New Developments: Sequence Analysis

Ignace Glorieux
Vrije Universiteit Brussel – Research Group TOR

Time Use Data Access System Workshop
Morrill Hall and LeFrak Hall University of Maryland
June 29-July 1, 2016

Vrije
Universiteit
Brussel

Parameters of social time

- Duration – How long?
- Tempo – How much?
- Timing – When?
- Sequence – In what order?

In time-use studies mostly only durations are studied intensively: durations are added, subtracted, … just like social time is a homogeneous flux as conceptualized in Newtonian time in natural sciences

Vrije Universiteit Brussel

Social time

- The flow of the day is NOT a succession of identical moments
- The ‘quality’ of time is related to the parameters of time
- Time-use data provide a wealth of details (context) … we need statistical techniques to deal with this complexity and to do justice to the ‘social’ quality of time

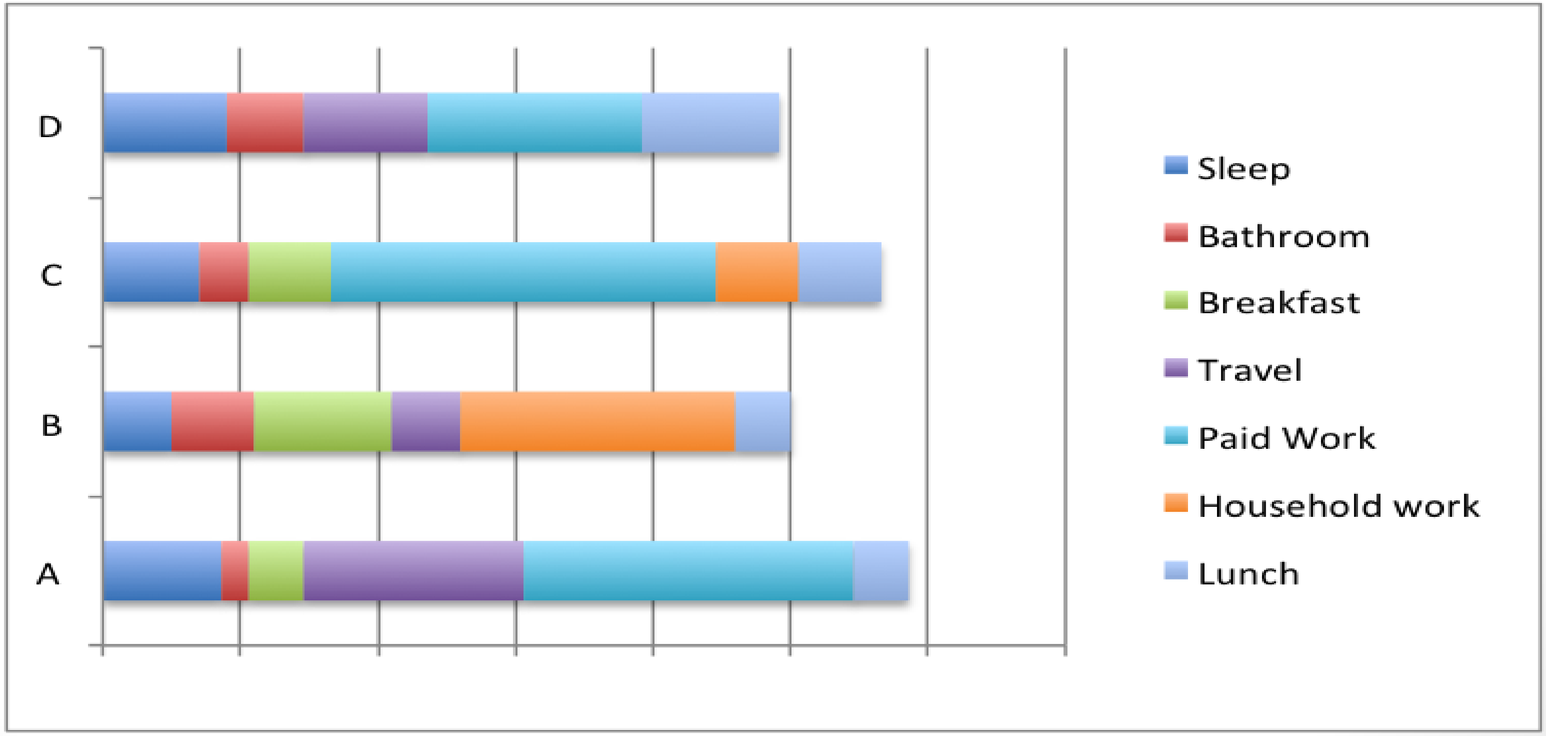
Optimal Matching Analysis

- Optimal Matching Analysis (OMA) comes from molecular biology and was aimed at the decryption of DNA sequences
- The technique was introduced in the social science by Andrew Abbott …
- ... and introduced in time-use research bij Laurent Lesnard (2004, 2006, 2008).
- The first part of this presentation is mainly based on Lesnard 2004

