Using optimal matching analysis in sociology: Cost setting and sociology of time

Laurent Lesnard

To cite this version:

Laurent Lesnard. Using optimal matching analysis in sociology: Cost setting and sociology of time. 2014. hal-00973029

HAL Id: hal-00973029
https://sciencespo.hal.science/hal-00973029

Preprint submitted on 3 Apr 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

SciencesPo. | OSC CNRS

Using Optimal Matching Analysis in Sociology: Cost Setting and Sociology of Time

Laurent Lesnard

Notes & Documents

n° 2014-01 Février 2014

Résumé :

Ce texte est une réflexion sur les conditions sous lesquelles les méthodes d'appariement optimal (*Optimal Matching Analysis, OMA)* peuvent être utilisées en sociologie. Le succès de ces méthodes en biologie ne tient pas à une compatibilité particulière avec des processus biologiques mais plutôt à la manière dont des hypothèses théoriques en biologie sont mobilisées pour déterminer les coûts relatifs à ces méthodes d'appariement. Étant donné que les séquences en sociologie sont le plus souvent composées d'événements et de temps, la détermination des coûts dans les OMA appliqués en sociologie devrait être guidée par des éléments théoriques en sociologie du temps. Après une discussion de la signification et des conséquences des coûts en OMA, ce texte revient sur des éléments théoriques émanant de la sociologie du temps (Durkheim, Elias et Bourdieu) et montre comment on peut les intégrer dans une variante des méthodes d'appariement optimal appelée *Dynamic Hamming*. À ce titre, ce texte illustre comment la théorie sociologique peut aider à la détermination des coûts lorsque les méthodes d'appariement optimal sont utilisées en sociologie.

Pour citer ce document :

Lesnard, Laurent (2014). « Using Optimal Matching Analysis in Sociology: Cost Setting and Sociology of Time », Notes & Documents, 2014-01, Paris, OSC, Sciences Po/CNRS.

Pour une version électronique de ce document de travail et des autres numéros des Notes & Documents de http://www.sciencespo.fr/osc/content/notes-documents-de-l-osc

Abstract:

This paper is a reflection on the conditions required to use Optimal Matching Analysis (OMA) in sociology. The success of OMA in biology is not related to any supposed similarity of the method with biological processes but comes from setting costs in OMA in accordance with biological theory. As sequences in sociology are made of events and time, the determination of costs should be guided by sociological theories of time. After a discussion of the sociological meaning and consequences of costs, this paper comes back on the Dynamic Hamming Distance and the body of social theories of time (Durkheim, Elias, Bourdieu) from which it is derived as an example of how sociological theory can inform cost setting in using OMA in sociology.

Readers wishing to cite this document are asked to use the following form of words:

Lesnard, Laurent (2014). "Using Optimal Matching Analysis in Sociology: Cost Setting and Sociology of Time", Notes & Documents, 2014-01, Paris, OSC, Sciences Po/CNRS.

For an on-line version of this working paper and others in the series, please visit the OSC website at: http://www.sciencespo.fr/osc/content/notes-documents-de-l-osc

1. Optimal Matching Analysis and Biology

As with any new statistical method in the social sciences (Desrosières 1993) the use of Optimal Matching Analysis (OMA) was much debated at the end of the 1990s, especially with regard to the sociological interpretation of transformations and the role of costs in the findings (Levine 2000; Wu 2000). OMA was widely used in biology, where the three operations (insertion, deletion and substitution) supposedly are reflected in biochemical processes. Yet OMA was not created by biologists, but rather originated from research in computer science that Richard Hamming and Vladimir Levenshtein conducted in the 1950s and '60s (Hamming 1950; Levenshtein 1966). In OMA terms, Hamming's distance only uses substitution operations whereas the first distance put forward by Levenshtein uses all three operations, each with a unit cost (see Table 1). Thus the operations used in OMA have nothing to do with biochemical transformations; many of the latter, such as the transposition of sub-sequences, would be missing (Abbott 2000). While it is tempting to attribute the method's success in biology to these similarities, it is in fact through the interpretation and determination of costs that biologists were able to adapt OMA to their data.

