A Life History Graph Approach to the Analysis and Comparison of Life Histories*

Sean Fitzhugh^{i,1}

3151 Social Science Plaza,
University of California,
Irvine, CA, USA, 92697

sean.fitzhugh@uci.edu

Carter Buttsⁱⁱ

3151 Social Science Plaza,
University of California,
Irvine, CA, USA, 92697

buttsc@uci.edu

Joy Pixley^c

3151 Social Science Plaza,
University of California,
Irvine, CA, USA, 92697

jpixley@uci.edu

ⁱ Department of Sociology, University of California, Irvine

ⁱⁱ Institute for Mathematical Behavioral Sciences, Department of Statistics, Department of Electrical Engineering and Computer Science, University of California, Irvine

^c California Institute for Telecommunications and Information Technology, University of California, Irvine

¹ To whom correspondence should be addressed: sean.fitzhugh@uci.edu, +16613104997

* This work was supported by ONR award N00014-08-1-1015, ARO award W911NF-14-1-0552, and NSF award IIS-1251267.

ABSTRACT

Life histories are complex, multidimensional entities composed of a series of ongoing, often interrelated contexts, activities, and relationships. Each life domain contributes to the context within which events in other domains are shaped, and is in turn shaped by activities or events in other domains. Capturing these two key aspects of life histories—simultaneous dependence, and multidimensionality—is an important challenge for gerontology and life course research. In this paper we introduce a graph theoretic technique for exploratory analysis of life history data that speaks to this challenge. Using the notion of life history graphs, we represent a single life history as a series of spells (activities with specified start and end points) linked by their patterns of temporal overlap. Given multiple individuals, their associated life history graphs can be directly compared using distance-based methods that assess differences in the topologies of their respective spell structures while preserving qualitative differences between life domains (e.g., non-exchangeability of job and marital spells). Predicated on overlaps across life domains, our approach is designed to incorporate both the multidimensional structure of detailed life history data and the importance of simultaneity. Given a sample of individuals from a target population, our techniques can also be used to identify and test hypotheses regarding the factors that differentiate life histories within the population in question. Using data from the United States, England, China, and Vietnam, we demonstrate how our methods allow us to make comparisons both within and across populations.

Keywords: life course, life histories, interval graph, networks

Abbreviations: LHG, life history graph: a network-based representation of a life history where nodes represent spells and edges represent temporal overlap of those spells

INTRODUCTION

Recent decades have seen significant advances in our ability to represent, elicit, and analyze life history data (Laub and Sampson, 1998; Settersten and Mayer, 1997; Mayer, 2009; Blossfeld and Rohwer, 1997; Aisenbrey and Fasang, 2010). Nevertheless, important challenges remain. Among these is a need for methods that more precisely represent life course concepts (Alwin, 2012), and that place life events within a broader temporal and societal context (Hendricks, 2012; Billari, 2001; Settersten and Mayer, 1997). In this paper, we attempt to respond to these challenges by introducing a novel technique for the exploratory analysis of life histories that allows for the identification and quantitative analysis of population-level patterns in life history structure from high-dimensional (i.e., multiple domain) data.

Systematically representing and analyzing life histories has been an ongoing methodological challenge within the social sciences. Because of the complexity of life history data---and the wide range of questions that can be asked of it---no one approach is well-suited to all analysis tasks. The most common current methods for studying life histories, event history analysis and sequence analysis, are particularly well developed for cases involving discrete events or state sequences where the number of distinct event types is small and states cannot occur simultaneously. Classic examples of contexts in which such methods have been successfully employed include employment history, family history, fertility history, and health trajectories (Blair-Loy, 1999; Abbot and Tsay, 2000; Aisenbrey and Fasang, 2010; Ben-Shlomo and Kuh, 2002; Kuh, et al., 2003; Lynch and Smith, 2003), with more recent applications including complex phenomena such as diffusion (Diekmann, 1989; Nazio and Blossfeld, 2003; Billari and Philipov, 2004), communication dynamics (Butts, 2008a), and classroom behavior (DuBois, et al., 2013). The great strength of these approaches is in capturing patterns of sequential dependence among well-ordered events, either by continuous time stochastic process models (event history analysis) or by exploratory multivariate methods (sequence analysis).

Both sequence analytic and event history techniques become more difficult to apply in the context of spell data, particularly where spells involve large numbers of qualitatively distinct activities or relationships that may overlap in complex ways. The multidimensionality induced by these overlaps poses serious challenges for sequence analysis techniques. Inclusion of multiple life domains (and, further, the potential for multiple overlapping spells within a domain) leads to a combinatorial explosion of the set of discrete states on which sequence analytic representations are based (Pollock, 2007), requiring the use of either complex data reduction methods (MacMillan and Eliason, 2003) or separation of distinct domains (and hence loss of contextual information (Settersten and Mayer, 1997)) to render analysis feasible. Simultaneity rapidly increases the number of distinct states (i.e., the “alphabet”) and although one could merge similar states in the alphabet or prune unusual or uncommon states, such approaches discard potentially important information. One must have a strong understanding of the data to know which states are meaningful and which can be discarded; this is a potentially tenuous assumption given that sequence analysis is frequently employed for exploratory data analysis. A

recently proposed technique by Bürgin and Ritschard (2014) addresses a traditional shortcoming of sequence analysis by rendering simultaneous events through an exploratory data analysis approach, but becomes difficult to apply to alphabets larger than ten or so states. Likewise, event history models for multiple simultaneous spells across multiple domains face a proliferation of event types (i.e., onset and termination events for each potential spell) and potential sources of dependence, making model parameterization and selection difficult without extensive prior knowledge of the potentially relevant social mechanisms within the study population (Marcum and Butts, 2015). Given that life course trajectories within a given domain (e.g., employment history) may shape and be shaped by trajectories in other domains (e.g., educational history, marital status, fertility history, etc. (Germaine, 1975; Elder, 1998; Quick and Moen, 1998; Schneider and Stevenson, 1999)), the space of potential models is typically quite large. As a precursor to a model-driven analysis, it is thus desirable to have exploratory techniques that can reveal the primary correlates of differences in life trajectories within a population; to be useful, these methods must be viable in the presence of multiple domains, and should be sensitive to patterns of temporal overlap (so as to capture the life course context in which spells occur).

In prior work, Butts and Pixley (2004) introduced a framework for representing multi-domain life history data via *life history graphs*. As discussed below, we use the term “graph” here to denote the formal representation of network structure (Wasserman and Faust, 1994); in particular, life history graphs should not be confused with the notion of a *life history chart*, to which they are only loosely related. Life history graphs encode the patterns of overlaps (and, optionally, temporal ordering) among spells, providing a holistic representation of individual life histories that preserves the life course context in which activities and relationships are performed. The contexts in which these spells occur are of key importance (Gerson, 1985; Spain and Bianchi, 1996; Elder, 1995). For example, the job spell of a young, married, childless individual differs markedly from the same job spell performed by a middle-aged, unmarried individual with three children (Moen, 1992; Voydanoff, 1988). The work-family balance and financial security of these two individuals differ substantially and we begin to grasp these differences by understanding the life course contexts in which spells occur. A natural extension of this argument is to examine temporal overlap, derived from a representation of life history as a series of time intervals. A wealth of information is embedded in the timing and overlap of life events, but untangling and making sense of this information is an ongoing challenge, particularly at the population level (Butts and Pixley, 2004; Elder, 1998). In this paper, we build on prior work by providing a method that uses life history graphs as a basis for comparative analysis of multiple life histories. This exploratory data analytic method uses information on coterminousness of activities and relationships to form a foundation for systematic comparison across life histories at the population level. By examining the topology of spell overlaps, we can quantitatively differentiate life histories from one another and identify patterns at the population level that may be suitable for subsequent modeling and analysis (using, e.g., event history methods).

The remainder of the paper proceeds as follows. We begin with a review of the life history graph concept, and some of its connections to substantive concepts from the life course literature. We then discuss the quantitative comparison of life history graphs using structural distances, and the use of these distances with exploratory multivariate methods. This is then followed by an illustrative application of our methodology to life history data from four nationally defined populations (the United States, China, England, and Vietnam). The results within each nation---and stark contrasts across nations---demonstrate the utility of the method and illustrate typical results that provide considerable insights into life histories within a population.