Vrije Universiteit Brussel

Main idea

- OMA is basically an algorithm that calculates a distance matrix between a set of sequences



OMA in a nutshell

- OMA is a way to measure the degree of dissimilarity between any pair of sequences
- In OMA, the dissimilarity between sequences is the cost required to make the two sequences identical with the help of **three basis operations**: *in*sertion, *del*etions (*indel* operations) and substitution
- Each operation is associated with a cost and the dissimilarity produced by OMA is the minimum total cost required to match two sequences

Vrije Universiteit Brussel

OMA in a nutshell

Example:

X: A A A A B

Y: A B B B

OMA in a nutshell

We can transform X in Y by 3 deletions and 2 insertions:

X: A A A A B

X: ~~A A A~~ A B (3 deletions)

X: ~~A A A~~ A B B B (3 deletions & 2 insertions)

Y: A B B B

Vrije Universiteit Brussel

OMA in a nutshell

We can transform X in Y by 1 deletion and 3 substitutions:

X: A A A A B

X: A A A A ~~B~~ (1 deletion)

X: A B B B ~~B~~ (1 deletion & 3 substitutions)

Y: A B B B

OMA in a nutshell

- If a cost is associated with each operation then it is possible to determine the dissimilarity (as the minimum cost to achieve sequence matching)
- Traditionally each indel operation costs one unit
- The choice of the substitution cost depends on the interpretation of replacing one state by another; if the transitions do not have a meaning, the substitution cost is usually set to two units (Abbott, 2000)

Vrije Universiteit Brussel

OMA in a nutshell

- In time-use studies transitions do have a meaning; the transition from work to non-work is quite common between 5 and 6 pm, quite rare between 1 and 2 am
- Indel-operations are neither neutral in time-use studies: indel operations tend to separate events from their timing of occurrence thereby warping the temporal structure

Vrije Universiteit Brussel

Using the collective rhythm to determine the substitution costs

- Since indel-operations do warp the temporal structure, it is better to avoid them in time-use studies
- Since a single transition matrix does not take the temporal variations into account that are the essence of social time – sequences and transitions between activities are an inherent aspect of collective rhythms – Lesnard (2004) proposed to derive substitution costs from the observed transitions between states

Vrije Universiteit Brussel

Using the collective rhythm to determine the substitution costs

Consequently, we have a transition matrix for each time slot in the sequences we are comparing (f.i. 144 ten-minute intervals in the day cycle):

- The rarer the transition shift between two states in a single time slot, the higher the distance between these states
- The distance at every moment between two individuals depends on the probablity that this transition occurs at that moment in the entire population

Vrije Universiteit Brussel

Example: working time patterns in Belgium 1966-1999

Glorieux, I., I. Mestdag & J. Minnen (2008) The Coming of the 24-hour Economy? Changing work schedules in Belgium between 1966-1999. *Time & Society*, 17,1: 63-83.

Goal: studying evolutions in working time patterns between 1966 - 1999

Example: working time patterns in Belgium 1966-1999

Belgian data of the Multinational Comparative Time-Budget Research project 1966 (Szalai'66) and the TUS-survey of Statistics Belgium (TUS'99)

- Szalai'66, N=2077
  19 to 65 years - 1 diary day
- TUS'99, N=8382
  12 to 95 years - 1 week day and 1 weekend day

