Exploring sequences with VISUAL-TimePAcTS

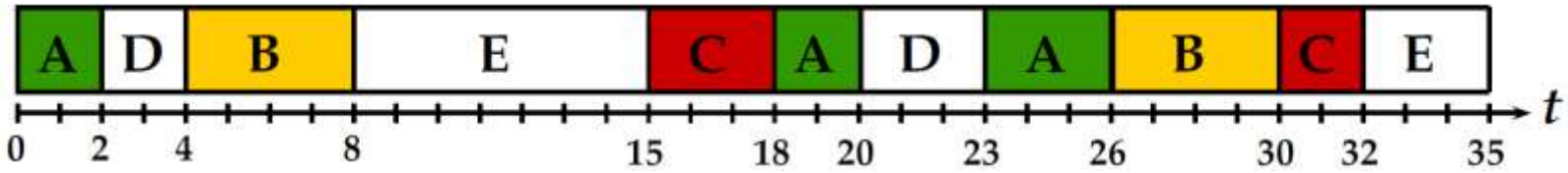
Katerina Vrotsou
katerina.vrotsou@liu.se

LiU expanding reality

Focus on sequences

Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

- Event-sequence data
- Composed of sequences of ordered events
- Exact or relative time



time use data

CRESEARCH

LiU

Sequences in everyday life

Introduction

Background

VISUAL-TimePAcTS

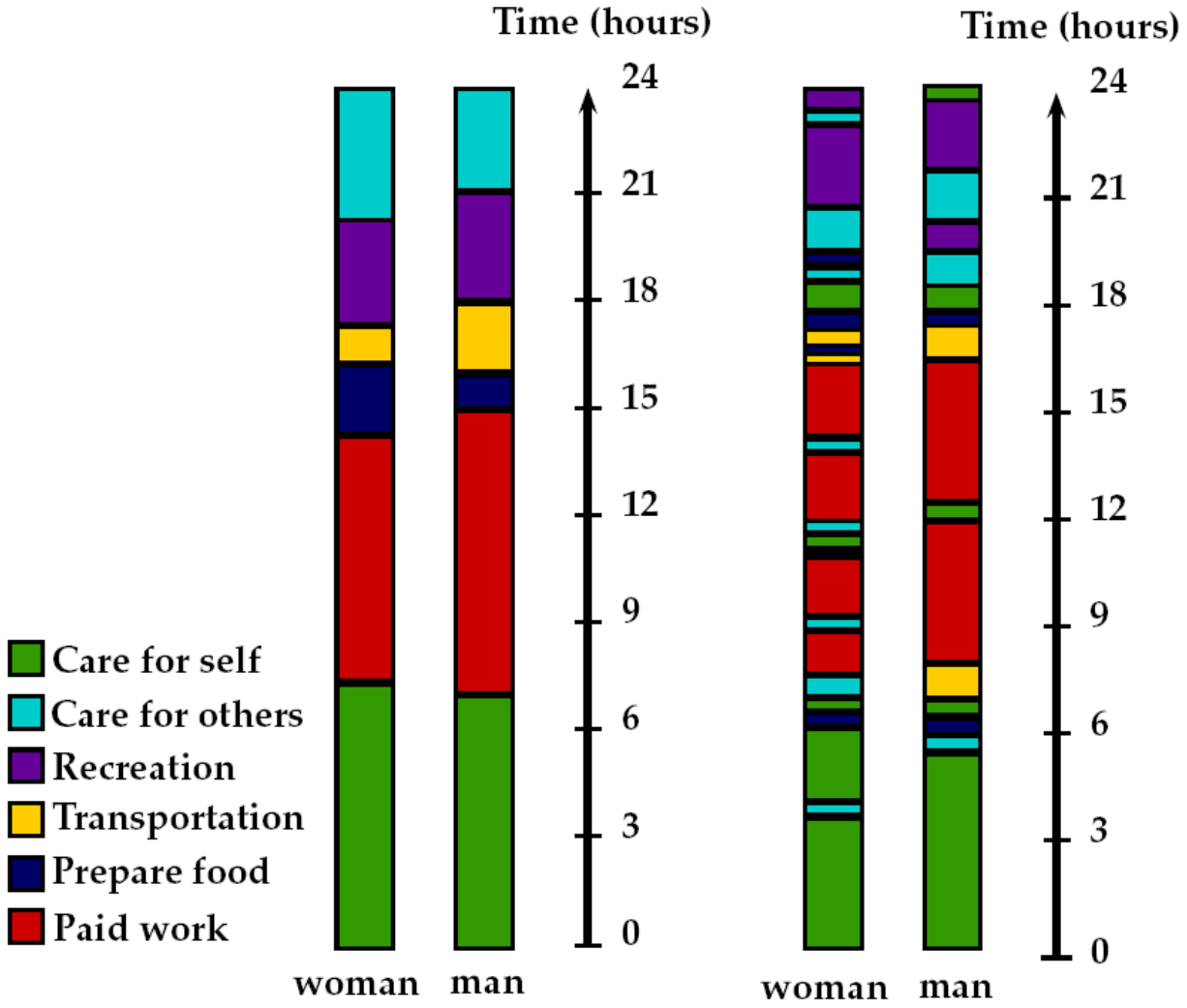
Highlighting

Mining

Exploring

Clustering

Conclusions



CRESEARCH

LiU

Analysis task

Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

- Communicate data
- Compare sequences
  - Content, structure
- Identify patterns
  - Combinations of activities exhibiting interesting behaviour
  - Frequent, evenly distributed, repetitive, outliers
- Compare and visually analyse pattern distribution
- Classify, predict

CRESEARCH

LiU

VISUAL-TimePAcTS

Introduction

Background

VISUAL-TimePAcTS

Highlighting

Mining

Exploring

Clustering

Conclusions

- **VISUAL**ization
- **Time**
- **P**lace
- **Ac**tivities
- **T**echnologies
- **S**ocialization