METHODS

Life History Graphs

Life histories can be conceptualized in terms of a series of overlapping intervals, or *spells*, each representing a period in which an individual is involved in a specific activity, role, relationship, or context. Spells may in turn be viewed as belonging to a smaller number of *life domains* representing closely related activities, roles, or relationships, such as schooling, work, marriage, childcare, etc. The structure of spells within a life history provides insight into the numbers, types, and coincidences of roles experienced by an individual over the life course (e.g., unmarried student, working parent, widowed retiree). The timing and coterminousness of spells have important implications on the timing and likelihood subsequent spells (Elder, 1998; Heinz, 2003; Mayer, 2004; Settersten, 2003); an individual's currently active spells influence the types of spells he or she may engage in next. These transitions from one role to another are rarely a linear sequence of events, however. Instead, such life course transitions are shaped by and embedded in multiple trajectories (Elder, 2000; Elder and Johnson, 2003).

The life course is thus a multidimensional entity, in which activities from one domain influence spells in other domains. Employment histories of American women in the 20th century have been a classic example of this phenomenon (Germaine, 1975; Quick and Moen, 1998). Normative expectations for women in their roles as mothers during this period often conflicted with expectations for their roles as workers (Giele, 1998), leading some women to leave the workforce while raising children (Drobnic, et al., 1999; McRae, et al., 1996). Although many eventually returned to the workforce (McRae, et al., 1996), these women experienced interruptions in their employment histories while raising young children. These interruptions, in turn, led in some cases to lowered occupational attainment (McRae, 1993) and/or a later age at retirement (Pienta, 1999; Hank, 2004). To continue the above example, the timing and number of childbirths (and hence parental spells) is influenced both by marital and educational history (Drobnic, et al., 1999; McRae, et al., 1996); but these, in turn, are themselves influenced by

fertility (Jacobs and King, 2002; Moore, 1978), with the entire set of spells comprising a complex system of interdependent domains. These complexities motivate the need for methods that can account for multiple domains simultaneously (Alwin, 2012; Hendricks, 2012; Settersten and Mayer, 1997). By deriving information from overlaps across spells and domains, the life history graph formalism is able to integrate information from multiple trajectories and events.

A *graph* is a formal representation of network structure composed of a *vertex set* (representing potentially interacting entities) and an *edge set* (representing interactions among pairs of entities). A life history graph (Butts and Pixley, 2004) is a type of network in which nodes or vertices correspond to spells, and ties between nodes (i.e., between spells) correspond to temporal overlaps. Life history graphs (LHGs) are thus examples of *interval graphs*, a class of networks frequently used in computer science and operations research to represent e.g., scheduling problems. The life history graph's vertex set consists of the spells in which an individual participates over the course of his or her life. These nodes become tied to one another by temporal overlap, or *coterminousness*. In this paper, we specifically consider two spells to be tied if and only if there exists some point in time at which they are both active. (Obviously, each spell in the life history graph is coterminous with itself and therefore has a reflexive tie.) Applying this principle to an entire life history leads to the corresponding LHG, as shown in Figure 1. Figure 1 depicts an illustrative case of an individual with spells in four domains (education, work, marriage, and childcare), schematically displayed as a life history chart (Figure 1A). By defining spells as adjacent if they overlap in time (thin vertical lines), we construct the edge structure of the corresponding life history graph. We represent this life history graph as an adjacency matrix in Figure 1B. Here the i, j cell is 1 if the spells overlap in time; otherwise the value is 0. In Figure 1C we represent the graph structure of the life history graph. Each spell is represented by a node and nodes are tied if they are coterminous. Elaboration of the set of included life domains (i.e. increased multidimensionality) and/or finer definitions of spell structure allow for increasingly detailed LHGs; there is in principle no limit to the number of domains or spells that may be included a LHG, a property that makes them especially attractive in an exploratory context.

[Figure 1 here]

A

B

	School 1	School 2	School 3	Work 1	Work 2	Marriage	Child 1	Child 2
School 1	1	0	0	0	0	0	0	0
School 2	0	1	0	0	0	0	0	0
School 3	0	0	1	0	0	1	0	0
Work 1	0	0	0	1	0	1	1	0
Work 2	0	0	0	0	1	1	1	1
Marriage	0	0	1	1	1	1	1	1
Child 1	0	0	0	1	1	1	1	1
Child 2	0	0	0	0	1	1	1	1

C

Figure 1: Demonstration of how we use spell overlaps (A) to generate a life history graph. We represent the life history graph as an adjacency matrix (B) and sociogram (C).

Although we employ life history graphs predicated on coterminousness, one may use any number of measures for ties between spells, as described in Butts and Pixley (2004). For instance, if one were interested in the *duration* of overlaps between spells, one could use *valued* life history graphs. In such graphs, two spells are tied if they overlap in time and the weight of that tie is proportional to the duration of the overlap. In practice, however, these valued life history graphs tend to be strongly correlated with simple life history graphs and yield qualitatively similar results. Additionally, one could employ spell precedence graphs to preserve temporal order (Butts and Pixley, 2004). The life history graph approach is flexible enough to enable multiple classes of graph-based representations of the life history. Which representation one chooses depends on which aspects of the life history are most important to the analysis (timing of overlaps, duration of overlaps, order of spells).

Comparative Analysis of Life History Graphs

To compare LHGs across multiple lives, we construct a life history graph to represent each member of the population of interest (or of a sample therefrom) via the procedure outlined above. A typical source of such data is the retrospective life history study, where respondents are asked about their past spells within a fixed set of domains (Freedman, et al., 1988). Having constructed an LHG for each sampled individual, we may then analyze the resulting collection of graphs using any of a range of relevant techniques from the social network analysis literature (Wasserman and Faust, 1994; Butts, 2008b); here, we focus on distance-based methods that are a natural analog of distance-based sequence comparison methods familiar to life course researchers (Han and Moen, 1999; Abbot and Tsay, 2000).

Distance-based multivariate analysis of LHGs begins with the construction of a distance matrix for all graphs in the sample, D , such that D_{ij} is a measure of the dissimilarity between the LHGs of individuals i and j . We here employ the *structural distance* of Butts and Carley (2005) for this purpose, a family of metric distance measures that assess the structural similarity of graphs based on correspondences between their edge sets. Intuitively, the structural distance between two graphs (G_1 and G_2) is equivalent to the minimum number of additions and/or deletions to G_1 's edge set to make it identical to the edge set of G_2 , with the proviso that vertices from G_1 may be matched to any vertex of G_2 to which they are functionally equivalent. The calculation of the structural distance is a problem whose computational complexity scales with $O(n!)$, although simulated annealing methods reduce calculation time and are rather simple to implement (Butts and Carley, 2005). These structural distances are symmetric ($d_{i,j}=d_{j,i}$), reflexively zero ($d_{i,i}=0$), and satisfy the triangle inequality; hence, they are metrics, and may be used as inputs to clustering, scaling, or other methods that depend upon this assumption.

In the context of life history graphs, the structural distance thus refers to the minimum number of spell overlaps we must add and/or remove to make any given pair of life histories identical to one another, where spells are allowed to be matched only to other spells of the same type. To allow for calculation of structural distance between life history graphs of different sizes we add “phantom spells” into the smaller graph as an accounting device to give it the same number of vertices as the larger graph. Unlike regular spells, these have no reflexive ties or ties to other nodes, and serve only to ensure that comparisons are always well-defined. Without reflexive ties, these phantom spells do not affect distances between graphs, as the number of ties in the graph remains the same when phantom spells are introduced.

Structural distance computation is closely analogous to optimal matching (Abbot and Tsay, 2000), and may be summarized as follows. First, let us consider a hypothetical case of two LHGs in which there is only one substantively meaningful mapping of vertices between the two graphs: i.e., each vertex is of a distinct type, having only one counterpart in the other graph. Such a case is illustrated in Figure 2, which shows two small LHGs with one spell in each of several domains. By representing each graph in adjacency form, and sorting the rows and columns of each matrix so that the type of spell associated with each row/column for one matrix matches that of the other (Figure 2C), we can easily “map” the cells of the first LHG onto those of the other. The number of overlap changes needed to take one graph into the other (the Hamming distance (Hamming 1950; Butts and Carley, 2005), d_H) is then simply the sum of elementwise differences between the two sorted adjacency matrices. If we were using valued life history graphs, however, the difference would be equal to the sum of *values* of overlaps needed to make the two graphs identical. The mapping, or ordering of one set of vertices relative to the other is referred to as a *labeling* of the two graphs. In the case where there is only one meaningful mapping of vertices between graphs, we say that the graphs are *fully labeled* with respect to each other, and the resulting *labeled distance* (here, the Hamming distance) is the structural distance. In substantive terms, this distance expresses the total difference in structure between the two life histories.

