Sequence analyses that reveal country-specific typologies of start-up processes and their institutional foundations

Andrea M. Herrmann
Lukas Held
Tomás Santos Mendes

Document Identifier

Annex 4 to D5.1 Internationally comparative dataset on start-up processes and their institutional foundations in Germany, Italy, the UK and the US

Version

1.0

Date Due

M36

Submission date

xx

WorkPackage

5

Lead Beneficiary

UU

Grant Agreement Number 649378

Change log

Version[1]	Date	Amended by	Changes
1.0			-
1.1			

[1]Please start with version 0.1. All minor changes will lead to a new number (0.2, 0.3, 0.4 etc.). The first complete draft will get the number 1.0. Again all minor revisions will lead to a new decimal number (1.1, 1.2, 1.3 etc.). A major revision will become 2.0 etc. etc. Until there is a final version which will be called ‘final’.

1. Executive summary

Sequence analyses can be conducted in different forms. Optimal matching approaches are one of the most widely used forms of sequence analyses. Ever since optimal matching (OM) analyses have been used to decode the human genome, they have become an established method in scientific disciplines as diverse as biology, computer science, and sociology. Given that OM analyses can identify patterns of trajectories characterized by timely ordered events, they are an ideal tool for longitudinal analyses of venture creation processes. Remarkably, though, such studies hardly exist to date. To pave the way, we discuss how sequence analyses can be used in order to analyse venture creation – in particular with a view of country-specific differences. More specifically, we use a unique dataset of 351 venture creation processes to illustrate how founder involvement within venture creation differs across the institutional environment of Germany and the US – also in comparison to other venture characteristics.

2. Sequence analyses that reveal country-specific typologies of start-up processes and their institutional foundations

1. Introduction

Analyses of organizational processes are of paramount interest to organization researchers. Just take the examples of logistic flows, work processes, market developments, selection of job applicants, or venture creation processes. Are there particularly successful or efficient ways of organizing such processes? And, if so, is there a finite set of successful, or unsuccessful, processes – and how do they look like? Answers to such questions are of highest relevance, because they reveal typologies of organizational processes, as well as their timing and drivers. Ultimately, this makes it possible to optimize the respective processes.

To answer questions about how organizational processes unfold over time, longitudinal large-N analyses are needed. One of the most opportune methods to this end are sequence analyses (SA) based on optimal-matching (OM) algorithms. Contrary to traditional quantitative methods investigating temporal dynamics, OM techniques make it possible to treat one sequence of events as one unit of analysis, which enables the calculation of similarity degrees between sequences. Based on their respective similarities, sequences are then grouped into clusters of resembling processes. In this way, the entity of all processes observed is distilled into the most representative set of process types – based on the length, order, and duration of activities taking place. This substantially distinguishes SA from other quantitative dynamic methods in general, and its most viable alternative, event-history analysis in particular, because these methods typically consider one process as multiple stochastically generated events (Abbott, 1995). OM techniques therefore constitute a particularly valuable tool for longitudinal research into process typologies (Aisenbrey & Fasang, 2010).

Despite their potential for analyzing organizational processes, and despite their wide-spread use in sociology (see, for example, Abbott & Hrycak, 1990; Stovel, Savage, & Bearman, 1996; Han & Moen, 1999; Brzinsky-Fay, 2007; Lesnard, 2008), OM techniques have to date hardly been used in organizational research. This "limited application of OM analysis in the management area" has, most importantly, been attributed to "the fact that researchers are often unfamiliar with OM and its potential" (Biemann & Datta, 2013: 52).

Seeking to address this research gap, we illustrate how OM can be applied to the study of venture creation processes. Based on a dataset of 351 venture creation processes, we illustrate how founder involvement in setting-up ventures differ between the different institutional environments of Germany and the US. In order to highlight these differences, we contrast our results to the differences of venture creation approaches between industries, types of goods developed and degree of innovativeness. Our main contribution thereby consists in illustrating how sequence analyses work and how they allow to discern distinct venture creation approaches.

To illustrate our arguments, the remainder of the paper is organized as follows. Section 2 presents the state of the art of today's SA literature and explains how OM sequence analyses work. Section 3 illustrates how OM analyses can be applied to the study of founder involvement in venture creation. Section 4 presents the results. Section 5 concludes by summarizing our findings and critically discussing the potential of OM analyses for entrepreneurship research.

2. Literature Review: Towards an Application of OM Analyses To Venture Creation Processes

Over the past three decades, OM analyses have become an established method in scientific disciplines as diverse as biology, computer science, and sociology (Aisenbrey & Fasang, 2010). Their potential was first proven in biology at the end of the last century, when OM techniques were used to decode the human genome. Briefly afterwards, several OM variants were applied in computer science, where comparisons of character strings are still at the basis of today's spell checkers – to name just one example. Abbott and Forrest (1986) pioneered the application of OM techniques to social science data by analyzing sequences of dance patterns. Ever since, OM applications have gained momentum in sociological research, where they were chiefly used to investigate career paths and life-course trajectories (see, for example, Abbott and Hrycak (1990), Blair-Loy (1999), Han and Moen (1999), Salvato et al. (2012)).

Despite their initial success, the early applications of OM analysis in sociology did not remain without criticism. Led by Levine (2000) and Wu (2000), the OM opponents criticized that the adoption of this natural-science method to the social sciences was based on limited or unfounded theoretical assumptions.

Challenged by this criticism, the OM proponents further refined sequence analyses (see Brzinsky-Fay, Kohler, Halpin, Lesnard, Aisenbrey, Fasang, Elzinga, Anyadike-Danes, & McVicar, 2010; Liao, 2015). These methodological refinements have had two major consequences. First, a standard way of running OM analyses has crystallized (see Han & Moen, 1999; Abbott & Tsay, 2000; Stark & Vedres, 2006; Biemann & Datta, 2013).