CRESEARCH

VISUAL-TimePAcTS

Introduction

Background

VISUAL-TimePAcTS

Highlighting

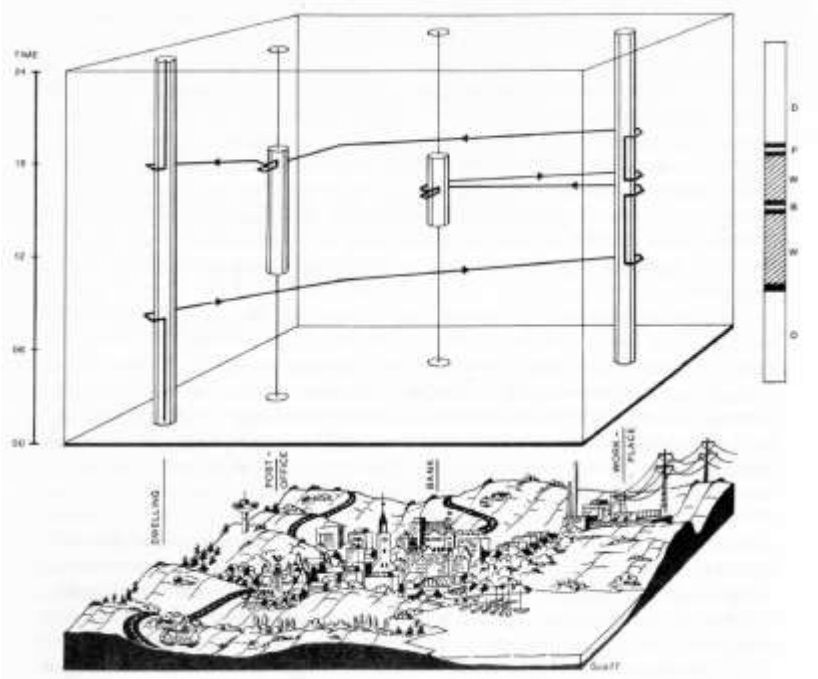
Mining

Exploring

Clustering

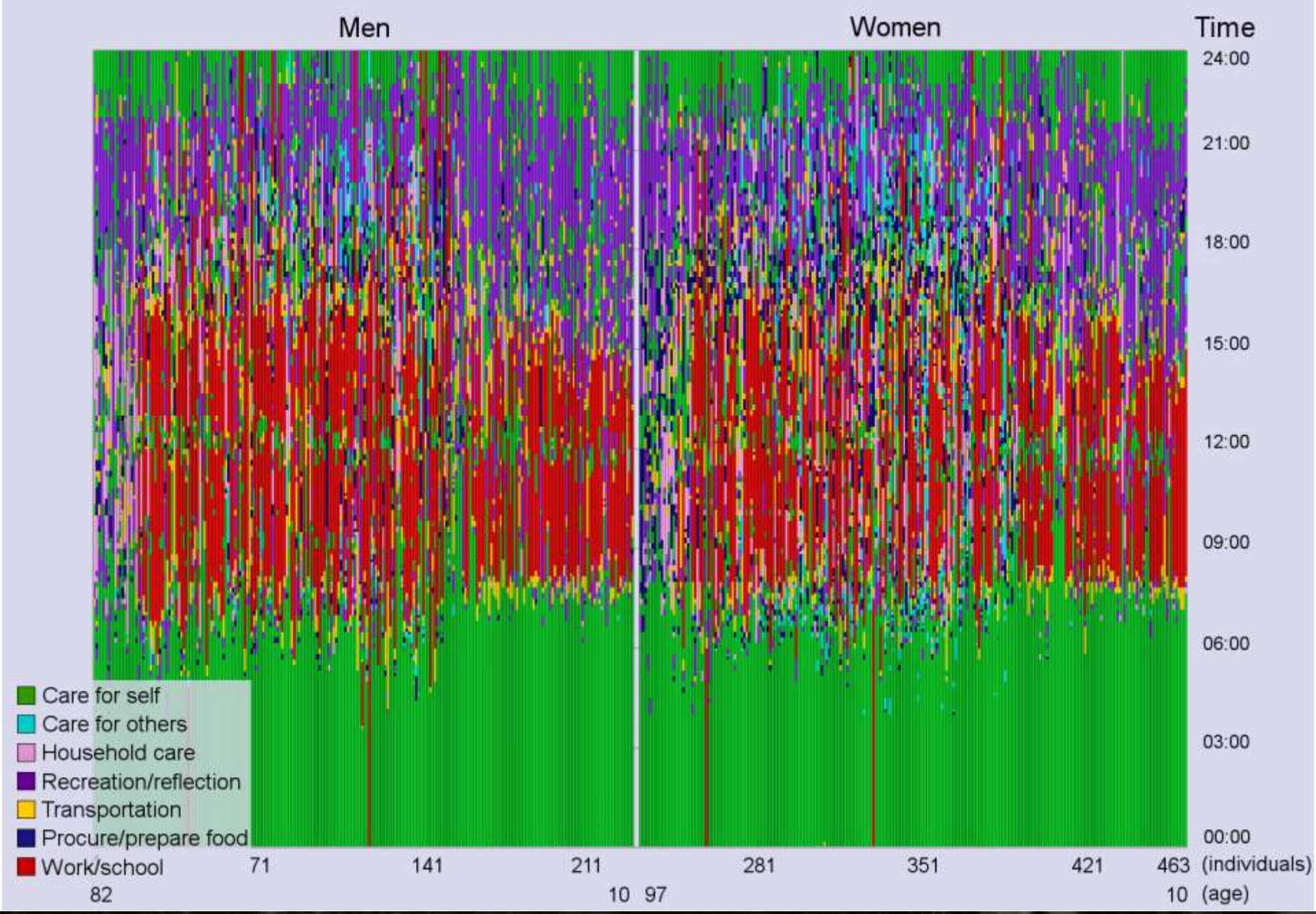
Conclusions

- *Algorithmically guided visual analysis* tool for
  - exploring event-sequences; time use data and
  - detecting sequential patterns
- Time-geographic approach and representation
  - Focus on individual
  - Activity path
- Testbed for research



CRESEARCH

LiU


Men
Women
Time
24:00
21:00
18:00
15:00
12:00
09:00
06:00
03:00
00:00
Care for self
Care for others
Household care
Recreation/reflection
Transportation
Procure/prepare food
Work/school
71
141
211
281
351
421
463
(individuals)
82
10
97
10
(age)