Table 1 – Hamming and Levenshtein Distances (Lesnard 2010)

	Transformations used	
Distance	Substitution	Insertion and deletion
Hamming	Yes (cost = 1)	No
Levenshtein I	Yes (cost = 1)	Yes (cost = 1)
Levenshtein II	No	Yes (cost = 1)

The applications of OMA in biology are quite different from its use in the social sciences since in the former case the goal is not to create typologies, but more often to match a group of sequences with unknown properties to other sequences with known properties (Durbin et al. 1998). Substitution costs are determined on the basis of phylogenetic hypotheses and statistical results, while insertion and deletion costs are set more arbitrarily. Indeed, substitution costs must reflect the probability of change and are empirically estimated from sample sequences with established or probable phylogenetic links. For example, the PAM (Point Accepted Mutation) matrixes are estimated on the basis of proteins that are experimentally or hypothetically related phylogenetically (Dayhoff et al. 1978).

2. The importance of costs

The main lesson that can be drawn from the use of OMA in biology is the importance of focusing on the interpretation and determination of costs rather than seeking an interpretation in the operations themselves. The interpretation of substitution costs is key in this respect since they need to reflect the proximity between the different states in which the sequences unfold. In biology this proximity between states is interpreted in phylogenetic terms and the costs are therefore derived from samples based on phylogenetic knowledge and hypotheses. In the social sciences this probability must express the degree of sociological proximity between states. But what is meant by sociological proximity?

It is precisely this issue that other critiques have targeted. In fact, the first researchers to use OMA in sociology often expressed doubts as to their choice of costs and the sensitivity of their results to these choices. As Katherine Stovel and her colleagues stated, “The assignment of transformation costs haunts all optimal matching analyses” (Stovel et al. 1996, p. 394). If the results are too sensitive to costs it means that the costs, and not the data, are producing the results (Wu 2000). Conversely, if the results are not sensitive to costs, how could OMA be valid (Levine 2000)?

2.1. The empirical consequences of costs

Before delving into the interpretation and determination of costs, it is worth noting that in instances where the analysed data shows temporal regularities, costs can only have minor effects on the results. This is hardly surprising. In terms of classification analysis these are instances of strong clusters or stable groups, that is, groups of very similar individuals that will end up being grouped together no matter what distances and classification methods are used. (Lebart et al. 2006, p. 254). Classification specialists generally seek to identify these strong clusters. By definition, a group of virtually identical sequences needs very few transformations to make them identical, and any assigned cost will have very little influence on this group. The greater the differences between the compared sequences the greater the importance of costs.

When the data include several strong clusters, costs are important at two levels. First, costs play a role in cases that do not fit into stable groups as easily as others do. Depending on their nature, costs assigned these atypical sequences to one or another of the strong clusters, and, if need be, to a more heterogeneous group including rare sequences. Second, costs also have an impact on the distance between the strong clusters and, in the case of an ascending hierarchical classification, affect the way in which they might be aggregated. Costs therefore play a role, albeit not always a clear one since it often involves variations in the number of sequences assigned to each stable group by the classification or partitioning algorithm.

Determining costs first requires understanding and interpreting them with regard to the processed data and purpose of the analysis. In the social sciences the order of states is not tied to a biochemical combination as it is in biology, but rather to time. When a state is inserted in, or deleted from a sequence to make it identical to another, a time shift occurs between these two sequences. The more these two operations are used, the more the elements to be compared will be distant in the original sequences.

Insertion and deletion operations therefore warp the position of the elements of the sequence. On the other hand, substitution operations respect the position of sequence elements, the trade-off being that states are replaced. Once insertion and deletion operations have been used substitution operations do not introduce additional shifts; however, they only recognise the shifted positions of the states to be substituted.