[Figure 2 here]

Figure 2: Comparison between two life history graphs highlights differences between adjacency matrices (A, C) and sociograms (B, D).

Now, let us consider the more general case in which there is more than one spell of a given type, and hence more than one way to align (“label”) the graphs. For expository purposes, it is useful to think of labelings in terms of the ordering of rows and columns in the graph adjacency matrix of each graph (see Figure 3), and we shall use that description here. Following Butts and Carley (2005), we introduce a *labeling function*, ℓ , such that $\ell(G)$ represents the adjacency matrix of LHG G in a specific row/column ordering. Since each row and column of G 's adjacency matrix reflects a spell of a particular type, we also introduce a *coloring function*, κ , that returns the vector of spell types (referred to generically as “colors”) for any given labeling of G . For instance, in the default labeling ℓ of the LHG in Figure 3A, (School 1, School 2, School 3, Work 1, Work 2, Marriage, Child), $\kappa(\ell(G))$ would be (School, School, School, Work, Work, Marriage, Child). A labeling that permuted e.g. the second and third spells of this graph would result in the same value of κ ; however, for an alternate labeling ℓ' that permuted the second and fourth spells of G (while leaving the others in the same order) would yield the substantively distinct $\kappa(\ell'(G)) = (\text{School, Work, School, School, Work, Marriage, Child})$. Choice of labeling is arbitrary when dealing with a single graph, but is important when we map one graph to another: in particular, a substantively meaningful comparison demands that each spell in one graph be mapped to a spell in the other graph having the same type (“color”). This constraint is equivalent to preservation of coloring—i.e., for two LHGs G_1, G_2 , a respective labeling ℓ_1, ℓ_2 is only valid when $\kappa(\ell_1(G_1)) = \kappa(\ell_2(G_2))$. As noted above, the Hamming metric d_H

provides a natural basis for assessing differences between graphs when the mapping from G_1 to G_2 is unique. Although alternatives to the Hamming metric as the base distance can also be employed, they are less interpretable in the present context. In the general case, the structural distance is defined as the minimum of this distance across all valid labelings:

$$D_{ij} = \min_{\ell_1, \ell_2: \kappa(\ell_1(G_1)) = \kappa(\ell_2(G_2))} d_H(\ell_1(G_1), \ell_2(G_2))$$

Taking the minimum distance ensures that measured differences between graphs result only from differences in the underlying graph structure (including both overlap patterns and spell counts), and not on arbitrary properties of how the graphs are represented (e.g., order in which spells were entered into the data set). The resulting distance has a number of appealing properties. First, it is a true metric (i.e., it is reflexive, symmetric, and obeys the triangle inequality), which facilitates interpretation and use with techniques such as multidimensional scaling. Second, in the case of LHGs, the structural distance between two graphs is always greater than or equal to the sum of differences in number of spells between the two graphs, by type; thus, it preserves the intuition that the distance between two life histories should increase as the difference in their composition increases. Finally, the measure preserves substantive distinctions among spell types in a manner that is selectable by the researcher. By choosing κ to allow more or less refined distinctions, the researcher can “focus” the analysis on differences resulting from very general differences among life domains (e.g., employment versus romantic relationships) versus finer details (e.g., full-time versus part-time employment).

[Figure 3 here]

A	School 1	School 2	School 3	Work 1	Work 2	Marriage	Child
School 1	1	0	0	0	0	0	0
School 2	0	1	0	1	0	0	0
School 3	0	0	1	0	0	0	0
Work 1	0	1	0	1	0	1	1
Work 2	0	0	0	0	1	1	1
Marriage	0	0	0	1	1	1	1
Child	0	0	0	1	1	1	1

B	School 1	School 2	School 3	Work 1	Work 2	Marriage	Child
School 1	1	0	0	0	0	0	0
School 2	0	1	0	0	0	0	0
School 3	0	0	1	1	0	0	0
Work 1	0	0	1	1	0	1	1
Work 2	0	0	0	0	1	1	1
Marriage	0	0	0	1	1	1	1
Child	0	0	0	1	1	1	1

Figure 3: Adjacency matrices representing two life histories. By permuting the labeling, and therefore the adjacency matrix rows and columns, of A from (School 1, School 2, School 3, Work 1, Work 2, Marriage, Child) to (School 1, School 3, School 2, Work 1, Work 2, Marriage, Child), we find identical life history graphs across (A) and (B). The resulting distance between them is thus 0.

Given the calculation of distances among all pairs of graphs in our sample, we can proceed with any standard multivariate or other distance-based analytical technique. These include methods such as hierarchical clustering (Johnson, 1967) and multidimensional scaling (Cox and Cox, 2000; Kruskal and Wish, 1978), which are widely used for the exploratory analysis of multivariate data. Hypotheses regarding differences in life histories between or among groups can also be investigated using non-parametric procedures such as permutation tests, or by model-based analysis. Below we illustrate several techniques for identifying and analyzing patterns of differentiation across life histories, applying the above methods to samples from four modern populations.

RESULTS

Application to Life Histories in Four Modern Populations

As an illustrative application of our approach, we consider life history data from four modern populations: China, England, the United States, and Vietnam. The Life Histories and Social Change in Contemporary China survey provides life history data (N=6,090) from two samples, one urban and one rural, that are respectively representative of their populations age 20-69 (Treiman and Walder, 1996). The English Longitudinal Study of Aging (N=12,100) uses a nationally representative sample of English residents over the age of 50. The American life history data come from the Wisconsin Longitudinal Study (N=10,317), whose sample is the cohort of Wisconsin's high school graduation class of 1957. Finally, the Vietnam Life History Survey (N=921) uses stratified sampling of urban and rural residents. The survey has two urban samples and two rural samples, all of which are similar in size. Each of these four surveys gathers retrospective life history data, where respondents provide start and end dates for a variety of spells throughout their lives.

To facilitate comparison, we separate respondents' life histories into five domains that are common to all of our data sources: education, employment, military participation, marriage, and childcare. The five spell types correspond to the “colors” we defined in the previous section. Marriage, military participation, and childcare spells are operationalized in a similar manner across all three surveys. In each survey informants report the end dates of all their spells except childcare spells. To address this we had to estimate the length of childcare spells. Each childcare spell begins with the birth each child. To estimate dates when children leave the household, we used existing data from each country on the average age of departure from the parental household. We ended childcare spells at these respective ages. We tried multiple estimates and found the results to be quite robust to different lengths of childcare.

The surveys differ slightly in their approaches to collecting data on education and employment spells. The Vietnamese and Chinese education spells are most straightforwardly collected, with respondents providing the start and end points of each school spell (primary, secondary, university, etc.). The Wisconsin Longitudinal Survey defines post-high school education spells as any period of time during which the respondent enrolled in a course for college credit, not including vocational, technical, or on-the-job training. Finally, the English sample specifies education spells as full-time school attendance. All four surveys operationalize employment to include both part-time and full-time work. The length of time one must hold the job varies across the surveys, however. The Chinese data is least restrictive and records all employment spells noted by the respondent. Both the English and American surveys record all employment spells lasting longer than six months. The Vietnam Life History Survey records all spells of primary employment, defined as the job at which the respondent worked the most hours. However, it only records secondary, tertiary, etc. jobs that last longer than one year. The other surveys have no additional restrictions on recording additional employment spells beyond

the respondent's primary job. Given the scales at which we are analyzing life histories, we are not particularly concerned with these differences in definitions for employment spells. If one were using the life history graph approach and using coterminousness at a finer temporal level, such as monthly or weekly, these differences would likely be a bigger obstacle. One must always be cautious when comparing life histories, as definitions of spells (especially employment) may vary from survey to survey. However, there is enough overlap among the definitions of these spells across these four surveys that we do not expect differences in the operationalization of spells to produce artifacts that would substantively alter our findings.