Second, OM analyses have become an established tool in sociology as they offer the following advantages (see Aisenbrey & Fasang, 2010): (I) OM analyses make it possible to map event- and trajectory-based theories. (II) They allow for quantitative measurements of sequence similarity. (III) They are exploratory to the extent that distributional assumptions prior to the analyses are not necessary. (IV) They provide a comprehensive description of process typologies that are part of a broader population. (V) Finally, OM results can be combined with traditional statistical tools, such as multinomial regression analyses, in order to assess which factors influence the process typologies identified.

While OM applications in venture creation research offer the same advantages, they are still exceptional. To pave the way for OM analyses in entrepreneurship studies, we start with briefly illustrating how OM works. In doing so, we follow the dominant approaches used today (see Han & Moen, 1999; Abbott & Tsay, 2000; Stark & Vedres, 2006; Biemann & Datta, 2013). Accordingly, we determine the OM cost functions on the basis of frequency transitions and use Ward's minimum variance method combined with normalized-gamma, silhouette and Calinski-Harabasz validation indexes (Halkidi, Batistakis, & Vazirgiannis, 2001) in order to determine sequence clusters.

Before applying OM techniques to founder involvement in venture creation processes, it seems opportune to briefly explain how OM work: To identify sequence typologies, the majority of OM analyses proceeds in four steps.[2] **First**, the processes under investigation are coded into categorical values in order to describe them as sequences of states and events. A *state* refers to the activity that is taking place at a given moment during the observed process. A state thus indicates, for each time unit, in which situation the process is in. Take the example of a venture creation process, where product development, the acquisition of finance, or the recruitment of employees are examples of states. An *event* is a change of state; for example, when a start-up company ends the period of financial acquisition and begins with product development.

To report real-world processes as sequences of states and events, each state must be characterized by a code that belongs to a set of predefined categorical values, called alphabet. Consider the following example of an alphabet describing venture creation processes: Imagine that, during the start-up phase, venture founders would exclusively develop new products (=P), seek investment (= I), and hire employees (= L). Then, the alphabet of codes reporting this universe of venture creation activities is: *L*, *I*, *P*. Accordingly, *P P I I L* describes a start-up process in which the venture founders focus on product development during the first two time periods (e.g. months), then seek investment for the next two time periods, and finally hire employees during the last start-up period.

In a **second step**, a cost function needs to be determined in order to assess the degree of similarity between two sequences. To this end, optimal matching algorithms determine the costs of converting one sequence into a second one, whereby the states of the first sequence are substituted, deleted, or newly inserted until they resemble the second sequence. Take the example of the two sequences *L I P* and *L L I*. By deleting the initial *L* of the second sequence and inserting *P* at its end, the second sequence becomes equal to the first.

[2] As with virtually all quantitative and qualitative analyses, several variants of running OM analyses exist. A comprehensive overview over these variants has been provided by Aisenbrey and Fasang (2010). Since it is our major aim to illustrate how organization researcher can identify the most opportune coding approaches on the basis of their raw-data characteristics, we base our illustrations on today's standard way of running OM. Given that detailed illustrations of this standard procedure has been provided by Biemann and Datta (2013), we here only sketch the most important steps in order to allow novices to this method to follow our arguments.

Alternatively, the states *I* and *P* of the first sequence could be substituted with the states *L* and *I*. Importantly, the algorithm always chooses the least costly way of transforming sequences or, rather, assessing their difference. This leads us to the question of how 'expensive' substitutions, insertions and deletions are. Should one substitution cost more than another? Should an insertion or deletion cost the same as a substitution?

While several approaches to determining substitution costs exist, Rohwer's (1997) data-driven approach is currently seen as the most valid one and has therefore become today's standard. Rohwer suggests to set costs based on empirical grounds, namely according to the frequency with which events, i.e. transitions from one state into another, occur within a dataset. Accordingly, substitutions of events, occurring more frequently, are less costly – and vice-versa.

Costs of insertions and deletions are set in relation to substitution costs and determine which operation is less costly and, hence, favoured. While substitutions give more weight to the temporal position of states, because they may destroy event patterns but preserve the moments at which processes are in a certain state, insertions and deletions (indels) give more weight to event patterns but not to the moment at which they occur. Consider, for example, the sequences *P P I I L L* and *L L P P I I*. If one indel operation is equally expensive (i.e. if one insertion and one deletion is about half as expensive) as one substitution, the two sequences become equal by applying just two indel operations.[3] If, however, one indel operation is multiple times more expensive than one substitution, then only substitutions are possible to transform one sequence into the other. In this case, the distance between the two aforementioned sequences would be maximal, because each state would have to be substituted with the one of the other sequence.

Consequently, low indel costs motivate the preservation of event patterns by facilitating the addition or removal of states, which may also lead to an increase or decrease in sequence length. Sequence patterns are thus matched regardless of their position in the sequence. High indel costs, on the contrary, favor substitutions and therefore emphasize the temporal position of states. In order to stress the occurrence of events rather than the temporal position of states, we here set indel costs to the minimum value that satisfies the triangular inequality, meaning that deleting and inserting a state is never cheaper than the equivalent substitution.

Having defined the respective transformation costs, the **third OM step** consists in determining the distance between all sequence pairs within a dataset (see Abbott, 1995). To this end, the OM algorithm iteratively minimizes the costs of transforming sequences one into another through substitutions and indel operations. This process ultimately leads to the development of a distance matrix, which reports - for each sequence - how similar, or different, it is from any other sequence in the dataset.

[3] For example, the last two *L* states could be deleted from the first sequence and then inserted at its beginning.