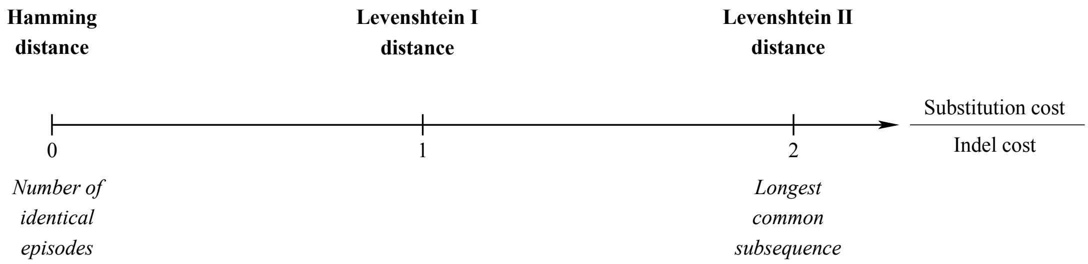

Figure 1 – Patterns corresponding to costs in Optimal Matching Analysis (Lesnard 2010)

In sum, insertions and deletions shift the sequences in order to identify identically coded sub-sequences, while substitutions do not shift the sequences but replace states with others. In the first case the time structure of the sequences is altered, and in the second the states of which they are made undergo change (Lesnard 2010). When the substitution cost is double the insertion-deletion cost or more, substitution operations are no longer used since it is equivalent or less costly to insert the desired state and delete the other than it is to substitute them (Kruskal 1983). Conversely, when substitution costs are very low compared to insertion-deletion costs, the latter operations are no longer used. As a result, any cost combination falls within the distances ranging between Levenshtein 2 and Hamming (see Figure 1). In the first case (Levenshtein 2) the algorithm seeks to shift the sequences in order to compare identically coded, but not necessarily contemporaneous sub-sequences. In the second, the algorithm only takes identical sub-sequences into account (that is, a series of identical states located in the same place in the sequences). The Levenshtein 1 distance is a sort of compromise between these two poles. The choice of costs thus depends on the time scale and its importance for the analysis. The greater the importance of respecting the time scale, the more imperative it is for costs to approach or reach the Hamming distance.

2.2. Multiple substitution costs

By definition, in a state space the constituent states are considered to be different and to refer to situations that are not, from an analytical viewpoint, of the same nature. Several substitution costs are generally used and either empirically or theoretically defined to reflect the fact that some states are considered to be closer than others. Thus, in their study of intergenerational social mobility, Halpin and Chan define substitution costs according to a nomenclature of social classes and their theoretical distance (Halpin and Chan 1998). Frequencies of transitions between states, calculated on the basis of all episodes, have also been used to define substitution costs. This solution is equivalent to assigning costs of substitution between states that decrease with higher levels of observed transitions between them. However, this practice has been challenged because substituting a state with another is theoretically not a transition and substitution costs must reflect the proximity of states (Abbott and Hrycak 1990; Halpin 2010).

The use of only one substitution cost therefore presumes that no state is a priori closer than another. Using a matrix of substitution costs introduces information on the closeness of states into the analysis. To set lower substitution costs for certain states consists of positing that these states are close but different, allowing for the grouping of these sequences if they are sufficiently similar. In other words, if the differences between two sequences only concern certain states with low substitution costs, then the total minimum cost will be low and they will show close general proximity. In the context of sequence comparison in the social sciences, this close proximity between states can be interpreted to mean that even if a difference appears in two sequences at a given moment this difference is almost negligible, and that if the differences are of this nature then the algorithm should consider them to be unimportant.

Indeed, a low substitution cost is only of interest if it is much lower than the combined cost of a deletion and an insertion. A low substitution cost will cause the algorithm not to shift the sequences to look for an identical sub-sequence, while a high cost will prompt a search for identical but shifted sub-sequences. Low substitution costs lead to the search for contemporaneous, identical sequences, thereby implying that the sequences are also close. *In other words, to claim that certain states are close is equivalent to positing that the presence of these states in two sequences is the marker of their proximity as a whole, and that they might belong to very similar social rhythms*. Conversely, if two sequences have two states with high substitution costs the implication is that the sequences are distant and that it is more useful to look for identical but shifted sequences. This also holds true in biology since substitution costs between two elements of a DNA sequence decrease the closer they are phylogenetically – a proximity that extends to all the sequences in which they appear.

In the time frame characterizing the use of OMA in the social sciences, substitution costs also relate to the whole sequence: does the presence of two different states point to two very different social rhythms? With a single substitution cost there is no reason to presume the answer to this question and the facts can speak for themselves: social rhythms appear in the typology. A matrix of substitution costs suggests the existence, be it theoretical or empirical, of certain social rhythms – combinations of states that signal closer proximity. Typologies created with a matrix of substitution costs present two types of groups. The first includes sequences where some parts are close but time-shifted while the second includes sequences where identical or very similar sub-sequences are contemporaneous.