Given our choice of domains, we construct a life history graph to represent each individual in each of the samples and then find the structural distances between all graphs within each sample using the approach described above. The life history graph approach is particularly well suited for this task of differentiating life histories by calculating distances between individual life history graphs. If we had more life domains with spell data, we could make even finer distinctions between individual life histories. By contrast, with only these five domains, a sequence analysis approach struggle to accommodate alphabets of the size we observe. With five domains we will observe up to 2^5 combinations of spell overlaps if we neglect cases with multiple ongoing spells from the same domain (which is unreasonable, as it is quite common for people to raise multiple children simultaneously or work multiple jobs). Once we recognize multiple spells from the same domain, we observe a combinatorial explosion of overlaps we would need to represent with our alphabet. To construct an alphabet representing all observed combinations of spell overlaps, we would need an alphabet of size 333 in our China sample, 629 in our England sample, 346 in the United States sample, and 137 in our Vietnam sample. Existing SA approaches cannot accommodate alphabets of this size and would further struggle with a larger number of domains. We use metric multidimensional scaling (MDS) to represent our distance matrix in a visualizable (Euclidean) space (Kruskal and Wish, 1978). We visualize the differences among life histories by plotting the first two dimensions of the MDS solution. Although additional dimensions could reveal additional patterns, we find that the first two dimensions are sufficient to capture most of the variation in life histories, as illustrated in Figure 4. (All analyses were performed using the R statistical computing environment (R Core Team), with network analytic functions supplied by the `sna` library (Butts, 2008c).)

[Figure 4 here]

Figure 4: Eigenvalues demonstrate the variation explained by each additional dimension.

In the visual representation of the MDS solution, points that are close together represent individuals who have similar overlaps in their life histories, while those further apart have more dissimilar life history graphs. Although the sizes of our samples vary, this does not directly impact the differentiation we observe in the structural distances or in the MDS solution. However, larger samples give us more opportunities to observe individuals whose life histories are outliers from the population and, correspondingly, more opportunities to observe those with typical life histories. As we show, applying simple visualization techniques (e.g., varying coloration, point scaling, etc. by known covariates) to the MDS solution allows us quickly to identify the primary factors differentiating life histories within each country, and to compare patterns of life history differentiation across countries.

As a first step to identifying correlates of life history differentiation, we use coloration to distinguish among individuals with different numbers of spells in a particular domain. Where

differences in numbers of spells are aligned with total life history distance (as revealed by the distances between points in the metric MDS solution), the domain in question can be seen to act as a major axis of differentiation within the population as a whole; by contrast, spell counts not strongly associated with life history differences indicate domains whose differential participation is not strongly associated with large-scale patterns of life history differentiation within the broader population. For instance, in Figure 5 we show the two-dimensional MDS solution for life history differences in the English sample, with individual respondents colored by either their total number of childcare spells (A) or job spells (B). As the plot reveals, the childcare spells are associated with strong, qualitative differences between life histories, visible as a well-defined band structure ranging from those without children (black) to those with eight or more (gray). Within each band, variation is strongly associated with number of job spells (with increasing differentiation between those with more jobs, and increasing differentiation by childcare spells among those with a greater total number of job spells). It is interesting to note that while childcare spells are associated with a discretely banded structure, differences associated with job spells are more continuously varying; it would appear, then, that while both number of children and number of jobs are strongly associated with differences in life histories, the two operate in distinct ways.

[Figure 5 here]

Figure 5: Metric MDS solution represents the differences among individual life histories within the England sample. Individuals are colored by the covariates that primarily drive differentiation, where the colors represent the number of spells of that type in which the individual has participated.

By contrast, in Figure 6 we find much less differentiation among those with different histories of military participation (A) and education (C), a finding we confirm statistically in Table 2. Although these spells may be important for particular individuals, they are not associated with the major patterns of differentiation in English society as a whole. We find mild differentiation (B) among those who have not been married and those with one or more marriages. Based on the patterns in Figure 5, however, this differentiation appears to be a reflection of the difference between those with and without children. Many of the individuals without childcare spells also lack marriage spells.

[Figure 6 here]

Figure 6: Other spells play less of a role in differentiating English life histories.

In the next several figures we examine particular domains and use color to distinguish individuals with different numbers of spells in that domain. Whether the domain plays a key role in differentiating members of the population is revealed by the extent to which these colors reflect distinct patterns in the population. We examine four domains that are associated with the largest variation across life histories in these four populations: fertility history, marital history, military participation, and employment history.

In Figure 7 we color nodes based on the number of child care spells they report. In all four samples we observe distinct bands associated with numbers of child care spells. These illustrate clear differences among individuals with different numbers of children, a sign that fertility history is a clear axis along which life histories are differentiated. While the bands are quite distinct in Vietnam (D) and, to a lesser extent, the United States (B), we observe less homogeneity among those with 5 or more children in England (A) and 4 or more children in China (C).

[Figure 7 here]

Figure 7: MDS solutions illustrate how fertility history differentiates life histories across our four populations.

Figure 8 distinguishes among individuals based on their marital histories. In England (A), the United States (B), and Vietnam (D) we find distinct groupings of individuals without a marriage spell. These findings in tandem with those from Figure 7 suggest that there is little differentiation among the life histories of those without children and spouses. However, in China (C) we find no clear patterns associated with marital histories.

[Figure 8 here]

Figure 8: A comparison of the effect of marital histories on differentiation across life histories.

In Figure 9 we use coloration to distinguish between those who have and have not joined the military at some point during their lives. We do not find any clear differentiation between those with differences in military participation in England (A), the United States (B), and China (C). Vietnamese life histories (D) generally demonstrate dissimilarity among those with and without a history of military service.

[Figure 9 here]

Figure 9: Differentiation of life histories associated with participation in the military.

Figure 10 illustrates distances among life histories based on differences in employment history. We find clear patterns of differentiation in England (A), the United States (B), and Vietnam (D). In both the U.S. and England we find a continuum between those with 0 jobs and those with 8 or more jobs. This accounts for a large amount of variation in life histories, as those with employment histories differing by multiple employment spells tend to have dissimilar life

histories. In both the United States (B) Vietnam (D) we find a handful of individuals without past employment who are clear outliers from the rest of the population.

[Figure 10 here]

Figure 10: Employment histories and their effects on differentiating life histories.

Exploratory analysis with multidimensional scaling yields useful insights, but is both qualitative in nature and difficult to extend to the high-dimensional case. Where an MDS representation lacks sufficient clarity or where we are interested in quantitative measurement and/or hypothesis testing, we can use additional techniques such as regression analysis of distances to examine the distance structure. This exploratory data analysis provides insight into the factors differentiating life histories and lays a foundation for testing of theoretically informed mechanisms to explain *why* these patterns of differences emerge.

Regression Analysis

Linear regression and ANOVA are examples of very familiar analytical tools that can be easily applied to the analysis of structural distances. In Table 1 we model the graph distances (log-transformed) between all *pairs* of individuals as a function of differences in number of spells in our domains--education, military service, work, marriage, and childbirth--as well as differences in demographic measures of gender and age. The coefficients associated with each effect indicate the relationship of a one-unit change in the similarity of the indicated property across individuals with the corresponding distance between their life history graphs, holding all other factors constant. A large effect for a given attribute thus indicates that any given pair of individuals differing in this attribute will have, on average, a correspondingly large difference between their overall life histories. As Table 1 shows, the factors associated with larger differences in life histories vary markedly across countries. For instance, for two otherwise similar individuals, a one-spell difference in military participation (i.e. one was in the military and one was not) is associated with a 10% increase in distance in China, a 6% increase in England, and a 15% increase in the United States, while in Vietnam this rises to a substantial 54% increase in distance. This clearly tells us that spells of military service are associated with tangible differences between individual life histories in Vietnam, while the difference is smaller in China, England, and the United States.

[Table 1 here]

Table 1: Linear models illustrate the effects of spell counts on pairwise log-transformed differences between individual life histories.

