomly selected dyads, is to be repeated with Z number of repeats of the simulation, with each simulation computing the 1,000 number of randomized dyads. For the simulation reported here, I set $Z = 30$ to be the total number of repeats of conducting the simulation; because there is little difference in the patterns of results between just a single simulation and 100 repeats of the simulation, 30 repeats should be sufficient. Keep in mind that for each repeat, 1,000 randomly assigned dyads were generated. Thus, $1,000 \times 30 = 30,000$ randomized dyads serve as the basis for the computation reported in the figures. Figures 4 and 5 present the simulated U and the V statistics based on Hamming distance, respectively.

Each figure contains five panels of boxplots, each for a specific sample size from 50 to 1,000 pairs of dyadic sequences for assessing any potential effects of sample size. In each panel, nine sets of simulations are reported, from 1 percent to 9 percent shared states in the paired dyadic sequences. The plotted values are the output from the program seqpolyads. The U_i values indicate the Hamming distance between "observed" dyadic sequences and the average of randomly combined dyadic sequences. That is, the greater a U_i value, the greater the average distance between two randomly paired dyadic members than between two "observed" dyadic members.[5] The V_i values in Figure 5 measure the confidence probability/proportion of linked dyadic sequences, which reflects how well each of the "observed" dyadic sequences outperforms—having smaller intradyadic distance than—randomly assigned dyadic sequences.
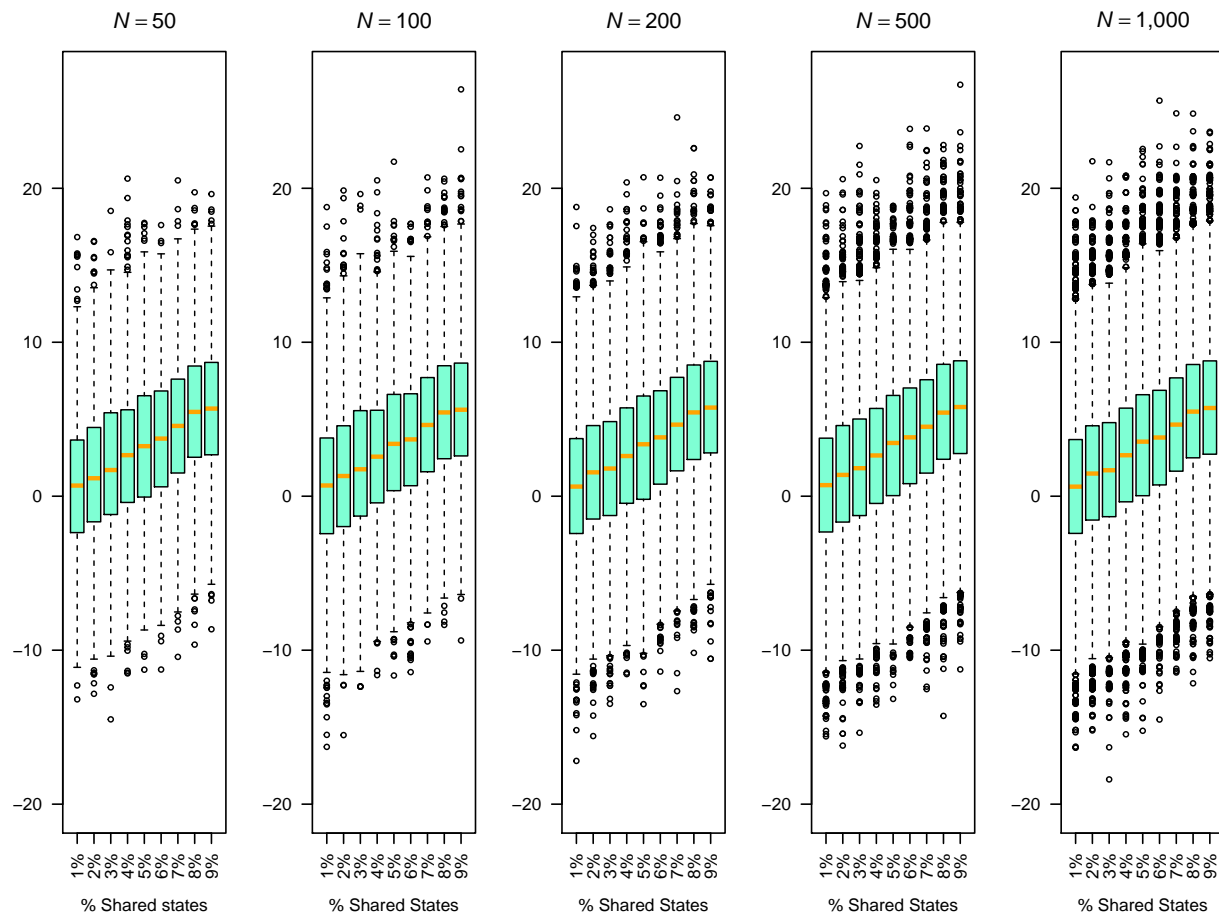
Figure 4: Boxplots of U_i assessing difference between "observed" and 1,000 simulated dyadic sequences with sequence length = 100 and sample size = 50, 100, 200, 500, and 1,000 for percent shared states = 1 percent to 9 percent, 30 repeats of the simulation.

We can make three observations about the simulation results in these figures. First, the performance of the proposed method shows a nice linear pattern. That is, the increase in the size of U_i or V_i corresponds exactly to the proportion of shared states in the paired dyadic sequences. Second, the method is rather insensitive to or robust for sample size variation. As is obvious from the figures, the shape of the distribution of the boxplots is almost identical across all sample sizes in all plots. Finally, the V_i statistic covers a wider range within the range of [0,1], whereas the U_i statistic concentrates in a much narrower band, relative to its minimum and maximum values. The smaller variation of U_i is evidence for its reliance on the averaged R_t.

