

# Sequence Analysis: Where Are We, Where Are We Going?



Gilbert Ritschard and Matthias Studer

## 1 Sequence Analysis: Optimal Matching and Much More

Sequence Analysis (SA) has gained increasing importance in the field of social sciences since the pioneering contributions of Andrew Abbott (e.g., Abbott 1983; Abbott and Forrest 1986) and has become a popular tool with the release of powerful dedicated software (Brzinsky-Fay et al. 2006; Gabadinho et al. 2011; Halpin 2014). The increasing availability of longitudinal data sources such as panel and retrospective surveys also contributed to the rise of interest in SA. In recent times, SA has garnered a central role in life course studies to appraise, for example, occupational careers, cohabitation pathways, or health trajectories. In addition, it is effectively used in domains such as time use, spatio-temporal development of economic activities, and historical evolution of political institutions. In fact, sequences are a convenient way of coding individual narratives into a form suitable for quantitative analysis.

Briefly, SA primarily provides a comprehensible overall picture of sets of individual categorical sequences—the retained coding of the narratives—and involves using this overall picture for objectives such as discovering the characteristics of a set of sequences, identifying possible atypical or deviant individual trajectories, and comparing trajectory patterns among groups such as sexes, birth cohorts, or regions.

Abbott and Tsay (2000) describe the typical SA as a three-step program: (1) code the narratives as sequences, (2) compute pairwise dissimilarities between sequences, and (3) analyse the sequences based on their dissimilarities. The coding stage involves the selection of a suitable alphabet for the states or events to be considered, and the definition of the timing scheme to be used to time align the sequences.

---

G. Ritschard (✉) · M. Studer

NCCR LIVES and Geneva School of Social Sciences, University of Geneva, Geneva, Switzerland  
e-mail: [gilbert.ritschard@unige.ch](mailto:gilbert.ritschard@unige.ch); [matthias.studer@unige.ch](mailto:matthias.studer@unige.ch)

© The Author(s) 2018

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

Life Course Research and Social Policies 10,

[https://doi.org/10.1007/978-3-319-95420-2\\_1](https://doi.org/10.1007/978-3-319-95420-2_1)

The computation of the dissimilarities implies the choice of a suitable dissimilarity measure. The analysis itself typically involves building a typology of the trajectories by applying a clustering algorithm that takes the dissimilarities as input, even though Abbott and Tsay (2000) also consider multidimensional scaling.

The optimal matching (OM) distance borrowed from signal processing (Levenshtein 1966; Hamming 1950) and biology (e.g., Needleman and Wunsch 1970; Sankoff and Kruskal 1983) and popularized in the social science by Abbott and Forrest (1986) was typically used to compare the sequences. OM was so intimately connected with SA that the expression ‘optimal matching analysis’ was—and still is—often used as a synonym for SA, and this even when no or non-optimal matching distances are used.

In fact, during the second wave of SA development (see Aisenbrey and Fasang 2010) most methodological developments focused on the measurement of the dissimilarities between sequences and more specifically on OM. Several new variants for defining the OM-costs, but also new non-OM-based distance measures were proposed to address criticisms raised in the literature. This scattered development was recently summarized by Studer and Ritschard (2016) in their comparative review of dissimilarity measures.

More recently, the use of SA has seen developments in several other directions. The visualisation of sequences started veritably with the release of software dedicated to SA (Brzinsky-Fay et al. 2006; Gabadinho et al. 2011). The multiple possibilities to render a set of sequences with easily interpretable colourful plots undoubtedly boosted interest in SA. The most popular of these plots are index plots (Scherer 2001) that render the individual sequences and their diversity, and chronograms, which display the evolution of the cross-sectional distribution at successive time points. The cluttered aspect of the former and the over simplification of the latter—which completely overshadows the diversity of the sequences—naturally called for lightened forms of index plots. The search and plot of representative sequences by Gabadinho and Ritschard (2013) and the relative frequency sequence plot of Fasang and Liao (2014) offer solutions to this challenge. The decorated parallel coordinate plot (Bürgin and Ritschard 2014) focuses on sequencing within trajectories and is useful for identifying the most typical sequencing patterns while rendering the diversity of the entire set of observed trajectories at the same time.

