**A Methodological Essay on the Application of Social Sequence Analysis to the Study of Creative Trajectories***

Giovanni Formilan
Management Department
University of Bologna
giovanni.formilan2@unibo.it
&
University of Warwick
Centre for Interdisciplinary Methodologies

Simone Ferriani
Management Department
University of Bologna
simone.ferriani@unibo.it
&
Cass Business School
City University London

Gino Cattani
Department of Management & Organizations
Stern School of Business – NYU
gcattani@stern.nyu.edu

To appear in abbreviated format in

Formilan G., Ferriani, S. and Cattani, G. (2019). A methodological essay on the application of social sequence analysis to the study of creative trajectories. In Dörfler, V. and Stierand, M. (eds.). *Handbook of Research Methods on Creativity*, Edward Elgar Publishing.

**Acknowledgments**
The authors gratefully acknowledge financial assistance from the University of Bologna (Alma Idea Funding Scheme), the MIUR-PRIN ("Creativity, Audiences and Social Evaluation" [Grant: 2015LJXRXJ] and the European Research Council ("Blindspot" [Grant: 695256]).

## Abstract

In this essay, we present and illustrate a few applications of social sequence analysis (SSA) to the study of creativity. Focusing on complete sequences of events rather than on localized situations, SSA enables the analytical treatment of creativity as a process that unfolds over time, offering a fuller representation of temporal dynamics of creativity than is typically possible with other methods such as event history analysis, repeated measures, or panel design methods. We suggest that SSA holds great promise for research on creative industries, as it is particularly well suited to detect similarities among diverse creative trajectories while at the same time preserving their singularities. To substantiate our suggestions we employ data from the underground electronic music to examine trajectories of stylistic variation and illustrate how to implement sequence methods to augment and/or complement other research designs. Our purpose is to stimulate interest in SSA and encourage its application to the study of creativity at the individual, organizational and industry level.

> *"From the age of six, I was in the habit of drawing all kinds of things. Although I had produced numerous designs by my fiftieth year, none of my work done before my seventieth is really worth counting. At the age of seventy-three, I have come to understand the true forms of animals, insects and fish and the nature of plants and trees. Consequently, by the age of eighty-six, I will have made more and more progress, and at ninety, I will have got significantly closer to the essence of art. At the age of one hundred, I will have reached a magnificent level and at one hundred and ten, each dot and each line will be alive. I would like to ask you who outlive me to observe that I have not spoken without reason"* (Katsushika Hokusai, quoted in Dormandy 2000, 105).

The opening quote by Hokusai, the most distinguished Japanese printmaker and creator of world-famous iconic images like *The Great Wave* and *The Red Fuji*, offers a striking view of creativity and lifespan development. His remarks suggest that creativity is shaped by the experiences as well as the context of unique meaning that the creator builds up over his or her career (Kozbelt and Durmysheva 2007). It also underscores the interdependence between early choices/outcomes and later ones. A rich tradition of scholarly inquiry concerned with the unfolding of creativity across time, domains and/or other categorical boundaries confirms the importance of studying creativity in its wider context of production, i.e., considering the impact of the creator's personal and professional trajectory on the making of particular works (White 1993; Simonton 1997). Conceptually, viewing creativity through the lens of career trajectories resonates with established perspectives that see creativity as stemming from the interplay of people efforts at producing novelty, the domain of work and the social arrangements through which individuals' creative efforts are channeled and assessed (Gardner 1993; Csikszentmihalyi 1999).

The analytical treatment of such conceptualization, however, poses a few challenges. If, in fact, one embraces the notion of creativity as the manifestation of unique careers and insofar as careers can be understood as "evolving sequences of … experiences over time" (Arthur, Hall, and Lawrence 1989, 8), then an interesting duality emerges. On the one hand, there are trajectories that make some sequences of work experiences comparable to others; on the other hand, each person's creative trajectory is emblematic of a highly idiosyncratic journey. In other words, as suggested by Anand, Peiperl and Arthur (2002), while there are recognizable patterns to the way people flow through structures over the course of their creative careers, there is "a distinction between a person's work history, reflecting a publicly observable sequence of job experiences, and the subjective sense a person makes of those experiences" (p. 3). Tackling this duality invites analytical strategies designed to reconstitute uniqueness and similarity without the loss of sequence information that is typical of methods based on stochastic modeling of individual transitions or the well-known generalizability limits of approaches more sensitive to process questions. Theoretical interests inform the choice of methods, and methods in turn shape theoretical questions (Ragin 1987; Abbott 1988). In Ragin's (1987) terms, most scholarship concerned with the unfolding of creativity over time is either "variables-oriented" and quantitative or "case-oriented" and qualitative. In the compromise between specificity and similarities, variable-based strategies leverage generalizable similarities, while case-based approaches explain specificity. Variable-based statistical strategies are uniquely suited to tease out probabilistic relations between variables in large populations. Yet such strategies commonly must rely on particular assumptions about the underlying stochastic processes that produce the observed data. Assumptions that are typically unsatisfactory in market settings fraught with extreme levels of uncertainty on both the supply and the demand side (Caves 2000); such is the case with most domains of creative enterprise.[1] Qualitative, case-based research on creativity journeys over time generally focuses on a case of particular interest based on a small non-random sample of respondents interviewed in depth, or the analysis of detailed archival case studies. These approaches are particularly helpful for process questions (e.g., "How did the sequence of activities that constitute a given trajectory unfold?") and, more generally, for uncovering unique combinations of individual and or social forces that change or reproduce social processes. However, they are unsuitable for making general statements of empirical similarities about large populations.

In this essay, we discuss the analysis of social sequences (henceforth SSA) as a recently developed methodological tool that is especially sensitive to the duality of creativity trajectories making it possible to study patterns of social processes over time in an eventful way similar to historiography while retaining social scientific abstraction (Abbott 1990; Stark and Vedres 2006). Indeed, one of the most distinctive features of SSA resides in its logic that accounts for uniqueness and similarities at the same time. As Naomi Waltham-Smith posed it in her essay on musical sequences, "A sequence is a bipolar machine for transforming identity into difference and difference into identity" (Waltham-Smith 2015, 1). Rather than considering discrete moments in a sequence of events, SSA takes complete trajectories as unit of analysis, and makes no assumption about the data-generating process behind them. SSA therefore considers the uniqueness of diverse trajectories and, through algorithmic procedure, investigates the similarity among a set of trajectories by rearranging them into clusters of "sequential equivalence" (Han and Moen 1999). Another key benefit of this approach is that it allows for the treatment of events not as cases independent from each other but as sequences of concatenated activities, thus providing a powerful toolkit for leveraging the constitutive temporality of the creative trajectories under investigation (Stark and Vedres 2006), as Hokusai's opening quote alludes to.

To illustrate the opportunities that the SSA methodology affords in studying temporal patterns of creativity, we offer some insights from the analysis of stylistic variation sequences in the underground electronic music field. Specifically, we first present and discuss the stylistic trajectories of two worldwide acclaimed electronic music artists: Paul Kalkbrenner (Germany) and Four Tet (UK). We then extend the sample and use a dataset composed of the first 10 records published by 579 artists in the same music genre. As we shall see, these careers exhibit varying progression and significant movement across categorical and labor market boundaries, allowing us to probe complex social processes over time rather than aggregating and reducing them into a few categorical shifts. Using these examples we demonstrate how sequence methods can be leveraged to represent and analyze temporal dynamics of style in a way that may be inaccessible to other traditional analytical approaches such as event history or panel design methods.

[1] A well-known corollary to this phenomenon is the so-called "nobody knows" property of the creative industries, that is, the impossibility of predicting which product or person will receive market recognition (De Vany and Walls 1996).

**Social Sequence Analysis: the methodology**

SSA is a methodology originally borrowed from sequence analysis and the study of DNA sequences in biology. In the social sciences, it has become a methodology used primarily for the analysis of life course data (Abbott and Hrycak 1990; Chan 1995; Stovel, Savage, and Bearman 1996; Abbott 1995; Han and Moen 1999). SSA has a holistic approach as it turns the study of localized transitions into the analysis of comprehensive trajectories (Aisenbrey and Fasang 2010) and offers complementary tools to event history and survival analysis (Mayer and Tuma 1990; Yamaguchi 1991; Hosmer and Lemeshow 1999). The application of the SSA methodology to the realm of social sciences has experienced significant improvements since the 1990s (Cornwell 2015). In its original introduction to the social sciences (Abbott 1995; Abbott and Hrycak 1990), SSA has been discussed in terms of optimal matching (OM) algorithms (Sankoff and Kruskal 1983). For the purposes of this chapter, we largely focus on the OM procedure because it constitutes the basis for most sequence analytic approaches and the blueprint for successive developments.[2]

We provide suggestions for how to use sequence methods to study creativity and conduct sequence analysis. To make such suggestions more concrete, we develop a detailed example based on underground electronic music data that highlights the potential for sequence methods to analyze the evolution of musical style, a quintessential manifestation of creativity in fields of artistic production. In particular, we show how a researcher might look for distinguishable sets of creative trajectories across artists. One could then examine the factors influencings these groupings, or link them to relevant individual level outcomes such us critical acclaim or market appeal (Herndon and Lewis, 2015). We organize the discussion into three main phases. First, we introduce two artists that have received considerable recognition from the electronic music scene and summarize some idiosyncratic features of their music style. We then discuss how to manipulate stylistic features and make them suitable for SSA, and present a number of ways to describe and visualize sequential material. Second, we extend this simple case by including two additional artists that have similar stylistic features. With the resulting four-observation data, we show how to determine the parameters for OM analysis and hierarchical clustering. Finally, we broaden the scope of the analysis by considering a larger sample composed of 579 randomly selected artists. With this larger sample, we develop two types of sequences (i.e., trajectories of style, and trajectories of market success), and discuss how the outcome of SSA can inform subsequent data analysis and research questions. Conclusions point to the use of SSA as a methodology that can enhance our understanding of trajectory-level dynamics of creativity.