To evaluate the performance of these two statistics for a greater amount of shared states, Figures 6 and 7 present the nine sets of simulated results of 10 percent to 90 percent shared states in the paired dyadic sequences, using the same procedure as generated Figures 4 and 5. Here the U_i statistics continue the covered range left
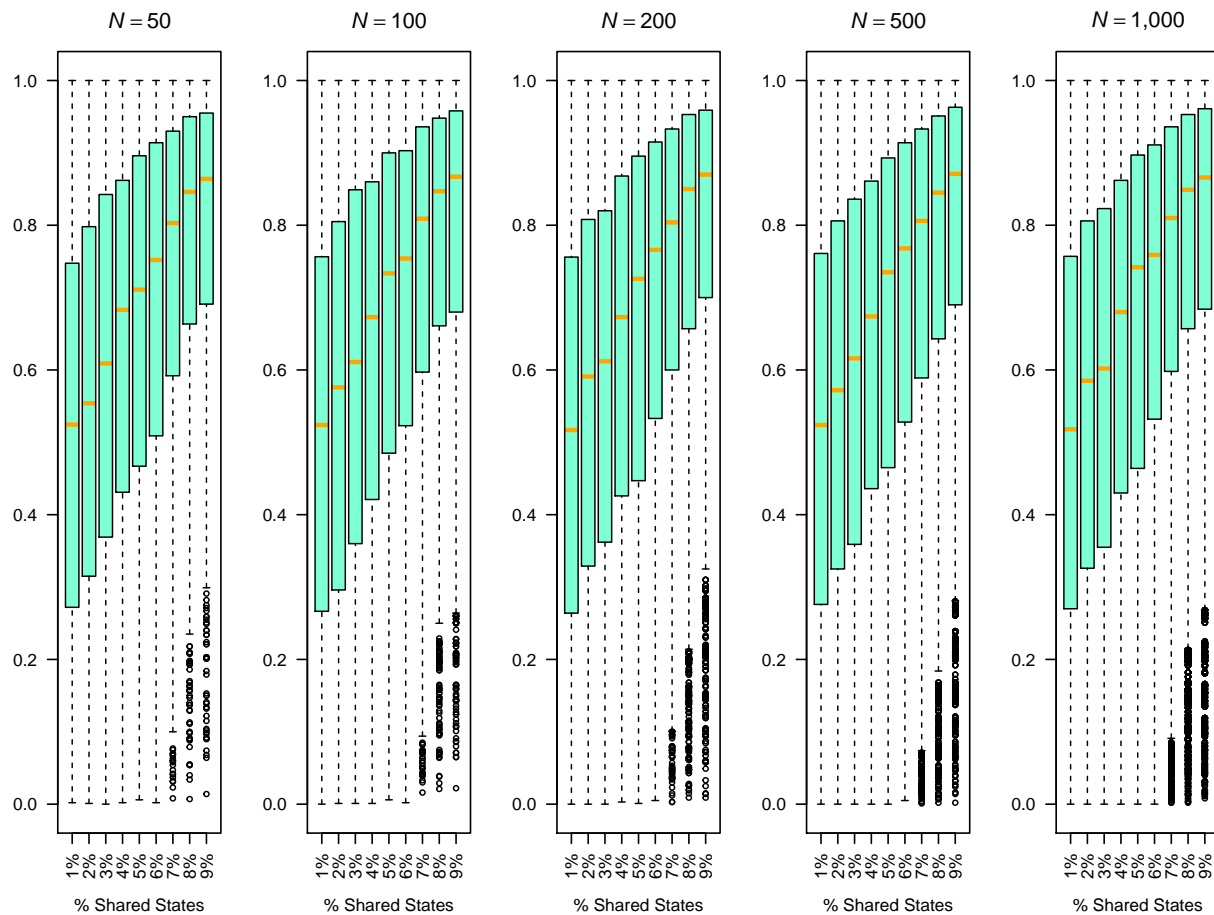
Figure 5: Boxplots of V_i assessing difference between "observed" and 1,000 simulated dyadic sequences with sequence length = 100 and sample size = 50, 100, 200, 500, and 1,000 for percent shared states = 1 percent to 9 percent, 30 repeats of the simulation.

off in Figure 4, all the way to a U_i value of about 60, again in a linear fashion and without variation across different sample sizes. Therefore, the entire range of U_i covered in Figures 4 and 6 should provide an idea of the performance of U_i when applied to the Hamming distance measuring dyadic similarity.

The patterns of V_i distributions, although consistent across sample sizes, do not show a linear progression when the percentage of shared states increases. When the degree of shared states is above 20 percent, dyadic linkage as measured by V_i shows a typical value of 0.95 or above. In comparison, U_i values are below 30 for dyads with 20 percent to 30 percent shared states, much lower than the results from those with 90 percent shared states. The difference between the two statistics demonstrates the usefulness of both, with U_i showing greater ability to differentiate shared common states and V_i demonstrating a randomization test significance (or in this case, confidence probability) with the null hypothesis stating no difference
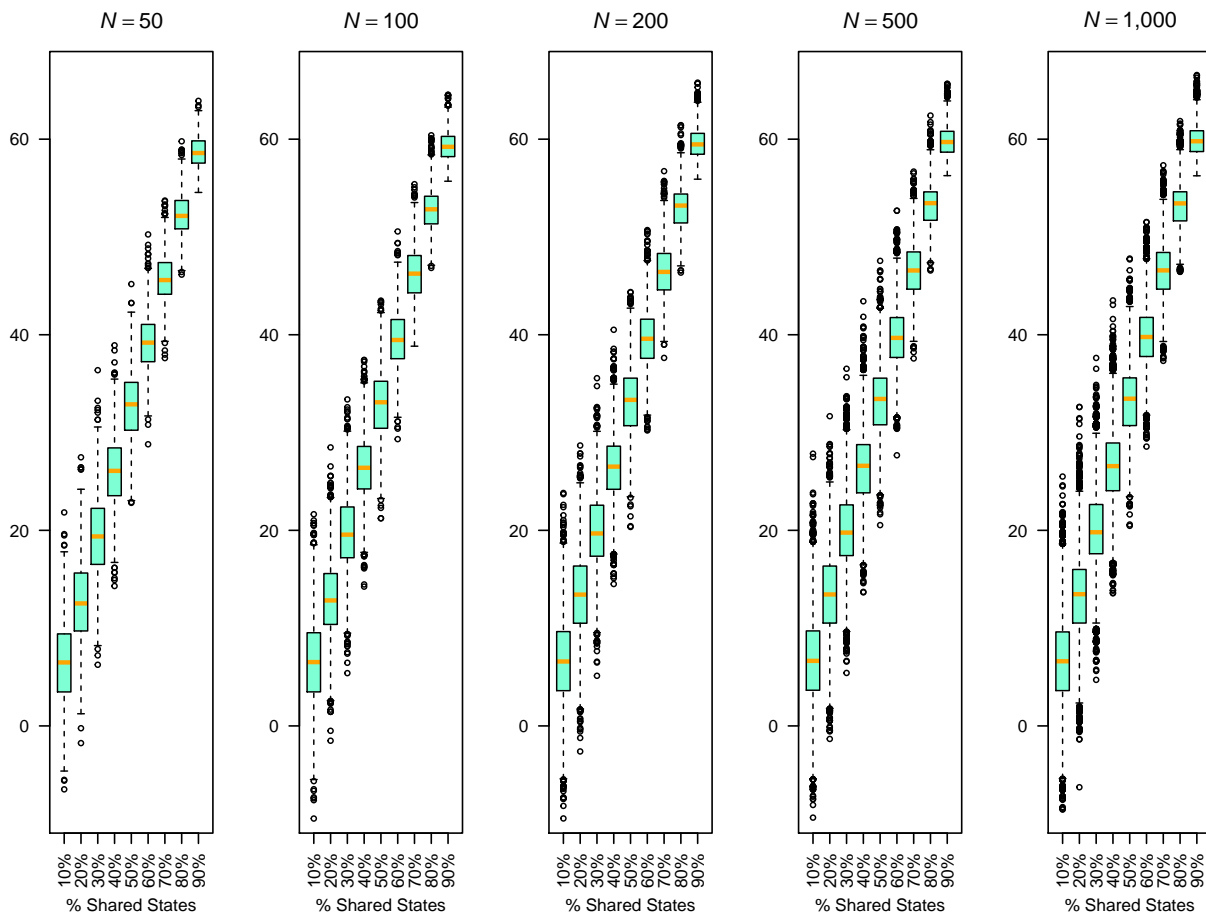
Figure 6: Boxplots of U_i assessing difference between "observed" and 1,000 simulated dyadic sequences with sequence length = 100 and sample size = 50, 100, 200, 500, and 1,000 for percent shared states = 10 percent to 90 percent, 30 repeats of the simulation.