In a **last step**, sequences are grouped into clusters in such a way that compactness and separateness are maximized, meaning that sequence similarity is highest within each cluster (compactness), while sequence difference is highest between clusters (separateness). In order to group sequences, a clustering algorithm is applied. This algorithm uses the distance matrix, established in the third OM step, as an input and places similar sequences with low distance scores into the same cluster, while sequences with high distance scores are placed in different clusters (Witten, Frank, & Hall, 2011). While the choice of 'the best' cluster algorithm is a debate in its own right, Ward's hierarchical clustering method has become the standard approach in OM analyses.

Given that the number of clusters needs to be determined *ex ante*, cluster validation techniques are typically used in combination with cluster analyses to determine the optimum cluster solution. Amongst the broad variety of cluster validation techniques, the normalized-gamma, silhouette and Calinski-Harabasz indicators are part of the most frequently used validation indices (see Halkidi, Batistakis, & Vazirgiannis, 2001). Accordingly, we here use Ward's hierarchical clustering techniques to group the dataset into 2 to 10 clusters and then determine the optimum cluster number by calculating the best average ranking score on the basis of the normalized-gamma, silhouette and Calinski-Harabasz rankings.

3. Methodological Approach: How To Investigate Founder Involvement in Venture Creation Processes

To illustrate how OM techniques can be applied to venture creation research, we here focus on the involvement of founders during the start-up phase of their ventures. When setting up a venture, founders face two fundamental questions: First, they need to decide about how much time they are willing and able to invest in venture creation. It is thereby especially important to find the right moment for giving up a previous job in order to change from a part-time (PT) to full-time (FT) commitment. Second, they need to decide whether it is opportune to set-up the new venture alone or together with a team of co-founders who provide a more diverse skill basis. Naturally, this decision implies that the founder cannot retain the complete control over all start-up decisions, but needs to share it with his co-founders.

Obviously, there is not 'one best way' of founder contribution to venture creation. There are many (Gartner, Shaver, Carter, & Reynolds, 2004). Importantly, previous research illustrates that these decisions are influenced by the circumstances in which venture creation occurs, most notably the flexibility of a country's labour-market institutions (Baughn, Sugheir, & Neupert, 2010). Accordingly, founders in the US with its flexible labour-market institutions were shown to approach venture creation in different ways than founders in Germany with its regulated institutional environment (Held, Herrmann, & van Mossel, forthcoming 2018). Additional factors that were found to influence founder involvement are a venture's industry (Dencker, Gruber, & Shah, 2009), the type of product or service developed (see Gartner, 1985), and the venture's degree of innovativeness (Beckman, 2006). Accordingly, we here use OM analyses to illustrate how founder involvement differs across the institutional setting of different countries and – in comparison – between industries, the type of good produced, and its degree of innovativeness.

Our analyses are based on the 'perfect timing' database as this dataset currently offers the most complete insights into how successful entrepreneurs proceeded when starting their own ventures. The database was collected between 2011 and 2014 at the Innovation Studies Group of Utrecht University in collaboration with Columbia University (New York). The dataset contains 423 cases overall, including start-up processes in Germany, and the US. Amongst all 423 start-up cases included in the dataset, our analyses are based on those 351 success cases of start-ups that achieved sustainable profits at the end of the venture creation. Unsuccessful cases of bankruptcy were excluded for the purpose of our analyses. While all 351 success cases are included in the analyses, we graphically report venture creation processes only for the first 75 months of venture creation for practical reasons.

To illustrate different team formation approaches during the venture creation process, we proceeded as illustrated in section 2: Accordingly, we conducted sequence analyses based on optimal matching techniques, whereby we determine substitution costs on the basis of transition matrices. This allows us to determine how similar each process is to every other process in the sample. These sequence analyses are combined with cluster analyses, whereby we use the standard clustering approach, i.e. Ward's hierarchical clustering combined with normalized-gamma, silhouette and Calinski-Harabasz validation indices. With regard to cluster analyses, we determined the optimum number of clusters on the basis of the entire dataset. Thereby, we only report clusters of more than 10 cases. Clusters including less than 10 cases are considered to be outliers and, thus, not reported.

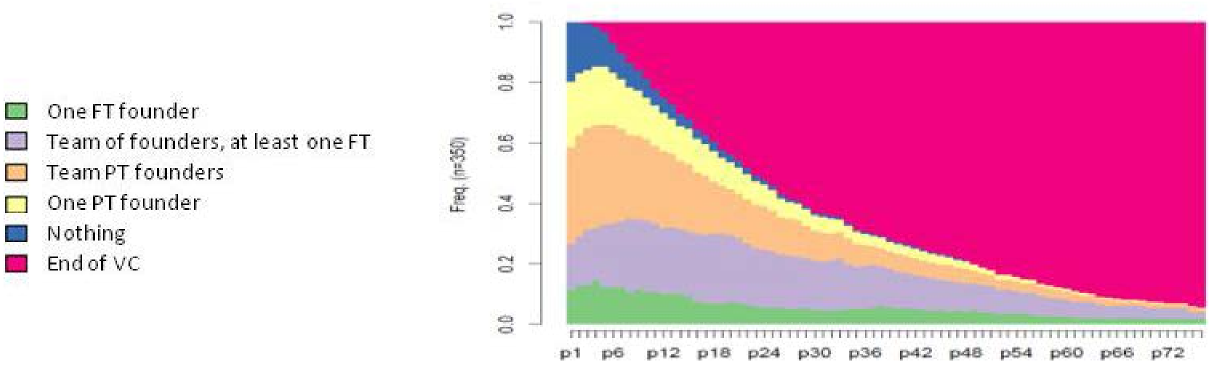
4. Results: Differences in Founder Involvement in Venture Creation Processes Between Countries and Other Venture-Creation Circumstances

The results of the aforementioned analyses are reported in the form of graphs. These graphs display the length of the venture creation process on the x-axis, whereby the time units report the number of months that have passed since the start of venture creation. This start is defined as the first time that one of the founders discussed the idea to set up the venture. The corresponding end of the venture creation process is defined as the first time that the new venture made sustainable profits. However, as already mentioned above, we depict venture creation processes only until month 75 in order to improve readability.