	China	England	United States	Vietnam
Intercept	2.840***	2.890***	2.336***	1.597***
Δ College	0.269***	-0.014***	0.136***	0.372***
Δ Military	0.099***	0.059***	0.145***	0.539***
Δ No. Jobs	0.219***	0.133***	0.195***	0.139***
Δ No. Marriages	0.007***	0.094***	0.186***	0.145***
Δ No. Children	0.070***	0.203***	0.311***	0.404***
Δ Gender	0.008***	0.020***	-0.012***	-0.031***

Δ Age (per 10 yrs.)	0.028***	-0.019***	0.041***	0.015***
DF	20,680,597	30,674,584	1,861,477	412,678
Adj. R2	.369	.55	.443	.557

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

While some effects are important in more than one country, none of the four countries examined here show identical patterns across all life domains. In China we find that the largest distances between individual pairs of life histories are associated with differences in employment history and college attendance. Employment history and college attendance also play a strong role in differentiating individuals lives in the United States, but differences in fertility history are associated with large differences between pairs of individuals. Fertility history also plays an important role in differentiating individual life histories in England, as does employment history. Fertility history and college attendance are major differentiators in Vietnam, but military participation is the largest differentiator between individual Vietnamese life histories. Factors traditionally associated with differences between individuals, such as gender and age at interview, become minor once we account for the number of domain-specific spells in our model. That is, these effects are not associated with major differences in life histories above and beyond the spells we include in the model. We also find differences across the intercepts for each model. These intercepts reflect the level of differentiation we expect to see between two individuals who are identical across the covariates in our models. By itself this is a particularly interesting finding that reflects that the typical differences among Chinese, English, and American life histories are considerably larger than the baseline differences we see among Vietnamese life histories during the study period. This may be reflective of a larger number of common life trajectories in China, England, and the United States; these kinds of findings are precisely what this kind of exploratory data analysis is intended to elicit in order to provoke additional, more refined research questions.

Working with the distance matrix in a linear modeling framework increases our risk of exposure to row/column autocorrelation, as each individual will appear multiple times in these pairwise comparisons. Such autocorrelation does not alter the first-order properties of regression statistics (Krackhardt, 1988) but can deflate standard errors. A Durbin-Watson test (Durbin and Watson 1950; Durbin and Watson, 1951) indicates that we have weak, positive autocorrelation of our residuals. Our DW statistics ranged between 1.5 and 1.75 in our models, which is a small enough deviation from 2 (which indicates no autocorrelation) to be of minimal concern (Wang and Jain, 2003). The DW statistic was significant, however, and in an abundance of caution we used QAP semi-partialling (Dekker, et al., 2003) as implemented in the *sna* software package (Butts, 2008c) to replicate our results and demonstrate their robustness to this mild autocorrelation.

While certain factors may have large effects in our regression model, those factors may not always account for most of the population variance explained (Kutner, et al., 2005; Cohen, et al., 2013). In Table 2, we employ an ANOVA on the models of Table 1 to determine how much *population-level variance explained* is associated with each variable in our models, conditional on those added previously. While difference in military service, for example, has a large mean effect on the distance between any given pair of individuals in Vietnam, it accounts for only a small fraction of the total variation in the population. The total sum of squares thus reflects a combination of the variation within the sample on the characteristic in question and the extent to which the characteristic is associated with distinct life trajectories within a given society.

[Table 2 here]

Table 2: ANOVA demonstrates how much variation between log-transformed distances at the population level is associated with differences in spells.

	DF	China		DF	England	
		Sum Sq.	Mean Sq.		Sum Sq.	Mean Sq.
College	1	236,695	236,695	1	967	967
Military	1	36,883	36,883	1	15,096	15,096
Jobs	1	2,548,267	2,548,267	1	3,131,529	3,131,529
Marriages	1	1,509	1,509	1	115,645	115,645
Children	1	106,291	106,291	1	2,440,342	2,440,342
Gender	1	382	382	1	4,385	4,385
Age	1	17,579	17,579	1	5,890	5,890
Residuals	20,680,597	5,038,689	0.7754	30,674,584	5,089,560	0.1659

	DF	U. S. A.		DF	Vietnam	
		Sum Sq.	Mean Sq.		Sum Sq.	Mean Sq.
College	1	10,186	10,186	1	6,814	6,814
Military	1	9,905	9,905	1	38,704	38,704
Jobs	1	139,706	139,706	1	7,057	7,057
Marriages	1	26,723	26,723	1	11,085	11,085
Children	1	360,269	360,269	1	380,129	380,129
Gender	1	61	61	1	95	95
Age	1	132	132	1	78	78
Residuals	1,861,477	689,175	0.3702	412,678	353,361	0.8563

As the above suggests, the effects that account for the largest fraction of variation at the population level do not always match those associated with the strongest effects on differences between individuals. The strongest effects on the pairwise distances between individuals are employment history and education level in China, number of children and number of jobs in England, number of children, marriages, and jobs in the United States, and military participation,

education, and number of children in Vietnam. However, employment history counts for much of the population variation explained in life histories in China, England, and the United States, while fertility history plays a major role in differentiating lives in England, the United States, and Vietnam. Even though in some cases we see large pairwise differences associated with college attendance and military participation, these generally account for relatively little of the population-level variance in life histories. Together, linear regression and ANOVA demonstrate where we observe notable effects between individual life histories where we factors explain much of the population variance, respectively. This provides insight into the patterns of differentiation at both the individual- and population-level and serves as a foundation for further analysis. Having recognized the primary correlates of differentiation, one could further analyze the characteristics of life histories that contribute to these trajectories.

Network Analysis

The novelty of the life history graphs approach extends beyond the systematic representation and comparison of life histories. The network representation of life histories enables further investigation into life histories with an extensive suite of methods in network analysis (Wasserman and Faust, 1994; Butts, 2008b). Harnessing the network representation of life histories provides a highly versatile foundation for analysis, one which cannot be reproduced with any other technique for representing life histories. Furthermore, analyzing individual life histories with network analysis enables a principled approach to person-centered analysis, a holistic perspective on representing and comparing narratives of life histories (Singer, et al. 1998). In this section we illustrate how measures of centrality lend help to highlight critical spells in one's life history.

Measures of centrality are among the most commonly used descriptive statistics in network analysis. Centrality indices capture the extent to which nodes occupy central roles in a network (Freeman, 1979). Several centrality measures exist to capture different definitions of centrality and here we use *betweenness* centrality, which measures the extent to which a node sits on many shortest paths between other nodes in the network. Nodes with high betweenness centrality connect groups of nodes that are otherwise distantly connected or disconnected. In the context of life histories, spells with higher betweenness centrality link together distant or disconnected elements of the life course. Additionally, spells with high betweenness may indicate bridges between life epochs. Betweenness centrality helps to illustrate spells in the life history that link together multiple periods of the life history or serve as transitional periods.

Patterns of betweenness within a population of life histories illustrate spells that play important roles in connecting life epochs. For each of the life histories in our four samples we calculated the normalized betweenness of all spells; we then average these centrality scores across domains (e.g. average betweenness of all work spells in a life history). We illustrate those domain-specific centrality scores in Figure 11. We find that marital spells have the highest

betweenness in England and the United States while military spells have the highest betweenness in China and work spells have the highest betweenness in Vietnam.

[Figure 11 here]

Figure 11: Average betweenness centrality for each spell domain highlights domains that link life epochs or serve as transitions in the life course.

We highlight a handful of examples of life histories whose life history graphs represent the typical pattern of betweenness from Figure 11. In these figures we illustrate both a life history chart and a life history graph, where nodes are color-coded by spell and their size is proportional to their betweenness. In Figure 12 we illustrate the life history chart and its corresponding life history graph of an individual whose betweenness closely resembles the typical pattern of betweenness in Chinese life histories. Military and work spells have the highest betweenness in this case while marriage and child care have the lowest betweenness. The military spell bridges the educational and working periods of the life course while the work spells bridge the education/military periods and the more stable work-marriage-parenthood period of the individual’s later life.

[Figure 12 here]

Betweenness Centrality Exemplar: China

Figure 12: Life history chart and life history graph of a Chinese male with typical spell betweenness distribution.

In Figure 13 we highlight a woman whose life history reflects typical betweenness patterns in England, where marriage spells have high betweenness, work spells have moderate betweenness, and education, military participation, and parenthood spells have low betweenness. In the illustrative example the marriage spells link three distinct periods of the life history. Early life in she attended school and then worked multiple jobs simultaneously, with each job lasting a short period of time. Not long after her first marriage she left the workforce to have two children and later returned to the workforce when her children reached school age. After a divorce and her children's transition to adulthood, she enters the third epoch of her life in which she remarries and leaves the workforce for two years before attaining another job. Her marriage spells acquire high betweenness for linking three otherwise disconnected epochs while her employment spells link together much of her life history and gain moderate betweenness.