**Example data: underground electronic music**

We use discography-level data collected from *discogs.com*. For each music record, our data include the

[2] Indeed, the OM procedure constitutes the standard approach to social sequence analysis and is widely conceived as a synonym of sequence analysis in general terms (Elzinga 2003; Aisenbrey and Fasang 2010).

stylistic features (stylistic categories) and the market sales of each record (Formilan and Boari 2018). Although it is beyond the scope of this essay to provide detailed evidence of sequential patterns in the field of electronic music, it is worth mentioning some of the features that make our data well suited to discuss SSA. Underground electronic music is a contemporary genre that developed as a machine-generated recombination of other genres – especially funk, rhythm'n'blues, and soul (Reynolds 1998). Electronic music was thus a style-recombinant genre since its inception, and continuous attempts to innovate the sound are central to the field. However, stylistic consistency is a valuable trait in creative and cultural fields (Althuizen and Sgourev 2014), and this peculiarity therefore raises questions about whether more or less experimental artists tend to outperform their colleagues, and if diverse stylistic trajectories represent an advantage or disadvantage in terms of visibility and success. Through a sequence analytic approach, we can outline trajectories of stylistic innovation that are differently rich, and relate them to diverse pattern of market performance. Specifically, we focus on these characteristics and construct two discography-level sequences: 1) the sequence of stylistic variations from one record to the following one, and 2) the sequence of successful and average-performing records. For the sake of exposition, we analytically discuss the sequence of stylistic variations only. The second sequence will be used later to show possible uses of the outcomes of sequence analysis.

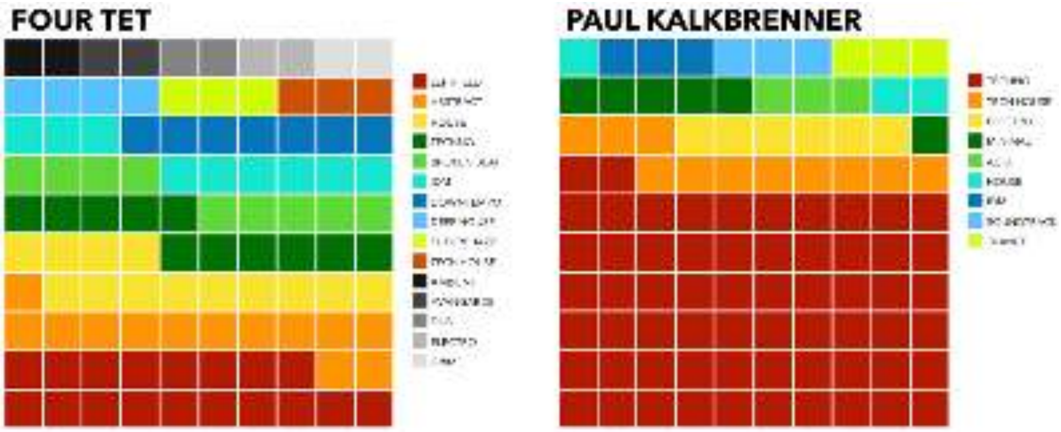
To determine and analyze social sequences, we used the open-source statistical environment R (R Core Team 2013) and the packages "TraMineR" (Gabadinho et al. 2011) and "cluster" (Maechler et al. 2018). As the primary purpose of SSA is to aggregate sequences into clusters of similarity, the OM-based SSA can be summarized into three main phases: 1) definition of the state space and the costs to align different sequences; 2) computation of pairwise distances between sequences; and 3) aggregation of sequences into clusters.

### *1. State Space and Costs*

Let us consider the stylistic trajectories of two worldwide acclaimed electronic music artists: Paul Kalkbrenner (Germany) and Four Tet (UK). Both artists have been included in the Top-100 ranking of the underground scene, compiled by the influential community *residentadvisor.net*.

Paul Kalkbrenner and Four Tet have markedly different style trajectories. Paul Kalkbrenner is unanimously considered one of the most successful representatives of the Berlin minimal techno, consecrated also by his protagonist role in the German movie *Berlin Calling*. The sound of Paul Kalkbrenner has remained relatively consistent over time, with stylistic traits that recur throughout his entire career. By contrast, Four Tet is famous for his eclectic style that comprises club-oriented techno music, improvisational jazz works, psychedelic and ambient records, and collaborations with several artists, including the drummer Steve Raid and Thom Yorke, leader of the worldwide acclaimed band Radiohead. His solo discography comprises a wide variety of styles, and his music often recombines traits from multiple stylistic domains. Figure 1 offers a visual representation of the main styles of the "album" and "single" records released by each artist from the beginning of their careers till 2018. The two images, consisting of 100 colored squares each, show the percentage incidence of each style on the whole artists' discographies. 62% of Paul Kalkbrenner's records are dominated by a single style (techno), while the discography of Four Tet is much more stylistically varied, with the most frequent style (leftfield) characterizing only the 18% of his music production.

Figure 1. *Percentage incidence of music styles in Four Tet's and Paul Kalkbrenner's discographies.*



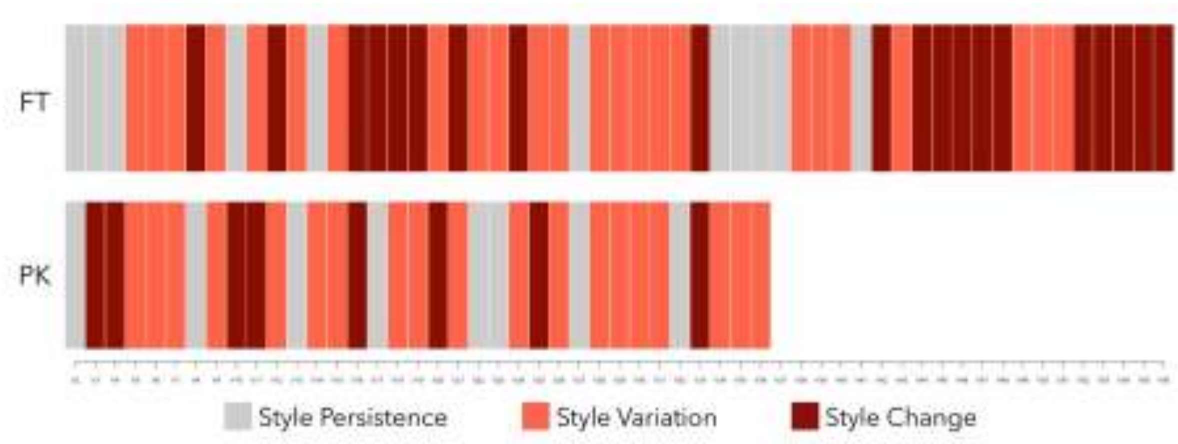

To proceed with SSA, we first have to identify the variable of interest and determine the so-called *state space*, defined as the set of possible events that can occur in a sequence. In our example, we are interested in detecting the stylistic dynamics of electronic music artists as a manifestation of their creative journeys. Accordingly, we recoded the stylistic feature of each artist's records by defining the style of every record relative to the previous record. Drawing from the stylistic categories to which the *discogs.com* contributors assign each release, we constructed a three-level categorical measure of stylistic variation by focusing on the main style assigned to a focal record (which plays a major role in the definition of the properties of objects; Gregan-Paxton, Hoeffler, and Zhao 2005) and the co-occurrence-based distance among its whole set of styles (Kovács and Hannan 2015; Leung 2014). We named the levels *style persistence* (P), *style variation* (V), and *style change* (C), and computed them analytically for each artist *i* using the following formula:

$$Stylistic\ Variation_{i,t} = \begin{cases} P, \text{if } intensity_{i,t} \leq \mathrm{M_{cooc}} + \partial \mathrm{SD_{cooc}} \wedge\ distance_{i,t} \in [distance_{i,t-1} - \beta SD_{cooc}\ , distance_{i,t-1} + \beta SD_{cooc}] \\ V, \text{if } intensity_{i,t} \leq \mathrm{M_{cooc}} + \partial \mathrm{SD_{cooc}} \wedge\ distance_{i,t} \notin [distance_{i,t-1} - \beta SD_{cooc}\ , distance_{i,t-1} + \beta SD_{cooc}] \\ V, \text{if } intensity_{i,t} > \mathrm{M_{cooc}} + \partial \mathrm{SD_{cooc}} \wedge\ distance_{i,t} \in [distance_{i,t-1} - \beta SD_{cooc}\ , distance_{i,t-1} + \beta SD_{cooc}] \\ C, \text{if } intensity_{i,t} > \mathrm{M_{cooc}} + \partial \mathrm{SD_{cooc}} \wedge\ distance_{i,t} \notin [distance_{i,t-1} - \beta SD_{cooc}\ , distance_{i,t-1} + \beta SD_{cooc}] \end{cases}$$

where $intensity_{i,t}$ is the co-occurrence distance between the main style assigned to the focal release and the main style of the previous release, $M_{cooc}$ and $SD_{cooc}$ are, respectively, the mean level and standard deviation of the co-occurrence distance among all the styles appearing in the discography of artist $i$, and $\partial$ and $\beta$ are two coefficients that take on the values (0.5; 1), respectively, in the examples with two and four artists, and (2; 0.5) in the extended-sample case. We chose arbitrarily these coefficients to enhance the visualization and meaningfulness of the output of SSA.