More precisely, since the sequences that are not part of a strong cluster are most affected by costs, the sequences that are most difficult to classify will be assigned to the strong cluster to which they come closest. A matrix of substitution costs will then adjust this allocation by emphasizing either the closest contemporaneous sub-sequences (low substitution costs) or shifted identical sub-sequences (high substitution costs). Is it legitimate to assign the sequences to strong clusters in different ways, according to the proximity defined in the matrix of substitution costs? Rare sequences are, by definition, difficult to classify, and given that the goal of taxonomy is to classify there are only two possible solutions: either assign them to strong clusters or relegate them all to an unclassifiable category.

There is a need here to distinguish between rare sequences with sub-sequences that approach strong clusters, and very rare sequences that are difficult to move toward a stable group. Empirically, what most often appears is a catchall category that cannot be interpreted. So the question is whether it is desirable to assign rare sequences that share several characteristics with strong clusters. If the goal is to find the most homogeneous groups (outside of the one that brings together the unclassifiable sequences), then all of the costs must be high in order to increase the distance between them and the rare sequences. If less homogeneous classes are acceptable, then a matrix of substitution costs allows for more flexibility in managing allocations to stable groups.

From this perspective transitions provide partial information on social rhythms and it therefore seems reasonable to use this information; indeed, the use of a low substitution cost in the social sciences implies that the difference between the two states does not indicate that the sequences belong to two different groups of rhythms. Using transitions to determine substitution costs is equivalent to substituting diachronic proximity for synchronic proximity, which recognizes that certain states link together more often than others over time, pointing to a greater proximity of these states. However, using transitions between states to set substitution costs raises the issue of circularity because transitions are generally calculated from sequences that are then analysed with OMA. Insofar as this method is descriptive it is not really a problem, especially since transitions are aggregated while sequences are individual.

3. Optimal matching methods and social theory of time

A large portion of my research using OMA focuses on time use surveys (Lesnard 2008; Lesnard 2009; Lesnard 2010; Lesnard and Kan 2011). These surveys feature a time diary in which respondents must describe activities over one or several days. The temporal dimension of the sequences derived from the time use surveys is therefore very strong. In this respect the analysis of time use through OMA aims to go beyond the so-called “time budget” approach, which reduces time uses to durations, by reintroducing chronology into the study of daily lives. It is therefore necessary to limit time distortions and curb the use of insertion and deletion operations that emphasize identical sub-sequences – that is, in terms of time use, the time-consuming identical activities that take place at different times of the day – and thereby focus on duration rather than timing.

This is also true for other sequences that include a time dimension: when shifts are emphasized the similarities highlighted are series of identical states, meaning the duration of identical subsequences is emphasized to the detriment of timing. Sequences in the social sciences most often have a time dimension and the goal of OMA is precisely to underscore possible temporal regularities. The risk in intensively using insertions and deletions is to obtain results that are highly dependent on similarities in duration to the detriment of similarities in timing. It is perfectly legitimate to value duration over timing, but in this case solutions that are simpler and faster than OMA should be favoured, such as a cluster analysis directly and exclusively applied to duration.

Thus the analysis of time use surveys using OMA reveals more general problems with the use of this type of method on sequences in the social sciences; these are especially important in the case of time use surveys. Indeed, in the even more specific case of research on the work hours of individuals and couples, where the point of using OMA is to identify different types of distributions of work hours, the warping that any insertion or deletion introduces into the subject of study itself is particularly problematic (Lesnard 2008). Consequently, only substitution operations should be used. The state space is highly simplified: two for individual work hours (work and non-work), four for couples (nobody works, the two spouses work, one of the spouses works but the other does not and vice-versa). The issue I faced was whether to use one cost – Hamming's distance – or several. The Hamming distance effectively amounts to counting the number of discordant pairs in two sequences: any gap, even if it is minimal, increases the distance. Even though insertion-deletion operations are undesirable, I felt it was necessary to somewhat mitigate the radicalism of this parameterisation.