The y-axis, in turn, represents the share of ventures which find themselves in a certain state, i.e. undertake a specific activity, at a given point in time. These states are represented by a color code for each of the three areas of team formation. In doing so, two states (color codes) are present in any of the following graphs: namely the "end of the venture creation" and "nothing" happening, meaning that the founder is not undertaking any activity in the respective area of team formation after having discussed the start-up idea with another person.

The data about founder involvement in the start-up process [Graph 1] reveals that about half of all observed ventures start out as a team effort (*purple*, *orange*). Out of these ventures founded by a team, more than 2/3 are started by a team of PT founders (*orange*), whereas in the rest of the cases at least one member of the team works FT on the venture (*purple*). A similar pattern is observed amongst the ventures created by a single founder (*green*, *yellow*). Individual efforts make up 30% of the overall sample, out of which 2/3 worked exclusively PT (*yellow*) on their venture. The remaining 20% of ventures are not being actively developed by their founders after the latter had discussed the start-up idea with another person (*blue*). Over the course of the first 6 months, the share of inactive ventures and those managed by individuals in FT (*green*) decreases, whereas a larger percentage of team efforts, both PT and FT, can be observed.

Graph 1: Founder Involvement – All Ventures Observed (N=350)



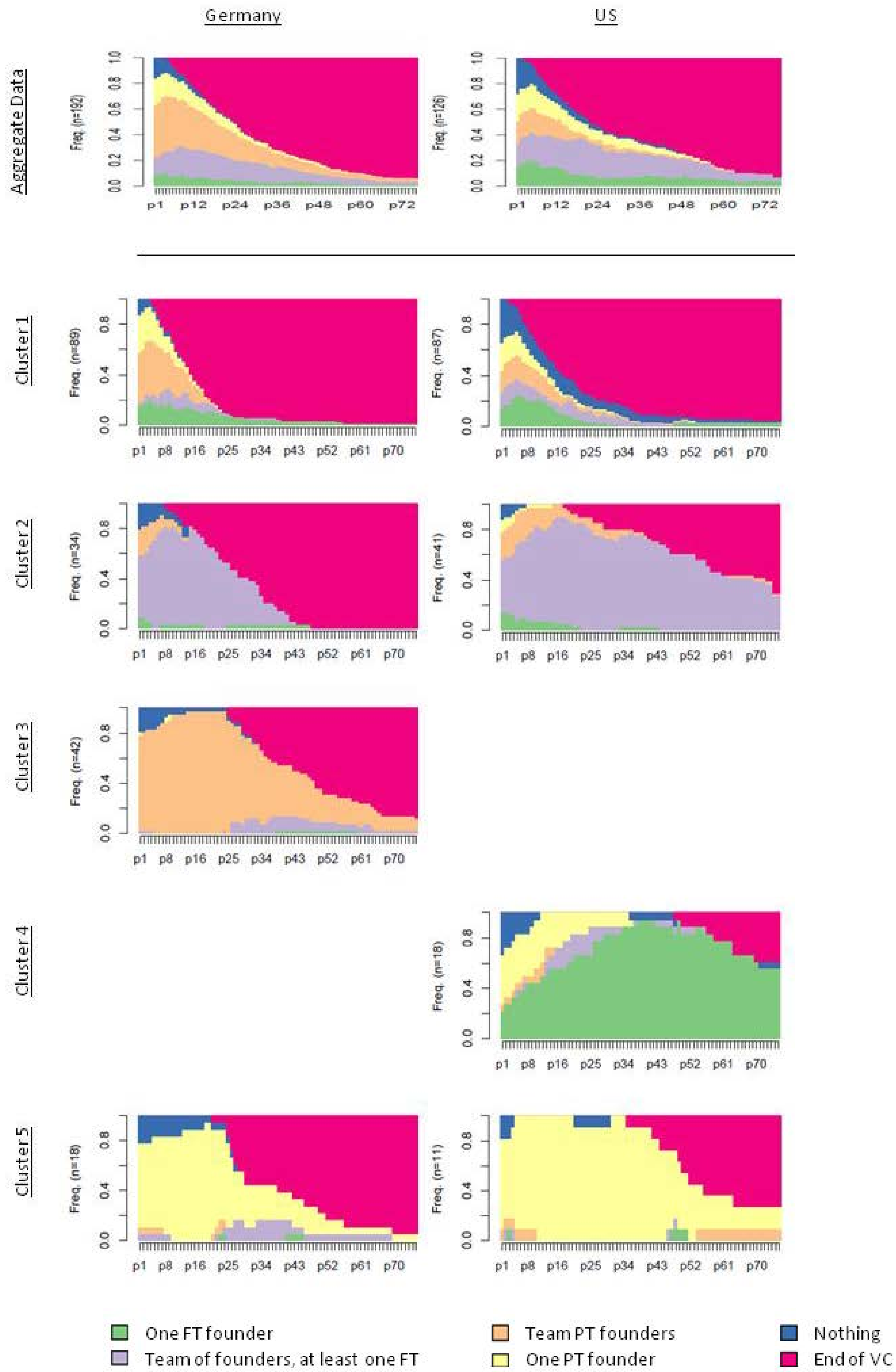
When we break this data down **by country** [Graph 2, Aggregate Data], clear differences between the two countries emerge. In comparison to the overall sample [see Graph 1], ventures in Germany are more likely to be created by a team of PT founders. This can be interpreted to the extent that Germany's rigid labour-market institutions make founders reluctant to engage in venture creation full-time at an early stage: Founders in Germany are only willing to give up their previous job and engage in venture creation full-time once they see good chances for their venture to generate sustainable profits. In contrast, the flexible labour-market institutions of the US drive founders to start their venture in full-time early on.