Multichannel sequences, i.e., the joint analysis of narratives of different domains such as linked lives, or occupational and cohabitation trajectories, also received increasing attention. Here, one difficulty is the exploding size of the alphabet that results from the combination of different dimensions. For the specific case of OM, some authors (e.g., Pollock 2007) proposed tricks for defining the substitution costs between different combinations of states from costs set for each individual dimension. This dramatically reduces the number of parameters that need to be specified. The measurement of the strength of association between channels is probably more interesting and promising to study the relationship or interaction between dimensions such as familial and professional pathways. The contributions of Piccarreta and Elzinga (2013) and Piccarreta (2017) are path-breaking in that respect. The graphical rendering of multichannel sequences also requires special attention and effective solutions are, for example, provided by Helske and Helske (2017) and their `seqHMM` package.

There has been some work on exploiting the pairwise dissimilarities between sequences differently from clustering and multidimensional scaling. Studer et al. (2011) showed how to conduct an ANOVA-like analysis and to grow regression trees for sequence data. Gabadinho and Ritschard (2013) used the dissimilarities to find representative sequences such as the most central sequences or sequences with the densest neighbourhood.

With regards to clustering, there are attempts to get rid of explicit dissimilarity measures by resorting to latent class approaches (Vermunt et al. 2008; Barban and Billari 2012), or more or less similarly hidden Markov models (e.g., Helske and Helske 2017; Bolano et al. 2016), to clustering the sequences. Markov-based approaches may also contribute to understanding the dynamics that drive the unfolding of the sequences. However, the difficulty to synthesize the outcome of Markov-based transition models—especially when more realistic models with order greater than one are considered—negatively affected the extension of their usage. The graphical rendering of hidden Markov models (HMM) by Helske and Helske (2017) (see also Helske et al. 2018, in this bundle) as well as the use and rendering of probabilistic suffix trees (Gabadinho and Ritschard 2016) for sequence analysis should facilitate the access to such probabilistic approaches and shed light on how the current situation is linked to the history of previous situations.

More or less independently of classical SA, graph and network approaches (e.g., Butts and Pixley 2004; Bison 2014; Cornwell and Watkins 2015) have proven to provide useful summaries of individual sequences as well as synthetic views at the population level. See also Cornwell (2018) and Hamberger (2018) in this bundle.

Alongside the discovery of characteristics of the set of sequences, such as the diversity among sequences and typologies of trajectories, there are works concerned with summary numbers of individual sequences. More specifically, here, the aim was to complement simple indexes such as the sequence length, the number of different states visited, and the number of transitions, with measures of internal diversity and complexity of an individual sequence. Contributions in that direction have been made, for example, by Brzinsky-Fay (2007), Elzinga and Liefbroer (2007), Elzinga (2010), and Gabadinho et al. (2010).

## 2 Towards Stronger Interaction with Related Approaches

The short survey included above covering what has been done in SA is certainly incomplete. In particular, we did not mention two important issues that already received some attention in the literature: the handling of censored sequences and more generally of missing values, and the possibility to study the relationship of sequences with time-varying covariates. These issues have no definitive solution thus far and deserve further research. While different schemes for imputing missing values have been proposed (e.g., Halpin 2015; Gabadinho and Ritschard 2016), there remains the question of the maximal proportion of missing values that are appropriate to impute in a sequence. Moreover, the real impact of such imputations

on the SA outcome remains to be investigated. Regarding time-varying covariates, Studer et al. (2018a,b) proposed procedures combining SA with event history analysis in order to study the influence of such time-dependent covariates on a trajectory in a semi-holistic perspective. A possible completely holistic solution lies in the multichannel approach and the joint analysis (Piccarreta 2017) of the dependent channel with those defined by the history of values of each time-varying covariate. Here again, further investigation seems necessary. Therefore, there is still room for further development in classical SA. However, we think that the future of SA is intimately linked with the development of its interaction with other—more inferential and/or probabilistic—methods for longitudinal data.