To this end I developed a variant of this distance: the dynamic Hamming distance (DHD) (Lesnard 2010). Like the Hamming distance, DHD only uses substitution operations in order to best preserve the timing of the sequences. Its specificity is in the use of costs, which depend on transitions between states and which change over time. The distance between states varies over time and decreases when the transitions before and after the considered date are strong between the state and the one for which it is to be substituted.

For example, in the 1998-99 time use survey the transition from no work to work accounted for 22% of the working population between 8:00 and 8:10 but only 3% between 10:40 and 10:50. With DHD the two states would be considered close at 8:00 and far at 10:40. For the vast majority of employed people the early morning is the beginning of the workday, so it would not be very costly to make identical two sequences that contain different states, given that differences at that time of day are not important. Mid-morning, however, such differences indicate different working rhythms that will be reflected in the distances between their respective sequences. DHD thus takes limited gaps into account without resorting to insertions and deletions that would shift the sequences.

The transitions between the different possible states at any given moment illuminate certain aspects of social rhythms. However, the transition matrixes only provide macro-social information by date without links. With OMA these matrixes can be expressed in terms of distances between individual sequences, allowing for the creation of taxonomy. In other words, by integrating transitions between states in OMA, DHD individualizes them and connects them to one another in order to better identify the different social rhythms that underlie the aggregate figures of transition matrixes.

An arithmetically identical number of years (or minutes) may refer to substantial social differences. This is a classical result from the sociology of time pioneered by Hubert. Using his research on the links between time, magic and religion, Hubert showed that the calendar is underpinned not by a quantitative time, but rather a qualitative time made of discontinuous and heterogeneous episodes (Hubert 1905). The linearity of calendars and clocks in no way implies that time is linear. By considering sequences in their entirety, OMA, and DHD in particular, help place the various elements into their context and thus restore this heterogeneity.

3.1. Sociology of time

The process of deriving substitution and transition costs and allowing them to vary over time also needs to be linked to the sociology of time program initiated by Hubert and Durkheim and further developed by Norbert Elias and Pierre Bourdieu. Building on Hubert's seminal work, Durkheim outlined an agenda to develop the sociology of knowledge. For Durkheim categories of understanding are not transcendental (that is to say innate, constitutive of human nature) but rather derive from the organization of life in society (Durkheim 1912). In an analysis of ethnographic writings and materials on the Aboriginal, Durkheim showed that time in these societies is a religious symbol that separates the sacred from the profane, forming two homogeneous, alternating periods of time that punctuate the Aboriginal calendar. This alternation also reflects the ebb and flow of community life: during sacred times the group gathers, whereas profane times are more isolating, focused on individual and family survival. The homogeneity of, and conceptual difference between the two times thus refer to distinct collective practices and temporalities, or relations to time: in other words the time category of understanding is interdependent with the Aboriginals' binary social rhythm.

One of the consequences of this first cornerstone of Durkheim's sociology of time is that a "calendar expresses the rhythm of collective activity while ensuring that regularity" (Durkheim 1912, p. 15). The history of calendars and other systems for determining time is also that of social rhythms and of temporality. Indeed, as Pierre Bourdieu put it, time is a symbolic system that is structured – by the collective rhythm, embodied by religion in less differentiated societies – and structuring – as a temporality or form of temporal knowledge (Bourdieu 2001, p. 201). Time, and more generally all symbolic systems, follows "logical conformism" (Bourdieu 2001, p. 204), meaning that it both accepts and validates the social order. A system for determining time, as a symbol, cannot be used if its use is not known and recognized by the whole group. In less differentiated societies that consist of small autonomous communities with a limited division of labour and somewhat interchangeable members, given that identity is not very individualised but rather dominated by the group, social order is based on respect for the group and its rhythm, through religion. In these societies, to follow the collective rhythm is to alternate between the sacred and the profane, and between collective time and more individual time.

Calendars and, in general, any system to determine time such as the clock, were not intended as time measuring instruments, but as a means to mark time: "the institution of calendars is not intended solely, and probably not even primarily, to measure time as a quantity. It does not derive from the idea of a purely quantitative time, but rather the idea of a qualitative time that is made of discontinuous and heterogeneous parts and that constantly turns on itself (Hubert 1905). Calendars crystallize and stabilize collective rhythms and activities (Durkheim 1912, p. 15): "Units of time are not units of measurement but of a rhythm where oscillation between alternatives periodically leads back to the similar" (Hubert 1905).