The vector (P, V, C) represents the state space of our sequences and consists of three states that collect information on the dimension of interest. A trajectory within the state space (P, V, C) depicts the extent to which the style of an artist remains consistent or evolves over time. Figure 2 shows the sequences of style evolution of Paul Kalkbrenner (PK) and Four Tet (FT). Each bar represents a record, and its color reflects the stylistic properties of the record in the state space (P, V, C).

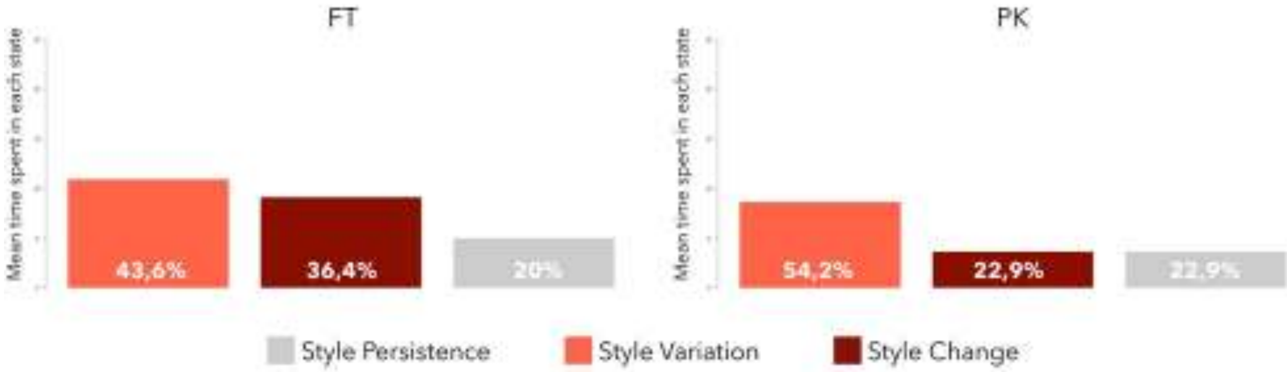
Figure 2. *Four Tet's and Paul Kalkbrenner's sequences of style evolution.*




Four Tet's sequence starts with three records that have similar styles, followed by three records that introduce minor stylistic variations, and a record markedly different from the previous one in the seventh position. It is also possible to visually appreciate two main periods of stylistic experimentation, at the second, fourth and the end of Four Tet's discography. The sequence of Paul Kalkbrenner, on the other hand, starts with a short period of stylistic experimentation (time 2 and 3) and is punctuated almost regularly by single records that introduce stylistic changes, preceded and followed variously by minor variations and persistence of style.

Besides the different length of the two artists' sequences, one can also see how Four Tet has a higher number of records that marked a style change compared to Paul Kalkbrenner. If we consider the mean time each artist spent in each state and the percentage incidence of each state over the sequences reported in Figure 2, marked stylistic changes have occurred more often in the discography of Four Tet than in the career of Paul Kalkbrenner (Figure 3).

Figure 3. *Plot of mean time spent in each state by the artists.*




After having identified the relevant dimensions to include in the analysis and determined the state space of the sequences, the second step in the preparatory phase involves the definition of the transformation costs required to align the states of a sequence to those of another sequence. As the goal of SSA is to detect sequences that are similar, this aspect is of crucial importance and has raised a number of criticisms and solutions from social sequence analysts (Brzinsky-Fay 2007; Stark and Vedres 2006; Elzinga 2003; Stovel and Bolan 2004). Limiting our discussion to the general approach of OM, there are two primary costs involved in the alignment of sequences. The first cost is the *substitution cost*, i.e., the cost of substituting one state with a different state. For example, a sequence P-P-V can be aligned to P-V-V by substituting the second P with V in the first sequence, or alternatively changing the second V with P in the second sequence. If the substitution cost is set equal to 1, the overall cost of alignment between the two sequences will be 1. It is worth noting, however, that the substitution costs between different states, in principle, are not the same. The magnitude of the substitution cost between any two states largely depends on theoretical considerations. For instance, we might assume that the transition between P and V can occur at a relatively low cost (a minor variation does not amount to betraying audience's expectations), while the cost of transitioning from P to C might be higher, given the efforts needed to deeply change one's style and face the risk of adverse reaction from the market. As we shall see, a common way to define the substitution costs is to rely on the actual transitions occurring in the empirical data.[3]

The second cost is the *indel cost* (insertion/deletion). Two sequences, P-P and P-V-P, can thus be aligned by inserting a V state between the two P states in the first sequence, or removing the V state in second sequence. Like in the case of the substitution costs, indel costs might differ from state to state depending on the characteristics of the empirical setting in which the sequences are investigated. However, most SSA applications employ a fixed indel cost because of the uncertainty that surrounds sequences – in principle, any state is as likely as any other state. Following Abbott's (1990) seminal work, we adopt this

[3] Another possibility is to hold the substitution cost constant. The decision to rely on variable or constant substitution cost has implications for the analysis and should be strongly supported by theoretical arguments. For instance, having a constant substitution cost would be poorly informative in correctly specifying the transition between levels of musical education, since it is very unlikely to take solfège class after having acquired a solfège degree at the conservatory.

approach in our study.

As we mentioned previously, a widely used approach to compute the substitution cost for a given state space is to look at the transition rate between each pair of states. In general terms, the transition rate between two states, $s_i$ and $s_j$, is computed as:

$$tr(s_i|s_j) = \frac{\sum_{t=1}^{L-1} n_{t,t+1}(s_i, s_j)}{\sum_{t=1}^{L-1} n_t(s_i)}$$

where *L* is the maximal length of the observed sequences, and *n* is the number of sequences that have a given position ($s_i$, $s_j$, or both) at time *t*. In the case of Paul Kalkbrenner and Four Tet, the transition rate matrices between the states appearing in each sequence independently are the following:

Table 1. *Transition rate matrices for Four Tet's and Paul Kalkbrenner's sequences.*

**Four Tet**

| | –> V | –> C | –> P |
|---|---|---|---|
| –> V | 0.500 | 0.333 | 0.167 |
| –> C | 0.368 | 0.579 | 0.053 |
| –> P | 0.455 | 0.091 | 0.455 |

**Paul Kalkbrenner**

| | –> V | –> C | –> P |
|---|---|---|---|
| –> V | 0.500 | 0.222 | 0.278 |
| –> C | 0.625 | 0.250 | 0.125 |
| –> P | 0.625 | 0.250 | 0.125 |

From the transition matrix of Paul Kalkbrenner, for instance, we can see that 62,5% of the times a record with a minor stylistic variation (first column) follows a record that introduced a marked stylistic change (second row). In general terms, the transition rate matrix should be interpreted as the likelihood of transition from any given state to any other state within a set of sequences. If we combine the two transition rate matrices, we can compute the substitution cost for the object (PK, FT) that reflects how often the transition between any two states is likely to occur. The transition rate-based substitution cost between two states is typically defined as the reciprocal of the rate of transition between the two states. For the object (PK, FT), the substitution cost between each pair of states is reported in Table 2.

Table 2. *Transition rate-based substitution cost for the sequence object (PK, FT)*

| | –> V | –> C | –> P |
|---|---|---|---|
| –> V | 0 | 1.270 | 1.259 |
| –> C | 1.270 | 0 | 1.768 |
| –> P | 1.259 | 1.768 | 0 |

We can see, for instance, that the cost of introducing a marked stylistic change is higher than the cost of introducing a stylistic variation after a record that had the same persistent style of the previous one (third row). The diagonal of the substitution cost matrix is 0 because it assumes that remaining in the same state has no cost. Moreover, the matrix is symmetrical because, in the alignment of any two sequences, the cost of substituting a P with a C in one sequence has to be the same as the cost of substituting a C with a P in the other sequence.

***2. Pairwise Distance***

After defining the state space and the transformation costs, we can compute the dissimilarity between each pair of sequences. In general terms, the pairwise dissimilarity measure is calculated using the following formula:

$$d(Seq_x, Seq_y) = A(Seq_x, Seq_x) + A(Seq_y, Seq_y) - 2A(Seq_x, Seq_y)$$

where $A(Seq_x, Seq_y)$ is the count of common states between sequences $Seq_x$ and $Seq_y$. If $A(Seq_x, Seq_y)=0$, the two sequences have no state in common, and the dissimilarity is thus maximal, $d(Seq_x, Seq_y)=1$. The general dissimilarity formula is then implemented in different ways by different distance measures, the most widely used being the ones reported in Table 3.