Overview

Introduction

Background

VISUAL-TimePAcTS

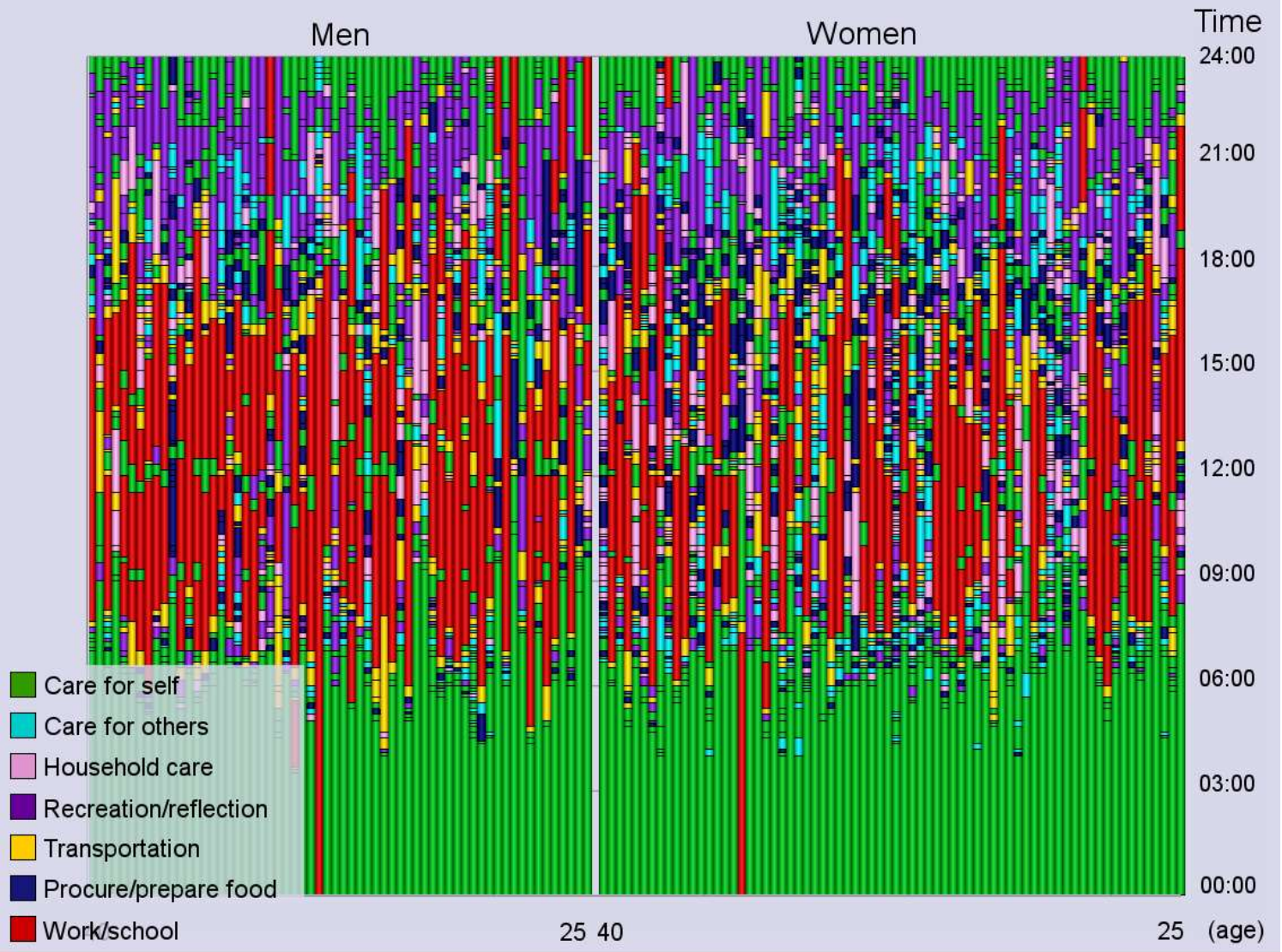
Highlighting

Mining

Exploring

Clustering

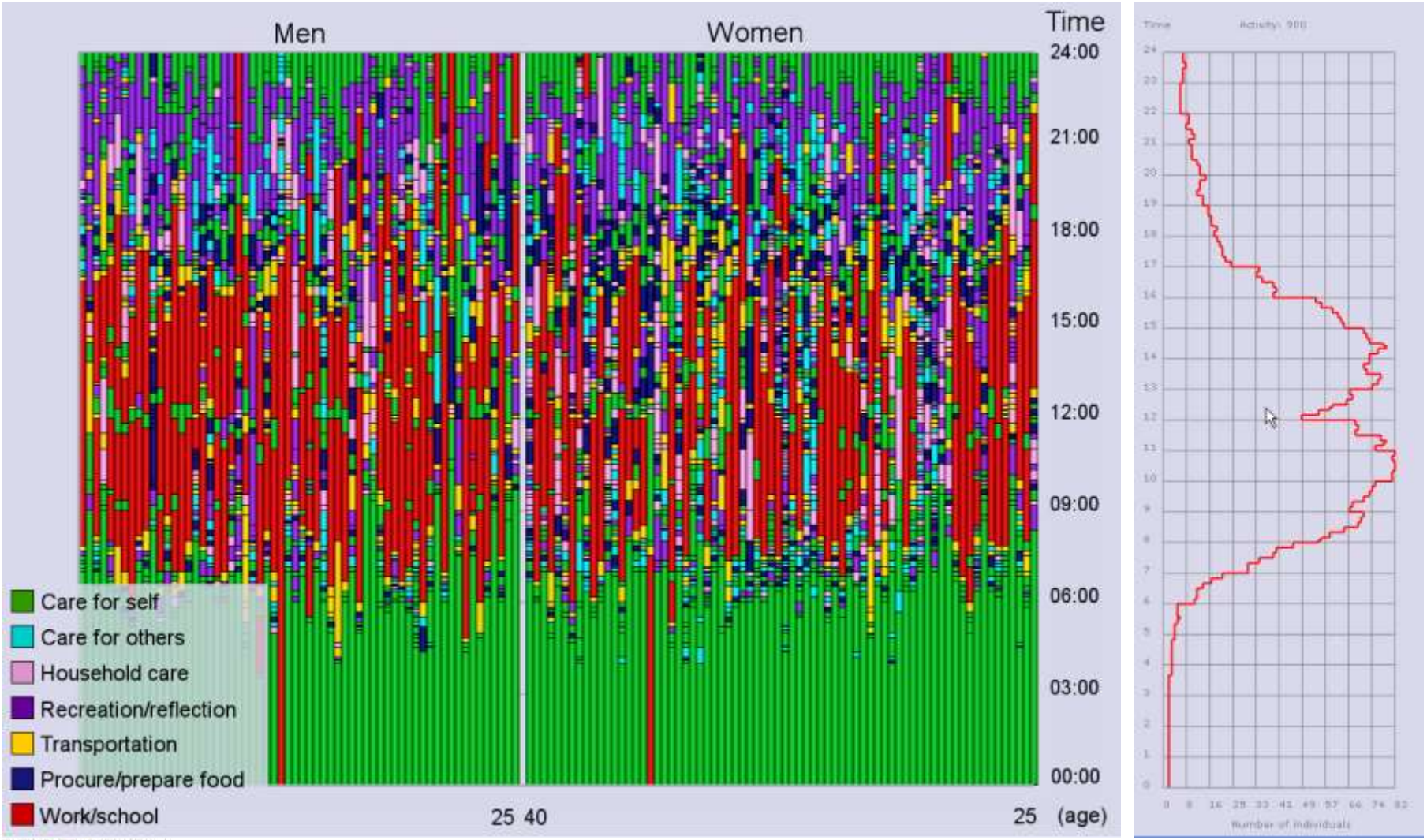
Conclusions



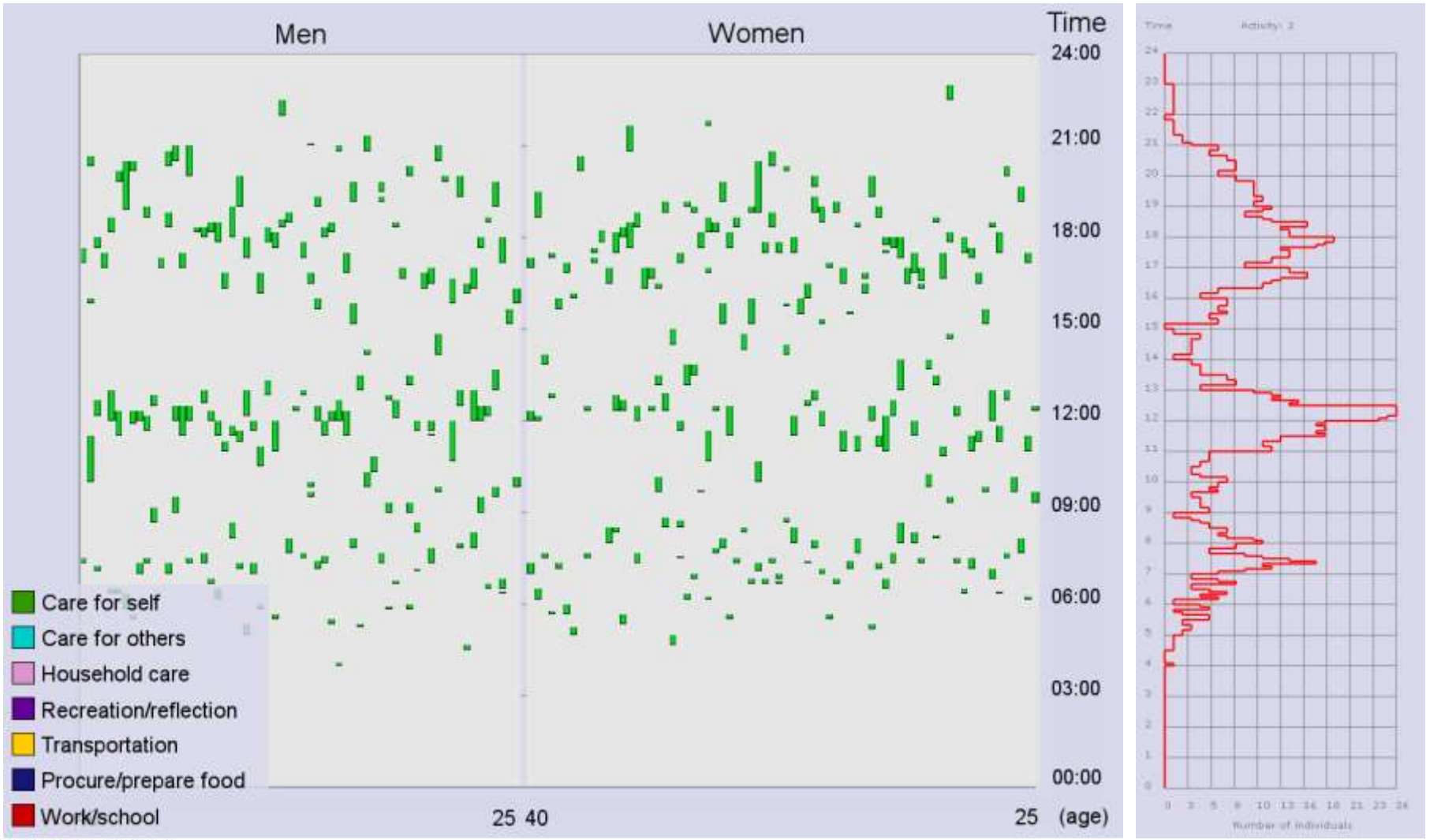
CRESEARCH

LiU

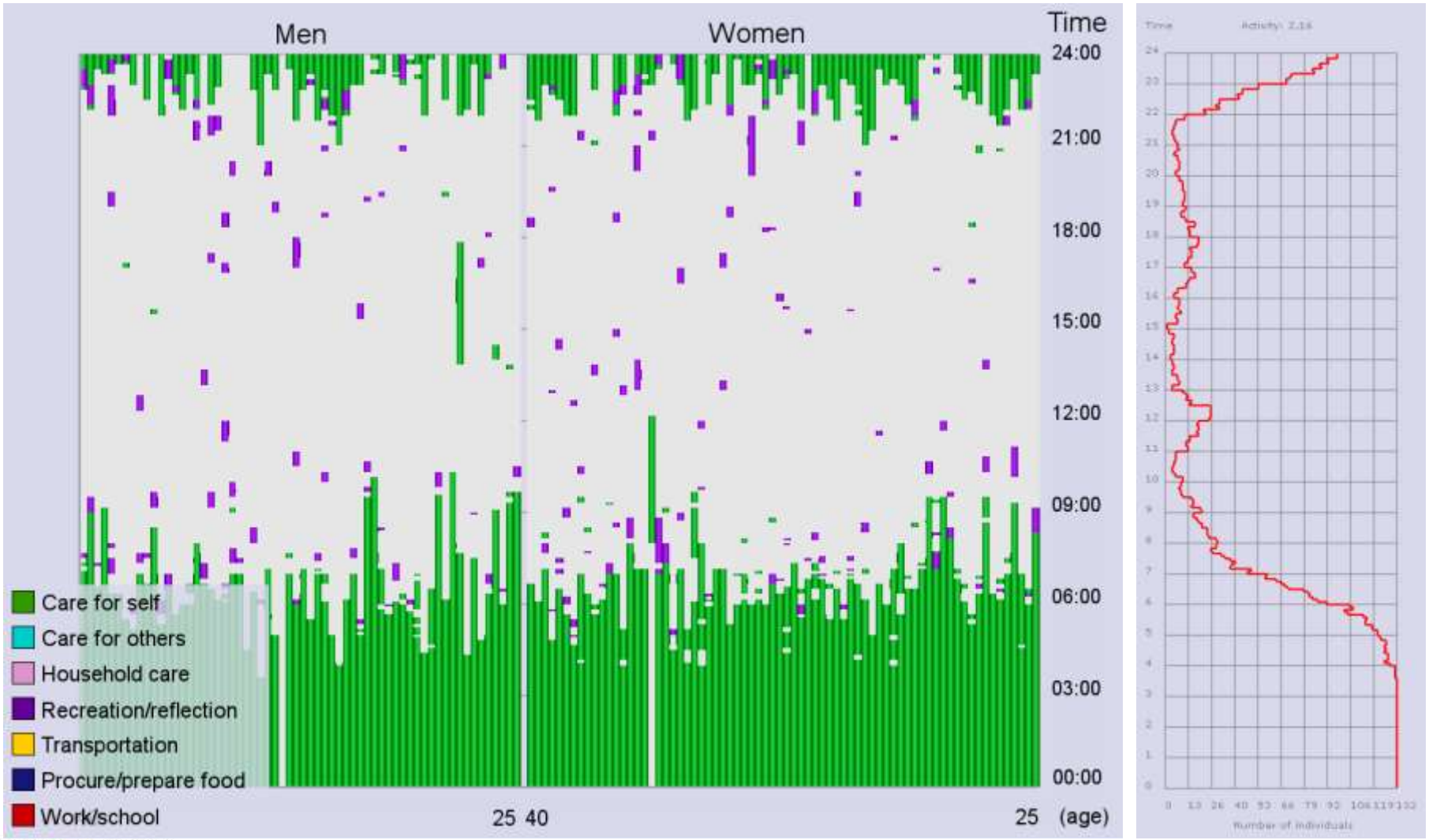
Highlighting



Highlighting