It is therefore the tools developed in relation to the social symbol of time that bear the traces of shared transformations in the organization of society and time, that is, of collective rhythms and temporalities. To paraphrase Hubert, the history of calendars reflects the evolution in the "code of the qualities of time”[1], the direction and degree of diversification of collective life and also the “level of synthesis”, as defined by Elias (Elias 1991), of the symbol of time.

In this respect, the appearance of bell clocks in European towns in the Middle Ages was a key step in the process of social and temporal differentiation. The different rhythms of urban collective life in the Middle Ages[2] created synchrony issues throughout the day that were resolved by using multiple acoustic and optical signals: bands, flags, pennants, and especially bells. At first canonical hours – some marked by ringing church bells – were used as time references in several areas of collective life: for example, the market’s opening coincided with the bell sounding the end of mass, while the end of the workday for day labourers was marked by the bell for the compline.

Beginning in the 14th century the development of cities created the need for a wider variety of time signals that church bells alone could no longer handle. Cities decided to acquire bells to convene meetings, indicate payment due dates for interest and taxes, assemble each neighbourhood, set market hours, etc. This inflation of urban signals culminated with work bells that set working hours for salaried workers. The multiplication of urban signals also reflected the increasing division of labour since a number of them were no longer aimed at the community as a whole, but rather at particular groups (employees, members of various councils, travellers, merchants, etc.).

The increasing need to synchronize certain groups fed into a complex system of acoustic and optical signals in big cities that probably reached its climax at the beginning of the 14th century. Growth in the number of acoustic signals could in large part be attributed to the coordination needs of certain groups, as opposed to the urban communities as a whole. As long as collective life was punctuated by a series of events involving the entire community, a single signal could ensure coordination, as was the case in the sounding of canonical hours. The proliferation of bells testified to the division of labour, and therefore the pluralisation of time: in the towns of the Middle Ages, time was no longer alternating between the sacred and the profane, but differentiating with the development of functional interdependence. However, when the bells started sounding the temporal rhythms for certain segments of the population in addition to those for the community as a whole, they became less effective at finely synchronising sub-groups of the population. Today, it is hard to imagine the background noise that the bell signals created on a daily basis: “Everyday life was temporally structured through and through by bell signals, with hardly a day resembling another. Apart from signals for proclamations, prohibitions and ordinances, the inhabitants of the city also received a wealth of acoustic information about important public civic events” (Dohrn-van Rossum 1992, p. 217).

---

[1] “The calendar is the periodic order of rites. Its history also teaches us that it is the code of the qualities of time. The first calendars were almanacs that recorded, day by day, magico-religious forecasts and prescriptions” (Hubert 1905).

[2] Meetings of bourgeois councils or courts to deal with the affairs of the town, market gatherings, the beginning and end of work for day labourers, the opening and closing of city gates, and more unusual events such as gatherings to respond to threats of fire or of war (Landes 2000, p. 76; Dohrn-van Rossum 1992, p. 206).

The introduction of bell clocks in the 14th century unified and simplified the various acoustic signals. The clock bells did not chime for anyone in particular. Rather, different groups or segments of the population could use the signal to synchronize. Because clocks can only ring for *equal* hours without introducing too many technical complications, the system of canonical hours, and attendant hold of religion over time, collapsed and liberated other collective times from their ties to religious time (Landes 2000, p. 81).

Bell clocks substituted a regular and neutral tempo for the profusion of specific and muddled signals. However, hours did not mark any particular event; it was up to the local community's members to associate them to collective or private activities they wished, or were required to attend. Before bell clocks were introduced the relationship to time was theoretically still largely an external constraint that elicited poor anticipation. Everyone knew that it was time for a lunch break when they heard the work bell ring a second time. One only needed to recognize the sound of different bells to situate oneself in time. With bell clocks hours had to be linked to events: time constraints were no longer explicitly marked by a particular bell, but rather implicitly signalled. Each hour chime only evoked the constraints that one chose to associate with it.