between observed and random dyads. Therefore, the two statistics complement each other.

Application

In this empirical application, I focus on intergenerational dyadic life course data from the United States, using data from the LSOG analyzed in Fasang and Raab's (2014) study. The LSOG sequences record family trajectories of middle-class parents born around 1920 to 1930, whose family formation took place approximately from 1935 to 1960, and the trajectories of their children whose family formation took place between 1955 and 1990. Analyzing life course sequence data, Fasang and Raab (2014) made two contributions: they conceptualized family formation holistically instead of focusing on isolated events, and they identified three types of intergenerational family formation patterns instead of estimating average transmission effects. Using
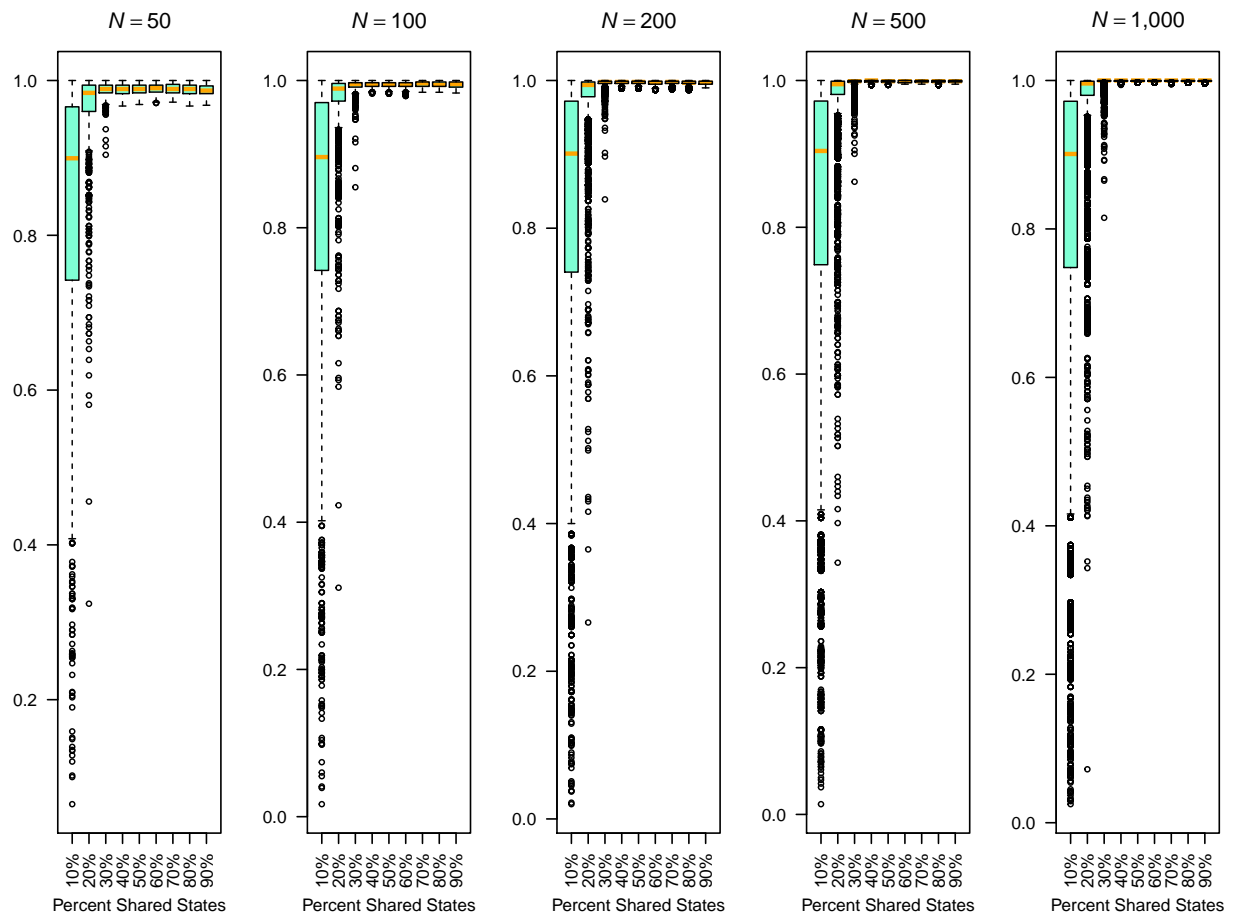
Figure 7: Boxplots of V_i assessing difference between "observed" and 1,000 simulated dyadic sequences with sequence length = 100 and sample size = 50, 100, 200, 500, and 1,000 for percent shared states = 10 percent to 90 percent, 30 repeats of the simulation.

the method proposed in this article, I move beyond Fasang and Raab's (2014) study by implementing a more detailed measure of intergenerational transmission and by dissecting the three dimensions of life course timing, duration, and order.

The LSOG provides complete family formation sequences of parents and their children between ages 15 and 40. Fasang and Raab's (2014) research represents the first attempt to fully exploit the unique intergenerational and longitudinal information on family formation of the LSOG. Like them, I use data from two generations: (1) the parent generation, the "silent generation" born in the 1920s and 1930s, and (2) their children, the Baby Boom generation born in the late 1940s and 1950s. The data set for the analysis to follow has 461 parent-child dyads. The parent-child dyads belong to four types of gender constellations: mother-daughter, mother-son, father-daughter, and father-son. For further details on the data, see Fasang and Raab (2014).