Table 3. *Popular measures of Pairwise Distance*

| **Distance** | **Sequence Types** | **Considered Costs** | **Rationale** |
|---|---|---|---|
| *Optimal Matching (OM)* | Equal-length | Substitution and indel | The minimal cost of transforming one sequence into the other one. |
| *Generalized Hamming distance (HAM)* | Equal-length | Substitution only | The number of positions at which two equal-length sequences differ. |
| *Dynamic Hamming distance (DHD)* | Equal-length | Substitution only | The number of positions at which two equal-length sequences differ (with position-dependent costs). |
| *Longest common subsequence (LCS)* | Unequal-length; Equal-length | Substitution only | The number of common positions (also non-consecutive) within the sequences. |
| *Longest common prefix (LCP)* | Unequal-length; Equal-length | Substitution only | The number of successive common positions starting from the beginning of the sequences. |
| *Reversed longest common prefix (RLCP)* | Unequal-length; Equal-length | Substitution only | The number of successive common positions starting from the end of the sequences. |

From the pairwise distances we can compute the distance matrix that collects all pairwise distances between each sequence entering the sequence object. In our example, the distance matrix is a 2-by-2 table with symmetrical values. The rationale behind the computation of pairwise distances is to minimize the cost of transforming one sequence into another, so a lower cost will always be preferred over a higher one. Specifically, if the indel cost of a given pair of states is higher than their substitution cost, the latter will be used by the OM algorithm to manipulate the sequences. If, for the sake of exposition, we reduce the length of the sequences to a shorter length (t=3), it is possible to appreciate how different specifications of the indel and substitution costs lead to different distances between the trajectories of Paul Kalkbrenner (P-C-C) and Four Tet (P-P-P). Table 4 reports two different distance matrices between the two reduced sequences.

Table 4. *Distance matrices between PK and FT, according to different specifications of the indel cost.*

**Low Indel (0.667)**

| | PK | FT |
|---|---|---|
| PK | 0 | 2.667 |
| FT | 2.667 | 0 |

**Default Indel (2)**

| | PK | FT |
|---|---|---|
| PK | 0 | 3.333 |
| FT | 3.333 | 0 |

In the first case, the indel cost is set at a very low level (defined here arbitrarily as 1 below the minimum substitution cost), while in the second case the indel cost is higher than the substitution cost (2, which is the default value in TraMineR and in most SSA software). In both cases, the transition rate-based substitution cost between C and P is 1.667. The algorithm can align the two sequences P-C-C and P-P-P in two ways: it can substitute the two C with two P in the second sequence (and vice versa; two substitution operation), or delete the two C and insert two new P state (and vice versa; four indel operations). When the indel cost is lower than the substitution cost, the distance is minimized using the indel cost (0. 667*4 = 2.667 < 1.667*2 = 3.33). When the indel cost is higher than the substitution cost, the substitution cost will be preferred instead. The definition of the indel and substitution costs is crucial for a meaningful treatment of sequential data. Setting a particularly high substitution cost, for instance, would cause the OM algorithm to value the order of states over their timing, while prohibitive indel costs would favor substitution and therefore maintain the position of those states that need no change.

### *3. Clustering*

The main purpose of computing pairwise dissimilarity between sequences is to produce groupings of similar sequences (typologies of sequences) and identify recurrent patterns. Clustering is probably the most used analysis to explore data and find homogeneous groups of observations (Kaufman and Rousseeuw 2005). A common and widely used criterion for hierarchical clustering is the minimum variance method developed by Ward (1963). This method selects the pair of single-observation clusters to merge into a hierarchically higher-level cluster based on the optimization of an objective function (the Euclidean distance, in the original formulation, or any preferred distance). This criterion, therefore, minimizes the total within-cluster variance. Operationally, the hierarchical clustering distance for two sequences $Seq_x$ and $Seq_y$ is given by:

$$HC_{D_{x,y}} = d(\{Seq_x\},\{Seq_y\}) = \|Seq_x - Seq_y\|^2.$$

Since it would be meaningless to exemplify the clustering procedure with only two sequences, we added two additional sequences to our illustrative case. We purposefully selected two artists exhibiting stylistic features similar to Four Tet's and Paul Kalkbrenner's ones. These artists are Guy Gerber (GG, stylistically eclectic) and Ben Klock (BK, stylistically consistent). Figure 4 shows the distribution of different styles in the discographies of the four artists, grouped by level of stylistic eclecticism, while Figure 5 depicts the whole sequences.

Figure 4. *Percentage incidence of music styles in the four artists' discographies.*

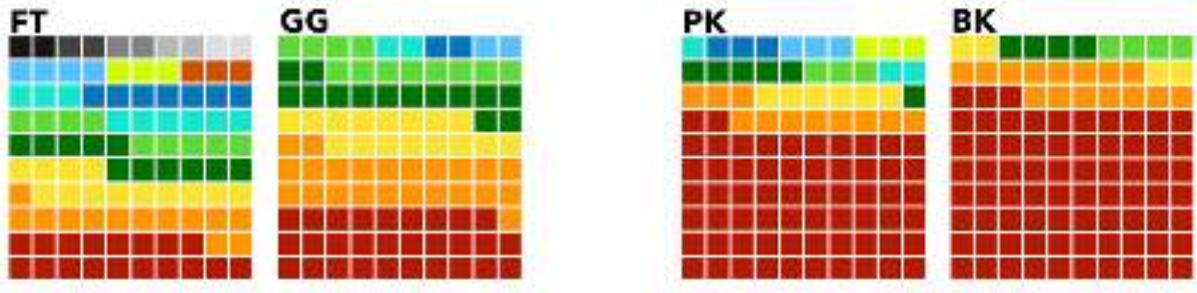
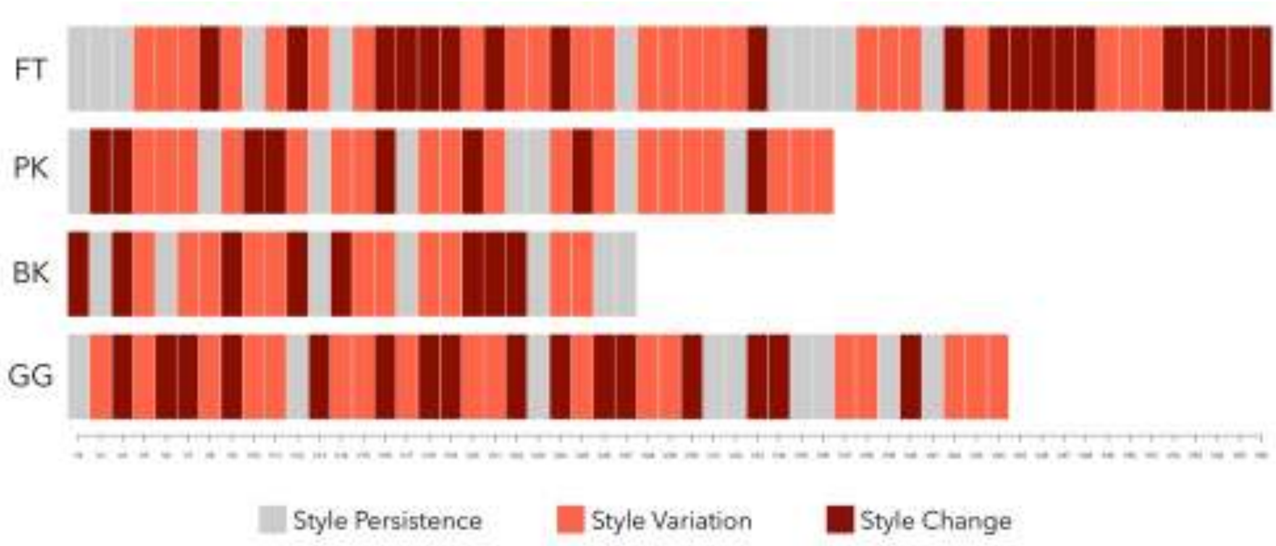
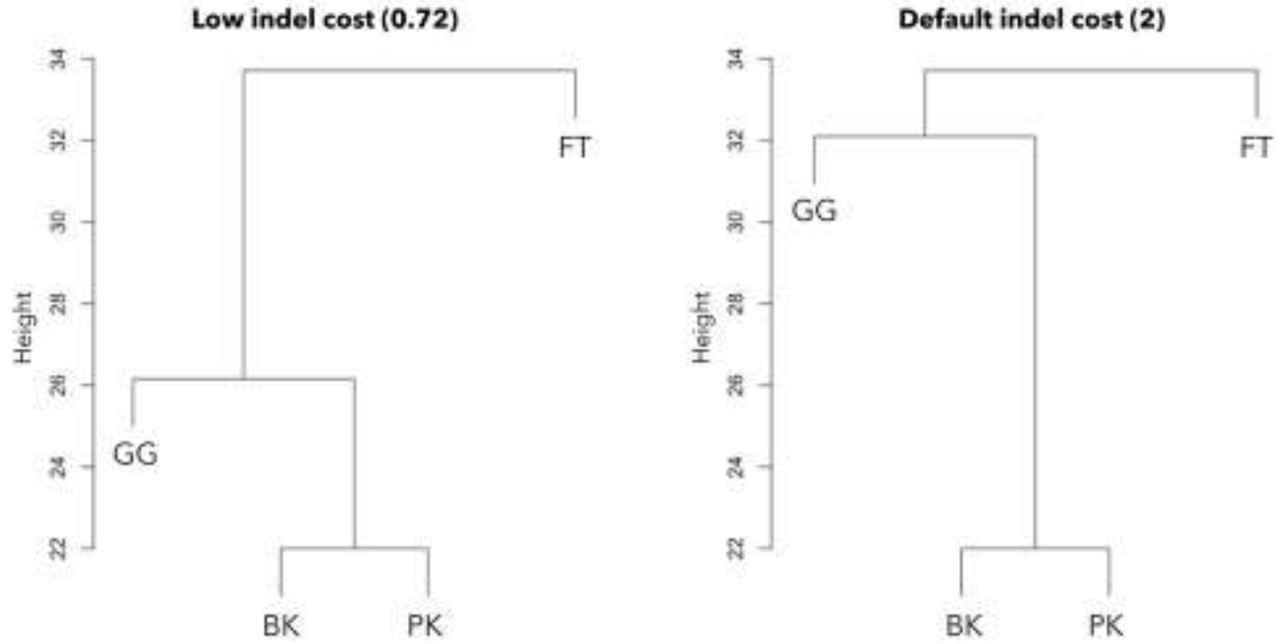



Figure 5. *Sequences of style evolution of the four artists.*




As we have seen, the OM algorithm produces different distance matrices depending on the specification of the indel and substitution costs. Clustering techniques use these matrices to group artists into clusters of structural equivalence. Figure 6 shows the dendograms that reflect slightly different groupings depending on two different distance matrices that use indel cost of 2 (standard) and 0.72 (0.5 below the lowest substitution cost), respectively[4].