[Figure 13 here]

Betweenness Centrality Exemplar: England

Figure 13: Life history chart and graph of an English female with typical spell betweenness distribution.

Figure 14 represents an American woman with high marital betweenness and low betweenness for childcare, work, and education spells. In this case, her marital spell links together distant aspects of her life history. After marriage she has a child and starts working when her child turns eleven. She briefly exits the workforce and then returns shortly before her child leaves the household. She then continues working until the present day. The betweenness measure in this life history helps to illustrate the marital spell that links together a variety of other activities in her life.

[Figure 14 here]

Betweenness Centrality Exemplar: United States

Figure 14: Life history chart and graph of an American female with typical spell betweenness distribution.

In our final example in Figure 15 we illustrate a male with typical betweenness patterns in Vietnam. His work spells have the highest betweenness and link his marital history and fertility history in later life to his education and military participation in early life. Once again, the betweenness measure helps to illustrate which spells sit on many shortest paths between other spells in the life history; in this case the pattern seems to indicate a transition from early life to later life.

[Figure 15 here]

Betweenness Centrality Exemplar: Vietnam

Figure 15: Life history chart and graph of a Vietnamese male with typical spell betweenness distribution.

The above cases illustrate the insights provided by one of the many centrality indices that can be employed to quantify spell-level features of life history structures. Closeness centrality, degree centrality, eigenvector centrality, and other centrality measures will highlight other aspects of the life history. Beyond centrality, other network properties such as cohesive subgroups and structural equivalence offer further tools for extracting information from life history graphs. The breadth and depth of network analytic techniques available to analyze life history graphs, and the richness of the graph theoretic representation of life history data, are unique benefits of this approach.

DISCUSSION

We have demonstrated the simplicity and utility of employing life history graphs for the representation and exploratory analysis of life histories. The method is well suited for identifying and measuring population-level patterns in life history data, particularly when large numbers of domains must be considered simultaneously. Our approach yields insights at multiple levels of analysis: on the one hand, we can determine what is associated with population-level variation in life histories, while on the other we can also identify the covariates that predict large differences in the life histories of any two individuals. The intuitions gained from this method are immediate and provide us a foundation for further analysis of the study population. Once we know what factors are at work in differentiating life histories---both between individuals and across populations---we can begin to refine our questions about the population of interest. With such refinement we can begin formulating and testing hypotheses using model-based approaches.

Obviously, life history graph analyses of the type described here are not the only tools a social scientist should wield when studying life histories. While distance-based methods work well for data exploration, they do not e.g. allow for direct representation of social processes or dynamics. The findings from such methods do, however, provide the researcher with insights that can form the basis for model building. Knowing what accounts for substantial differences between individuals, or what accounts for population-level distances, can provide important clues as to the factors that should be included in event history or other models (or, alternately, factors that are unlikely to prove pivotal). Likewise, observed patterns of differentiation at the social level can serve as important *targets* for model adequacy checking, particularly in demographic or policy settings (in which prediction beyond the individual level is of heightened importance). Additionally, one might be interested in determining whether groups identified by differences at the population level are associated with particular outcomes, such as stress (Pearlin, et al., 1981) or health outcomes (Meadows, et al., 2008; Ben-Shlomo and Kuh, 2002; Kuh, et al., 2003; Lynch and Smith, 2003). Cluster analytic techniques are of particular utility in this regard.

Although we illustrate our method with individual life histories, we can extend this method to any kind of temporal data composed of overlapping events. Whether working with short-term temporal data (e.g. composed of spells occurring within a single day or week) or traditional life history data, this method identifies patterns of differentiation at individual and population levels.

Conclusions

Predicated on the analysis of multiple, simultaneous life trajectories and utilizing a precise, systematic method for quantifying and comparing differences across life histories, the life history graph methods introduced here address several of the chief concerns surrounding exploratory analyses of life histories. Analysis of structural distances between LHGs enables a quantitative comparison of typical differences between individual life histories and factors that contribute to population-level differences in life histories. Used with visualization tools such as multidimensional scaling, this method enables simple pattern detection among large populations of life histories. Furthermore, network analysis enables further investigation into the features of individual life histories. While the use of methods that encompass multiple domains is a methodological development still in its early stages, recent developments provide a foundation for social scientists to begin to incorporate multiple, simultaneous domains in their analyses (Butts and Pixley, 2004; MacMillan and Eliason, 2003; Gauthier, et al., 2010; Pollock, 2007). With these tools we can develop a more complete understanding of the life course, as opposed to studying lives as isolated, fragmented sequences of events.

REFERENCES

Abbot, Andrew and Tsay, Angela. (2000). "Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect." *Sociological Methods and Research*. Vol. 29, Issue 3.

<http://dx.doi.org/10.1177/0049124100029001001>

Aisenbrey, Silke and Fasang, Anette E. (2010). "New Life for Old Ideas: The 'Second Wave' of Sequence Analysis Bringing the 'Course' Back Into the Life Course" *Sociological Methods Research*. Vol. Vol. 38, No. 3. pp. 420-462.

<http://dx.doi.org/10.1177/0049124109357532>

Alwin, Duane F. (2012) "Integrating Varieties of Life Course Concepts" *Journals of Gerontology, Series B: Psychological and Social Sciences*. Vol. 67B, No. 2. pp. 206-220

<http://dx.doi.org/10.1093/geronb/gbr146>

PMid:22399576 PMCID:PMC3307990

Banks, David and Carley, Kathleen M. (1994). "Metric Inference for Social Networks." *Journal of Classification*. Vol. 11, No. 1. pp. 121-49

<http://dx.doi.org/10.1007/BF01201026>

Ben-Shlomo, Yoav and Kuh, Diana. (2002) "A Life Course Approach to Chronic Disease Epidemiology: Conceptual Models, Empirical Challenges, and Interdisciplinary Perspectives" *International Journal of Epidemiology*. Vol. 31, No. 2. pp. 285-293.

<http://dx.doi.org/10.1093/ije/31.2.285>

PMid:11980781

Billari, Francesco C. (2001) . "The Analysis of Early Life courses: Complex Descriptions of the Transition to Adulthood" *Journal of Population Research*. Vol. 18, No. 2. pp. 119-142.

<http://dx.doi.org/10.1007/BF03031885>

Billari, Francesco C. and Philipov, Dimiter. (2004). "Women's Education and Entry into a First Union: A Simultaneous-Hazard Comparative Analysis of Central and Eastern Europe." *American Journal of Sociology*. Vol. 104, No. 5. pp.1346-97.

Blair-Loy, Mary. (1999). "Career Patterns of Executive Women in Finance: An Optimal Matching Analysis." *American Journal of Sociology*. Vol. 104, No. 5. pp.1346-97.

<http://dx.doi.org/10.1086/210177>

Blossfeld, Hans-Peter and Rohwer, Götz. (1997). "Causal Inference, Time and Observation Plans in the Social Sciences" *Quality and Quantity*. Vol. 31, Issue 4. pp. 361-384.

<http://dx.doi.org/10.1023/A:1004289932598>

Bürgin, Reto and Ritschard, Gilbert. (2014). "A Decorated Parallel Coordinate Plot for Categorical Longitudinal Data" *The American Statistician*, Vol. 68, Issue 2. pp. 98-103.