The dyadic sequence data contain nine family formation states: single, no child; single, one or more children; married, no child; married, one child; married, two children; married, three children; married, four or more children; divorced, no child; and divorced, one or more children. Note that there is a logical order to these nine family formation states. For example, the single state always happens first, and either of the two divorced states cannot immediately succeed the single state. Furthermore, the states with children are in a logical order such that a state with a higher number of children always succeeds a state with a lower number of children, which in turn succeeds having no child. Because of such logical orders of the sequence data, I chose random generation mechanism 1 to analyze the data.

Such typical family formation sequences possess the three distinctive characteristics of timing, duration, and sequencing (order). They represent the onset of a certain state, the time a person spends in a given state, and the sequencing of a particular state vis-à-vis other states, respectively. Because of these characteristics, I used four different distance measures for the descriptive analysis and three for the regression analysis. Three measures are more sensitive to one of the life course characteristics, and the last one has equal (in)sensitivity to all three: Hamming distance (for timing), the CHI2 distance with $K = 1$ (for duration), the SVRspell distance with a spell duration weight exponent of 0 and a subsequence length-weight exponent of 1 (for order), and the OMspell distance with an expansion cost of 0.5 and an indel cost of 2 (this measure is not dominated by timing, duration, or order variations). I analyzed the LSOG data of 461 dyadic sequence pairs, and Figure 8 presents the density plot of the four types of dyadic distances as measured by U. The x axis records the U values, that is, the differences between observed pairs of dyadic sequence distances compared with differences based on randomly chosen dyadic sequences (i.e., pairs of parent sequences and child sequences).

All four density curves concentrate in the middle range centered around zero (the mean for duration $U = 0.055$; the mean for order $U = 0.012$; the mean for timing $U = 0.238$; the mean for neutral $U = 0.381$). Although zero means no difference between observed and randomly assigned dyads, we cannot further compare the U curves based on different distance measure scales, though we can do so with the V curves in Figure 9. To interpret a density plot, we must understand that the total area under the curve is 1 or unity. At a given point on the x axis, the y axis value can be >1 because the width of a point on the x axis can be rather narrow, and the Y value is obtained by dividing the area for a given point of X by the width.

To see the behavior of life course linkage (V), I plotted the counterpart density curves, and present them in Figure 9. The curves (V) representing the degree of linked lives generally show a greater spread over a much shorter range than do the mean-based U curves, although the V curve representing order shows more of a multimodal distribution than do the others.

The LSOG dyads measured by the V focused on timing and (to a smaller degree) the V focused on duration appear to resemble each other more than the V focused on order, which has weaker resemblance, indicated by the gravitation of the curve more toward the higher-valued end (and by a higher mean value given below). The neutral distance measure (with a mean degree of linked life courses of 0.507) is a summary measure of the other three curves (timing focused, mean = 0.479; duration
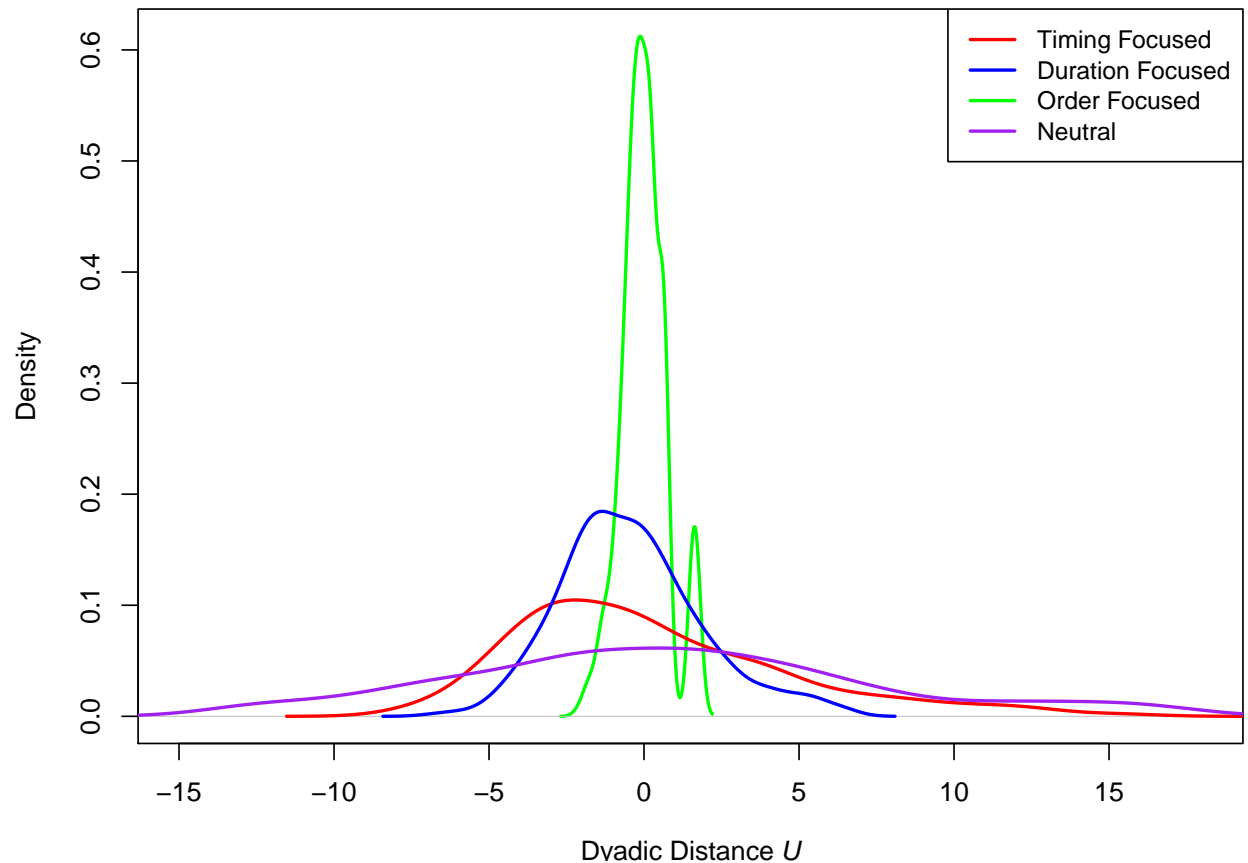
Figure 8: Density plots of the LSOG dyadic distances U, using 1,000 simulated dyads by random generation mechanism 1 ($N = 461$).

focused, mean = 0.488; and order focused, mean = 0.436). The green order-focused curve has two peaks at or above 0.5, confirming the finding from Figure 8. One observation not noticeable in Figure 8 is the higher concentration of the duration density curve below 0.5, which suggests that more than half the sample lacks a high degree of intergenerational transmission of family formation events in duration.