Highlighting



Sequential pattern mining

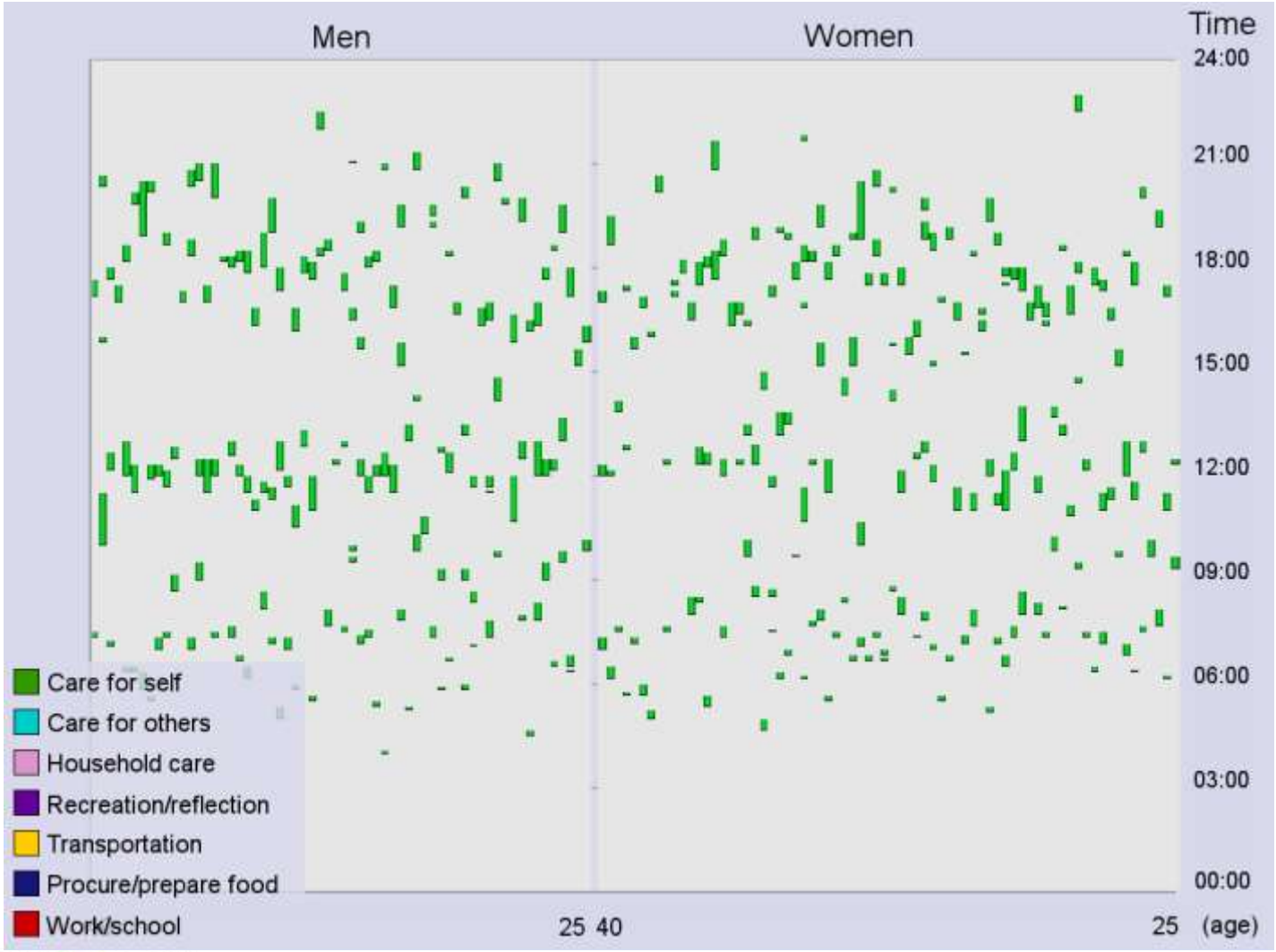
Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

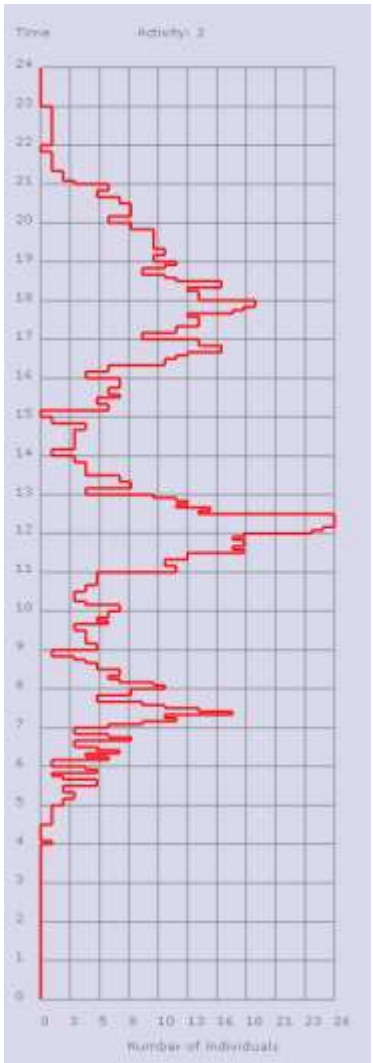
- *Sequential pattern*: interesting sub-sequence of activities
  - Activities belonging to same project or practice
- *Sequential pattern mining*: Given a database of time diaries (sequences of activities performed during the day), the problem of sequential pattern mining is to discover all sequential patterns that have a user-specified minimum support, defined as the number of customer-sequences that contain this pattern.
- Detect
  - Interesting behaviour
  - Trends