Despite few attempts to introduce inferential methods in SA with ANOVA-like analysis and Markov-based modelling of sequences, SA essentially remains exploratory and needs to be complemented with other modelling tools, especially when it comes to testing hypotheses or studying the dynamics that drive trajectories. The powerfulness of SA as an exploratory tool has been largely demonstrated by many substantive studies that use SA tools; moreover, most of these studies run SA in conjunction with other approaches, typically involving the use of the obtained typology of sequences either as an explanatory variable or as the response variable in a regression analysis. In these studies, the SA outcome serves as input for the regression stage, but we could imagine using regression outcome, e.g., the most contrasting profiles with respect to the regression response variable, to guide the SA analysis.

One advantage often discussed for SA is its holistic perspective (see e.g., Billari 2005) meaning that SA sheds light on the entire trajectory rather than, for example, on specific transitions in the trajectory. With this holistic perspective, sequences are considered as static objects, which is not suited for studying the process that generates the sequences. For investigating sequence dynamics, we require alternative methods such as probabilistic models of the occurrences of successive transitions in the sequence. As already mentioned above, an issue with such transition models is the difficulty to present their outcome synthetically. Here, SA could help rendering the outcome of the modelling phase.

As illustrated, by the abovementioned two examples, an intimate combination of SA and related methods seems necessary to achieve a better understanding of life course data. We describe below how the chapters of this book lead in that direction.

### **3 Directions for the Future: The Chapters of this Book**

In Part I, two chapters address the relationship between SA and other methods for analysing longitudinal data. In the first chapter, Daniel Courgeau (2018) examines four major approaches for longitudinal analysis in population science, namely, approaches based on sequences, duration between events, multiple levels, and networks. He first identifies the proper characteristics of each approach in terms of the mathematical tools involved and the considered statistical unit (e.g., event,

individual, group), among others. He then depicts a general robust program that could lead to the convergence of these different models. The next chapter by Mervi Eerola (2018) discusses three original ways of combining SA—in fact, the clustering of sequences—and probabilistic modelling. This is done through a short presentation of three case studies of life course analysis. With case study 1, Mervi Eerola considers the combined use of SA and prediction probabilities obtained with a marked point process—a kind of multistate model. In case study 2, SA is used to identify pathways to adulthood and is used in conjunction with a structural equation factor model of social and achievement strategies, and a model for transitional pathways accounting for the strategies. In case study 3, SA is used to identify the most vulnerable individuals among Finns between 18 and 25 years of age and Mervi Eerola addresses the use of either a latent transition model or a HMM for analysing their risk pattern, e.g., risk to be outside work force, to be living on social benefits, and to have the lowest educational attainment. These three case studies are good illustrations of the many possibilities of combining SA with modelling approaches.

Part II is devoted to the combination of SA and event history analysis or equivalently, survival analysis. The strength of the connection between SA and the survival models increases in each of the successive chapters. In the first chapter, Malin and Wise (2018) propose a study of gender differences in career advancement across occupations in West Germany. In this study, sequence visualization is used to provide an overall view of the data at hand, and the focus is then placed on the study of the time to a leadership position and time to leaving a leadership position by means of Kaplan-Meier (KM) survival curves and Cox regressions. The connection between sequence and survival analysis remains loose and no explicit SA outcome is used in the survival analysis. The second chapter by Lundevaller et al. (2018) studies the mortality of disabled and non-disabled individuals in nineteenth century Sweden. A classical clustering of life trajectories is realized—separately for women and men—and the obtained types are used as covariates in the survival analyses carried out with stratified KM and Cox regressions. Finally, the chapter by Rossignon et al. (2018) introduces an innovative method—called Sequence History Analysis (SHA)—where SA and event history analysis are more tightly entwined. The method consists of an event history analysis that accounts for the past trajectory at each time point. More specifically, SA is used to determine the type of past trajectory at each time point, which makes the past trajectory type a time-varying covariate. This method is applied to study how the risk of leaving home depends on the past co-residence trajectories in Switzerland.