1
2
3
4 Butts, Carter T. (2008a) "A Relational Event Framework for Social Action" *Sociological*
5
6 *Methodology*. Vol. 38, Issue 1. pp. 155-200.
7
8
9

10
11 Butts, Carter T. (2008b) "Social Networks: A Methodological Introduction." *Asian Journal of*
12
13 *Social Psychology*. Vol. 11, No. 1. pp. 13-41.
14
15
16 <http://dx.doi.org/10.1111/j.1467-839X.2007.00241.x>
17
18
19
20

21 Butts, Carter T. (2008c). "Social Network Analysis with sna" *Journal of Statistical Software*.
22
23 Vol. 24, No. 6. Pp. 1-51.
24
25
26
27

28 Butts, Carter T. and Carley, Kathleen M. (2001) "Multivariate Methods for Interstructural
29
30 Analysis" CASOS working paper, Center for the Computational Analysis of Social and
31
32 Organization Systems, Carnegie Mellon University.
33
34
35
36
37

38 Butts, Carter T. and Carley, Kathleen M. (2005) "Some Simple Algorithms for Structural
39
40 Comparison" *Computational Mathematical Organization Theory*. Vol. 11, No. 4 pp. 291-305.
41
42 <http://dx.doi.org/10.1007/s10588-005-5586-6>
43
44
45
46
47

48 Butts, Carter T. and Pixley, Joy E. (2004). "A Structural Approach to the Representation of Life
49
50 History Data." *Journal of Mathematical Sociology*. Vol. 28, No. 2. pp. 81-124.
51
52 <http://dx.doi.org/10.1080/00222500490448208>
53
54
55
56
57

58 Cohen, Jacob; Cohen, Patricia; West, Stephen G.; and Aiken, Leona S. (2013) *Applied Multiple*
59
60
61
62
63
64
65

1
2
3
4 *Regression/Correlation Analysis for the Behavioral Sciences*. Routledge.
5
6
7

8
9 Cox, Michael A. and Cox, Trevor F. (2000) *Multidimensional Scaling*. Chapman and Hall:
10
11 London, England.
12
13
14

15
16 Dekker, David; Krackhardt, David; and Snijders, Tom. (2003). "Multicollinearity Robust QAP
17
18 for Multiple Regression." *1st Annual Conference of the North American Association for*
19
20 *Computational Social and Organizational Science*. (pp. 22-25).
21
22
23

24
25
26 Diekmann, Andreas (1989). "Diffusion and Survival Models for the Process of Entry into
27
28 Marriage" *The Journal of Mathematical Sociology*. Vol. 14, Issue 1. pp. 31-44.
29

30
31 <http://dx.doi.org/10.1080/0022250X.1989.9990042>
32
33
34

35
36 Drobnic, Sonja; Blossfeld, Hans-Peter; and Rohwer, Gtz. (1999). "Dynamics of Women's
37
38 Employment Patterns over the Family Life Course: A Comparison of the United States and
39
40 Germany" *Journal of Marriage and Family*. Vol. 61, No. 1. pp. 133-146.
41

42
43 <http://dx.doi.org/10.2307/353889>
44
45
46

47
48 DuBois, Christopher; Butts, Carter T.; and Smyth, Padhraic. (2013) "Stochastic Blockmodeling
49
50 of Relational Event Dynamics" *Journal of Machine Learning Research: Workshop and*
51
52 *Conference Proceedings*. Vol. 31. pp. 238-246.
53
54
55

56
57
58 Durbin, James, and Geoffrey S. Watson. (1950) "Testing for serial correlation in least squares
59
60
61
62
63
64
65

1
2
3
4 regression: I." *Biometrika*. pp 409-428.
5
6
7

8
9 Durbin, James, and Geoffrey S. Watson. (1951) "Testing for serial correlation in least squares
10 regression. II." *Biometrika*. pp 159-177.
11
12
13
14

15
16 Elder, Glen H. Jr. (1995). *The Life Course Paradigm: Social Change and Individual*
17 *Development*. American Psychological Association.
18
19
20
21
22

23
24
25
26 Elder, Glen H. Jr. (1998). "The Life Course as Developmental Theory". *Child Development*.
27
28 Vol. 69, No. 1. pp.1-12
29
30 PMid:9499552
31
32
33
34
35

36 Elder, Glen H., Jr. (2000). "The Life Course" In E. F. Borgatta and R. J. V. Montgomery (Eds.),
37 *Encyclopedia of Sociology*. Vol. 3, 2nd edition. pp. 1614-1622. Macmillan Reference: New
38
39 York, NY.
40
41
42
43
44

45
46 Elder, Glen H., Jr. and Johnson, M. K. (2003). "The Life Course and Aging: Challenges,
47
48 Lessons, and New Directions" In R. A. Settersten (Ed.), *Invitation to the Life Course: Toward*
49 *New Understandings of Later Life*. pp. 49-81. Baywood: Amityville, NY.
50
51
52
53
54
55
56
57

58 Freeman, Linton C. (1979). "Centrality in Social Networks: Conceptual Clarification" *Social*
59
60
61
62
63
64
65

1
2
3
4 *Networks*. Vol. 1, No. 3. pp 215-239.
5
6
7

8
9 Freedman, Deborah; Thonton, Arland; Camburn, Donald; Alwin, Duane; Young-DeMarco,
10 Linda. (1988) "The Life History Calendar: A Technique for Collecting Retrospective Data"
11
12 *Sociological Methodology*. Vol. 18, No. 1. pp. 37-68
13
14

15
16 <http://dx.doi.org/10.2307/271044>
17

18
19 PMid:12282712
20
21

22
23 Gauthier, Jacques-Antoine; Widmer, Eric D.; Bucher, Philipp; and Notredame, Cedric. (2010)
24
25 "Multichannel Sequence Analysis" *Sociological Methodology*. Vol. 40, Issue 1. pp. 1-38.
26

27
28 <http://dx.doi.org/10.1111/j.1467-9531.2010.01227.x>
29
30

31
32
33 Germaine, Adrienne. (1975) "Status and Roles of Women as Factors in Fertility Behavior: A
34
35 Policy Analysis" *Studies in Family Planning*. Vol. 6, No. 7. pp. 192-200
36

37
38 <http://dx.doi.org/10.2307/1965271>
39
40

41
42
43 Gerson, Kathleen. (1985) *Hard Choices: How Women Decide About Work, Career, and*
44
45 *Motherhood*. Berkeley and Los Angeles: University of California Press.
46
47

48
49
50
51 Giele, Janet Z. (1998). "Innovation in the Typical Life Course" In J. Z. Giele and G. H. Elder
52
53 (Eds.), *Methods of Life Course Research: Qualitative and Quantitative Approaches*. pp. 231-
54
55 263. Sage Publications: Thousand Oaks, CA.
56
57
58
59
60
61
62
63
64
65

1
2
3
4 Hamming, Richard W. (1950). "Error Detecting and Error Correcting Codes" *Bell System*
5
6 *Technical Journal*. Vol. 29, Issue 2. pp. 147-160.

7
8
9 <http://dx.doi.org/10.1002/j.1538-7305.1950.tb00463.x>
10
11
12
13

14 Han, Shin-Kap and Moen, Phyllis. (1999). "Clocking Out: Temporal Patterning of Retirement"
15
16 *American Journal of Sociology*. Vol. 105, No. 1. pp. 191-236.

17
18
19 <http://dx.doi.org/10.1086/210271>
20
21
22
23

24 Hank, Karsten. (2004). "Effects of Early Life Family Events on Women's Late Life Labour
25
26 Market Behaviour" *European Sociological Review*. Vol. 20, Issue 3. pp. 189-198.

27
28
29 <http://dx.doi.org/10.1093/esr/jch017>
30
31
32
33

34 Hauser, Robert M. and Sewell, William H. (2010). Wisconsin Longitudinal Study (WLS):
35
36 1957-2005 Version 12.24. Madison, WI: University of Wisconsin-Madison, WLS.

37
38
39 <http://www.ssc.wisc.edu/wlsresearch/documentation/>
40
41
42
43
44
45

46 Heinz, Walter. R. (2003). "From Work Trajectories to Negotiated Careers: The Contingent
47
48 Work Life Course" In J. T. Mortimer M. J. Shanahan (Eds.), *Handbook of the Life Course* (pp.
49
50 185-204). Kluwer: New York, NY.
51
52
53
54
55
56
57

58 Hendricks, Jon. (2012) "Considering Life Course Concepts" *Journals of Gerontology, Series B:*
59
60
61
62
63
64
65

1
2
3
4 *Psychological and Social Sciences*. Vol. 67B, No. 2. pp. 226-231.

5
6
7 <http://dx.doi.org/10.1093/geronb/gbr147>

8
9 PMid:22391748

10
11
12
13
14 Hirschman, Charles; Lai, Tuong; San, Pham Bich; Loi, Vu Manh; Nguyen, Huu Minh. (1991).