CRESEARCH

LiU

Sequential pattern mining





Sequential pattern mining

Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

- *Apriori approach*
  - Bottom up
  - Candidate generation
  - Pruning
  - *n*-patterns iteratively
  - In VISUAL-TimePAcTS
- *Pattern growth approach*
  - Divide and conquer
  - Compress database into fp-tree
  - No candidate generation
  - Currently developed

CRESEARCH

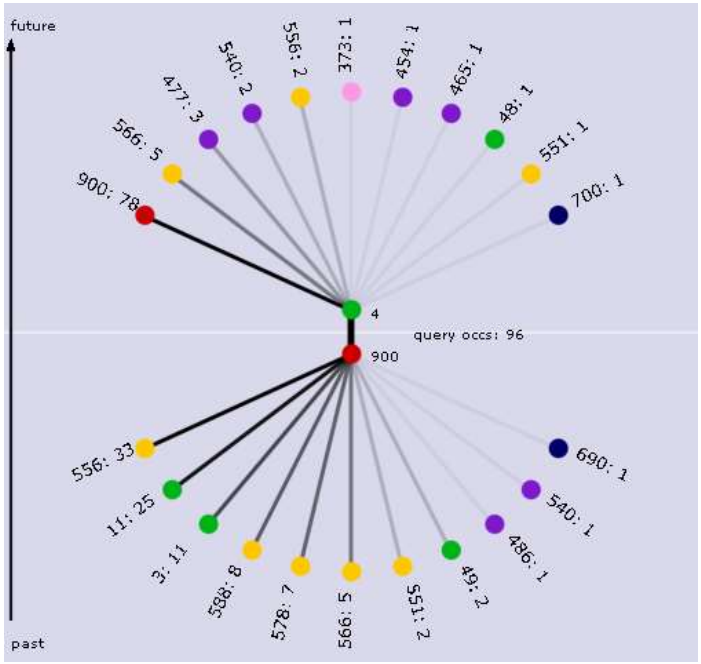
Interactive exploration

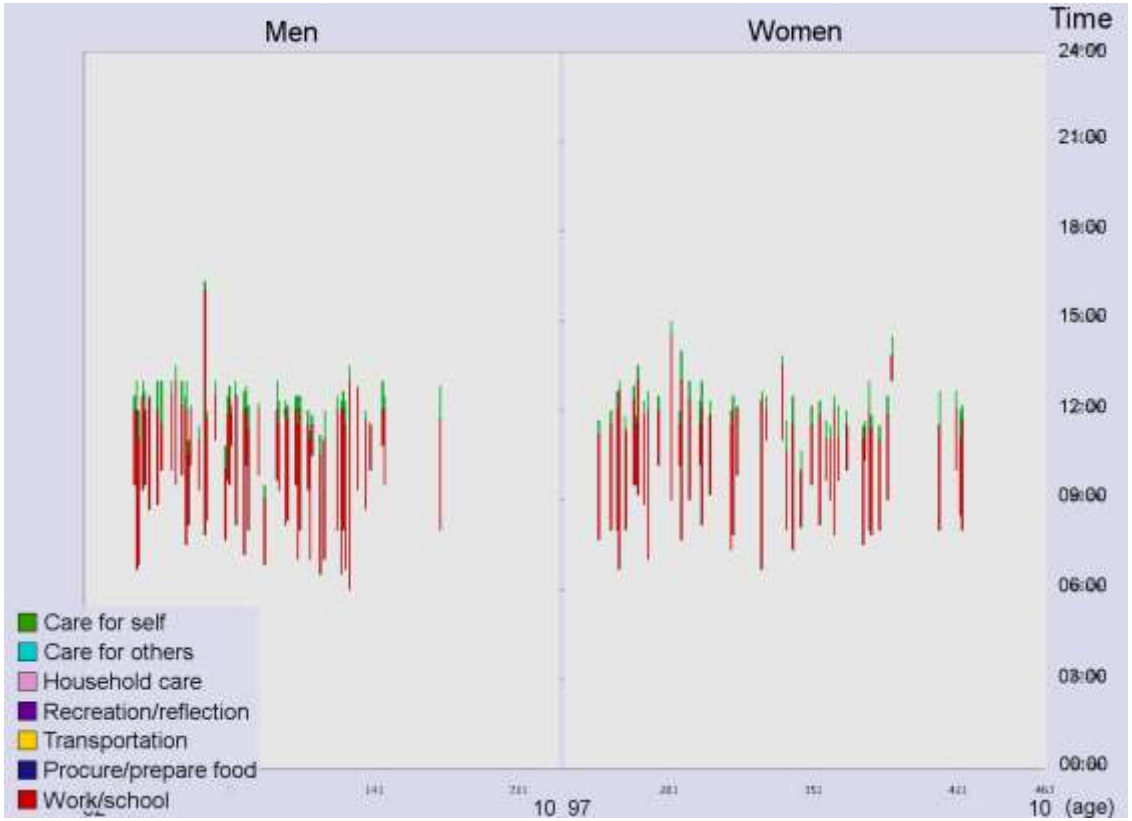
Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

- Tree inspired exploration tool
- Systematic identification of sequences
- Graph similarity approach
- Using algorithm for web searching
- Exploration controlled by user

