

Using *Sequence analysis* to study role trajectories in emerging adulthood:

How does *Sequence analysis* compare to
Latent class analysis and what benefits it
may provide?



Rimantas Vosylis

MRU, Institute of Psychology, Vilnius, Lithuania



MYKOLAS ROMERIS
UNIVERSITY

Transitions to adulthood

Events, such as graduating from school, leaving parental home, starting a career, entering marriage, and becoming a parent, are considered classical markers of becoming an adult (Settersten, 2007).

Timing of these transitions has changed, most notably these transitions became late and more diversified (Billari & Liefbroer, 2010).

Typology Analysis - aims to find a typology of life course (adult role) trajectories

The common type of life course trajectory is often referred to as a “**pathway**”

Life course trajectory

Life course trajectories can be described as the observation of a number of states and transitions between the states, throughout a specific interval of individuals life time

Life course trajectory to family roles can be **operationalized** as a sequence of states, e.g., **SSSSPPPPCCMMHH**, where,

S stands for “single”

P for “with partner”

C for “cohabiting”

M for “married”

H for “with a child”

Two strategies for analysing a typology of such sequences:

1. Latent class analysis
2. Sequence analysis with cluster analysis

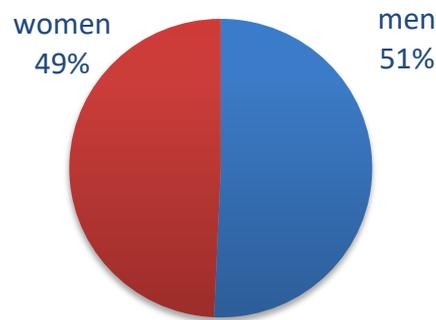
The current study

How do pathways investigated using Sequence Analysis approach compare to the ones investigated using Latent Class Analysis approach?

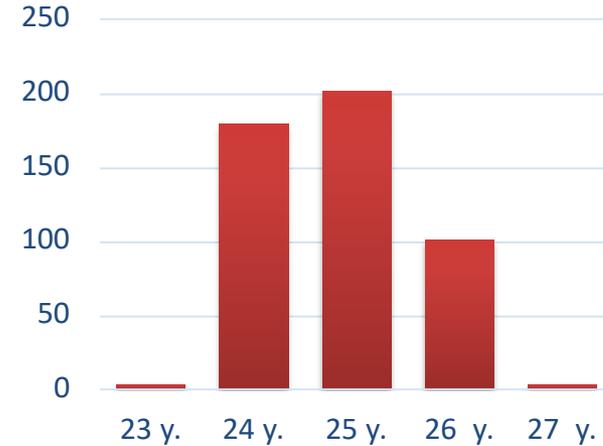
What kinds of transitional pathways do Lithuanian emerging adults follow in partnership-family domain?

Participants of the study

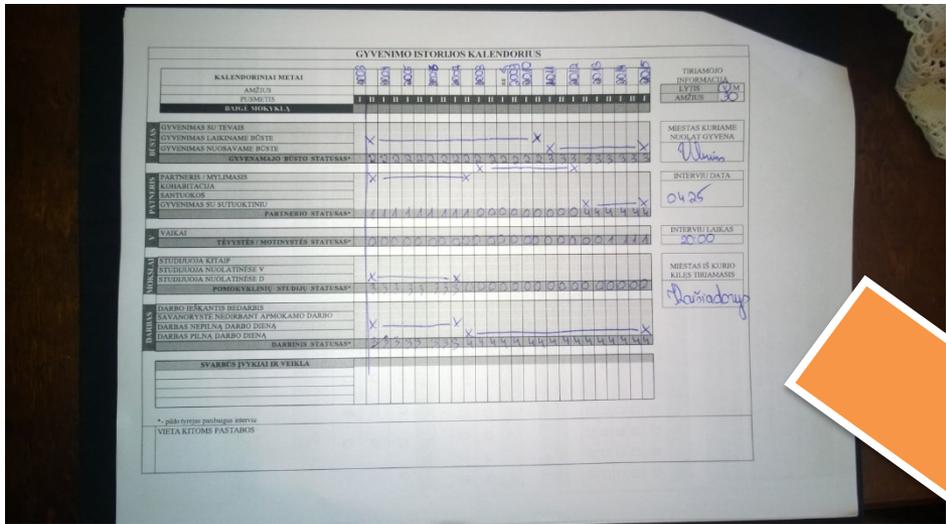
Quota based sample of 489 Lithuanian emerging adults