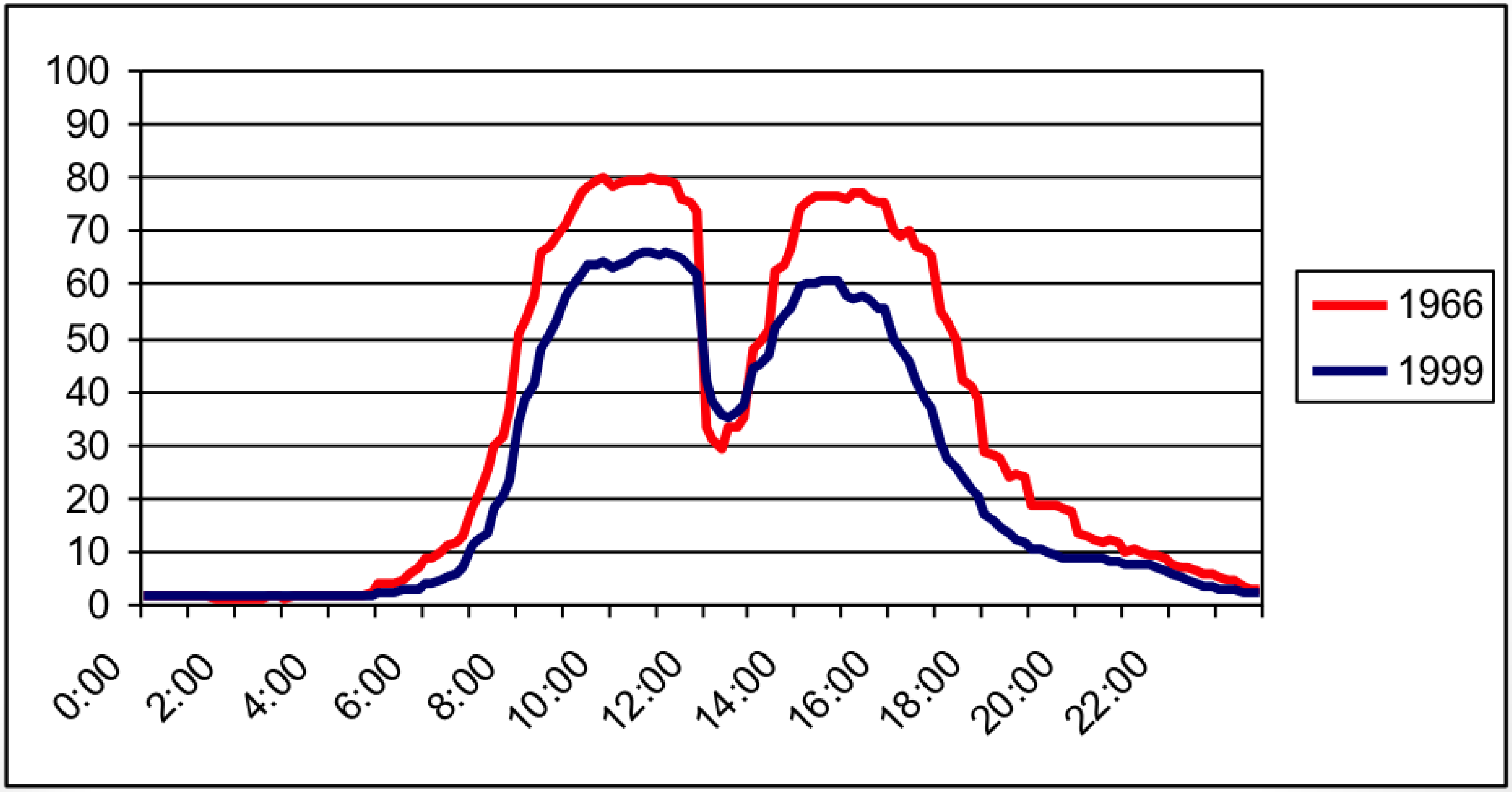
Pooled data for weekdays of both Time Use Surveys of the respondents between 19 and 65 years old who registered at least one hour of paid work

Example: working time patterns in Belgium 1966-1999

- When do Belgians work?
- Which evolutions occurred between 1966 and 1999 in working time patterns
- Which socio-demographic characteristics are related with specific working time patterns (in 1966 and 1999)?