CRESEARCH

Interactive exploration





CRESEARCH

LiU

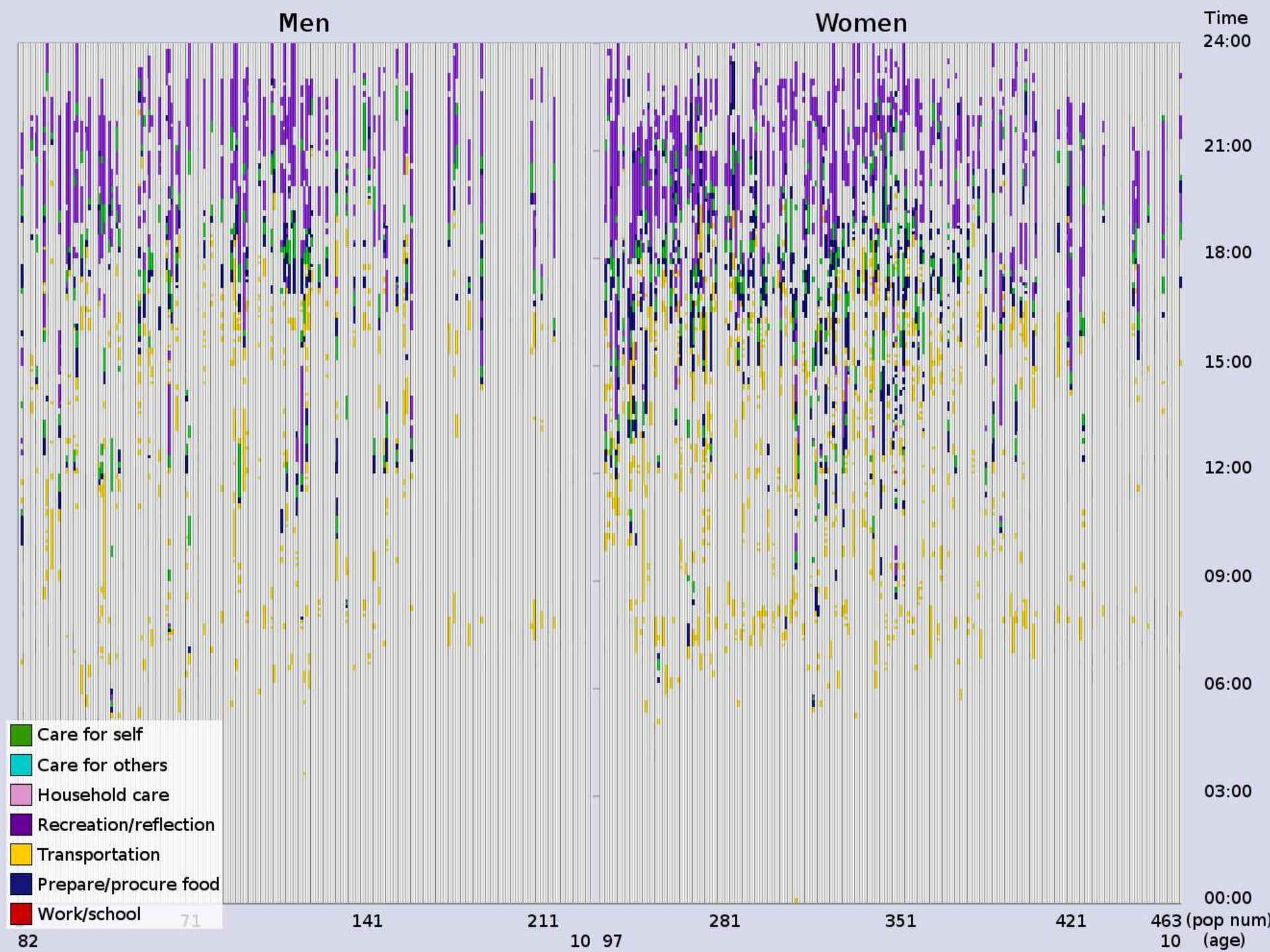
Clustering

Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

- Given an interesting sequence pattern
  - i.e. interesting behaviour, practice
- Identify similarly behaving groups
  - How practice is performed

CRESEARCH

LiU


Men
Women
Time
24:00
21:00
18:00
15:00
12:00
09:00
06:00
03:00
00:00
Care for self
Care for others
Household care
Recreation/reflection
Transportation
Prepare/procure food
Work/school
71
141
211
281
351
421
463 (pop num)
82
10 97
10 (age)

Clustering

Introduction

Background

VISUAL-TimePAcTS

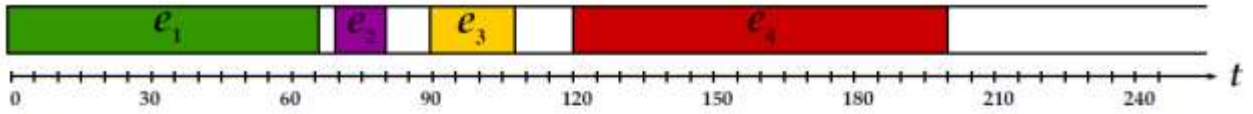
Highlighting

Mining

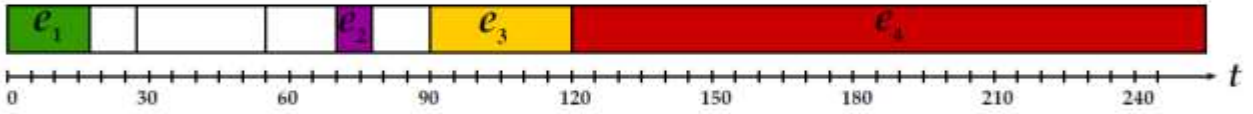
Exploring

Clustering

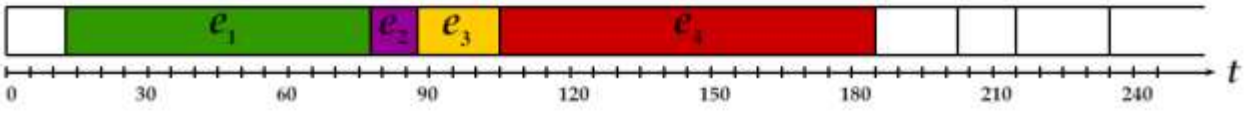
Conclusions



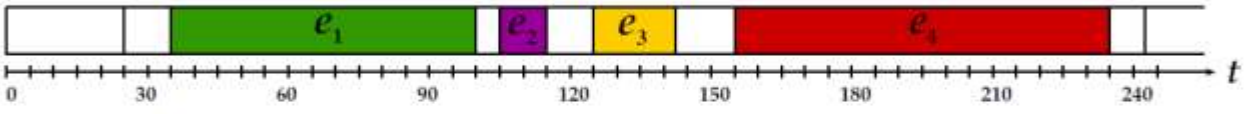
(a) Reference sequence containing the sub-sequence e_1, e_2, e_3, e_4 against which the other are compared.



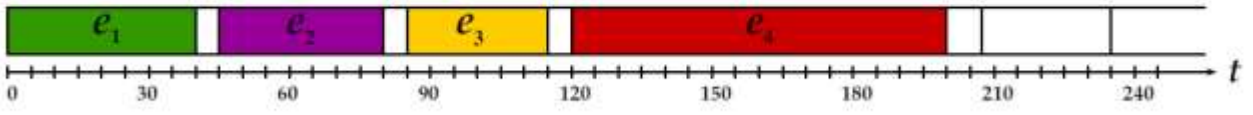
(b) The start times of the sub-sequence events are equal to the reference, their duration and intervening periods differ.



(c) The duration of each of the sub-sequence events is the same, their start times and the intervening periods differ.



(d) The sub-sequence is identical to the reference but is displaced in time.



(e) The total sub-sequence has equal duration to the reference, but the distinct events have different start times and durations.