Figure 6. *Dendograms resulting from low and default specifications of the indel cost.*




NOTE. The y-axis (Height) captures the distance between sequences as defined by the distance matrix.

In both cases, Ben Klock and Paul Kalkbrenner (the artists with the most consistent styles) are grouped together, while the default indel cost identifies a shorter distance between Four Tet and Guy Gerber compared to the solution with the low indel cost. While the two dendograms do not differ significantly, they nonetheless suggest how the definition of different indel and substitution costs influences the pairwise distance and, in so doing, determines diverse outcomes of grouping techniques. The accurate definition of the parameters involved in SSA is therefore central for a correct attribution of observational units to different similarity-based clusters, as well as the conclusions that can be drawn from the results of the analysis. In the following paragraph, we use a larger dataset to illustrate clustering results more extensively.

**Extended dataset: Data Preparation and Vocabulary Definition**

In this example, we are interested in the stylistic variations in the first 10 records released during electronic music artists' early careers. This restriction is imposed on the data in order to use the OM algorithm and facilitate graphical representation. Table 5 shows the percentage distribution of the three states in our sample. Following the same rationale for defining the state space used in the previous examples, we created the sequence of the first 10 records' stylistic variation.

[4] Given the different length of the sequences, we used a spell-adjusted distance measure (Halpin 2010). This is a modified version of the OM algorithm that scales the distance by considering how many identical states (so-called spells, exemplified by the V-V-V substring at the beginning of Guy Gerber's sequence) occur successively in each sequence (for details on modified versions of the OM algorithm, see Cornwell 2015, 5.6.4).

Table 5. *Descriptive statistics*

| State | N. | Percentage | Cumulative Percentage |
|---|---|---|---|
| *Style Persistence* | 3047 | 52.6% | 52.6% |
| *Style Variation* | 2415 | 41.7% | 94.3% |
| *Style Change* | 328 | 5.7% | 100% |

**Substitution and Indel Costs**

Since the timing feature of the sequences is relevant in our case, we adopted a transition rate-based substitution cost and an indel cost that is slightly higher than the minimum substitution cost (Stark and Vedres 2006). However, to show the differences in the construction of pairwise distances when the sample size increases, we present the clustering results of two extreme situations that use, respectively, a very high and very low indel costs[5].

The transition rates and the substitution costs of our sequence are reported in Table 6 and Table 7.

Table 6. *Transition rates for the extended sample*

| | —> Change | —> Persistence | —> Variation |
|---|---|---|---|
| Change —> | 0.3737 | 0.2458 | 0.3805 |
| Persistence —> | 0.0398 | 0.5583 | 0.4020 |
| Variation —> | 0.0498 | 0.3963 | 0.5540 |

Table 7. *Substitution Costs for the extended sample*

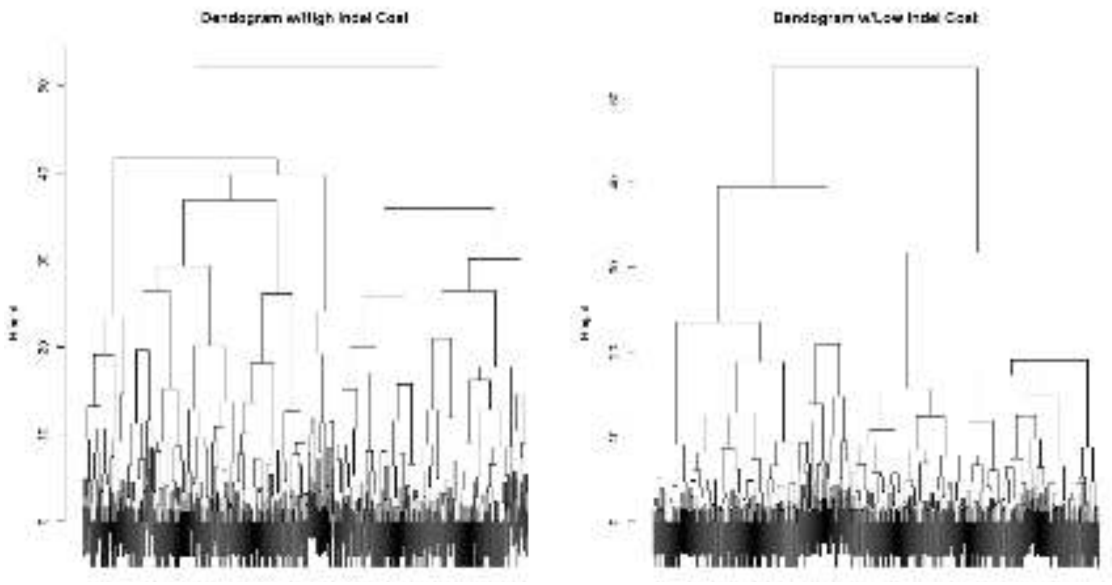
| | —> Change | —> Persistence | —> Variation |
|---|---|---|---|
| Change —> | 0 | 1.7144 | 1.5698 |
| Persistence —> | 1.7144 | 0 | 1.2018 |
| Variation —> | 1.5698 | 1.2018 | 0 |

**Pairwise Distance and Clustering**

Figure 7 shows the dendograms that result from the hierarchical clustering of the sequences based on the Ward's algorithm, according to the two situations of high and low indel cost.

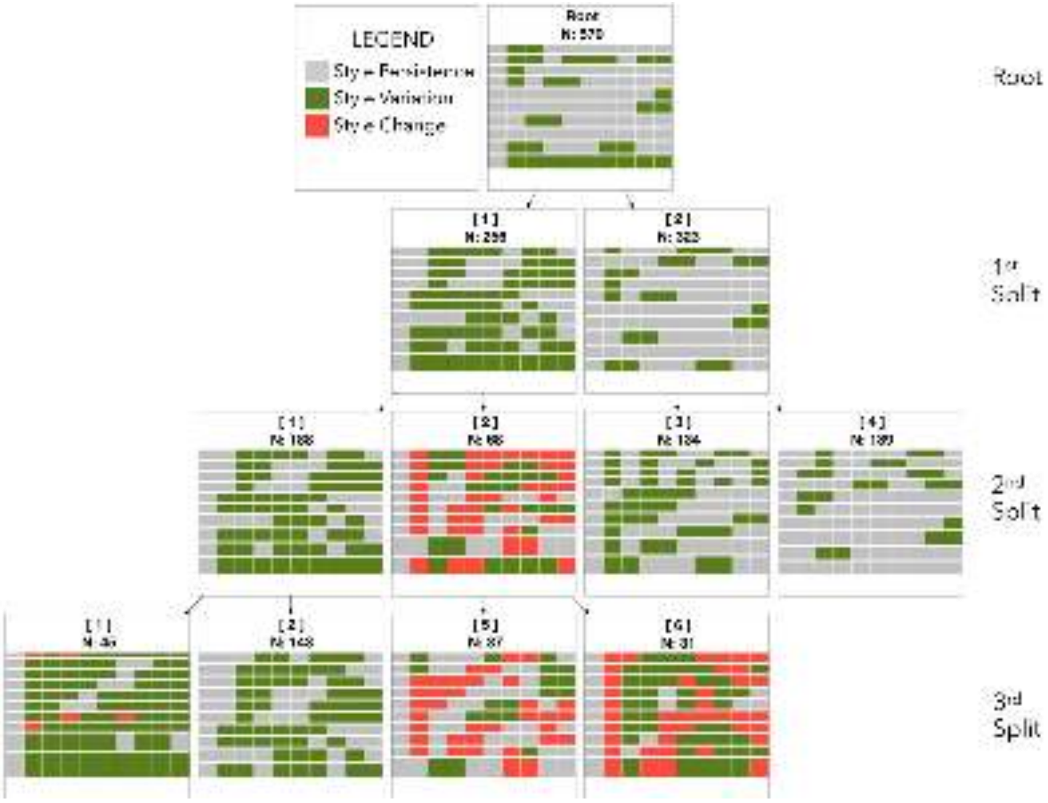
[5] The high indel cost is defined as the length of the sequences (10) divided by two times the maximum substitution cost for the state space (P, V, C), while the low indel cost is computed as 0.5 points below the minimum substitution cost.