I applied these two measures—dyadic distance (U) and degree of dyadic linkage (V)—in a reanalysis of the LSOG data reported in Fasang and Raab (2014). Fasang and Raab identified, via a cluster analysis, three types of family formation sequences of intergenerational transmission—strong transmission, moderate transmission, and no transmission (contrasting patterns)—before analyzing the three categories with a multinomial logit model. Instead of the three clusters, I allowed each dyad to take on a value of dyadic distance (U) and the degree of dyadic linkage (V), and I analyzed them in a series of regression models with robust standard errors using the same set of independent variables as in Table 2 in Fasang and Raab (2014). Table 2 reports the descriptive statistics of the variables used in the regression analyses.

Table 2 includes the descriptive statistics of the two sets of new outcome variables and those of the independent variables used in Fasang and Raab's (2014)
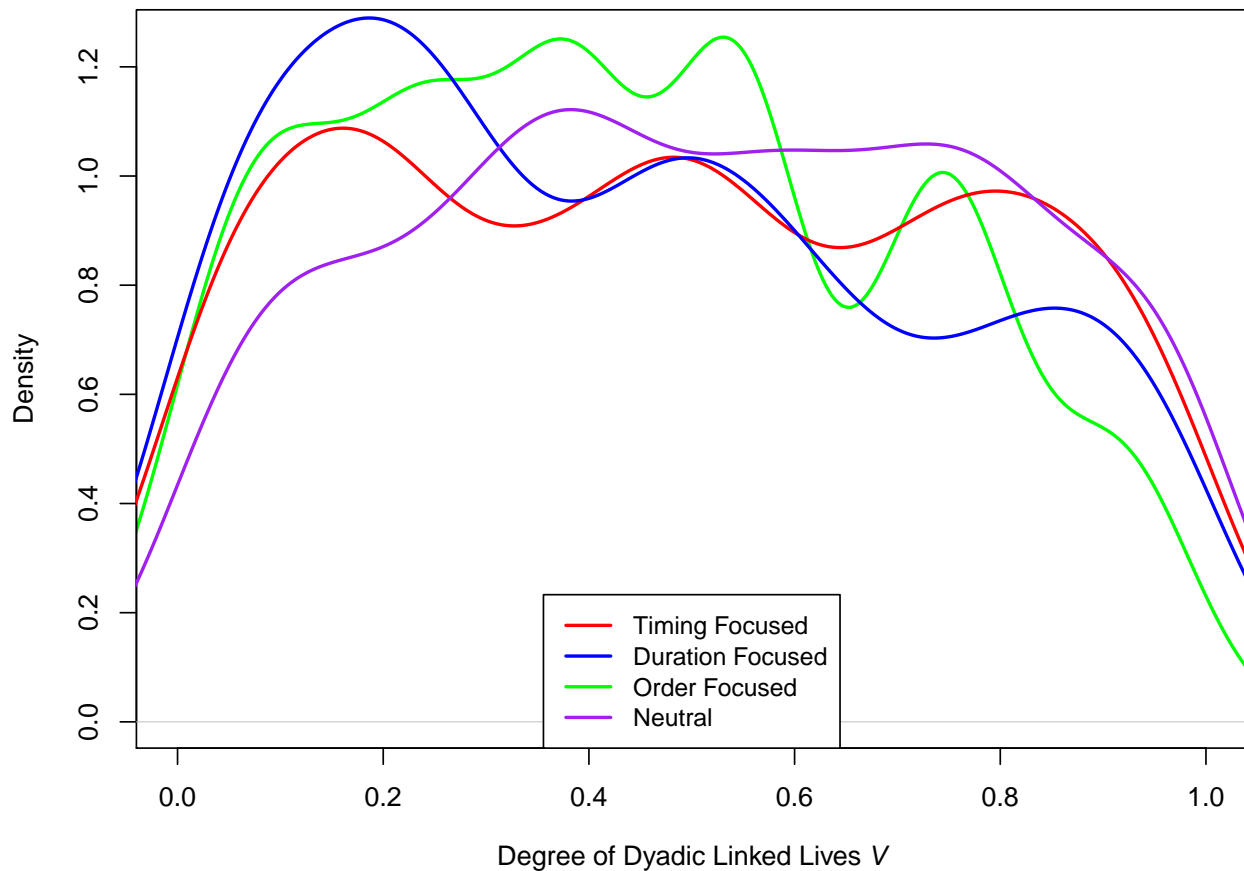
Figure 9: Density plots of the LSOG degrees of dyadic linked lives V, using 1,000 simulated dyads by random generation mechanism 1 ($N = 461$).

analysis. All these variables are measured at the dyadic level. Gender constellation represents the gender-specific combination of a dyad, with the mother-daughter combination as the reference category. Age difference records the difference between the parent's and the child's age in a dyad. Years of education for the parent and the child is measured by two variables, the difference between the two dyadic members and the average of the two members. Sibling position is the child's birth order. Affectual solidarity scale reflects the relationship quality between parents and children. For further details on these variables, see Fasang and Raab (2014).

I estimated a series of six linear regression models[6] with robust errors (for correcting dyadic clustering because some dyads may belong to the same family). For each dimension—timing, duration, and order—I estimated two models, one with dyadic distance (U) and the other with degree of dyadic linkage (V) as the outcome variable, and I report the coefficient estimates with t statistics in Table 3.