Age distribution



$M_{age} = 24.84,$
 $SD_{age} = 0.78$



Mplus 7.3 for LCA

R with Traminer package
for sequence analysis

| P001 | P002 | P011 | P012 | P021 | P022 | P031 | P032 | P041 | P042 |
|------|------|------|------|------|------|------|------|------|------|
| 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 1 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 0 |
| 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 2 | 2 |
| 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 1 | 1 | 1 | 1 | 1 | 1 | 2 |
| 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 1 |
| 1 | 0 | 1 | 0 | 0 | 0 | 1 | 4 | 4 | 4 |
| 2 | 2 | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 4 |
| 1 | 1 | 2 | 2 | 2 | 2 | 1 | 1 | 0 | 0 |

Latent class analysis

Strategy:

1. Submit partnership-parenthood statuses sequences to LCA
2. Run solutions with 1 to 8 classes
3. Inspect BIC and LMR-LRT test
4. Visualize results
5. Save classes for further analysis

Results:

1. BIC approaches minimum at 5 classes
2. LMR-LRT suggest 4 classes

BIC VALUE CHANGE ACROSS SOLUTIONS WITH DIFFERENT NUMBER OF CLASSES



Sequence similarity measures: **optimal matching** (OM) and other distance measures

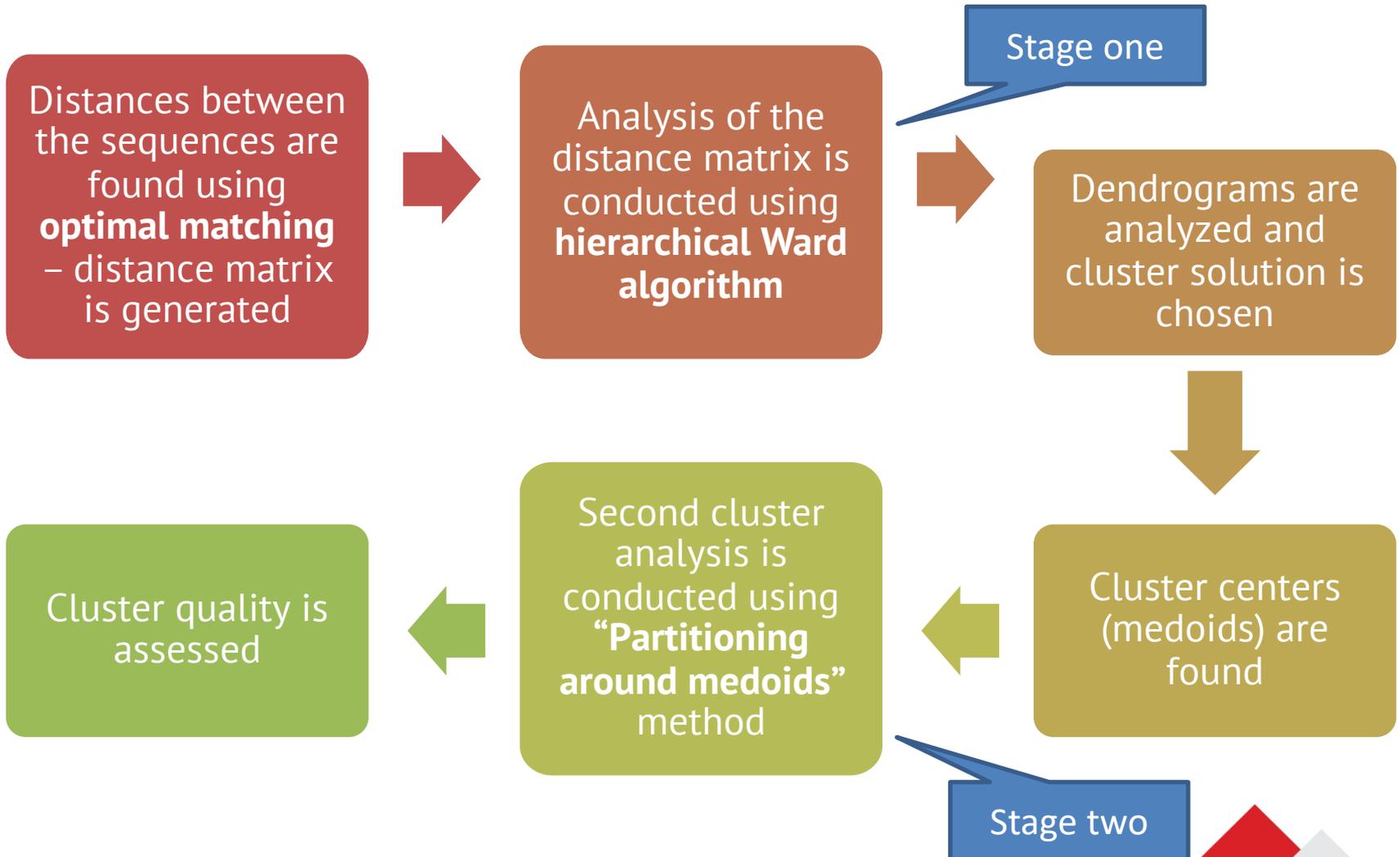
The basic idea of OM is to measure the dissimilarity between two sequences as a function of the number of steps it would take to transform one sequence into the other, i.e., the minimum number of transformations it takes to turn one sequence into the other.