Example: working time patterns in Belgium 1966-1999

Tempogram 'paid work' 1966-1999



Vrije Universiteit Brussel

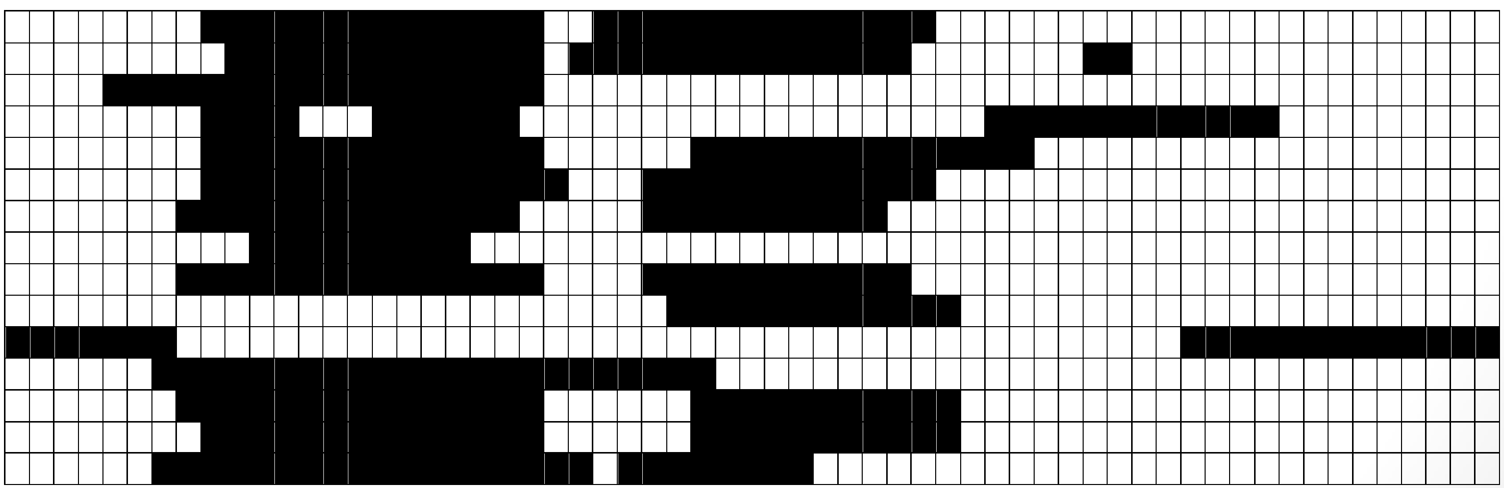
Example: working time patterns in Belgium 1966-1999

- Under the surface of an average tempogram, a variety of different work time patterns may be hidden
- Goal: the identification of different types of working time patterns by means of sequence analysis (OMA)

Example: working time patterns in Belgium 1966-1999

Sequence analysis

- assessing the difference between each pair of individual sequences, in this case individual work schedules (only two states: work – non-work)



Example: working time patterns in Belgium 1966-1999

Sequence analysis

- the analysis must take into account more than just the number of work to non-work substitutions necessary to equate work-day schedules since some substitutions are more probable than others, given their timing
- **Dynamic Hamming distances**: only substitutions – cost of substitutions based on the probability of this (f.i. transition from non-work to work rather 'normal' – low cost - between 8 and 9:30 am, rather 'unusual' – high cost - between 1 and 2 am)

Vrije Universiteit Brussel

Example: working time patterns in Belgium 1966-1999

Cluster analysis

- The dynamic Hamming distance procedure, performed by SAS or R, results in a dissimilarity matrix containing the distances between all possible pairs of sequences
- The matrix of distances between sequences is then used as input for a clustering algoritm
- Result: 12 clusters -> 5 general types of work schedules