While these are important differences of founder involvement at the country level, they tell us little about the most typical approaches of founder involvement *within* each country. In order to gain more in-depth knowledge on country-specific approaches, we cluster the data as described above. These analyses reveal that one approach is common to all three countries [Graph 2, Cluster 1]: namely a founding process characterized by a short process duration. However, the differences within this short process are quite pronounced between countries. While in the US a relatively high level of inactivity persists throughout the beginning of venture creation, founders in Germany set out to work right away after they discussed the start-up idea with another person. This indicates that Germany's founders tend to start the venture creation process only once they are convinced of the venture's eventual success.

In addition, US and German founders resemble each other in two other start-up approaches they take. As illustrated by Graph 2, cluster 2, one approach consists in co-founding the venture whereby at least one founder dedicates his full-time capacity to venture creation. Interestingly, though, the length of this process is decisively longer in the US than in Germany [Graph 2, Cluster 2]. The second approach consists in individual part-time founding [Graph 2, Cluster 5].

Finally, one country-specific start-up approach can be observed both among German founders and among their US counterparts: While German founders create their venture with a team of PT founders [Graph 2, Cluster 3], US founders have a tendency to set up their start-up company on their own and on a full-time basis. [Graph 2, Cluster 4]. This supports the idea that Germany's rigid labour-market institutions make founders more cautious with regard to early time commitments in venture creation, whereas labour-market flexibility motivates superior engagement in venture creation early-on.

Graph 2: Founder Involvement – Country-Specific Approaches

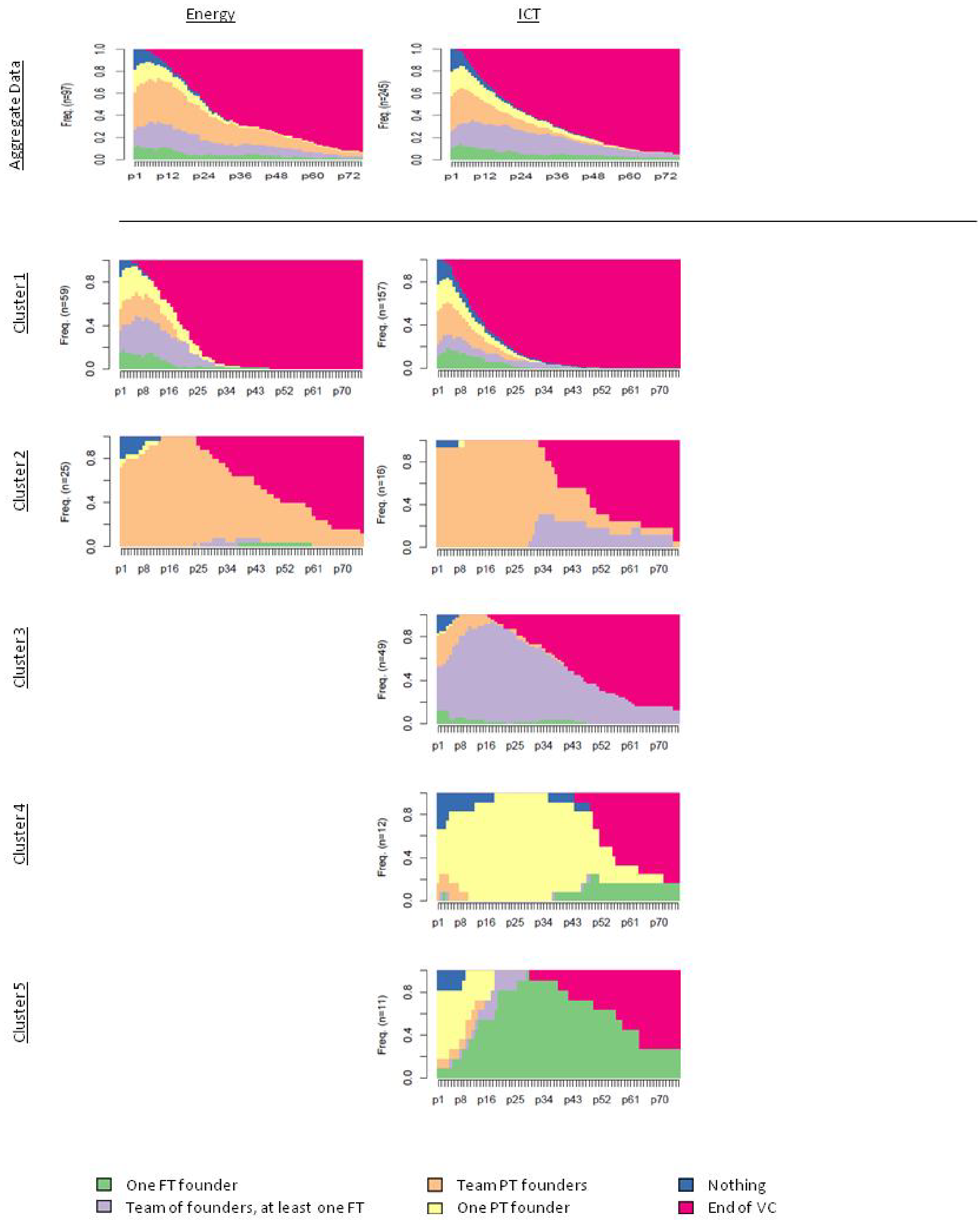


A split of the overall dataset **by industry** reveals that founder involvement differs fundamentally between the IT (information and communication technology) and the ET (environmental technology) sectors [Graph 3]. Founding processes in the ET sector end, on average, at a later point in time than ICT start-up processes [Graph 3, Aggregate Data]. This pattern is likely to be caused by the highly complex regulation of the ET sector, which entails that founders need to go through lengthy approval processes in order to obtain the ET permits needed for starting the new ventures. Accordingly, the ET sector is also characterized by a significantly larger share of ventures that are created by PT founders.