The overall patterns of effects (in terms of which variables are statistically significant) on intergenerational family formation transmission are largely consistent with those reported in Table 2 in Fasang and Raab (2014), other than those measuring gender constellation. In this analysis, the gender-specific combinations do

Table 2: Descriptive statistics for variables in the regression analysis ($N = 391$)

	Mean	SD
Timing U	0.238	4.434
Duration U	0.055	1.179
Order U	0.012	0.729
Timing V	0.479	0.298
Duration V	0.488	0.289
Order V	0.436	0.265
Gender constellation		
Mother-daughter	…	…
Father-son	0.176	0.381
Mother-son	0.215	0.411
Father-daughter	0.271	0.445
Dyad's age difference	25.321	4.421
Dyad's average years of education	14.670	2.067
Dyad's difference in years of education	1.375	3.127
Sibling position	1.666	0.819
Affectual solidarity scale	4.192	1.040

Note: 15.18 percent of the children have a missing value for the affectual solidarity scale, reducing the sample size to 391.

not distinguish themselves. Furthermore, new in the current analysis is that we can separate out the effects of timing, duration, and order. For example, the estimated timing and duration effects on V means a year's increase in a dyad's age difference (which suggests tradition) would increase the degree of resemblance in family formation between two generations in terms of timing and duration (another life course tradition) by 3.9 percent and 3.8 percent, respectively, compared with randomized dyads (because the outcome variable is measured in the range of [0,1]). This strong positive effect of age difference is also found in the contrast between the strong transmission and the different process patterns analyzed by Fasang and Raab (2014). It is possible that strong transmission of family formation is in part a byproduct of intergenerational transmission of status (Fasang and Raab 2014), and older parents tend to have more stable transmission of status. Education only matters in

Table 3: Regression estimates with robust standard errors correcting for dyadic family clustering ($N = 391$)

	Timing		Duration		Order	
	U	V	U	V	U	V
Gender constellation Mother-daughter (ref.)						
Father-son	−0.190 (−0.357)	0.002 (0.045)	−0.052 (−0.354)	0.002 (0.062)	0.053 (0.507)	0.016 (0.370)
Mother-son	−0.184 (−0.324)	−0.016 (−0.439)	−0.042 (−0.279)	−0.019 (−0.511)	0.086 (0.857)	0.034 (0.956)
Father-daughter	−0.063 (−0.131)	0.007 (0.216)	−0.026 (−0.197)	0.005 (0.176)	−0.028 (−0.367)	−0.014 (−0.544)
Dyad's age difference	0.525[†] (7.960)	0.039[†] (9.563)	0.129[†] (7.244)	0.038[†] (9.533)	0.046[†] (4.078)	0.017[†] (3.900)
Dyad's average years of education	−0.198 (−1.860)	−0.013 (−1.949)	−0.051 (−1.737)	−0.012 (−1.840)	0.003 (0.132)	0.004 (0.469)
Dyad's difference in years of education	−0.029 (−0.429)	−0.002 (−0.409)	−0.007 (−0.406)	−0.002 (−0.371)	0.033* (2.445)	0.011* (2.226)
Sibling position	−1.913[†] (−6.640)	−0.143[†] (−7.401)	−0.482[†] (−6.275)	−0.138[†] (−7.440)	−0.114 (−1.953)	−0.042 (−1.851)
Affectual solidarity scale (child-parent)	0.448 (1.931)	0.029* (2.028)	0.118 (1.863)	0.029* (2.056)	0.102* (2.554)	0.036* (2.469)
Constant	−8.664[†] (−3.757)	−0.197 (−1.381)	−2.104** (−3.247)	−0.181 (−1.309)	−1.482[†] (−3.687)	−0.146 (−1.016)
R^2	0.201	0.259	0.170	0.262	0.094	0.094

Note: t statistics are in parentheses. $* \, p < 0.05$, $† \, p < 0.01$.

dyadic members' difference in its effect on order (one year's increase in difference in years of education increases order resemblance by 1.1 percent, compared with randomized dyads, judged by the estimated effect on V). Sibling position matters for timing and duration. Take the effect on V for timing, for example: lowering birth order by one position would increase timing resemblance by 14.3 percent compared with randomized dyads. Finally, quality of parent-child relationships shows a moderate positive effect on intergenerational transmission of life course patterns, confirming the role model effect discussed by Schönpflug (2001). Of the three dimensions, the effect of affectual solidarity scale is consistently stronger on order (supported by the significance and size of the V estimate) than on timing or duration.

So far, I have focused on the estimated effects on V. The effects on U are similar to the effects on V except that they are consistently weaker. The distance measure of U quantifies how much an observed dyadic difference is smaller than the average of randomly generated dyadic distances. For example, when birth order increases by one position, dyadic distance decreases by about 1.9 units of Hamming's distance (or about 2 months). On the other hand, the degree of dyadic linkage measure quantifies the proportion by which the observed dyad outperforms the randomly generated dyads by having a smaller distance. We must bear these definitions in mind when making interpretations. In summary, interpretation of U relies on the definition of an actual distance measure, whereas interpretation of V is more intuitive because it suggests the percentage of change per unit change of an X variable for an observed dyad when compared with randomized dyads.

Because the proposed procedure relies on random selections, the issue of random seeds should be investigated. I conducted a sensitivity analysis of random seed selections (reported in the online supplement), and I draw the following conclusions. First, random seed selection can make a difference, based on density curve observations, especially between the seeds that generated the lowest and highest mean U and V value. Second, the U results, shown as density curves, tend to be more clustered together (relatively less variable) than their V counterparts. Finally, how sensitive are the empirical results reported in Table 3 to random seed selections? The supplement suggests that the results based on the dyadic distance U values are extremely consistent compared with those based on the dyadic linkage degree V values, which show some differences from those reported in Table 3. However, the random seed variations analyzed did not change any of the significance tests in any models of U and V. This is reassuring, and as a result, for most applications, the default seed can be used safely.

Conclusion

I have shown that the proposed U and V measures provide a useful method for measuring and analyzing dyadic/polyadic similarities and linkages, as illustrated with a simulation study, an empirical dyadic application, and a sensitivity analysis (reported in the online supplement). I will now summarize some conclusions about the general feasibility of the U and V measures, the potential applicability to polyads, and how best to use the U versus V measures.

First, as demonstrated through the simulation study, the proposed method provides a useful general way for analyzing linked life course trajectories. The method has the flexibility to implement all sequence distance measures, and the reader is advised to refer to the discussions provided by Ritschard and Studer (2016) for their usages. Thus, the method can be used with any distance measures available in `TraMineR`, the `R` package for sequence analysis.

Second, the proposed method of analyzing linked lives can be applied to the analysis of tetrads, pentads, hexads, and higher dimensional polyads. The section presenting the methodological procedure specified the general case of polyadic linked lives. The functionality is already programmed in the `seqpolyads` function (available in `R`'s `TraMineRextras` package). In the `R` program, a parameter not used

in the earlier dyadic application is that of weight. For example, in a three-member family, or triad consisting of two parents and a child, the importance of resemblance between the parents may not be the same as that between the father and the child or the mother and the child. We can capture such differential importance by differentially weighting their respective distances.

Third, because the behavior of the U measure is more stable than that of the V, as shown in the sensitivity analysis, it is advisable to apply U in actual empirical analyses that focus on the statistical significance of independent variables. The supplemented sensitivity analysis also suggests that the currently used default random seed value is a fine choice because it produces similar results to the average results produced using a sizable number of random seeds. This saves the extra trouble of relying on extra computations using a large set of different random seeds. If the data analyst is more interested in the degree of resemblance between polyadic members, however, then the V measure can provide an easier interpretation because the V measure, with its normed range of [0, 1], is directly comparable between an application of different distance measures, whereas the U measure is not. When both U and V measures are applied, we may also view the V_i measure as a confidence probability for the U_i results.