Vrije Universiteit Brussel

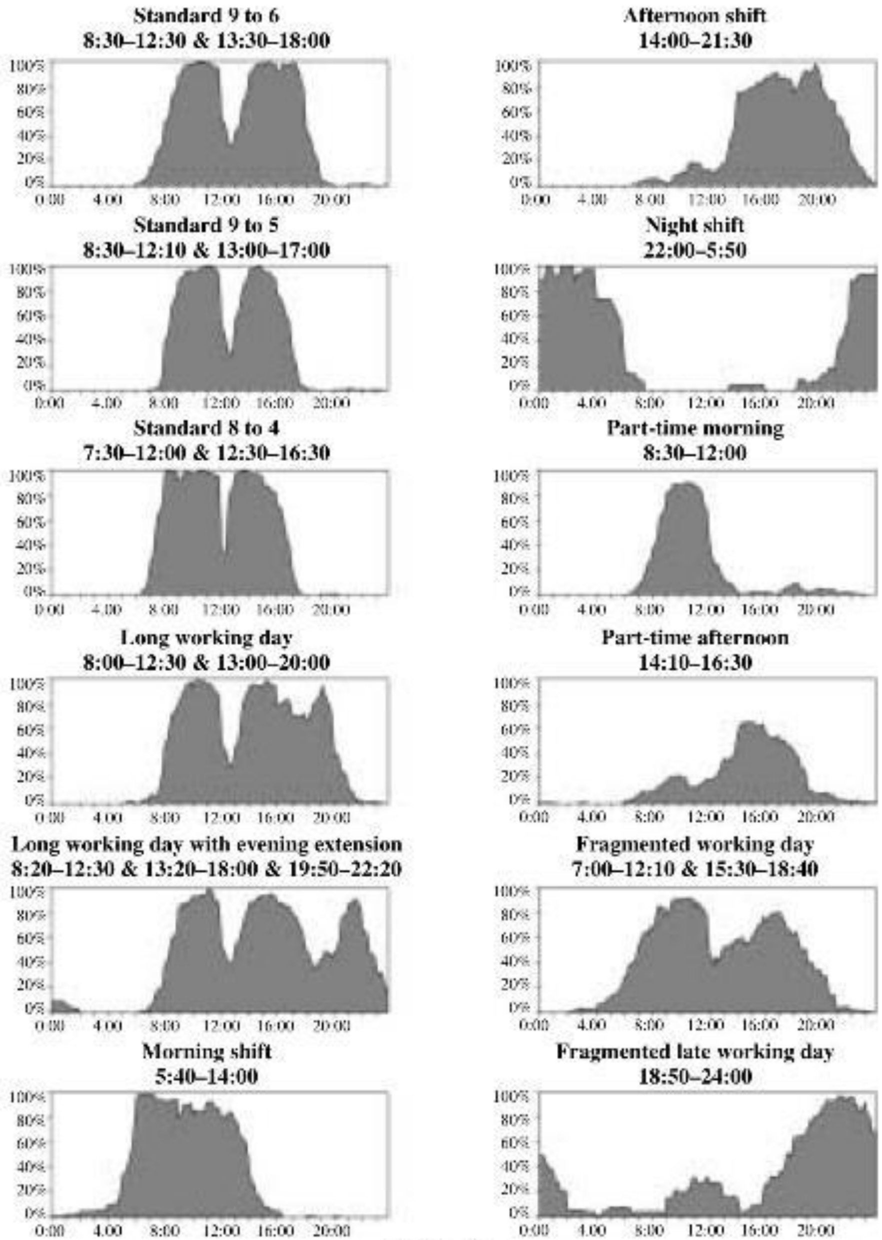


FIGURE 2
Tempograms and mean sequences for paid work carried out on weekdays for 1966 and 1999 for all 12 types of work schedules