Part III includes two papers concerned with the network approach in SA. In the first chapter, Benjamin Cornwell (2018) starts by explaining, with some detail, how a sequence can be represented as a network with states as nodes and time-adjacency between states as links. He then shows how a series of network concepts—such as network density, centralization, and homophily—prove useful for characterizing the structure of individual sequences and to compare multiple sequence structures with each other. The approach is illustrated with an analysis of daily activities using data from the American Time Use Survey. The next chapter by Klaus Hamberger (2018) introduces relational networks and demonstrates their scope in a study of mobility

patterns in Togo. Relational networks are built from networks of kinship and mobility relations. The nodes of the relational network are the classes of (mobility) events obtained by classifying the events according to the type of relation—e.g., kinship, employer, friend—between the individuals involved. The arcs indicate the immediate succession of the events. Then, the author proposes two complementary ways of using the personal networks. First, he aggregates the individual networks into one network for women and one for men with node sizes and arc widths proportionate to the observed counts, which allows the visual identification of the gender differences in the mobility itineraries. Second, he orders the individual networks along a spanning tree. Here, again, we see that women and men occupy different areas of the tree and thus reveal gender differences.

Part IV is composed of three chapters that attempt to gain knowledge about the process behind the observed trajectories. The chapter by Thomas Collas (2018) suggests that life trajectories decompose into phases and follow a different logic—possibly characterized with a different alphabet—in each phase. He shows how such multiphase trajectories can be formalized and rendered, and proposes a dissimilarity measure that can account for this decomposition into phases. The method is illustrated with an application to competition trajectories of French pastry cooks with an explicit distinction between the junior and senior phases. The next chapter by Borgna and Struffolino (2018) combines SA with qualitative comparative analysis (QCA), which is a method related to the mining of association rules, to find out factor configurations that are ‘logically sufficient’ to be in employment or education at crucial time points in divided Germany. Discrepancy analysis, more specifically the analysis of the evolution of the discrepancy among sequences along the time frame, is used to identify the crucial turning points in the divergence between sequences. QCA is then applied at the identified turning points to find out the relevant ‘sufficient’ factor combinations. The third chapter by Helske et al. (2018) also considers that trajectories belonging to a same group share similar phases. Here, however, the phases are not predefined, but are associated to the hidden states of a HMM and thus probabilistically determined. In fact, this chapter presents a general framework for the combined use of SA and HMM to analyse multichannel sequences. SA is used to cluster the sequences and then the HMM is used to identify similar phases within each group. In addition, the authors propose two original compressed representations of a group of (multichannel) sequences, including a graph of the structure of the hidden states and the transitions between them, and plots of the most probable individual pathways predicted by the HMMs for each group. The method is illustrated with data from the German National Educational Panel Survey (NEPS).

Part V includes three chapters that present advances in the original task of SA, namely the clustering of sequences. The chapter by Taushanov and Berchtold (2018) proposes clustering sequences of continuous data by means of a Markov-based mixture model—the hidden mixture transition distribution (HMTD) model—and applying their method to Swiss data obtained from the internet addiction test (IAT). The clustering is achieved by setting the transition matrix of the hidden states as the identity matrix, which makes their model a mixture of Gaussian distributions. One

advantage of the proposed approach is the possibility of accounting for covariates at the clustering level. The authors compare the results provided by their method with those obtained by means of a growth mixture model (GMM). Note that this chapter is the only one in that volume that deals with continuous data. The next chapter by Matthias Studer (2018) investigates two original dissimilarity-based ways of clustering sequences: a divisive property-based method and fuzzy clustering. The former orders the splits of a discrepancy-based regression tree that provides classification rules defined in terms of covariates, and considers the partition that results from the splits up to a given, optimally chosen rank. Splits are ordered according to the overall share of reduction of discrepancy that each of them produces. In fuzzy clustering, each sequence may belong to more than one cluster and cluster membership may have different degrees. The method is especially useful when clusters are not well separated. Here, the author addresses a series of issues such as the graphical representation of fuzzy clusters and how to measure the effect of covariates on the individual strengths of membership. The methods are illustrated with the school-to-work transition data from McVicar and Anyadike-Danes (2002). The last chapter by Bison and Scalcon (2018) focuses on the measure of the dissimilarity between sequences. The authors decompose each sequence into basic binary sequences that indicate for each element of the alphabet whether it is active at successive time points. Then, they associate to each binary sequence two index numbers, the first one indicating the proportion of time spent in the concerned state and the second one synthesizing when the concerned state is actually observed. The dissimilarity between a pair of binary sequences is obtained as the Euclidean distance between the couples of index numbers, and the dissimilarity between the original—possibly multichannel—sequences as the sum of the distances between the underlying binary sequences. The method is applied to cluster data describing the time-use during a typical day of Italian dual earners.