There are three transformation operations:

- **substitution,**
- **insertion, and**
- **deletion.**

The initial step of an OM analysis is to calculate the dissimilarity of all possible sequence pairs in the sample. This results in a matrix of dissimilarities or distances, which is often subjected to cluster analyses.

Cluster analysis with optimal matching as distance measure



Some basic R functions for running sequence analysis

```
setwd("C:/Users/Rimantas/Desktop/LCA FOR SSEA/R/R duomenys/2015/SSEA  
2019")
```

```
library(cluster)
```

```
library(TraMineR)
```

```
library(RColorBrewer)
```

```
library(WeightedCluster)
```

This package is needed for
Ward clustering

This is the main package
for sequence analysis

This helps with coloring
states

This is needed for
clustering sequences as
well

```
lhc <- read.delim("C:/.../SSEA2019PVLCA.dat")
```

```
View(lhc)
```

```
names(lhc)
```

#labeling values

```
PV.labels <- c("single", "w partner", "cohabiting",  
              "marriage", "mar w child", "single w child")
```

```
PV.shortlab <- c("SNGL", "WPRT", "COH", "MARD", "MCHLD", "SCHLD")
```

```
alphabet.PV.seq <- c("1", "2", "3", "4", "5", "6")
```

#defining sequences

```
PV.seq <- seqdef(lhc, var= 2:15, label=PV.labels, states=PV.shortlab,  
  alphabet=alphabet.PV.seq, left = NA, gaps = NA, right = "DEL")
```

#calculating transition rates and creating distance matrix

```
seqtrate(PV.seq)
```

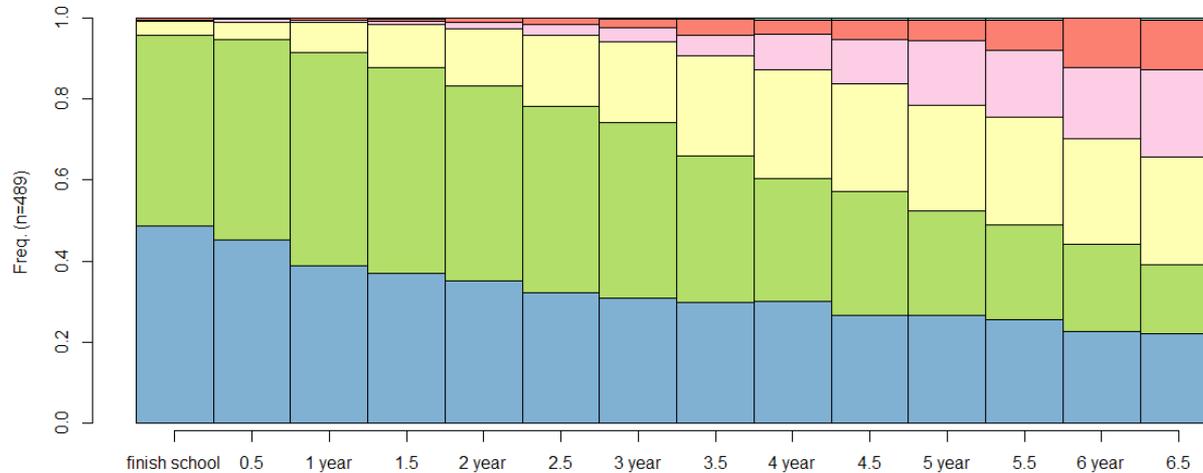
```
PV.trate <- seqtrate(PV.seq)
```

```
PV.seq.scost <- seqsubm(PV.seq, method = "TRATE", with.missing=T, miss.cost=2)
```

```
PV.sec.full.distOM <- seqdist(PV.seq, method="OM", norm=TRUE, indel=1,  
  sm=PV.seq.scost, with.missing=TRUE, full.matrix=TRUE)
```

```
View(PV.sec.full.distOM)
```

Partnership-parenthood status distribution at different intervals in time



```
seqdplot(PV.seq,
withlegend = F,
border = T,
xtlab=xvalues25)
View(PV.seq)
```

Single (no children)

Partner (in relationship with partner, without children)