CRESEARCH

LiU

Similarity measures

Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

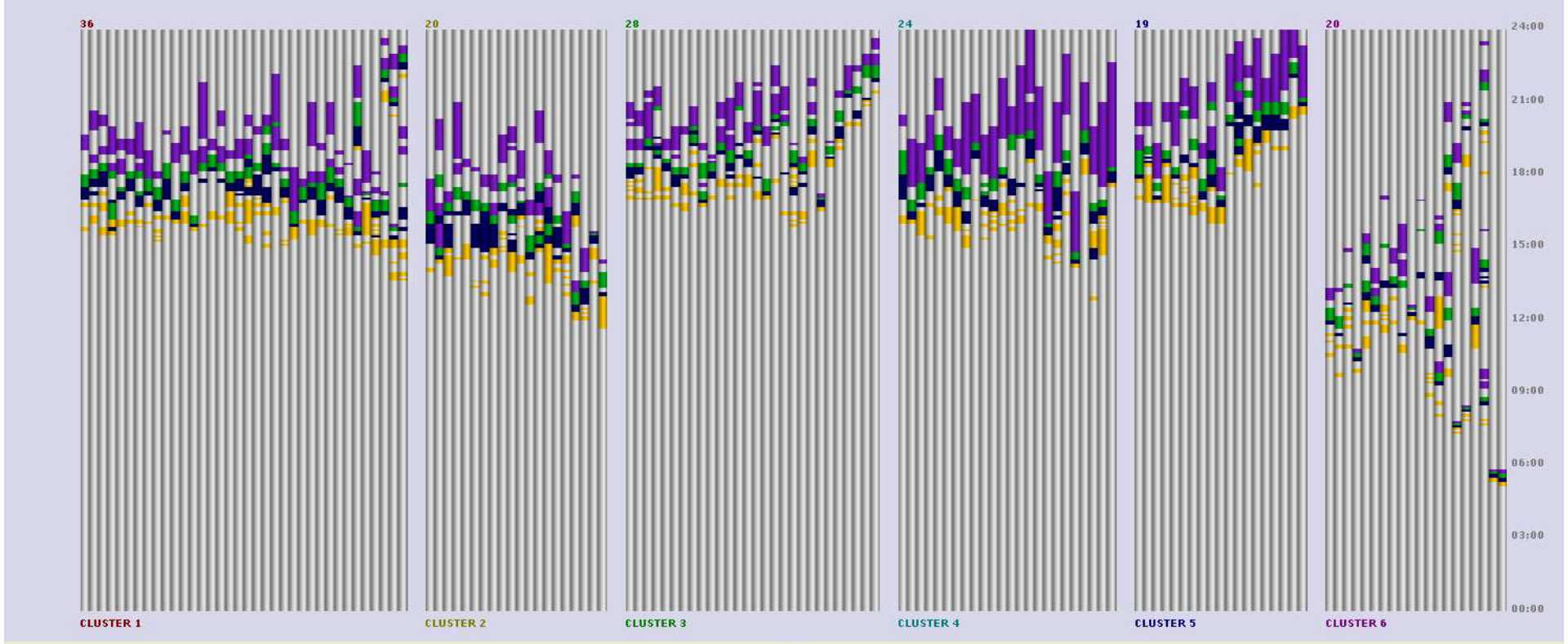
1. Fragmentation
2. Variation
3. Number of occurrences
4. Sequence start time
5. Sequence duration
6. Events duration
7. Gap size
8. Gap duration
9. Gap type

Record measures

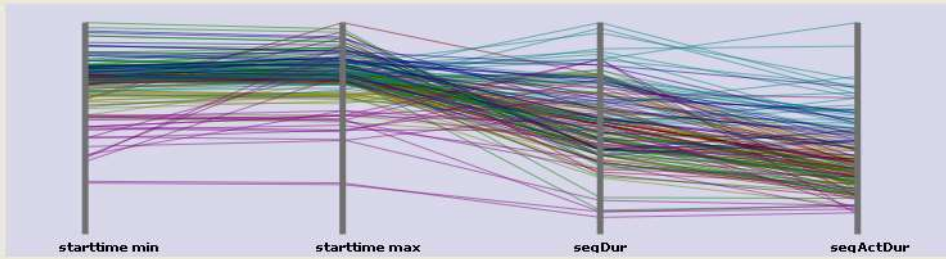
Query-match measures

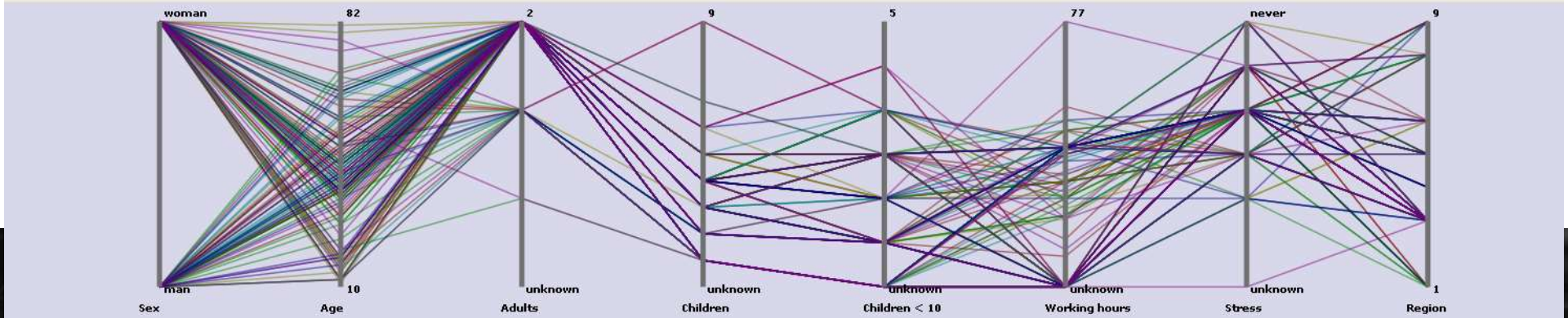
Gap measures

CRESEARCH


36
20
28
24
19
20
24:00
21:00
18:00
15:00
12:00
09:00
06:00
03:00
00:00
CLUSTER 1
CLUSTER 2
CLUSTER 3
CLUSTER 4
CLUSTER 5
CLUSTER 6


CL 6
CL 5
CL 3
CL 4
CL 2
CL 1
BAD
GOOD
starttime min
starttime max
seqDur
seqActDur


starttime min
starttime max
seqDur
seqActDur


woman
man
Sex
82
10
Age
2
unknown
Adults
9
unknown
Children
5
unknown
Children < 10
77
unknown
Working hours
never
unknown
Stress
9
1
Region

Currently ongoing project

Introduction
Background
VISUAL-TimePAcTS
Highlighting
Mining
Exploring
Clustering
Conclusions

- Data collection (PODD)
- Change in high-dimensional temporal data
  - Domain expertise in focus
- Interactive sequence mining
  - Domain expertise in focus
- Similarity of activities – activity landscapes

CRESEARCH

Thank you!

CRESEARCH

www.liu.se

LiU expanding reality