Finally, the book concludes with Part VI with two papers concerned with summary numbers of individual sequences. Both papers aim to measure the quality of sequences, i.e., to get an index that would allow the ranking of, for example, occupational sequences from the most negative—insecure, undesirable—to the most positive ones. The solutions proposed are, however, very different. In the first chapter, Manzoni and Mooi-Reci (2018) assume that each state of the alphabet can be classified as a success or a failure. Then, the proposed quality index is defined so as to increase with the proportion and recency of the successes. The index is applied to a study of the quality of employment career after a first spell of unemployment using data from HILDA, an Australian household survey. In the second chapter by Ritschard et al. (2018), the quality index—named precarity index—is defined based on the quality of the transitions rather than the states themselves. Assuming that the states of the alphabet can be (partially) ordered, the authors define the index as a complexity index corrected by a factor that depends on the proportion of upward and downward transitions. The scope of the index is illustrated with the school-to-work transition data from McVicar and Anyadike-Danes (2002) by showing the strong impact of the quality of the initial trajectory on the situation three years later.

## 4 Conclusion

To conclude, let us highlight how this volume traces the expected trend for the future of SA. First, we observe a shift in the methodological concern. The measure of dissimilarities between sequences that was the major central aspect of SA until recently, is the concern of only two papers (Collas 2018; Bison and Scalcon 2018) out of the fifteen.

The trend seems to be more oriented toward alternative approaches for SA and the combined use of dissimilarity-based SA with other related methods. Five papers address alternatives to the classical ‘compute dissimilarities—partition the set of sequences’ approach. These alternatives include feature-based and fuzzy clustering (Studer 2018) and non-dissimilarity-based methods such as those based on network representations of sequences (Cornwell 2018; Hamberger 2018), and Markov-based models (Helske et al. 2018; Taushanov and Berchtold 2018). Alongside the two general papers (Courgeau 2018; Eerola 2018) in Part I, five papers demonstrate the benefit of combining SA with other methods to grasp the dynamics that drive the trajectories. SA is combined with survival models (Malin and Wise 2018; Lundevaller et al. 2018; Rossignon et al. 2018), with QCA (Borgna and Struffolino 2018), and with hidden Markov models (Helske et al. 2018).

Finally, there seems to be an increasing interest in individual sequence summaries. Such summary numbers are central in Part VI (Manzoni and Mooi-Reci 2018; Ritschard et al. 2018), but also play an important role in two other chapters (Cornwell 2018; Bison and Scalcon 2018).

**Acknowledgements** The authors acknowledge the support of the Swiss National Centre of Competence in Research LIVES - Overcoming vulnerability: Life course perspectives, which is financed by the Swiss National Science Foundation (grant number: 51NF40-160590).

## References

- Abbott, A. (1983). Sequences of social events: Concepts and methods for the analysis of order in social processes. *Historical Methods*, 16(4), 129–147.
- Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History*, 16, 471–494.
- Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology, Review and prospect. *Sociological Methods and Research*, 29(1), 3–33. (With discussion, pp 34–76).
- Aisenbrey, S., & Fasang, A. E. (2010). New life for old ideas: The “second wave” of sequence analysis bringing the “course” back into the life course. *Sociological Methods and Research*, 38(3), 430–462.
- Barban, N., & Billari, F. C. (2012). Classifying life course trajectories: A comparison of latent class and sequence analysis. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 61(5), 765–784.