When breaking this aggregate data down for each industry with the use of cluster analyses, the results illustrate that two approaches are common to both industries: n8amely, first, a short process of founder involvement [Graph 3, Cluster 1]. The second approach consists in venture creation processes that are mainly driven by a team of part-time founders [Graph 3, Cluster 2]. Interestingly, the latter is much more common in the ET sector (26%) than among ICT ventures (7%).

In addition, three start-up processes exist that are particularly typical only for founders of ICT ventures. More concretely, these are founder-involvement processes centered around teams with at least one FT founder, processes advanced by a single PT founder, and processes driven by FT founders [Graph 3, Clusters 3-5]. Interestingly, founders of ET ventures do typically not pursue any of these three approaches when starting their ventures.

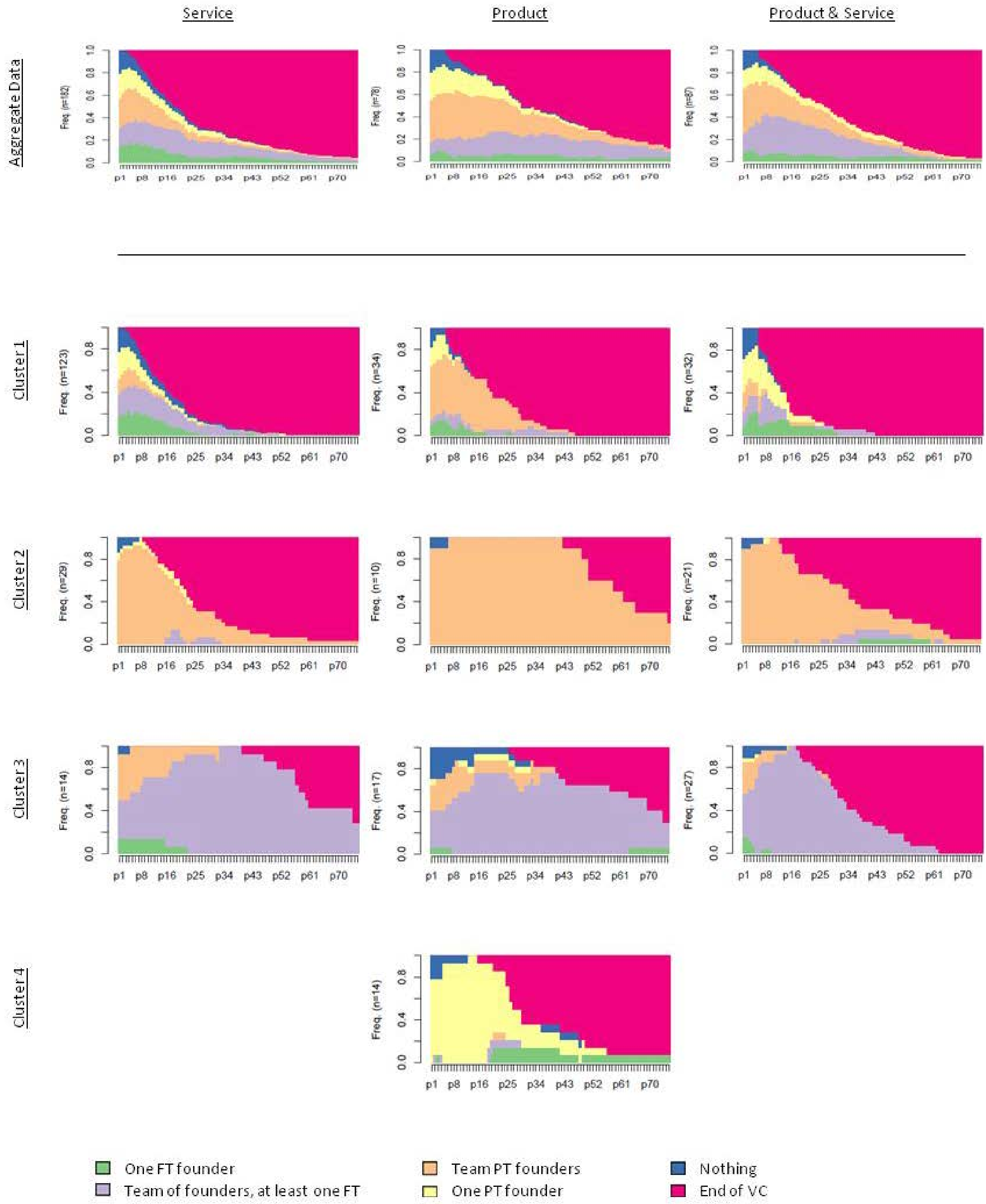
Graph 3: Founder Involvement – Industry-Specific Approaches



The ‘perfect timing’ database also allows us to distinguish founder involvement in venture creation processes depending on **the nature of the developed good** [Graph 4]. Hence, we are able to identify whether founder involvement differs for ventures that develop services, or products, or both. Our analyses of the overall dataset clearly show that, on average, the road to success is longer for ventures developing products, or both services and products, than for ventures developing services only [Graph 4, Aggregate Data]. The reason for this seems straight-forward: It, simply, takes longer to develop products than services. It is furthermore interesting to note that the percentage of ventures set-up by a team of PT founders is significantly higher in product than in service ventures. Service ventures, in contrast, are more frequently set up by individual FT founders.