Finally, although the current application involves family dyads, the method should be applicable to analyzing polyadic linkages defined by sources of data (Wahrendorf et al. 2019) as well as defined by other social groups, such as friendship networks, neighborhoods, companies, and birth cohorts or other types of cohorts and other forms of linked social groups. Furthermore, researchers can use the proposed method for analyzing sequence data in a range of (sub)disciplines. Although the current article is focused on linked family formation life courses, sequence analysis has recently gained popularity in business and organizational research as well as social network research (Cornwell 2015; Dinovitzer and Garth 2020; Heimann-Roppelt and Tegtmeier 2018; Ho et al. 2020; Nee et al. 2017). For family formation life course research, it is natural to define a polyadic social group that contains members of different generations in the same family or siblings in the same family. For business/organizational and social network research, a meaningful polyadic social group can be defined as a set of firms that conduct business closely together, thereby forming a network with ties. When data on firm-level or entrepreneurial-level attributes or qualities are collected over firms' life cycles, the proposed method can help researchers gain insight into similarity of within-group life cycle trajectories.

Notes

1 There has been an attempt in the form of an R package (Nightingale 2016) to use randomization for assessing household members' similarities by comparing how such members resemble one another compared to randomly generated data. There are two limitations of this package for analyzing life course sequences: First, the program produces a single statistic for the sample, yet a measure recording how linked the members' lives are in each dyadic cluster is desirable, like Liefbroer and Elzinga's (2012) method or the proposal in this article. Second, and more important, is the method for computing dif-

ferences between observed and randomly generated data. The number of state changes or transitions may be sufficient for capturing social behavior, such as migration, but insufficient for analyzing, more generally, family formation or other complex life course trajectories.

2 The data analyst can apply weight according to the substantive meaning of the relationships to arrive at an overall distance among the members of a polyad. For a dyad, $\binom{J}{2} = 1$, the single weight = 1; for a triad, $\binom{J}{2} = 3$, three weights can be assigned. For a five-generation data set, for example, the linkage between the first and fifth generations is rather weak, to the degree of nonexistance, so the researcher can assign a weight of close to 0.

3 However, for $Q = 3$, R_t can be reduced to a single vector of length T because the three pairwise distances can be summed with weights applied, yielding a single distance. For example, for a triad of father, mother, and child, if our main concern is intergenerational transmission, the weight for the distance between father and mother can be half of that between a parent and a child. The same principle applies to higher values of Q for polyadic sequence analysis.

4 The R program, `seqpolyads`, computes the two measures of polyadic distance (U) and polyadic linkage (V) of sequence data and a few other associated statistics. The user can choose a distance measure and its associated parameters, a random generation mechanism, a random seed number (for starting simulated data generation), and the number of simulated sequences. For polyads beyond dyads, the user can additionally choose a set of weights to assign to the distances between the members of a polyad.

5 "Observed" is in quotes because in the simulation study, these "observed" dyadic sequences are also simulated.

6 For dealing with potential conditional nonnormal distributions of the outcomes, I estimated a set of Gamma regression models. However, the substantive results differ little from ordinary least squares regression results, which are reported here.

References

Abbott, Andrew, and John Forrest. 1986. "Optimal Matching Methods for Historical Sequences." *Journal of Interdisciplinary History* 16(3):471–94. https://doi.org/10.2307/204500.

Aisenbrey, Silke, and Anette E. Fasang. 2017. "The Interplay of Work and Family Trajectories over the Life Course: Germany and the United States in Comparison." *American Journal of Sociology* 122(5):1448–4. https://doi.org/10.1086/691128.

Amory, Charles, Alexandre Trouvilliez, Hubert Gallée, Vincent Favier, F. Naaim-Bouvet, C. Genthon, Cécile Agosta, Luc Piard, and Hervé Bellot. 2015. "Comparison between Observed and Simulated Aeolian Snow Mass Fluxes in Adélie Land, East Antarctica." *The Cryosphere* 9:1373–83. https://doi.org/10.5194/tc-9-1373-2015.

Barban, N., X. de Luna, Emma Lundholm, Ingrid Svensson, and F. C. Billan. 2020. "Causal Effects of the Timing of Life-course Events: Age at Retirement and Subsequent Health." *Sociological Methods & Research* 49(1):216–49. https://doi.org/10.1177/0049124117729697.

Blanchard, Philippe, Felix Bühlmann, and Jacque-Antoine Gauthier, eds. 2014. *Advances in Sequence Analysis: Methods, Theories and Applications*. New York: Springer.

Brinberg, Miriam, Nilam Ram, Gizem Hülür, Timothy R. Brick, and Dennis Gerstorf. 2018. "Analyzing Dyadic Data Using Grid-Sequence Analysis: Interdyad Differences in In-

tradyad Dynamics." *Journal of Gerontology Series B* 73:5–18. https://doi.org/10.1093/geronb/gbw160.

Cornwell, Benjamin. 2015. *Social Sequence Analysis: Methods and Applications*. Cambridge, UK: Cambridge University Press.

Dijkstra, Wil, and Toon Taris. 1995. "Measuring the Agreement between Sequences." *Sociological Methods & Research* 24(2):214–31. https://doi.org/10.1177/0049124195024002004.

Dinovitzer, Ronit, and Bryant Garth. 2020. "The New Place of Corporate Law Firms in the Structuring of Elite Legal Careers." *Law & Social Inquiry* 45(2):339–71. https://doi.org/10.1017/lsi.2019.62.

Elder, Glen H., Jr. 1995. "The Life Course Paradigm: Social Change and Individual Development." Pp. 101–39 in *Examining Lives in Context: Perspectives on the Ecology of Human Development*, edited by Phyllis Moen, Glen H. Elder, Jr., and Kurt Luscher. Washington, DC: American Psychological Association. https://doi.org/10.1037/10176-003.

Elder, Glen H., Jr., Monica Kirkpatrick Johnson, and Robert Crosnoe. 2003. "The Emergence and Development of Life Course Theory." Pp. 3–19 in *Handbook of the Life Course*, edited by Jaylen T. Mortimer and Michael J. Shanahan. New York, NY: Plenum Publishers.

Elzinga, Cees H. 2003. "Sequence Similarity: A Nonaligning Technique." *Sociological Methods & Research* 32(1):3–29. https://doi.org/10.1177/0049124103253373.