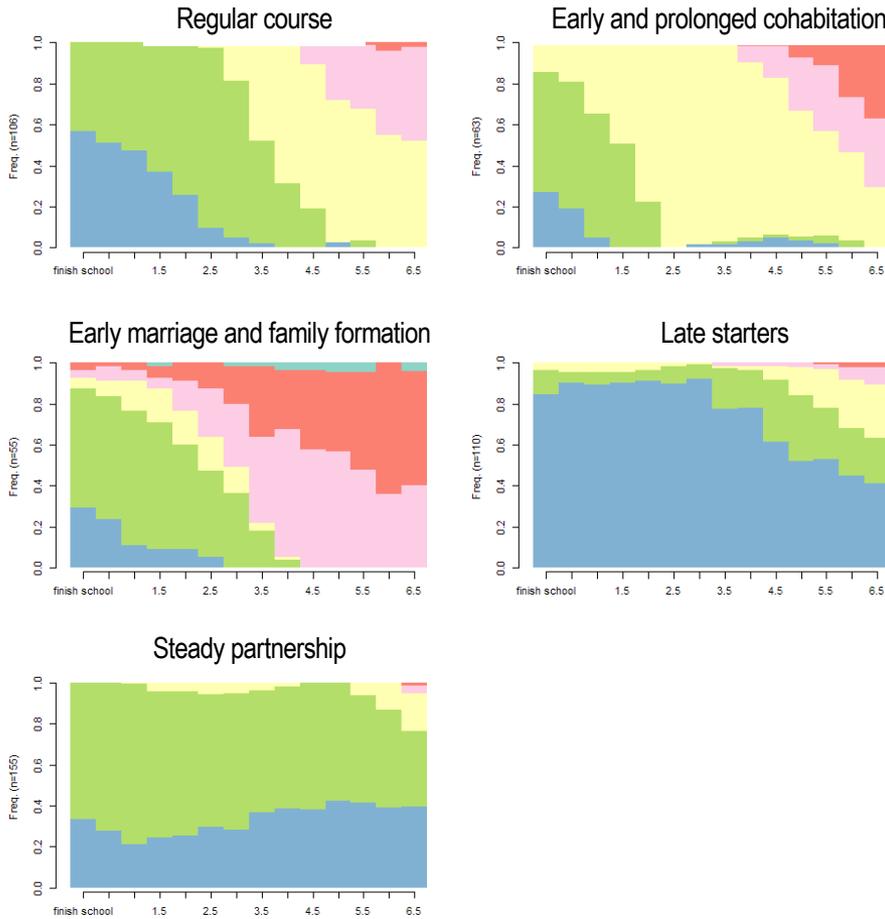
Cohabiting (living with a partner, without children)

Married (no children)

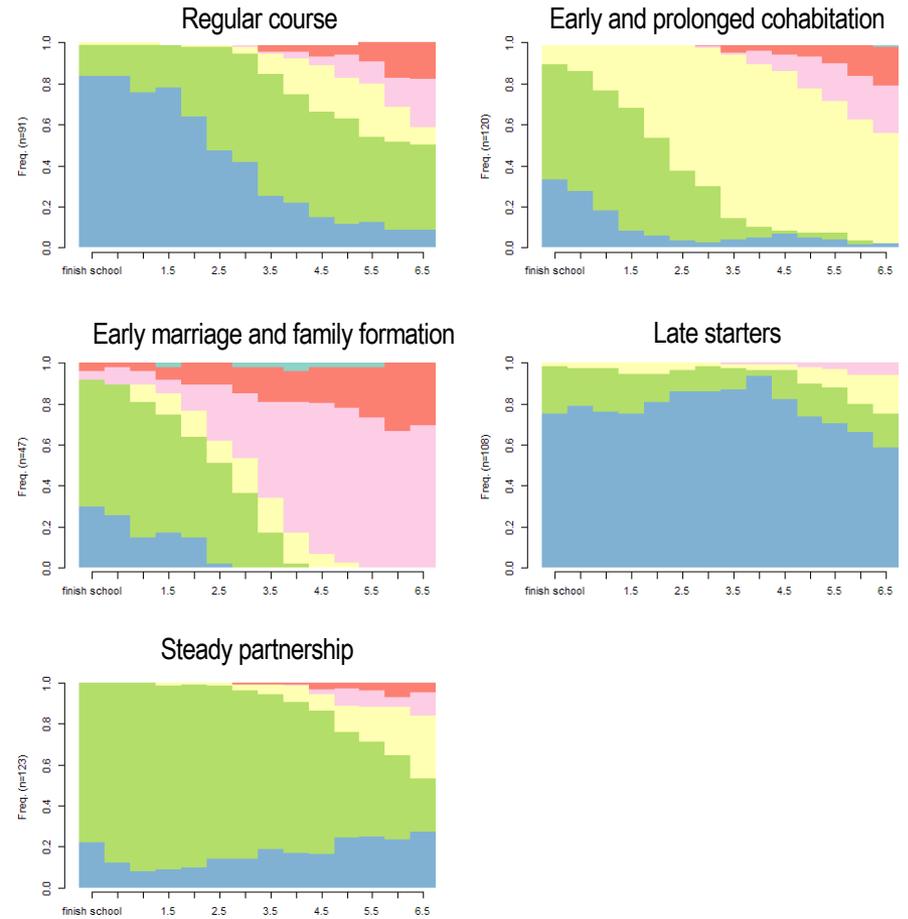
Family (with partner / cohabiting / married with children)

seqdplot(PV.seq, group=lhc\$LCAPV5, xlab=xvalues25, withlegend = FALSE, border = NA)

Latent class results:



OM Cluster analysis results:



How much two solutions match?