We, again, disaggregate the overall dataset with cluster analyses in order to discern the most typical founder approaches depending on the venture’s good developed. We find that three approaches of founder involvement are particularly typical for each type of good developed [Graph 4, Clusters 1 – 3]. They are: processes of short founder involvement [Cluster 1], processes that are chiefly driven by a team of PT founders [Cluster 2], and processes driven by a founder team with at least FT founder. Interestingly, founders of product developing ventures tend to pursue one additional approach to venture creation [Graph 4, Cluster 4], which consists in setting-up the new venture alone and on a part-time basis.

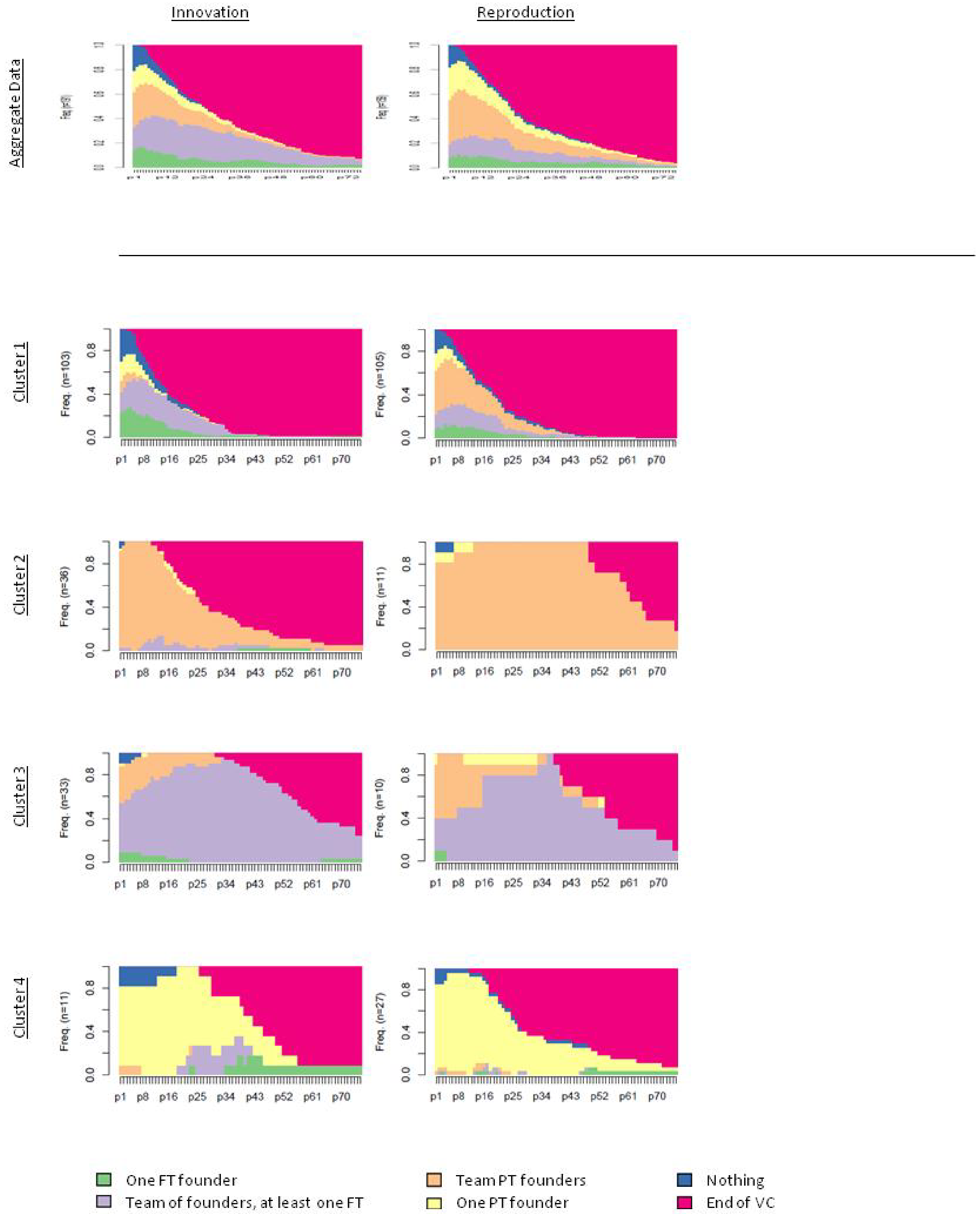
Graph 4: Founder Involvement – Approaches Specific to the Nature of Good Developed



Lastly, we differentiate between **reproducing and innovating ventures** [Graph 5]. At first sight, the analyses conducted on the basis of the overall dataset do not show substantial differences between ventures of either type with one exception: Innovative ventures are more frequently established by a start-up team involving at least on FT founder [Graph 5, Aggregate Data].

With the help of cluster analyses we obtain the same four approaches to venture creation for founders starting innovative and reproducing start-ups [Graph 5, Clusters 1 – 4]. Interestingly, though, the distribution of cases over these four clusters differs significantly: Innovative ventures are much more likely than reproducing ventures to be set up by a team of founders [Graph 5, Clusters 2, 3]. The respective share of cases grouped in founder team driven clusters 2 and 3 is 17% for innovative ventures compared to 7% reproducing ones. The focus on team founding processes by innovative start-ups can be explained by the fact that innovative products require a more diverse skill basis than the imitation of products. And the necessary skills are more readily found in teams. The opposite holds true for start-ups developing product imitations. Often, these ventures are set-up by individual PT founders (17% of cases in cluster 4), whereas this founding type is comparatively rare among innovative ventures (6% of cases in cluster 4).