- Billari, F. C. (2005). Life course analysis: Two (complementary) cultures? Some reflections with examples from the analysis of transition to adulthood. In R. Levy, P. Ghisletta, J.-M. Le Goff, D. Spini, & E. Widmer (Eds.), *Towards an interdisciplinary perspective on the life course* (Advances in life course research, Vol. 10, pp. 267–288). Amsterdam: Elsevier.
- Bison, I. (2014). Sequence as network: An attempt to apply network analysis to sequence analysis. In P. Blanchard, F. Bühlmann, & J.-A. Gauthier (Eds.), *Advances in sequence analysis: Theory, method, applications* (pp. 231–248). Cham: Springer.
- Bison, I., & Scalcon, A. (2018). From 07.00 to 22.00: A dual-earner couple's typical day in Italy. Old questions and new evidence from social sequence analysis. In Ritschard and Studer (2018) (this volume).
- Bolano, D., Berchtold, A., & Ritschard, G. (2016). A discussion on hidden Markov models for life course data. In *Proceedings of the International Conference on Sequence Analysis and Related Methods*, Lausanne, 8–10 June 2016.
- Borgna, C., & Struffolino, E. (2018). Unpacking configurational dynamics: Sequence analysis and qualitative comparative analysis as a mixed-method design. In Ritschard and Studer (2018) (this volume).
- Brzinsky-Fay, C. (2007). Lost in transition? Labour market entry sequences of school leavers in Europe. *European Sociological Review*, 23(4), 409–422.
- Brzinsky-Fay, C., Kohler, U., & Luniak, M. (2006). Sequence analysis with Stata. *The Stata Journal*, 6(4), 435–460.
- Bürgin, R., & Ritschard, G. (2014). A decorated parallel coordinate plot for categorical longitudinal data. *The American Statistician*, 68(2), 98–103.
- Butts, C. T., & Pixley, J. E. (2004). A structural approach to the representation of life history data. *The Journal of Mathematical Sociology*, 28(2), 81–124.
- Collas, T. (2018). Multiphase sequence analysis. In Ritschard and Studer (2018) (this volume).
- Cornwell, B. (2018). Network analysis of sequence structures. In Ritschard and Studer (2018) (this volume).
- Cornwell, B., & Watkins, K. (2015). Sequence-network analysis: A new framework for studying action in groups. In S. R. Thye & E. J. Lawler (Eds.), *Advances in group processes* (Vol. 32, pp. 31–63). Bingley: Emerald Group Publishing Limited.
- Courgeau, D. (2018). Do different approaches in population science lead to divergent or convergent models? In Ritschard and Studer (2018) (this volume).
- Eerola, M. (2018). Case studies of combining sequence analysis and modelling. In Ritschard and Studer (2018) (this volume).
- Elzinga, C. H. (2010). Complexity of categorical time series. *Sociological Methods & Research*, 38(3), 463–481.
- Elzinga, C. H., & Liefbroer, A. C. (2007). De-standardization of family-life trajectories of young adults: A cross-national comparison using sequence analysis. *European Journal of Population*, 23, 225–250.
- Fasang, A. E., & Liao, T. F. (2014). Visualizing sequences in the social sciences: Relative frequency sequence plots. *Sociological Methods & Research*, 43(4), 643–676.
- Gabadinho, A., & Ritschard, G. (2013). Searching for typical life trajectories applied to childbirth histories. In R. Levy & E. Widmer (Eds.), *Gendered life courses – Between individualization and standardization. A European approach applied to Switzerland* (pp. 287–312). Vienna: LIT-Verlag.
- Gabadinho, A., & Ritschard, G. (2016). Analysing state sequences with probabilistic suffix trees: The PST R library. *Journal of Statistical Software*, 72(3), 1–39.
- Gabadinho, A., Ritschard, G., Studer, M., & Müller, N. S. (2010). Indice de complexité pour le tri et la comparaison de séquences catégorielles. *Revue des nouvelles technologies de l'information RNTI, E-19*, 61–66.
- Gabadinho, A., Ritschard, G., Müller, N. S., & Studer, M. (2011). Analyzing and visualizing state sequences in R with TraMineR. *Journal of Statistical Software*, 40(4), 1–37.
- Halpin, B. (2014). SADI: Sequence analysis tools for Stata. Department of Sociology Working Paper Series WP2014-03, University of Limerick.