Vrije
Universiteit
Brussel

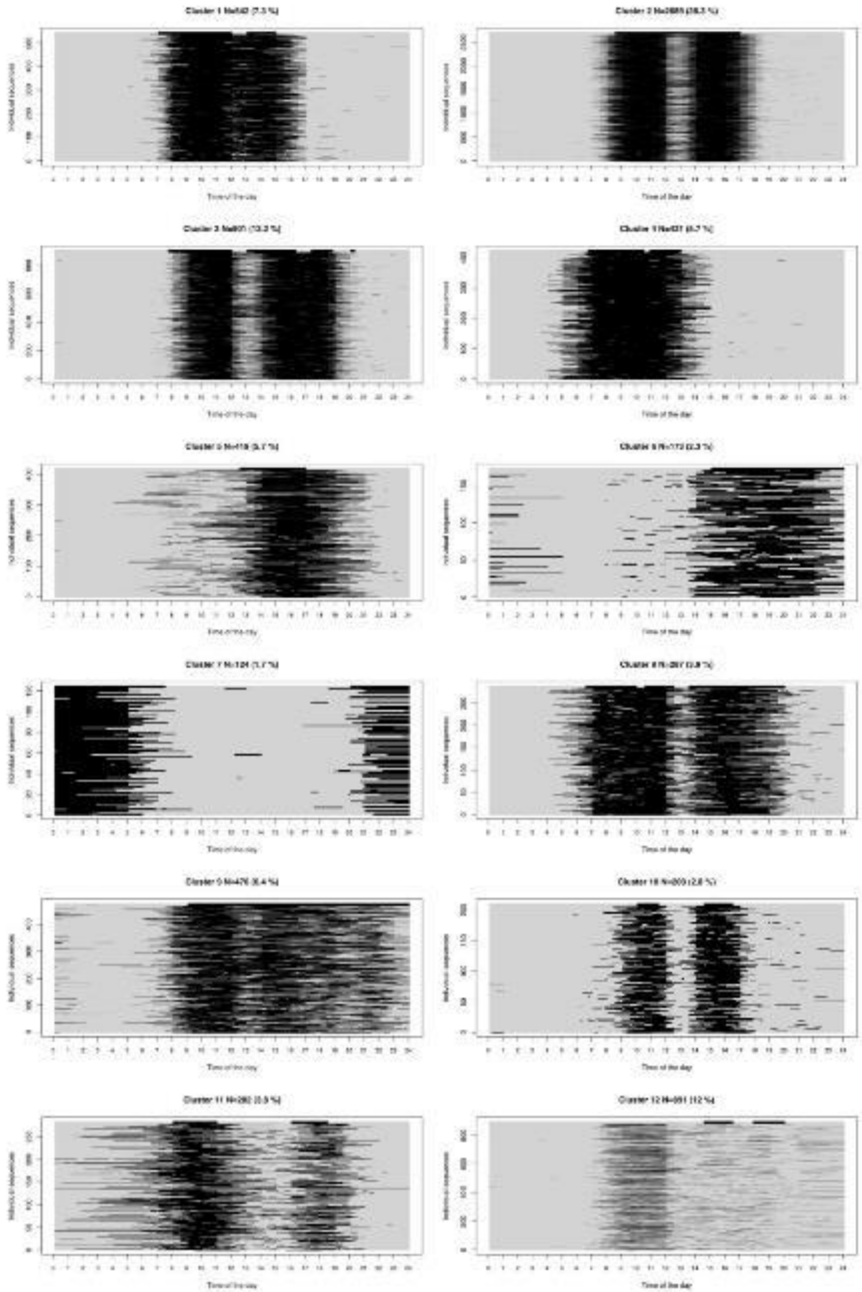


Figure 7 – Individual tempogram: individual sequences are represented horizontally. Black indicates work spells and light gray non-work spells.