Pearson $\chi^2 = 767.85, df = 16, p < .0000001$

Percent of agreement = 61%

Kappa = .50

Two analysis produce similar results, but not necessary the same

The more clusters are homogeneous the more likely solutions will be very similar

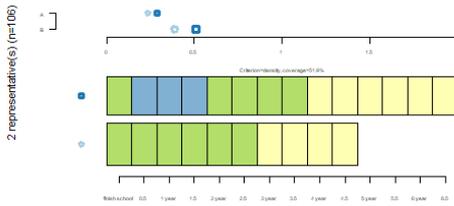
REPRESENTATIVE SEQUENCES

seqrplot(PV.seq, criterion = "density", group=lhc\$LCAPV5, dist.matrix = PV.sec.full.distOM, tsim = 0.25, trep = 0.50, xtlab=xvalues25, cex.plot = 0.5)

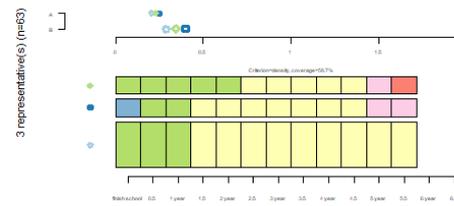
Latent class results:

OM Cluster analysis results:

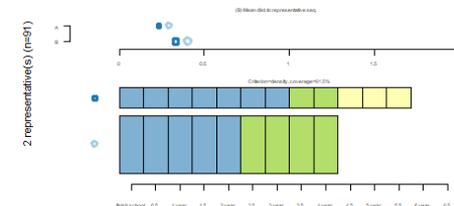
Regular course



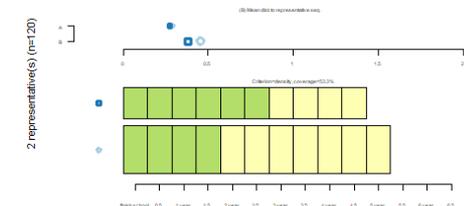
Early and prolonged cohabitation



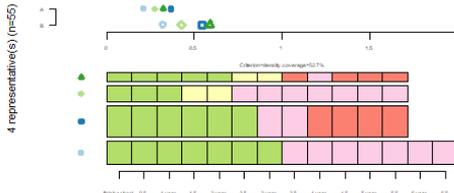
Regular course



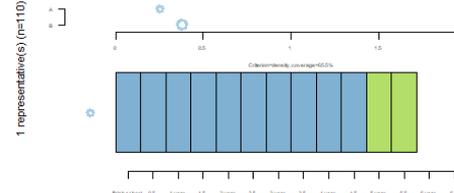
Early and prolonged cohabitation



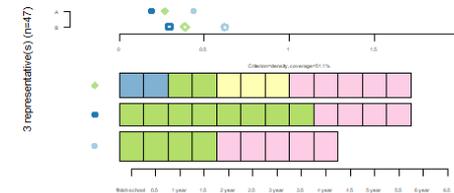
Early marriage and family formation



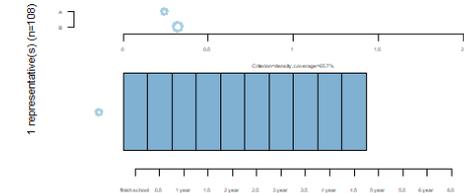
Late starters



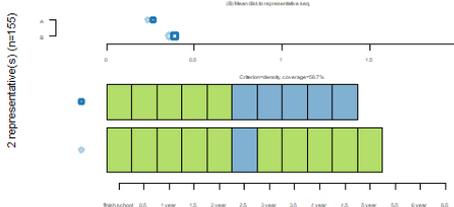
Early marriage and family formation



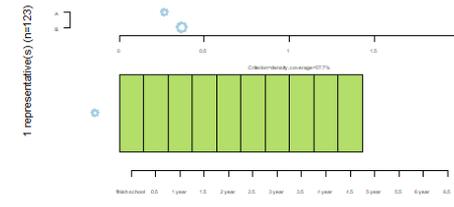
Late starters



Steady partnership



Steady partnership

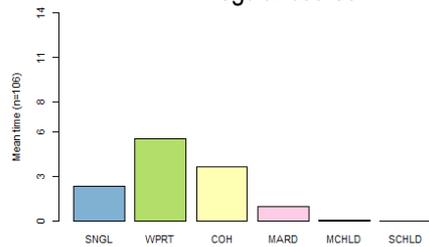


#mean time spent in each state

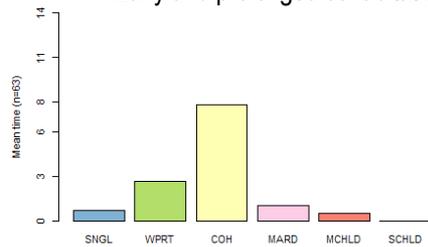
seqmplot(PV.seq, group = lhc\$CLUPV5, withtitle = F)

seqmplot(PV.seq, group = lhc\$LCAPV5, withtitle = F)