Figure 7. *Dendograms from Ward's clustering techniques, for high (right) and low (left) indel cost.*




With the extended dataset, the low indel cost condition delivers a much clearer clustering of the sequences, while the high indel cost situation keeps the differences between the sequences more pronounced. A low indel cost allows the algorithm to more deeply modify each sequence by inserting and deleting states regardless of their position and, therefore, is more flexible in the alignment of the sequences. However, as noted previously, a low indel cost can alter the ordering of states significantly and, inevitably, eliminate some of the intrinsic properties of the sequences. There are several tools that can sustain the definition of the meaningful number of clusters that should be produced and used in the analysis. For instance, by showing the progressive distance-based ramification of the original set of sequences into cluster branches, the clustering tree offers a visual support (Figure 8).

Figure 8. *Example of clustering tree of Ward's hierarchical clustering (low indel cost).*




Starting from the root cluster (which gathers the whole set of sequences), the clustering tree visually displays how the algorithm proceeds in splitting each cluster. For instance, cluster [2] identified at the 2nd split (that gathers 68 sequences with a strong presence of stylistic changes) is divided at the 3rd split into the clusters [5] and [6], that contain respectively 37 and 31 sequences. As visually noticeable, the algorithm distinguished eclectic sequences with a tendency towards persistence (cluster [5]) from eclectic sequences that involve style variations more frequently (cluster [6]).

Figure 9 displays the first 10 sequences of each cluster (with a number of clusters arbitrarily set equal to N=4), according to high and low indel costs. It is worth noting how the optimal matching procedure with low indel cost creates clusters that are of comparable size, while the high indel cost preserves diversity also in terms of cluster population.

Figure 9. *Sequences by clusters, at high (left) and low (right) indel cost.*

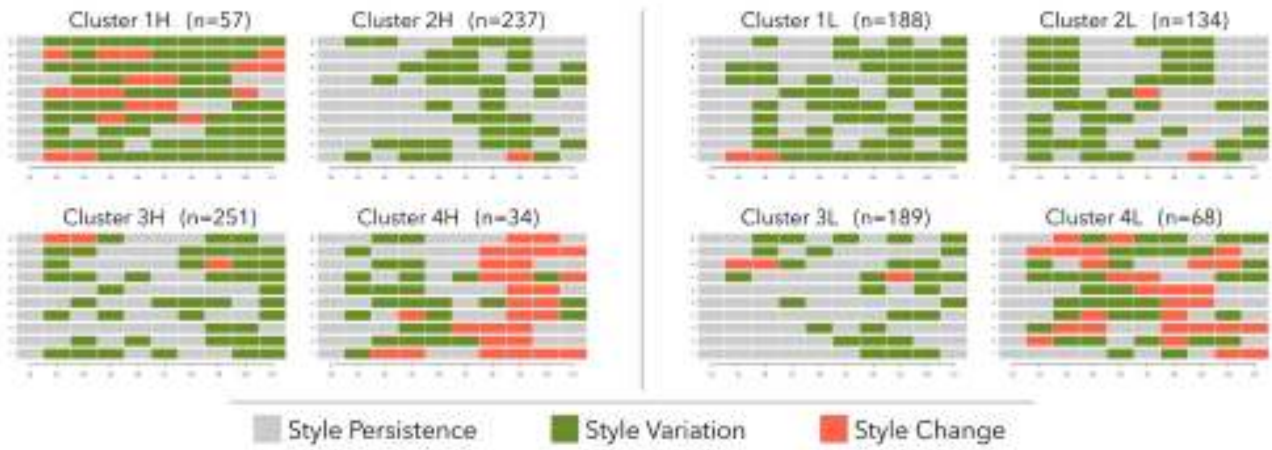


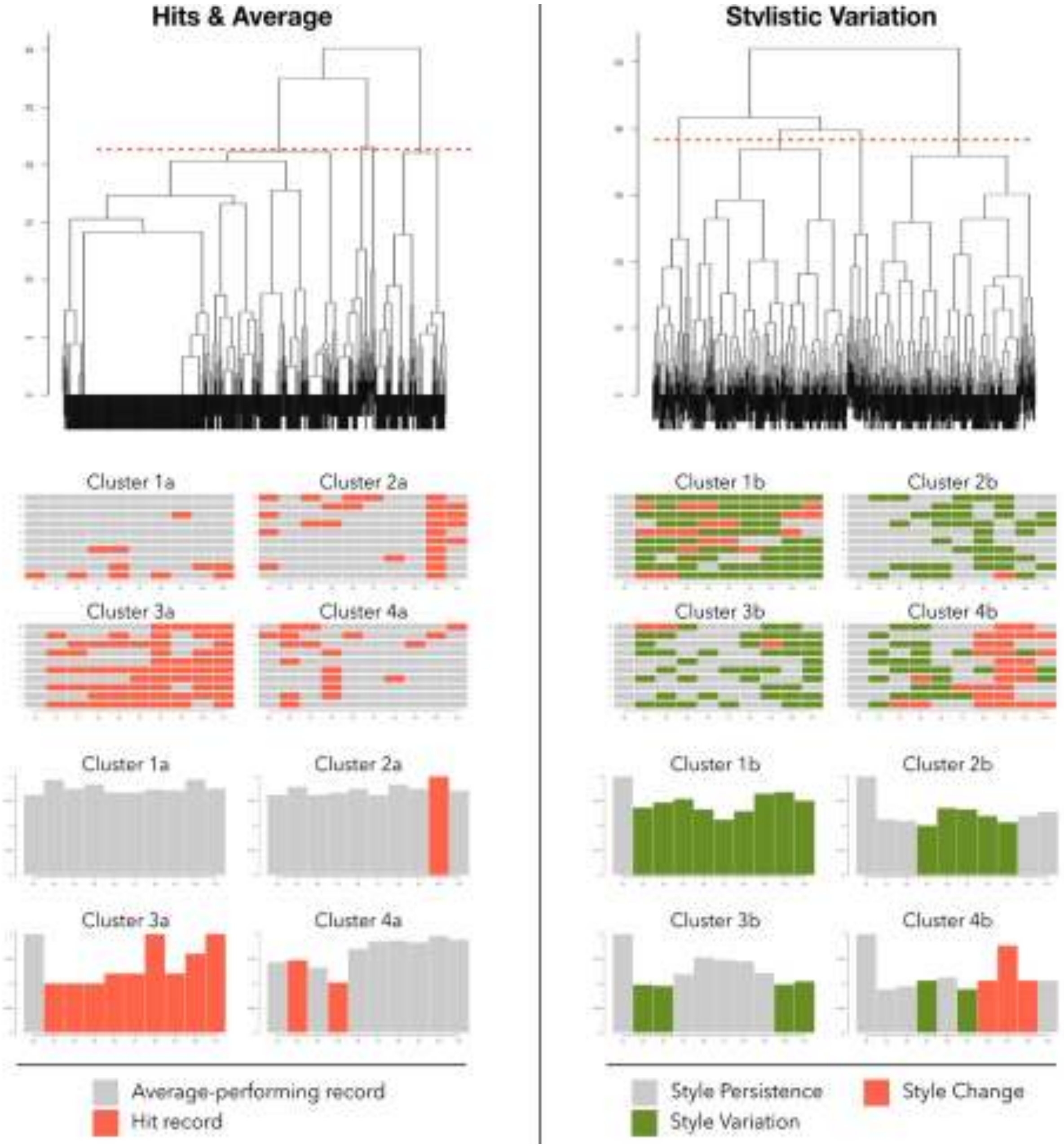

Figure 10 graphically depicts the modal state of each cluster's sequences according to high (left) and low (right) indel costs. If we were to name each cluster, the two indel costs would suggest different names. The clustering with high indel cost goes in the direction of having cluster 3H named "Mid-Way Persistent Artists": the modal trajectory in this cluster reflects stylistic persistence preceded and followed by stylistic variations. On the contrary, the situation with the low indel cost would suggest a "Pure Style Loyal" label for cluster 3L: no variation is involved in the modal trajectory of this cluster.

Figure 10. *Modal state sequence by cluster, at high (left) and low (right) indel cost.*


**Further uses of SSA's outcomes**

The results of SSA on one set of sequences can be combined with SSA on a different set of sequences to unveil patterns and correspondences between the same set of events observed from different angles. We followed the same logic that we employed in the construction and analysis of the previous sequence of style evolution, and formalized an additional manifestation of an artist's professional trajectory: the commercial performance sequence (expressed in terms of hits and average-performing records). Figure 11 shows the dendograms of the two sequence objects with the cutting points for the creation of the corresponding clusters, the representative sequences, and the modal states for each sequence object's clusters.

Figure 11. *Dendograms with cutting point, representative sequences, and modal states for the two sequences.*




We can now combine the clusters and create crosstabs in order to see if artists with a similar sequence of stylistic variation tend to display a similar sequence of commercial success (Table 8).

Table 8. *Crosstabs (nominal values; clusters' size in parentheses)*

| | | *Stylistic Variation* | | | | |
|---|---|---|---|---|---|---|
| | | Cluster 1b | Cluster 2b | Cluster 3b | Cluster 4b | |
| ***Hits Trajectory*** | Cluster 1a | 36 | 167 | 184 * | 28 | (415) |
| | Cluster 2a | 6 | 21 | 17 | 4 | (48) |
| | Cluster 3a | 1 | 2 | 7* | 0 | (10) |
| | Cluster 4a | 14 | 47 | 43 | 2 | (106) |
| | | (57) | (237) | (251) | (34) | (579) |

*: commented values.