Specific time constraints therefore had to be internalized by each individual. The introduction of the bell clock marked the beginning of the internalization of time constraints, or at least accelerated this process. With the clock striking hours, people had to get into the habit of adjusting their behaviour in response to the signal. The regularity of hours also allowed for a greater anticipation of the day's events, whereas the previous system was characterized by very low predictability. Thus the new system provided a greater opportunity to regulate one's behaviour, that is, to self-regulate. *Paradoxically the homogeneity of the movement of the clock's hands was precisely the sign of the heterogeneity of time*: autonomous social areas emerged along with their particular collective rhythms.

In the everyday life of individuals involved in a number of fields, time at the individual level alternated between the different times in various social fields (Halbwachs 1947, p. 167; Bourdieu 2003, p. 202). This meant that the rhythm of daily life was determined in large part by the rhythms associated with positions held in different social fields. Sequences of participation in social fields could come into conflict due to inconsistences in the rhythms associated with successively held positions.

Over the longer run, life courses can also be analysed as alternating participation in the principal social fields, which are the economic field and the family field (Bourdieu 1994, p. 135-45). In contemporary societies adult status is achieved through economic independence from one's original family and the formation of a so-called childbearing family. It therefore involves two forms of integration, in the productive and in the married and parental worlds, and is generally marked by events such as the end of studies, the first job, leaving the parental home, the first relationship and/or the first marriage and the birth of a first child (Modell et al. 1976). The study of early adulthood actually comes down to analysing access to the economic fields (directly or indirectly through marriage) and family field (with a new status, through the formation of a so-called childbearing family), as well as their interlinks.

A social field's time has two dimensions since it is both the rhythm of collective life for the group of agents involved, and a relationship to time and to the future in the field (temporality). Using all the transitions as substitution costs is a way to break down the average rhythm they represent into homogeneous collective sub-rhythms, and thereby describe in detail the collective life of the field under consideration. With DHD the structure of the state space (distance between the states) is no longer static but bends to the rhythm of collective life. In theory this adaptation of Hamming's distance goes further since the temporality of one field flows from the field's repeated exposure to the collective rhythm. The use of transitions allows for collective rhythms to be described in ways that make sense in terms of a field's temporality. The differences between states, which are variable over time, have a meaning for the agents in the given field. To work or not at 9:00 AM is not socially very significant, while the opposite would be true for 11:00 PM. Similarly, to transition from unemployment to employment at age 20 or 25 does not have the same significance as a first entry into the labour market at age 30.

This is what makes OMA and DHD in particular a type of descriptive method that is especially well suited for sequences in the social sciences. The mean is also descriptive and could be used to describe these sequences, but it does not take into account the heterogeneity of time and therefore imposes the linear framework of the clock on time; as a result, the collective rhythms disappear. DHD is a method for describing sequences that incorporates the heterogeneity of time in its very design, drawing on the sociological theories of Durkheim, Elias and Bourdieu. It includes key elements from each of these authors. DHD centres on the Durkheim's dual concept of time as it was reworked in the research of Elias and Bourdieu on social differentiation and the concept of fields.

Strictly speaking, the use of DHD implies the adoption of this general theoretical framework. This means considering the state space – a cornerstone of OMA that is often little discussed – in this framework. Discussion on the use of OMA in the social sciences has focused on transformation and cost issues and has relatively ignored the state space, which, unlike in biology and computer science, is not given but rather is a matter of choice. Field theory can help guide this choice: the state space must represent the simplified structure of a field. The different states should be able to capture the collective rhythm of the field or at least one of its aspects.

More generally, the inclusion of DHD in the theoretical basis of the sociology of time illustrates the links that can be made between theory and optimal matching methods. In the same way that regression models cast the functioning of society in terms of a "general linear reality" that is at odds with most sociological theories (Abbott 1988), the choice of costs for OMA is never methodologically neutral. In particular, the use of insertions and deletions focuses the analysis on duration and therefore ignores similarities in timing but also more generally the heterogeneity of time. This may be perfectly legitimate, even if in this case it would be easier to use a simple cluster analysis on the durations, but it must be justified. Although the effect of costs on the results is often subtle, use of OMA in sociology must be made in a rigorous methodological framework so that it no longer appears as the black box disparaged by some of its critics.

REFERENCES

Abbott, Andrew. 1988. "Transcending general linear reality." *Sociological Theory* 6 (2): 169–86.