Vrije Universiteit Brussel

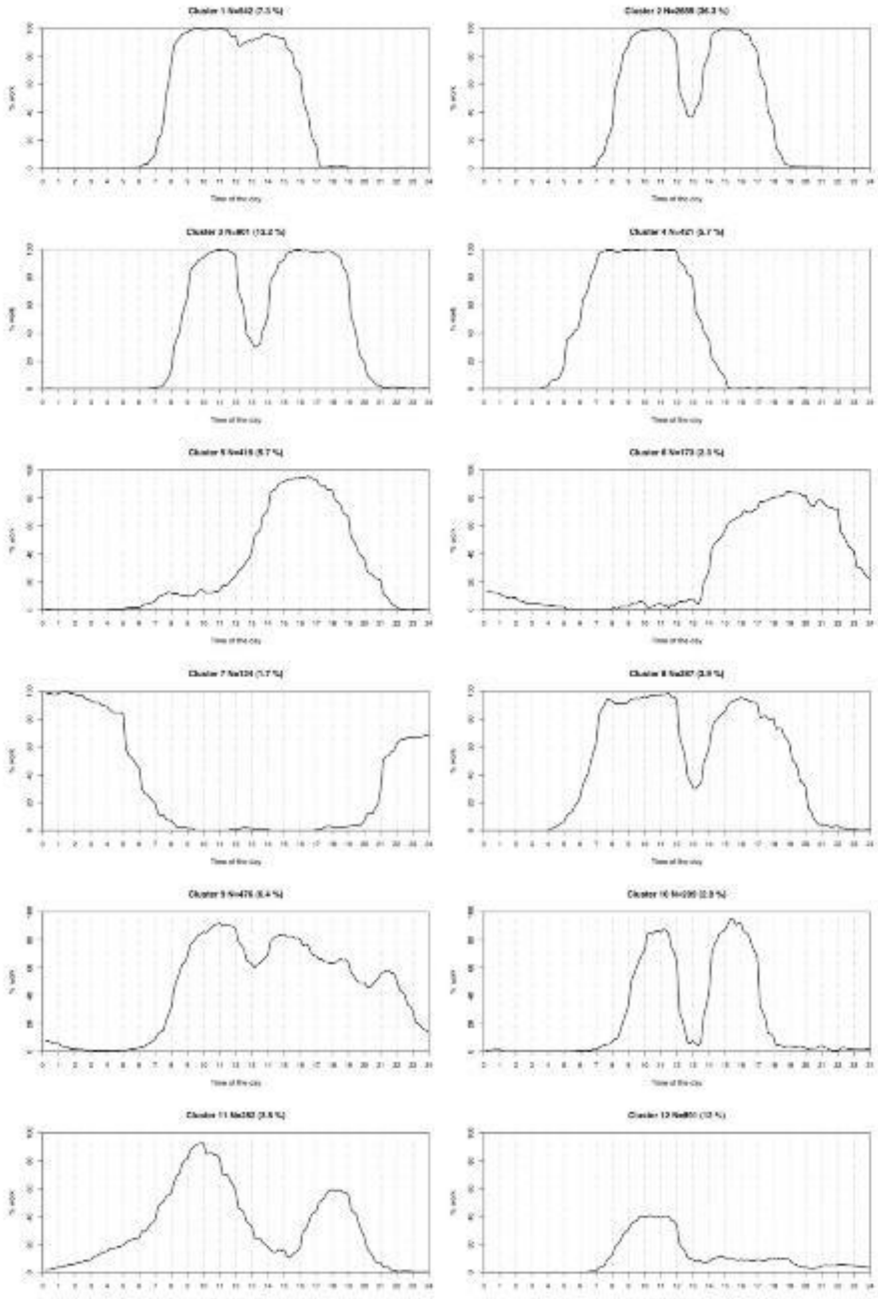


Figure 6 – Aggregate tempogram of the classification of work days

Example: activity patterns in Flanders, Belgium 1999-2004

Van Tienoven, T-P, I. Glorieux, I. Laurijssen & J. Minnen (2011) The social structure of time: optimal matching from time-use data. In: Carrasco Juan Antonio, Jara-Diaz Sergio & Munizaga Marcela (Eds.) , *Time Use Observatory*, Santiago de Chile: Grafica LOM: 141-157

Goal: discriminating activity patterns characterized by paid work and leisure

Example: activity patterns in Flanders, Belgium 1999-2004

- Data
  - Pooled Flemish time-use surveys of 1999 and 2004
  - 7-day diary registration → selection of one weekday
  - 2,285 individuals between 18 and 64 years old
- Structure
  - Each day = 144 ten-minute 'states'
  - Possible states
    - Paid work
    - Domestic work
    - Personal care
    - Watching TV
    - Leisure
    - Sleeping
    - Residual time

Vrije Universiteit Brussel

Example: activity patterns in Flanders, Belgium 1999-2004

Sequence analysis

- Hamming distances: only subsitutions
- Cost of substitutions asigned to each state individually: paid work = 4, domestic work = 3, personal care = 2, watching TV = 8, leisure = 8, sleeping = 1, residual = 1
- So, substituting paid work by leisure would cost 4 + 8 = 12, personal care for leisure: 2 + 8 = 10, etc.
- These costs were then devided by 16 (max. cost), so all costs lie between 0 and 1