Elzinga, Cees, Sven Rahmann, and Hui Wang. 2008. "Algorithms for Subsequence Combinatorics." *Theoretical Computer Science* 409(3):394–404. https://doi.org/10.1016/j.tcs.2008.08.035.

Fasang, Anette Eva, and Tim Futing Liao. 2014. "Visualizing Sequences in the Social Sciences: Relative Frequency Sequence Plots." *Sociological Methods & Research* 43(4):643–76. https://doi.org/10.1177/0049124113506563.

Fasang, Anette Eva, and Marcel Raab. 2014. "Beyond Transmission: Intergenerational Patterns of Family Formation among Middle-Class American Families." *Demography* 51:1703–28. https://doi.org/10.1007/s13524-014-0322-9.

Furman, Bradley T., Erin H. Leone, Susan S. Bell, Michael J. Durako, and Margaret O. Hall. 2018. "Braun-Blanquet Data in ANOVA Designs: Comparisons with Percent Cover and Transformations Using Simulated Data." *Marine Ecology Progress Series* 597:13–22. https://doi.org/10.3354/meps12604.

Gabadinho, Alexis, Gilbert Ritschard, Nicolas S. Müller, and Matthias Studer. 2011. "Analyzing and Visualizing State Sequences in R with TraMineR." *Journal of Statistical Software* 40(4):1–37. https://doi.org/10.18637/jss.v040.i04.

Gauthier, Jacques-Antoine, Eric D. Widmer, Philipp Bucher, and Cedric Notredame. 2010. "Multichannel Sequence Analysis Applied to Social Science Data." *Sociological Methodology* 40(1):1–38. https://doi.org/10.1111/j.1467-9531.2010.01227.x.

Heimann-Roppelt, Anna, and Silke Tegtmeier. 2018. "Sequence Analysis in Entrepreneurship Research: Business Founders' Life Courses and Early-Stage Firm Survival." *International Journal of Entrepreneurial Venturing* 10(3):333–61. https://doi.org/10.1504/IJEV.2018.093230.

Ho, Hillbun, Keng-Ming (Terence) Tien, Anne Wu, and Sonika Singh. "A Sequence Analysis Approach to Segmenting Credit Card Customers." *Journal of Retailing and Consumer Services*, first published on November 27, 2020 as doi:10.1016/j.jretconser.2020.102391.

Karhula, Aleksi, Jani Erola, Marcel Raab, and Anette Fasang. 2019. "Destination as a Process: Sibling Similarity in Early Socioeconomic Trajectories." *Advances in Life Course Research* 40:85–98. https://doi.org/10.1016/j.alcr.2019.04.015.

Liao, Tim Futing. 2002. *Statistical Group Comparison*. New York, NY: John Wiley & Sons. https://doi.org/10.1002/9781118204214.

Liefbroer, Aart C., and Cees H. Elzinga. 2012. "Intergenerational Transmission of Behavioural Patterns: How Similar Are Parents' and Children's Demographic Trajectories?" *Advances in Life Course Research* 17(1):1–10. https://doi.org/10.1016/j.alcr.2012.01.002.

Macmillan, Ross, and Ronda Copher. 2005. "Families in the Life Course: Interdependency of Roles, Role Configurations, and Pathways." *Journal of Marriage and Family 67*(4):858–879. https://doi.org/10.1111/j.1741-3737.2005.00180.x.

Nee, Victor, Lisha Liu, and Daniel DellaPosta. 2017. "The Entrepreneur's Network and Firm Performance." *Sociological Science* 4:552–79. https://doi.org/10.15195/v4.a23.

Nightingale, Glenna. 2016. "R Package Lifecourse." The CRAN R project repository. https://CRAN.R-project.org/package=lifecourse.

Onghena, Patrick. 2018. "Randomization Tests or Permutation Tests? A Historical and Terminological Clarification." Pp. 209–227 in *Randomization, Masking, and Allocation Concealment*, edited by Vance W. Berger. Boca Raton, FL: Chapman & Hall/CRC Press. https://doi.org/10.1201/9781315305110-14.

Piccarreta, Raffaella. 2017. "Joint Sequence Analysis: Association and Clustering." *Sociological Methods & Research* 46(2):252–87. https://doi.org/10.1177/0049124115591013.

Pollock, Gary. 2007. "Holistic Trajectories: A Study of Combined Employment, Housing and Family Careers by Using Multiple-Sequence Analysis." *Journal of the Royal Statistical Society, Series A: Statistics in Society* 170(1):167–83. https://doi.org/10.1111/j.1467-985X.2006.00450.x.

Raab, Marcel, Anette Eva Fasang, Aleksi Karhula, and Jani Erola. 2014. "Sibling Similarity in Family Formation." *Demography* 51(6): 2127–54. https://doi.org/10.1007/s13524-014-0341-6.

Schönpflug, Ute. 2001. "Intergenerational Transmission of Values: The Role of Transmission Belts." *Journal of Cross-Cultural Psychology* 32(2):174–85. https://doi.org/10.1177/0022022101032002005.

Studer, Matthias. 2013. WeightedCluster Library Manual: A Practical Guide to Creating Typologies of Trajectories in the Social Sciences with R. Working Paper, Institute for Demographic and Life Course Studies. http://doi.org/10.12682/lives.2296-1658.2013.24.

Studer, Matthias, and Gilbert Ritschard. 2016. What Matters in Differences between Life Trajectories: A Comparative Review of Sequence Dissimilarity Measures. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 179(2):481–511. https://doi.org/10.1111/rssa.12125.

Studer, Matthias, Gilbert Ritschard, Alexis Gabadinho, and Nicolas S. Müller. 2011. "Discrepancy Analysis of State Sequences." *Sociological Methods & Research* 40(3):471–510. https://doi.org/10.1177/0049124111415372.

Studer, Matthias, Emanuela Struffolino, and Anette E. Fasang. 2018. "Estimating the Relationship between Time-Varying Covariates and Trajectories: The Sequence Analysis Multistate Model Procedure." *Sociological Methodology* 48(1):103–35. https://doi.org/10.1177/0081175017747122.

Wahrendorf, Morten, Anja Marr, Manfred Antoni, Beate Pesch, Karl-Heinz Jöckel, Thorsten Lunau, Susanne Moebus, Marina Arendt, Thomas Brüning, Thomas Behrens, and Nico Dragano. 2019. "Agreement of Self-Reported and Administrative Data on Employment Histories in a German Cohort Study: A Sequence Analysis." *European Journal of Population* 35:329–46. https://doi.org/10.1007/s10680-018-9476-2.

Tim F. Liao: Department of Sociology, University of Illinois. E-mail: tfliao@illinois.edu.