—. 2000. "Reply to Levine and Wu." *Sociological Methods and Research* 29 65–76.

Abbott, Andrew, and Alexandra Hrycak. 1990. "Measuring resemblance in sequence analysis: an optimal matching analysis of musicians careers." *American Journal of Sociology* 96 (1): 144–85.

Bourdieu, Pierre. 1994. *Raisons pratiques : Sur la théorie de l'action*. Paris: Seuil.

—. 2001. *Langage et pouvoir symbolique*. Paris: Seuil.

—. 2003. *Méditations pascaliennes*. Paris: Seuil.

Dayhoff, Margaret O., Robert M. Schwartz, and B. C. Orcutt. 1978. "A model of evolutionary change in proteins." *Atlas of Protein Sequence and Structure* 5 (Supplement 3): 345–52.

Desrosières, Alain. 1993. *La politique des grands nombres. Histoire de la raison statistique*. Paris: La Découverte.

Dohrn-van Rossum, Gerhard. 1992. *Die Geschichte der Stunde*. Munich: C. Hanser.

Durbin, Richard, Sean R. Eddy, Anders Krogh, and Graeme Mitchison. 1998. *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge (UK), New York: Cambridge University Press.

Durkheim, Émile. 1912. *Les formes élémentaires de la vie religieuse*. Paris: Alcan.

Elias, Norbert. 1991. *The Symbol Theory*. London, Thousand Oaks, New Delhi: Sage.

Halbwachs, Maurice. 1947. "La mémoire collective et le temps." *Cahiers internationaux de sociologie* 3–31.

Halpin, Brendan. 2010. "Optimal matching analysis and life course data: The importance of duration." *Sociological Methods & Research* 38 (3): 365–88.

Halpin, Brendan, and Tak Wing Chan. 1998. "Class careers as sequences: An optimal matching analysis of work-life histories." *European Sociological Review* 14 (2): 111–30.

Hamming, Richard W. 1950. "Error-detecting and error-correcting codes." *Bell System Technical Journal* 29 (2): 147–60.

Hubert, Henri. 1905. "Étude sommaire de la représentation du temps dans la religion et la magie." *Annuaire de l'École Pratique des Hautes Études, section des sciences religieuses* 1–39.

Kruskal, Joseph B. 1983. "An overview of sequence comparison." *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison* 1–44.

Landes, David S. 2000. *Revolution in Time. Clocks and the Making of the Modern World*. Cambridge (MA): The Belknap Press of Harvard University Press.

Lebart, Ludovic, Marie Piron, and Alain Morineau. 2006. *Statistique exploratoire multidimensionnelle : visualisation et inférence en fouilles de données*. Paris: Dunod.

Lesnard, Laurent. 2008. "Off-scheduling within dual-earner couples: An unequal and negative externality for family time." *American Journal of Sociology* 114 (2): 447–90.

—. 2009. *La Famille désarticulée. Les nouvelles contraintes de l'emploi du temps*. Paris: PUF.

—. 2010. “Cost setting in optimal matching to uncover contemporaneous socio-temporal patterns.” *Sociological Methods & Research* 38 (3): 389–419.

Lesnard, Laurent, and Man Yee Kan. 2011. “Investigating scheduling of work: A two-stage optimal matching analysis of workdays and workweeks.” *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 174 (2): 349–68.

Levenshtein, Vladimir I. 1966. “Binary codes capable of correcting deletions, insertions, and reversals.” *Soviet Physics Doklady* 10 (8): 707–10.

Levine, Joel H. 2000. “But what have you done for us lately?: Commentary on Abbot and Tsay.” *Sociological Methods and Research* 29 (1): 34–40.

Modell, John, Franck F. Furstenberg, and Theodore Hershberg. 1976. “Social change and transitions to adulthood in historical perspective.” *Journal of Family History* 1 (1): 7–32.

Stovel, Katherine, Michael Savage, and Peter Bearman. 1996. “Ascription into achievement: Models of career systems at Lloyds Bank, 1890-1970.” *American Journal of Sociology* 107 (2): 358–99.

Wu, Lawrence L. 2000. “Some comments on “Sequences analysis and optimal matching methods in sociology: Review and prospects”.” *Sociological Research and Methods* 29 (1): 41–64.