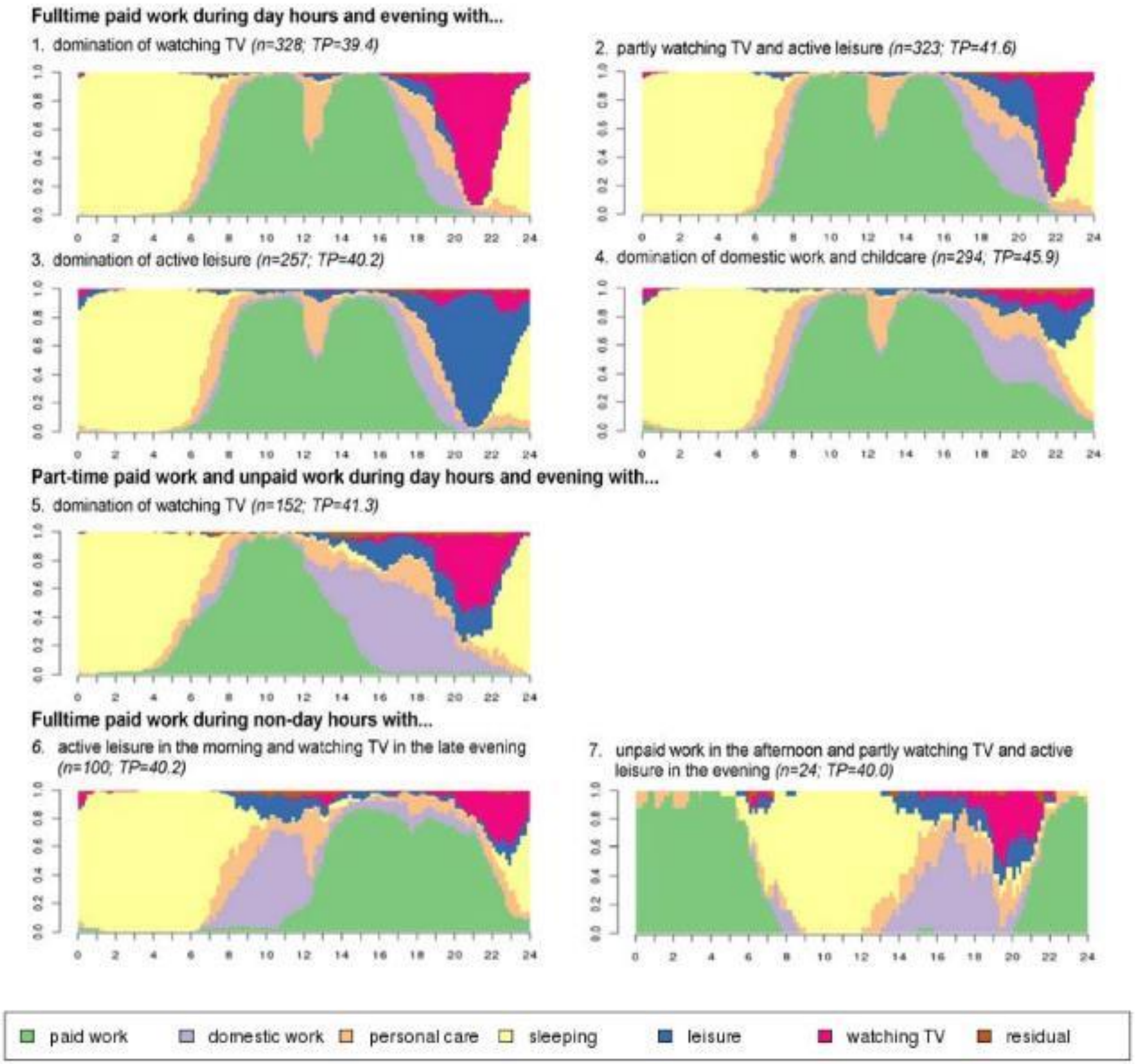
Vrije Universiteit Brussel

Example: activity patterns in Flanders, Belgium 1999-2004

	Paid work	Domestic work	Personal care	Watching TV	Leisure	Sleeping	Residual
Paid work	.000	.4375	.3750	.7500	.7500	.3125	.3125
Domestic work		.000	.3125	.6875	.6875	.2500	.2500
Personal care			.000	.6250	.6250	.1875	.1875
Watching TV				.000	1	.5625	.5625
Leisure					.000	.5625	.5625
Sleeping						.000	.1250
Residual							.000

Vrije Universiteit Brussel

Figure 2. Seven time-use patterns on weekdays based on paid work and leisure (TOR'99/TOR'04 - 18 to 64 year-olds) - n=respondents in pattern; TP=average score for time pressure within pattern.



Vrije Universiteit Brussel

References

- Glorieux, I., I. Mestdag & J. Minnen (2008) The Coming of the 24-hour Economy? Changing work schedules in Belgium between 1966-1999. *Time & Society*, 17,1: 63-83.
- Van Tienoven, T-P, I. Glorieux, I. Laurijssen & J. Minnen (2011) The social structure of time: optimal matching from time-use data. In: Carrasco Juan Antonio, Jara-Diaz Sergio & Munizaga Marcela (Eds.) , *Time Use Observatory*, Santiago de Chile: Grafica LOM: 141-157
- Lesnard, L. (2004) Schedules as sequences: a new method to analyze the use of time based on collective rhythm with an application to the work arrangements of French duel-earner couples. *electronic International Journal of Time Use Research*, 1, 63-88.
- Lesnard, L. (2006) *Optimal Matching and Social Science*. Working Paper No. 2006-01. Institut National de la Statistique et des Etudes Economiques, Paris.
- Lesnard, L. (2008) Off-scheduling within dual-earner couples: An unequal and negative externality for family time. *American Journal of Sociology*, 114, 447-490.
- Lesnard, L. and M.Y. Kan (2009) *Two-Stage Optimal Matching Analysis of Workdays and Workweeks*. Working Paper No. 2009-04. Department of Sociology University of Oxford.

Vrije Universiteit Brussel