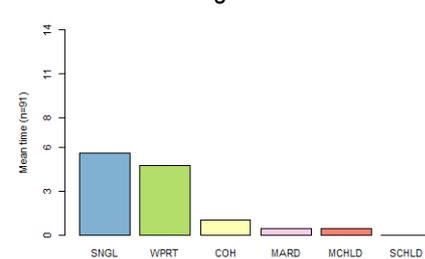
Regular course



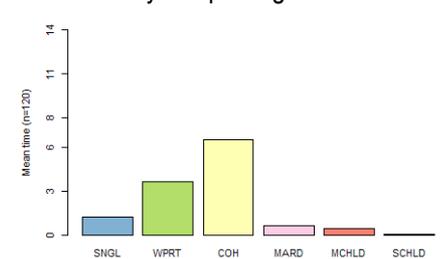
Early and prolonged cohabitation



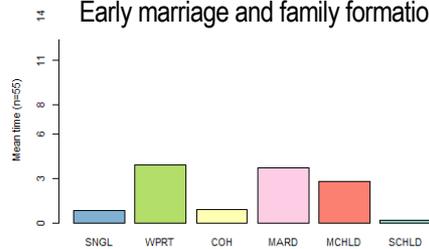
Regular course



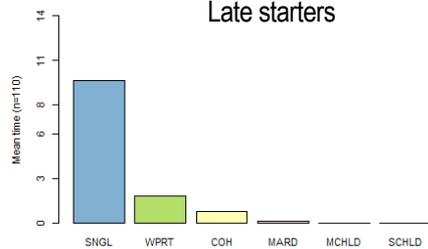
Early and prolonged cohabitation



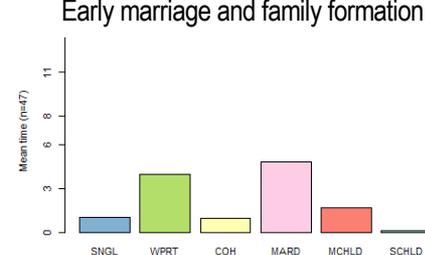
Early marriage and family formation



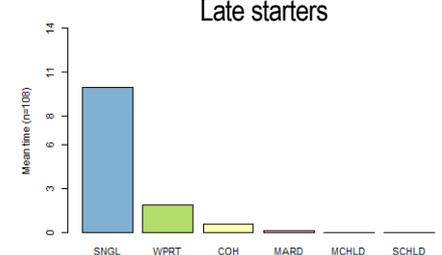
Late starters



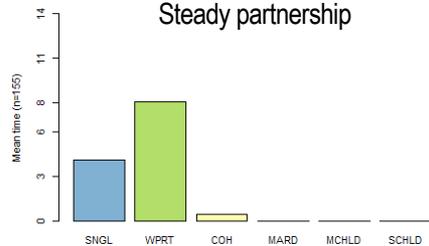
Early marriage and family formation



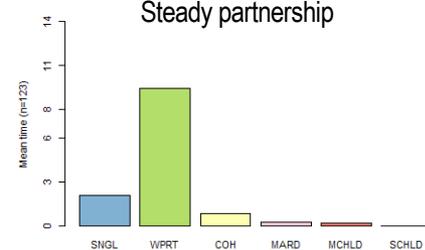
Late starters



Steady partnership



Steady partnership



Cluster validation...

as suggested by Han, Liefbroer & Elzinga (2017)

„**Typology can be viewed as a theoretical construct that is always part of a larger nomological network**, i.e. a set of relationships between the concept of interest (the typology) and other concepts.“

„If one has expectations about the statistical relationship between a typology and other concepts, one can measure the strength of these statistical relationships for each of the typologies. **The stronger these relationships, the more likely it is that the measure of a construct is valid.**“

Choice between LCA and Sequence analysis for sequence typology analysis may be carried out by trying to validate clusters by means of construct validity assessment.