15
16 Vietnam Life History Survey

17
18
19
20
21
22
23
24 Jacobs, Jerry A. and King, Rosalind B. (2002) "Age and College Completion: A Life History
25
26 Analysis of Women Aged 15-44" *Sociology of Education*. Vol. 75, No. 3. pp. 211-230.

27
28 <http://dx.doi.org/10.2307/3090266>

29
30
31
32
33
34 Johnson, Stephen (1967). "Hierarchical Clustering Schemes" *Psychometrika*. Vol. 32, No. 3. pp.
35
36 241-254.

37
38 <http://dx.doi.org/10.1007/BF02289588>

39
40
41 PMid:5234703

42
43
44
45
46 Krackhardt, David. (1988). "Predicting with Networks: Nonparametric Multiple Regression
47
48 Analysis of Dyadic Data." *Social Networks*. Vol. 10, No. 4. pp. 359-381.

49
50
51
52
53
54 Kuh, Diana; Ben-Shlomo, Yoav; Lynch, John; Hallqvist, Johan; and Power, Chris. (2003) "Life
55
56 Course Epidemiology" *Journal of Epidemiology and Community Health*. Vol. 57. pp. 778-783.

57
58 <http://dx.doi.org/10.1136/jech.57.10.778>

1
2
3
4 PMid:14573579 PMCID:PMC1732305
5
6
7

8
9 Kutner, Michael H.; Nachtsheim, Christopher J.; Neter, John; and Li, William (2005). *Applied*
10 *Linear Statistical Models*. McGraw-Hill Irwin: New York, NY.
11
12
13
14

15
16 Kruskal, Joseph B. and Wish, Myron. (1978) Multidimensional Scaling. *Sage University Paper*
17 *series on Quantitative Applications in the Social Sciences*. Number 07-011. Sage Publications,
18
19 Newbury Park, CA.
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34

35 Laub, John H. and Sampson, Robert J. (1998). "Integrating Quantitative and Qualitative Data" In
36
37 J. Z. Giele and G. H. Elder (Eds.), *Methods of Life Course Research: Qualitative and*
38 *Quantitative Approaches*. pp. 213-230. Sage Publications: Thousand Oaks, CA.
39
40
41
42
43
44

45 Lynch, John and Smith, George D. (2003) "A Life Course Approach to Chronic Disease
46
47 Epidemiology" *Annual Review of Public Health*. Vol. 26. pp. 1-35.
48
49 <http://dx.doi.org/10.1146/annurev.publhealth.26.021304.144505>
50
51

52 PMid:15760279
53
54
55
56

57 MacMillan, Ross and Eliason, Scott R. (2003). "Characterizing the Life Course as Role
58
59
60
61
62
63
64
65

Configurations and Pathways" *Handbook of the Life Course*. edited by Jeylan T. Mortimer and Michael J. Shanahan. Kluwer Academic/Plenum Publishers, New York: New York.

http://dx.doi.org/10.1007/978-0-306-48247-2_24

Marcum, Christopher S. and Butts, Carter T. (2015). "Constructing and Modifying Sequence Statistics for relevant Using informR in R" *Journal of Statistical Software*. Vol 64, Issue 5.

<http://www.jstatsoft.org/v64/i05>

Mayer, Karl U. (2004). "Whose Lives? How History, Societies, and Institutions Define and Shape Life Courses" *Research in Human Development*. Vol. 1, Issue 3. pp. 161-187

http://dx.doi.org/10.1207/s15427617rhd0103_3

Mayer, Karl U. (2009). "New Directions in Life Course Research" *American Sociological Review*. Vol. 35. pp. 413-433.

<http://dx.doi.org/10.1146/annurev.soc.34.040507.134619>

McRae, Susan. (1993) "Returning to Work after Childbirth: Opportunities and Inequalities" *European Sociological Review*. Vol. 9, No. 2. pp. 125-138.

McRae Susan; Joshi, Heather; and Dex, Shirley. (1996) "Employment after Childbearing: A Survival Analysis" *Work, Employment, and Society*. Vol. 10, No. 2. pp. 273-296

1
2
3
4
5
6
7 Meadows, Sarah O., McLanahan, Sara S., and Brooks-Gunn, Jeanne. (2008). "Family Structure
8 and Maternal Health Trajectories" *American Sociological Review*. Vol. 73. pp. 314-334.

9
10
11 <http://dx.doi.org/10.1177/000312240807300207>

12
13
14 PMid:20333277 PMCID:PMC2843941

15
16
17
18 Moen, Phyllis. (1992) *Women's Two Roles: A Contemporary Dilemma*. New York: Auburn
19 House.
20
21

22
23
24
25
26 Moore, Kristin A. (1978). "The Impact of an Early First Birth on Young Women's Educational
27 Attainment" *Social Forces*. Vol. 56, Issue 3. pp. 845-864.
28
29

30
31
32
33
34 Nazio, Tiziana and Blossfeld, Hans-Peter. (2003). "The Diffusion of Cohabitation among Young
35 Women in West Germany, East Germany, and Italy" *European Journal of Population*. Vol. 19.
36
37 pp. 47-82.
38
39

40
41 <http://dx.doi.org/10.1023/A:1022192608963>
42
43
44
45
46
47
48
49

50 Pearlín, Leonard I.; Menaghan, Elizabeth G.; Lieberman, Morton A.; and Mullan, Joseph T.
51
52 (1981). "The Stress Process" *Journal of Health and Social Behavior*. Vol. 22. pp. 237-256.
53
54

55 <http://dx.doi.org/10.2307/2136676>
56
57
58
59
60
61
62
63
64
65

Pienta, Amy. (1999). "Early Childbearing Patterns and Women's Labor Force Behavior in Later Life" *Journal of Women and Aging*. Vol. 11, Issue 1. pp. 69-84.

http://dx.doi.org/10.1300/J074v11n01_06

PMid:10323047

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*. Vol. 170, Issue 1. pp. 167-183.

<http://dx.doi.org/10.1111/j.1467-985X.2006.00450.x>

Quick, Heather E. and Moen, Phyllis. (1998) "Gender, Employment, and Retirement Quality: A Life Course Approach to the Differential Experiences of Men and Women" *Journal of Occupational Health Psychology*. Vol. 3, No. 1. pp. 44-64

<http://dx.doi.org/10.1037/1076-8998.3.1.44>

PMid:9552271

R Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Schneider, Barbara and Stevenson, David. (1999) *The Ambitious Generation: America's Teenagers, Motivated but Directionless*. R. R. Donnelley and Sons: Chicago, Illinois.

1
2
3
4 Settersten, Richard A., Jr. (2003). *Invitation to the Life Course: Toward New Understandings of*
5
6 *Later Life*. Baywood: Amityville, NY.
7
8
9

10
11 Settersten, Richard A. and Mayer, Karl Ulrich. (1997) "The Measurement of Age, Age
12
13 Structuring, and the Life Course" *Annual Review of Sociology*. Vol. 23. pp. 233-261.
14
15
16 <http://dx.doi.org/10.1146/annurev.soc.23.1.233>
17
18
19

20
21 Singer, Burton; Ryff, Carol. D.; Carr, Deborah; and Magee, William. J. (1998) "Linking Life
22
23 Histories and Mental Health: A Person-Centered Strategy" *Sociological methodology*, Vol. 28,
24
25 No. 1. pp. 1-51.
26
27
28
29

30
31 Spain, Daphne and Bianchi, Suzanne. (1996) *Balancing Act: Motherhood, Marriage, and*
32
33 *Employment Among American Women*. Russell Sage Foundation.
34
35
36
37

38
39 Treiman, Donald and Walder, Andrew. (1996) *Life Histories and Social Change in*
40
41 *Contemporary China*.
42
43
44
45

46
47
48 Voydanoff, Patricia. (1988) "Work Role Characteristics, Family Structure Demands, and
49
50 Work/Family Conflict." *Journal of Marriage and the Family*. pp. 749-761.
51
52
53
54

55 Wasserman, Stanley and Faust, Katherine. (1994) *Social Network Analysis: Methods and*
56
57 *Applications*. Cambridge University Press: Cambridge, UK
58
59
60
61
62
63
64
65

<http://dx.doi.org/10.1017/CBO9780511815478>

Wang, George C.S. and Jain, Chaman L. (2003) *Regression Analysis: Modeling and Forecasting*. Institute of Business Forecasting.