- Halpin, B. (2015). MICT: Multiple imputation for categorical time-series. Department of Sociology Working Paper Series WP2015-02, University of Limerick, Ireland.
- Hamberger, K. (2018). Relational sequence networks as a tool for studying gendered mobility patterns. In Ritschard and Studer (2018) (this volume).
- Hamming, R. W. (1950). Error detecting and error correcting codes. *Bell System Technical Journal*, 26(2), 147–160.
- Helske, S., & Helske, J. (2017). Mixture hidden Markov models for sequence data: The seqHMM package in R. Vignette of the seqHMM package, CRAN.
- Helske, S., Helske, J., & Eerola, M. (2018). Combining sequence analysis and hidden Markov models in the analysis of complex life sequence data. In Ritschard and Studer (2018) (this volume).
- Levenshtein, V. (1966). Binary codes capable of correcting deletions, insertions, and reversals. *Soviet Physics Doklady*, 10, 707–710.
- Lundevaller, E. H., Vikström, L., & Haage, H. (2018). Modelling mortality using life trajectories of disabled and non-disabled individuals in 19th-century Sweden. In Ritschard and Studer (2018) (this volume).
- Malin, L., & Wise, R. (2018). Glass ceilings, glass escalators and revolving doors: Comparing gendered occupational trajectories and the upward mobility of men and women in West Germany. In Ritschard and Studer (2018) (this volume).
- Manzoni, A., & Mooi-Reci, I. (2018). Measuring sequence quality. In Ritschard and Studer (2018) (this volume).
- McVicar, D., & Anyadike-Danes, M. (2002). Predicting successful and unsuccessful transitions from school to work using sequence methods. *Journal of the Royal Statistical Society A*, 165(2), 317–334.
- Needleman, S., & Wunsch, C. (1970). A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology*, 48, 443–453.
- Piccarreta, R. (2017). Joint sequence analysis: Association and clustering. *Sociological Methods & Research*, 46(2), 252–287.
- Piccarreta, R., & Elzinga, C. H. (2013). Mining for association between life course domains. In J. J. McArdle & G. Ritschard (Eds.), *Contemporary issues in exploratory data mining in the behavioral sciences* (Quantitative methodology, pp. 190–220). New York: Routledge.
- Pollock, G. (2007). Holistic trajectories: A study of combined employment, housing and family careers by using multiple-sequence analysis. *Journal of the Royal Statistical Society A*, 170(1), 167–183.
- Ritschard, G., & Studer, M. (Eds.) (2018). *Sequence analysis and related approaches: Innovative methods and applications*. (Life course research and social policies). Cham: Springer.
- Ritschard, G., Bussi, M., & O'Reilly, J. (2018). An index of precarity for measuring early employment insecurity. In Ritschard and Studer (2018) (this volume).
- Rossignon, F., Studer, M., Gauthier, J.-A., & Le Goff, J.-M. (2018). Sequence history analysis (SHA): Estimating the effect of past trajectories on an upcoming event. In Ritschard and Studer (2018) (this volume).
- Sankoff, D., & Kruskal, J. B. (Eds.) (1983). *Time warps, string edits, and macro-molecules: The theory and practice of sequence comparison*. Reading: Addison-Wesley.
- Scherer, S. (2001). Early career patterns: A comparison of Great Britain and West Germany. *European Sociological Review*, 17(2), 119–144.
- Studer, M. (2018). Divisive property-based and fuzzy clustering for sequence analysis. In Ritschard and Studer (2018) (this volume).
- Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society, Series A*, 179(2), 481–511.
- Studer, M., Ritschard, G., Gabadinho, A., & Müller, N. S. (2011). Discrepancy analysis of state sequences. *Sociological Methods and Research*, 40(3), 471–510.

- Studer, M., Liefbroer, A. C., & Mooyaart, J. E. (2018a). Understanding trends in family formation trajectories: An application of competing trajectories analysis (CTA). *Advances in Life Course Research*, 36, 1–12.
- Studer, M., Struffolino, E., & Fasang, A. E. (2018b). Estimating the relationship between time-varying covariates and trajectories: The sequence analysis multistate model procedure. *Sociological Methodology*. (First Published Online).
- Taushanov, Z., & Berchtold, A. (2018). Markovian-based clustering of internet addiction trajectories. In Ritschard and Studer (2018) (this volume).
- Vermunt, J., Tran, B., & Magidson, J. (2008). Latent class models in longitudinal research. In S. Menard (Ed.), *Handbook of longitudinal research: Design, measurement, and analysis* (pp. 373–385). Burlington, MA: Elsevier.

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

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