Mean level differences of **life satisfaction** across clusters

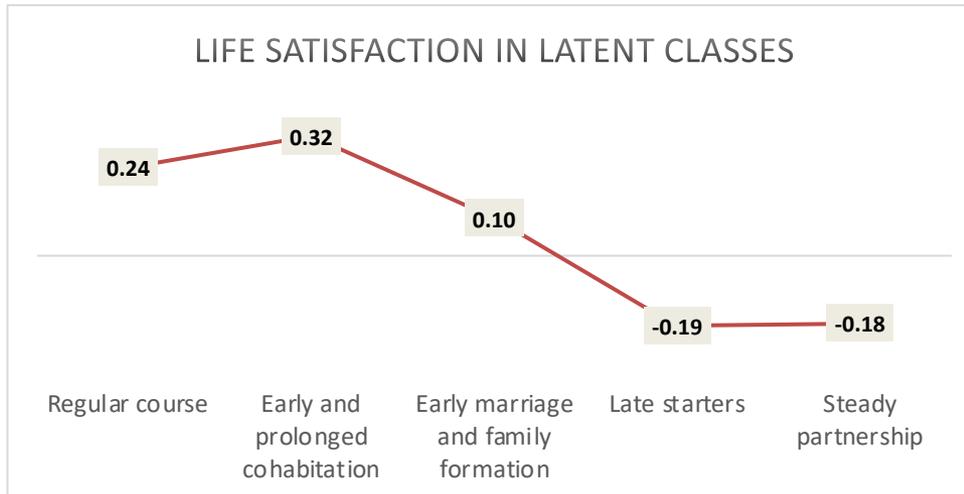
My choice of **life satisfaction** measure is based on developmental task theory (Havighurst, 1966):

“...a **task** which arises at or about a certain period in the life of an individual, **successful achievement of which leads to his happiness** and success with later **tasks**, while **failure leads to unhappiness** in the individual, disapproval by society, and difficulty with later **tasks**.”

Forming romantic relationships and family are considered life tasks for young adults, **therefore, partnership-parenthood state sequence typology should be related to life satisfaction**

Mean level differences of life satisfaction across clusters

LIFE SATISFACTION IN LATENT CLASSES



ANOVA $p < .001$

Eta = .225

LIFE SATISFACTION IN SEQUENCE ANALYSIS



ANOVA $p < .001$

Eta = .203

Latent classes are slightly more related to life satisfaction than sequence analysis clusters

In line with Han et al. 2007 results

So, when we might use sequence analysis instead of LCA?

Sequence analysis may be more preferable, if:

- You have a rather small sample (e.g., 200 or less)
- Your sequences consist of a large number of states (e.g., more than 20)
- When you want to visualize the results of Cluster analysis or LCA results
- When you want to extract representative sequences in LCA and Cluster analysis

LCA may be more preferable, if:

- You have large samples
- Your sequences are short

Thank You!

Contacts:

Rimantas Vosylis: rimantasv@mruni.eu

Mykolas Romeris University, Institute of Psychology, Vilnius, Lithuania



www.mruni.eu



MYKOLO ROMERIO
UNIVERSITETAS

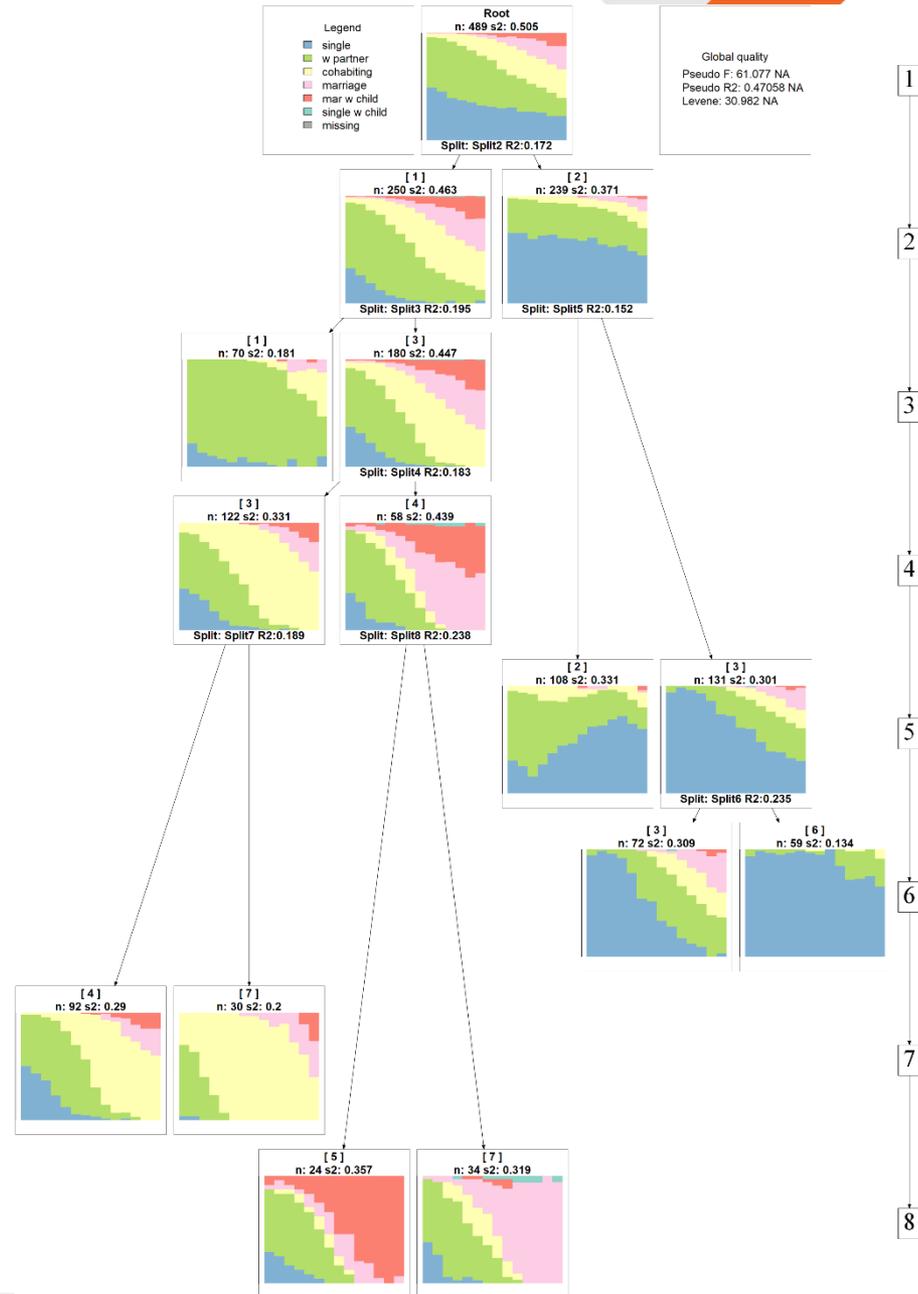
Hierarchical clustering tree

May be used to select the appropriate number (clearly distinct) of clusters

```
PV.wardCluster <-  
hclust(as.dist(PV.sec.full.distOM),  
method = "ward")
```

```
PV.wardTree <-  
as.seqtree(PV.wardCluster,  
seqdata=PV.seq,  
diss=PV.sec.full.distOM, ncluster=8)
```

```
seqtreedisplay(PV.wardTree, type =  
"d", border = NA, PVorder = NA,  
showdepth = TRUE)
```



Good sources on sequence analysis

<http://traminer.unige.ch/> - Main internet source for sequence analysis

Also:

- Dietrich, J., Andersson, H., & Salmela-Aro, K. (2014). Developmental psychologists' perspective on pathways through school and beyond. In P. Blanchard, F. Bühlmann, & J.-A. Gauthier (Eds.), *Advances in Sequence Analysis: Theory, Method, Applications* (pp. 129–150). Springer International Publishing
- Martin, P., & Wiggins, R. D. (2011). Optimal Matching Analysis. In M. Williams & W. P. Vogt (Eds.), *The SAGE Handbook of Innovation in Social Research Methods* (pp. 385–408). London: SAGE Publications Ltd.
- Martin, P., Schoon, I., & Ross, A. (2008). Beyond transitions: Applying optimal matching analysis to life course research. *International Journal of Social Research Methodology*, 11(3), 179–199. doi:10.1080/13645570701622025

What if you have multi-dimensional sequences?

Lets say you have a sequences of education-employment states and partnership-parenthood states and you are interested in using holistic approach:

1. Sequence analysis offers multi-channel sequence analysis
2. LCA offers hierarchical latent class analysis: focuses on investigating role configurations nested in holistic pathways

Examples of adult role transition sequence typology research

Robette, 2010: 8 pathways (e.g., spouses with children), sequence analysis

Salmela-Aro et al., 2011: 6 pathways (e.g., Fast starters), sequence analysis

Eliason et al., 2015: 5 pathways (e.g., tradition school-to-work, negligible family formation), hierarchical latent class analysis

Schwanitz, 2017: 7 pathways (e.g., late home leavers), sequence analysis
(multidimensional)

Osgood et al., 2005, Maggs et al., 2012, Schoon et al., 2012: 5-7 pathways, latent class analysis / role configurations