Looking at the modal states and the results of the crosstabs, it could be argued that those artists that maintain stylistic persistence in the middle part of their 20-year early career, after having introduced minor stylistic refinements in the early records (Cluster 3b), are more likely to experience a constant sequence of successes (Cluster 3a) compared to other stylistic trajectories. In fact, almost 3% of the artists in Cluster 3b (7 artists out of 251) display the performance trajectory represented in Cluster 3a, while such a performance trajectory regards only 1.7% and about 1% of the artists included in Cluster 1b and Cluster 2b, respectively. At the same time, however, a large number of artists follow the stylistic trajectory represented in Cluster 3b but still experience no success on the market (n=184).

We can also include clusters in parametric regression analysis and inspect the effects of cluster membership on some artist-level variable (Table 9). For instance, one could explore how different trajectories determine individual-level features like the mean level of atypicality[6] of an artist, or the average performance s/he experienced on the market. While no statistically significant result emerges from regressing different style clusters over the average performance (Model 2), the results of Model 1 suggest that the mean level of stylistic atypicality is higher for those artists who introduced minor variations after some stylistic persistence (Cluster 2b), or that maintained persistence in the middle of their early career (Cluster 3b). Still, as the significant coefficient of the intercept reveals, artists with Cluster 1b-type trajectories are the ones that have the highest average stylistic atypicality.

[6] Atypicality gauges the extent to which a combination of categories (here, music styles) is unusual to the field. In our case, atypicality reflects how much unusual is the level of stylistic experimentation of an artist's overall music production. For the sake of brevity, we don't report here the formula to compute the atypicality measure, but invite the interested reader to refer to Goldberg, Hannan, and Kovács (2016), page 224.

Table 9. *OLS models ($DV_{ATY}$: average atypicality; $DV_{PERF}$: average performance)*

| | Model 1<br>$DV_{ATY}$ | Model 2<br>$DV_{PERF}$ |
|---|---|---|
| Style Cluster 2b | 0.101***<br>(0.021) | -0.066<br>(0.098) |
| Style Cluster 3b | 0.058**<br>(0.021) | -0.027<br>(0.098) |
| Style Cluster 4b | -0.021<br>(0.031) | -0.016<br>(0.144) |
| Artist Release Size | -0.000<br>(0.000) | 0.000<br>(0.000) |
| Artist Career Length | -0.002*<br>(0.001) | 0.003<br>(0.004) |
| Constant | 0.351***<br>(0.023) | 4.539***<br>(0.106) |
| N. Observations | 573 | 573 |
| $R^2$ | 0.069 | 0.003 |

Significance codes: *** p<0.000, ** p<0.01, * p<0.05

While these results offer only preliminary evidence, they nonetheless reveal the usefulness of SSA as a method to study creative trajectories. As we have sketched in the previous paragraphs, SSA is a highly versatile methodology. It does not isolate single events over a temporal continuum, but rather consider these events "in their continuity" (Aisenbrey and Fasang, 2010, p. 441), and the results of sequence resemblance analysis can then be used as variables in a wide number of analytical designs (Herndon and Lewis, 2015). Despite these strengths, SSA presents some requirements that, when unattended, can impede a meaningful use of the methodology. One recurring problem when analyzing the corpus of production of diverse artists is the different number of objects produced by each artist, which determines sequences of different length. Social sequence analysts have developed several ways to cope with sequences of unequal length, e.g., the spell-adjusted distances (Halpin 2010) used in the 2-case example. Other solutions are the normalized distance (Abbott and Hrycak 1990), the Time Warp Edited Distance (Marteau 2009) and the Localized OM (Hollister 2009). Another problem is created by temporality. If we were interested in the allocation of working time during an artist's day, we might want to account for the fact that working at 6:00am is different from working at 11:00am. Therefore, one could reasonably try to define substitution costs that take into account the time period at which the substitution operation intervenes. Although several options exist to deal with both element (state) and position (temporality), the most widely used approach is the dynamic Hamming distance proposed by Lesnard (2010).[7]

[7] Cornwell (2015) discusses analytical details and solutions to the limits of OM algorithm. We highly recommend his book for an advanced understanding on the applications of SSA.

**Conclusions**

This chapter offers a concise overview of the key methodological features of SSA and a few exemplary applications in studying career trajectories in the domain of creativity. Specifically, we examined stylistic variation sequences in the field of underground electronic music. One of the chief strengths of sequence analysis is its ability to directly measure sequence resemblance. It provides a way of examining the temporal dynamics of life course outcomes by highlighting common sequential patterns in the data. As we showed such patterns are then amenable to further use either as dependent or as independent variables, depending on whether our further questions take the form of "Why do certain kinds of actors end up with certain kinds of career trajectories?" or "Why do certain kinds of past creative trajectories tend to lead to differing creative outcomes in the future?" These features give sequence analysis a great breadth of applications. One can categorize life-event sequences to see whether certain sequences characteristically lead to creative outcomes. For instance, at what stage of their lives are great innovators (whether scientists or artists) most creative? While Paul Cézanne saw creativity as a unique, temporarily consistent process, the artistic trajectory of Pablo Picasso exhibited a much more uneven path as he continuously experimented with different styles. Recent evidence suggests that professionals with more erratic career sequences increase their brokerage opportunities (Kleinbaum 2012) by developing a broader range of contacts through their movement between disparate groups. Future sequence oriented studies could explore the creative benefits that accrue to professionals with erratic job trajectories. Could artists or scientists with increasingly erratic histories offer more creative solutions by virtue of their greater brokering opportunities? Scientists' trajectory of inventiveness is another area of inquiry amenable to sequence-analytic approaches. One could for instance codify scientists' inventive productivity over time according to the technological classes in which their patents fall (Fleming, 2001). By proxying each patent as a string of classes one could then use sequence methodologies to expose the technological trajectories produced by subsequent patents and thus infer individual, organizational or industry level trajectories of innovation. Sequence analysis provides a rich toolbox and useful techniques to discriminate among diverse creativity sequences, unveil events in an innovator's career history that reinforced a progression along an existing path or marked the emergence of new paths, and explore to what extent do indeed innovators differ as to when, where, and how their creativity reached its pick (Accominotti 2009). Nor does one lose, in this analysis, the full sequence information that is neglected in most methods based on stochastic modelling of individual transitions (Abbott 1995).

One could follow Stinchcombe's (1978, 13–16, 89–97) injunction to develop ideal-typical sequences of historical development and study the developmental trajectory of creative organizations, artistic movements or project networks. For example, in such creative industries as advertising, musical, video game and film, careers are seen as sequences of projects that "could be compared in order to understand the success of a director or an actor and how one moves from periphery to core" (Maoret, Massa, and Jones 2011, 439). Previous studies tracing innovators' journeys from the periphery to the core or from the core to the periphery of an existing field have highlighted how these different career trajectories shape creative outcomes and reflect career choices that are sometimes deliberate and sometimes the result of events or forces over which innovators have no or only little control. This complex dynamic emerges vividly from the analysis of Coco Chanel's progression from being an outsider located at the margins of the French society, to being consecrated as an iconic and acclaimed figure within the world of haute couture (high-end fashion) in Paris during (Cattani, Colucci, and Ferriani 2016); or Stanley Kubrick's trajectory from the margins to the center of Hollywood and then back to its margins (Cattani, Ferriani, and Colucci 2015). Researchers might also be interested in examining the extent to which certain creative trajectories explain other trajectories, such as in the case of an earlier sequence predicting a later pattern within the same sequence-for example, predicting the entry into a certain stylistic phase as a result of a particular sequence of experiences of the artist up to that point. In these or other such cases, the use of SSA would afford a more nuanced understanding of what is truly unique and similar in innovators' creative journeys, and the sources of this uniqueness and similarity.

The uncertainty that characterizes the production, reception, and consecration of creative work in general makes any attempt to study temporal patterns of creativity especially challenging. Creativity is not merely the result of sudden bursts of inspiration, but emerges instead more gradually over time as individual actors "pursue new paths, not limited by precedent and traditions, but in response to changing needs and changing contexts" (Buck, Lee, and Madermid 2002, 77). From this perspective, a systematic study of creativity entails tracing a person's work history, the reception of this work by relevant field audiences over time (e.g., peers, critics, and users), as well as the impact of contextual factors (e.g., new artistic trends) on the type of creative work that is being produced and the changing orientation of the audiences evaluating such work. With its emphasis toward thinking about ''events in context'' instead of ''entities with variable attributes,'' SSA provides an analytical framework that is of extreme utility in this regard (Aisenbrey and Fasang 2010). In short, we believe sequence analysis holds great promise for the study of creativity. Most of our understanding of creativity implies evolutionary or interactional theories. SSA contributes to those theories directly by facilitating the identification of patterns of social processes over time – an important precondition that often receives little attention – before turning to the question of which mechanisms produce them (Abbott 1995). What the field needs now are scholars committed to exploring the untapped potential of sequence methods for improving our understanding of the temporal dynamics of creativity.