Graph 5: Founder Involvement - Approaches Specific to the Degree of Innovativeness



5. Conclusion

Our paper seeks to further pave the way for OM based sequence analyses in venture creation research. To this end, we have both explained the method and shown how it can be applied to the involvement of founders in venture creation processes. The most important lesson to be learned from our analysis is that founding processes are by no means homogenous. As demonstrated above, founder involvement heavily depends on a venture's specific setting. Hence, entrepreneurs looking for advice on how to best establish their own venture should first assess their specific circumstances of venture creation. Accordingly, our analyses have shown that important circumstances influencing the venture creation approaches chosen by founders include a venture's country, its industry, nature of good and degree of innovativeness.

Importantly, our findings also have implications for policy-makers. In particular, the insight that different ways to successful venture creation exist between countries serves as a reminder that the idea to recreate Silicon Valley in Europe might not be most effective to foster entrepreneurship. Rather, tailor-made policies are needed for each country to support founders in the respective countries to successfully start their ventures.

While mastering OM sequence analyses is certainly not without challenges, we wish to conclude by highlighting the potential of this method. Across the social sciences, calls for longitudinal, analyses have steadily increased over the past decades. While time series regressions, history-event analyses, and several other tools have been developed in answer to these calls, OM analysis still constitutes the only method that is able to assess processes in their entity. Such longitudinal OM assessments offer the additional advantage that they can be combined with traditional statistical tools, such as cross-tab analyses or multi-nomial logistic regressions in order to identify the determinants of the cluster results obtained (for an example, see Salvato's (2012) study of CEO careers). While the potential of OM analyses for venture creation studies is thus massive, the time investment necessary to master this method is manageable. We therefore expect OM analyses to become an integral part of the methodological toolkit of organizational researcher in general, and entrepreneurship scholars in particular.

References

Abbott, A. (1995). Sequence analysis: new methods for old ideas. *Annual Review of Sociology, 21*(1), 93-113.

Abbott, A., & Forrest, J. (1986). Optimal matching methods for historical sequences. *Journal of Interdisciplinary History*, 471-494.

Abbott, A., & Hrycak, A. (1990). Measuring resemblance in sequence data: An optimal matching analysis of musicians' careers. *American Journal of Sociology, 96*(1), 144-185.

Abbott, A., & Tsay, A. (2000). Sequence analysis and optimal matching methods in sociology: Review and prospect: Sequence analysis. *Sociological Methods & Research, 29*(1), 3-33.

Aisenbrey, S., & Fasang, A. E. (2010). New Life for Old Ideas: The ''Second Wave'' of Sequence Analysis Bringing the ''Course'' Back Into the Life Course. *Sociological Methods & Research, 38*(3), 420–462.

Baughn, C. C., Sugheir, J., & Neupert, K. E. (2010). Labor flexibility and the prevalence of high-growth entrepreneurial activity. *Journal of Small Business and Entrepreneurship, 23*(1), 1-15.

Beckman, C. M. (2006). The influence of founding team company affiliations on firm behavior. *Academy of Management Journal, 49*(4), 741-758.

Biemann, T., & Datta, D. K. (2013). Analyzing Sequence Data: Optimal Matching in Management Research. *Organizational Research Methods, 17*(1), 51-76.

Blair-Loy, M. (1999). Career Patterns of Executive Women in Finance: An Optimal Matching Analysis. *The American Journal of Socialogy, 104*(5), 1346-1397.

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., Halpin, B., Lesnard, L., Aisenbrey, S., Fasang, A. E., . . . McVicar, D. (2010). Special Issue on Sequence Analysis. *Sociological Methods & Research, 38*(3), 359-512.

Dencker, J. C., Gruber, M., & Shah, S. K. (2009). Individual and opportunity factors influencing job creation in new firms. *Academy of Management Journal 52*(6), 1125-1147.

Gartner, W. B. (1985). A conceptual framework for describing the phenomenon of new venture creation. *Academy of Management Review, 10*(4), 696-706.

Gartner, W. B., Shaver, K. G., Carter, N. M., & Reynolds, P. D. (2004). *Handbook of Entrepreneurial Dynamics: The Process of Business Creation*. Thousand Oaks: SAGE Publications.

Halkidi, M., Batistakis, Y., & Vazirgiannis, M. (2001). On Clustering Validation Techniques. *Journal of Intelligent Information Systems, 17*(2), 107-145.

Han, S.-K., & Moen, P. (1999). Clocking Out: Temporal Patterning of Retirement. *American Journal of Sociology, 105*(1), 191-236.

Held, L., Herrmann, A. M., & van Mossel, A. (forthcoming 2018). Team Formation Processes in New Ventures. *Small Business Economics*.

Lesnard, L. (2008). Off-Scheduling Within Dual-Earner Couples: An Unequal and Negative Externality for Family Time. *American Journal of Sociology 114*(2), 447-490.

Levine, J. H. (2000). But What Have You Done for Us Lately? Commentary on Abbott and Tsay. *Sociological Methods & Research, 29*(1), 34-40.

Liao, T. F. (2015). Symposium: Life-course Sequence Analysis. *Sociological Methodology, 45*(1), 1-100.

Rohwer, G. (1997). *TDA: User's Manual* Berlin: Max-Planck-Institut fur Bildungsforschung.

Salvato, C., Minichilli, A., & Piccarreta, R. (2012). Faster Route to the CEO Suite: Nepotism or Managerial Proficiency? *Family Business Review, 25*(2), 206-224.

Stark, D., & Vedres, B. (2006). Social Times of Network Spaces: Network Sequences and Foreign Investment in Hungary. *American Journal of Sociology, 111*(5), 1367–1411.

Stovel, K., Savage, M., & Bearman, P. (1996). Ascription into Achievement: Models of Career Systems at Lloyds Bank, 1890-1970. *American Journal of Sociology, 102*(2), 358-399.

Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Amsterdam and others: Elsevier.

Wu, L. L. (2000). Some Comments on" Sequence Analysis and Optimal Matching Methods in Sociology: Review and Prospect". *Sociological Methods & Research, 29*(1), 41-64.