**References**

Abbott, Andrew. 1988. "Transcending General Linear Reality." *Sociological Theory* 6: 169–186.
———. 1990. "A Primer on Sequence Methods." *Organization Science* 1 (4): 375–92.
———. 1995. "Sequence Analysis: New Methods for Old Ideas." *Annual Review of Sociology* 21: 93–113.
Abbott, Andrew, and Alexandra Hrycak. 1990. "Measuring Resemblance in Sequence Data: An Optimal Matching Analysis of Musicians' Careers." *American Journal of Sociology* 96: 144–85.
Accominotti, Fabien. 2009. "Creativity from Interaction: Artistic Movements and the Creativity Careers of Modern Painters." *Poetics* 37 (3): 267–294.
Aisenbrey, Silke, and Anette E. Fasang. 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): 420–62.
Althuizen, Niek, and Stoyan V. Sgourev. 2014. "Pièces de Résistance? Core and Casual Consumers' Valuations of Aesthetically Incongruent Artworks." *Psychology and Marketing* 31 (8): 604–14.
Anand, N., M. A. Peiperl, and M. B. Arthur. 2002. "Introducing Career Creativity." In *Career Creativity. Explorations in the Remaking of Work.*, edited by M. A. Peiperl, Michael B. Arthur, and N. Anand, 1–11. Oxford University Press.
Arthur, M. B., D. T. Hall, and B. S. Lawrence. 1989. "Generating New Directions in Career Theory: The Case for a Transdisciplinary Approach." In *Handbook of Career Theory*, edited by M. B. Arthur, D. T. Hall, and B. S. Lawrence, 7–25. Cambridge: Cambridge University Press.
Brzinsky-Fay, Christian. 2007. "Lost in Transition? Labour Market Entry Sequences of School Leavers in Europe." *European Sociological Review* 23 (4): 409–22.
Buck, Michelle L., Mary Dean Lee, and Shelley M. Madermid. 2002. "Designing Creative Careers and Creative Lives through Reduced-Load Work Arrangements." In *Career Creativity: Explorations in the Remaking of Work*, edited by M. A. Peiperl, Michael B. Arthur, R. Goffee, and N. Anand, 77–99. Oxford University Press: Oxford, UK.
Cattani, Gino, Mariachiara Colucci, and Simone Ferriani. 2016. "Chanel's Creative Trajectory in the Field of Fashion: The Optimal Network Structuration Strategy." In *Multidisciplinary Contributions to the Science of Creative Thinking*, edited by Giovanni Emanuele Corazza and Sergio Agnoli, 117–32. Springer, Singapore.
Cattani, Gino, Simone Ferriani, and Mariachiara Colucci. 2015. "Creativity in Social Networks. A Core-Periphery Perspective." In *The Oxford Handbook of Creative Industries*, edited by Candace Jones, Mark Lorenzen, and Jonathan Sapsed, Part II. Oxford University Press.
Caves, Richard E. 2000. *Creative Industries: Contracts between Art and Commerce.* Cambridge, MA: Harvard University Press.
Chan, Tak Wing. 1995. "Optimal Matching Analysis: A Methodological Note on Studying Career Mobility." *Work and Occupations* 22 (4): 467–90.
Cornwell, Benjamin. 2015. *Social Sequence Analysis. Methods and Applications.* New York: Cambridge University Press.
Csikszentmihalyi, Mihaly. 1999. "16 Implications of a Systems Perspective for the Study of Creativity." In *Handbook of Creativity*, 313–35. Cambridge University Press.
Dormandy, T. 2000. *Old Masters: Great Artists in Old Age.* New York: Hambledon and London.
Elzinga, Cees H. 2003. "Sequence Similarity: A Non-Aligning Technique." *Sociological Methods & Research* 32 (3): 3–29.
Fleming, L. 2001. Recombinant uncertainty in technological search. *Management Science* 47: 117-132.
Formilan, Giovanni, and Cristina Boari. 2018. "Do You Note Me? Social and Cognitive Dimensions of Categorization in the Evaluation of Category-Spanning Creative Products." In *Organizing in the Shadow of Power: Voices from the Italian Community of Organization Studies.*, edited by S. Consiglio, G. Mangia, M. Martinez, R. Mercurio, and L. Moschera. Studi MOA, Roma: Minerva Bancaria.
Gabadinho, A., G. Ritschard, N. S. Müller, and M. Studer. 2011. "Analyzing and Visualizing State Sequences in R with TraMineR." *Journal of Statistical Software* 40 (4): 1–37.
Gardner, H. 1993. *Creating Minds: An Anatomy of Creativity Seen through the Lives of Freud, Einstein, Picasso, Stravinskv, Eliot, Graham, and Gandhi.* New York: Basic Books.
Goldberg, Amir, Michael T. Hannan, and Balázs Kovács. 2016. "What Does It Mean to Span Cultural Boundaries? Variety and Atypicality in Cultural Consumption." *American Sociological Review* 81 (2): 215–41.
Gregan-Paxton, Jennifer, Steve Hoeffler, and Min Zhao. 2005. "When Categorization Is Ambiguous: Factors That Facilitate the Use of a Multiple Category Inference Strategy." *Journal of Consumer Psychology* 15 (2): 127–40.
Halpin, Brendan. 2010. "Optimal Matching Analysis and Life-Course Data: The Importance of Duration." *Sociological Methods & Research* 38: 365–388.
Han, Shin-Kap, and Phyllis Moen. 1999. "Clocking Out: Temporal Patterning of Retirement." *American Journal of Sociology* 105 (1): 191–236.
Herndon, B., & Lewis, K. (2015). Applying sequence methods to the study of team temporal dynamics. Organizational Psychology Review, 5: 318–332.
Hollister, Matissa. 2009. "Is Optimal Matching Suboptimal?" *Sociological Methods & Research* 38: 235–

264.
Hosmer, D. W., and Lemeshow S. 1999. *Applied Survival Analysis, Regression Modeling of Time to Event Data.* John Wiley & Sons, New York.
Kaufman, L., and P. J. Rousseeuw. 2005. *Finding Groups in Data.* John Wiley & Sons, Hoboken.
Kleinbaum, A. M. 2012. "Organizational Misfits and the Origins of Brokerage in Intra-Firm Networks." *Administrative Science Quarterly* 57 (4): 1–45.
Kovács, Balázs, and Michael T. Hannan. 2015. "Conceptual Spaces and the Consequences of Category Spanning." *Sociological Science* 2: 252–86.
Kozbelt, A., and Y. Durmysheva. 2007. "Lifespan Creativity in a Non-Western Artistic Tradition: A Study of Japanese Ukiyo-e Printmakers." *International Journal of Aging and Human Development* 65 (1): 23–51.
Lesnard, Laurent. 2010. "Cost Setting in Optimal Matching to Uncover Contemporaneous Socio-Temporal Patterns." *Sociological Methods & Research* 38: 389–419.
Leung, Ming D. 2014. "Dilettante or Renaissance Person? How the Order of Job Experiences Affects Hiring in an External Labor Market." *American Sociological Review* 79 (1): 136–58.
Maechler, M., P. Rousseeuw, A. Struyf, M. Hubert, and K. Hornik. 2018. *Cluster: Cluster Analysis Basics and Extensions.* R package version 2.0.7-1.
Maoret, Massimo, Felipe G. Massa, and Candace Jones. 2011. "Toward a Projects as Events Perspective." In *Advances in Strategic Management 28*, edited by Gino Cattani, Simone Ferriani, F. Täube, and L. Frederiksen, 427–44.
Marteau, P-F. 2009. "Time Warp Edit Distance with Stiffness Adjustment for Time Series Matching." *IEEE Transactions on Pattern Analysis and Machine Intelligence* 31 (2): 306–318.
Mayer, Karl Ulrich, and Nancy Tuma, eds. 1990. *Event History Analysis and Life Course Research.* Madison: University of Wisconsin Press.
R Core Team. 2013. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria. http://www.r-project.org/.
Ragin, Charles C. 1987. *The Comparative Method: Moving beyond Qualitative and Quantitative Strategies.* Berkeley and Los Angeles: University of California Press.
Reynolds, Simon. 1998. *Generation Ecstasy: Into the World of Techno and Rave Culture*. Psychology Press.
Sankoff, D., and J. B. Kruskal. 1983. *Time Warps, String Edits, and Macromolecules.* Reading, Massachusetts: Addison-Wesley.
Simonton, Dean Keith. 1997. "Creative Productivity: A Predictive and Explanatory Model of Career Trajectories and Landmarks." *Psychological Review* 104 (1): 66–89.
Stark, David, and Balàzs Vedres. 2006. "Social Times of Network Spaces: Network Sequences and Foreign Investment in Hungary." *American Journal of Sociology* 111 (5): 1367–1411.
Stinchcombe, A. L. 1978. *Theoretical Methods in Social History.* New York: Academic Press.
Stovel, Katherine, and Marc Bolan. 2004. "Residential Trajectories: Using Optimal Alignment to Reveal the Structure of Residential Mobility." *Sociological Methods & Research* 32 (4): 559–98.
Stovel, Katherine, Michael Savage, and Peter Bearman. 1996. "Ascription Into Achievement: Models of Career Systems at Lloyds Bank 1890-1970." *American Journal of Sociology* 102 (2): 358–99.
Vany, Arthur De, and David W. Walls. 1996. "Bose-Einstein Dynamics and Adaptive Contracting in the Motion Picture Industry." *The Economic Journal* 439 (106): 1493–1514.
Waltham-Smith, Naomi. 2015. "Sequence." In *The Oxford Handbook of Critical Concepts in Music Theory*, edited by Alexander Rehding and Steven Rings. Oxford University Press.
Ward, Joe H. Jr. 1963. "Hierarchical Grouping to Optimize an Objective Function." *Journal of the American Statistical Association* 301: 236–44.
White, Harrison C. 1993. *Careers and Creativity: Social Forces in the Arts.* Westview Press.
Yamaguchi, K. 1991. *Event History Analysis*. ASRM 28. Sage, Newbury Park and London.