Preferred citation style for this presentation

Schlich, R . (2001) Homogenous groups of travellers,
paper presented at the 10th International Conference on Travel Behaviour Research, Lucerne, August 2003.

Moving through nets:
The physical and social dimensions of travel
10th International Conference on Travel Behaviour Research
Lucerne, 10-15. August 2003

Homogenous groups of travellers

Robert Schlich

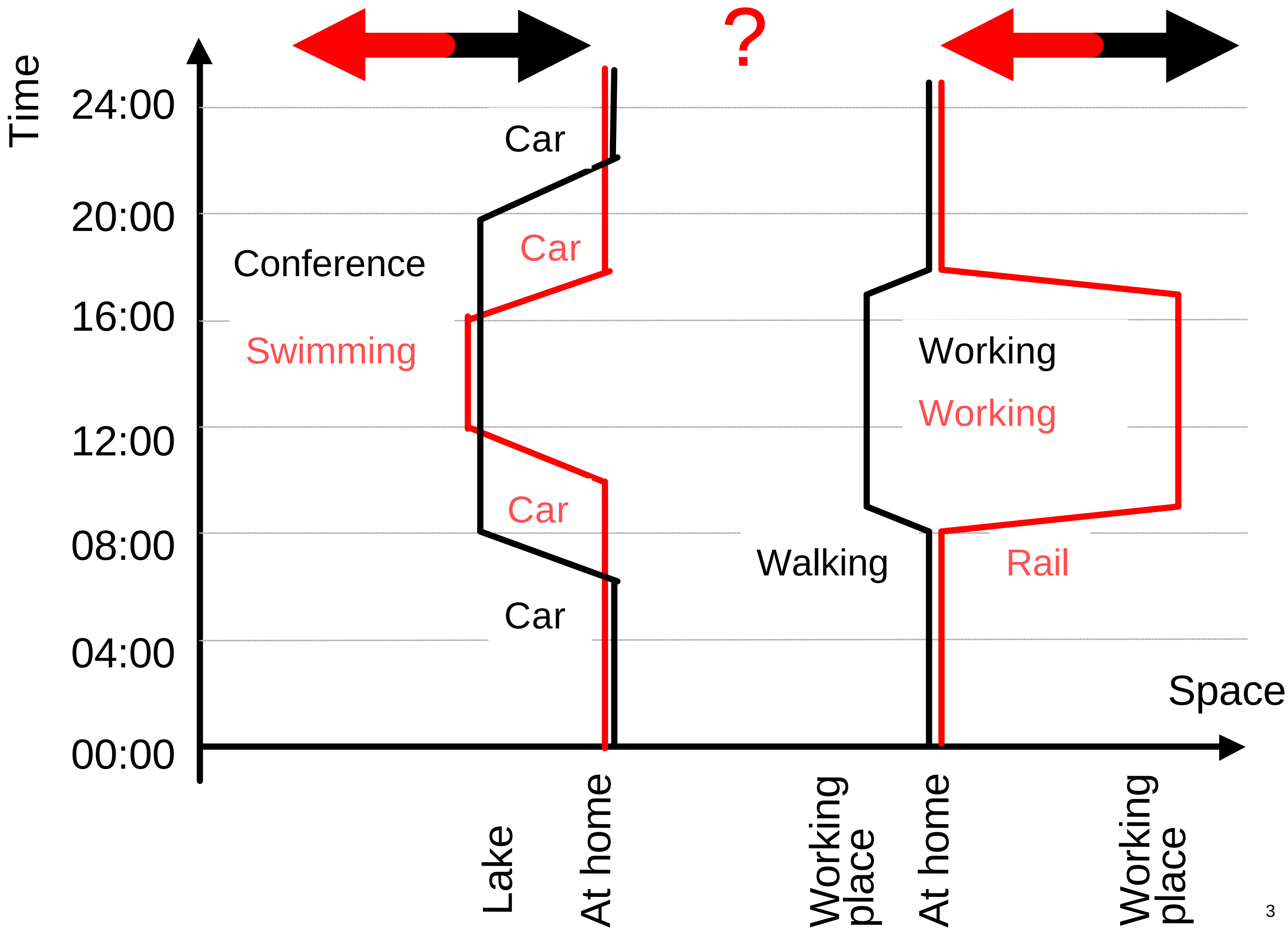
IVT
ETH
Zürich

August 2003

Institut für Verkehrsplanung und Transportsysteme
Institute for Transport Planning and Systems

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Measuring similarity



Theory of sequence alignment I

Measuring differences between two strings *s* $[s_1, s_2, \ldots s_n]$ and *g* $[g_1, g_2, \ldots g_n]$

$$d(s,g) = \sum_{i=1}^{n} f(x) \quad \text{and} \quad \begin{array}{l} f(x) = 1 \text{ if } s_i \neq g_i \\ f(x) = 0 \text{ if } s_i = g_i \end{array}$$

Example:

s=ABCDE

g=AFBCDE

d(*s*,*g*)=4

→ Problem of recognising sequential order or duration

Theory of sequence alignment II: Levenshtein

Similarity as total amount of effort to equalise $s[s_1, s_2,.....s_n]$ and $g[g_1, g_2,.....g_n]$

Four basic operation:

- Identity: $w_e(s_i,g_i)=0$
- Insertion: $w_i(Æ,g_i)=1$
- Deletion: $w_d(s_i, Æ)=1$
- Substitution: $w_s(s_i,g_i)=w_d(s_i,g_i)+w_i(s_i,g_i)=2$

Definition Levenshtein Distance:

Smallest sum of operation weighting values required to change $s[s_1, s_2,.....s_n]$ into $g\ [g_1, g_2,.....g_n]$

Theory of sequence alignment III: Trajectories

- Different possibilities to equal two strings
- Combination of operations are called trajectories

Example

s=CAMBRIDGE

g=CAMPING

1) substitute s_4(B:P), s_5(R:I), s_6(I:N), s_7(D:G) delete s_8(G), s_9(E) => d=10

2) substitute s_4(B:P), delete s_5(R), substitute s_6(D:N), delete s_8 (E) => d=6

Theory of sequence alignment IV: Problems

Different attributes of a trip are semi-dependent

- easiest possibility: Sum of „unidimensional“ sequence alignments across all attribute, not appropriate
- most exhaustive: calculate all possible trajectories across all attributes, not possible due to problems with computing times
- compromise: Optimum trajectory based sequence alignment (OT MDSAM) (Joh et al. 1999)

Software

Dana (C.H. Joh)

- Multidimensional
- Restricted number of allowed elements per string
- Restricted possibilities to change operation weights

ClustalG (C. Wilson, A. Harvey, and J. Thompson)

- Unidimensional
- Large strings allowed
- Better possibilities to change operation weights

Optimize, TDA

Dataset Mobidrive

- Reporting period: Six weeks
- Travel diary, weekly send out, mailed back and checked via phone
- Cities of Karlsruhe und Halle/Germany
- 162 households, 361 persons
- ca. 52.000 trips and 15.000 days reported September - November 1999 (Pretest: May-July 1999)

Comparison 1: Persons

Dimension	Variables chosen
Trip purpose	Share of leisure, school, work, shopping [%]
Timing	Share of trips in the morning [%] Share of trips at weekends [%]
Duration	Mean duration / trip[min]
Distance	Mean distance / trip[min]
Trip Mode	Share non-motorised, public transport, private motorised transport [%]
Frequency of trips and immobile days	Number trips/ day [N] Share of immobile days [%]
Intrapersonal variability	Levenshtein distance
Coupling constraints	Number of accompanying persons [N]

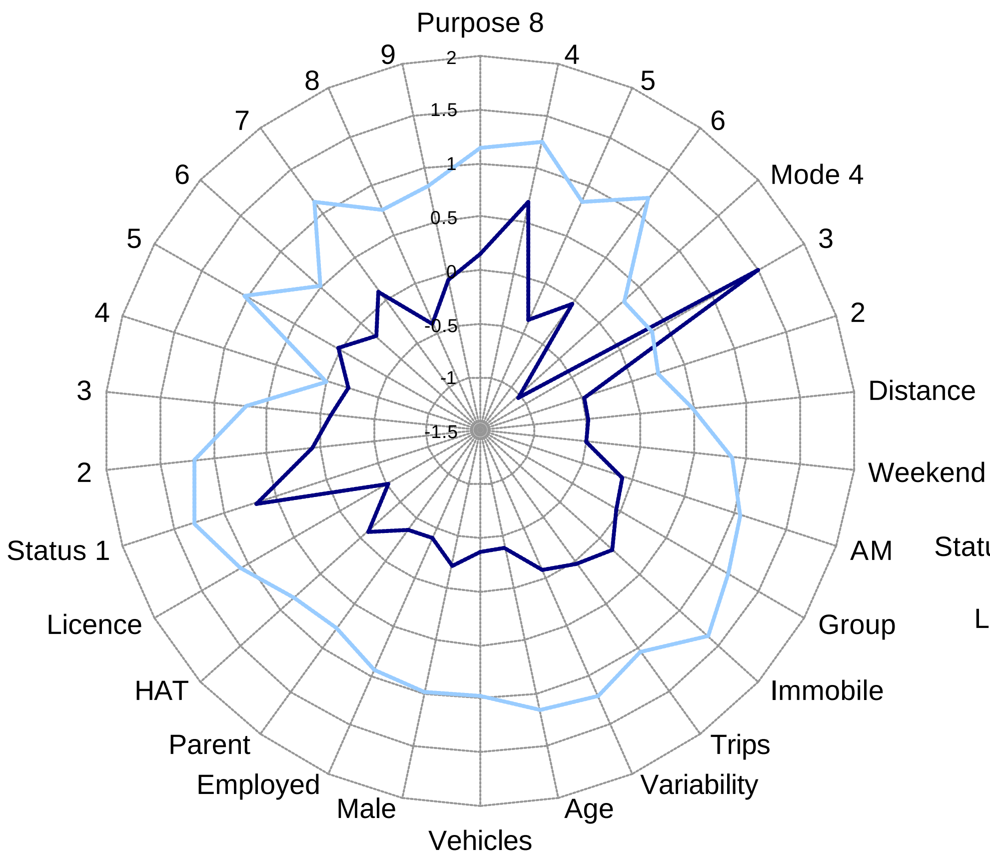
Comparison 2: Random days

Problem of OT MDSAM between all days: Computing time

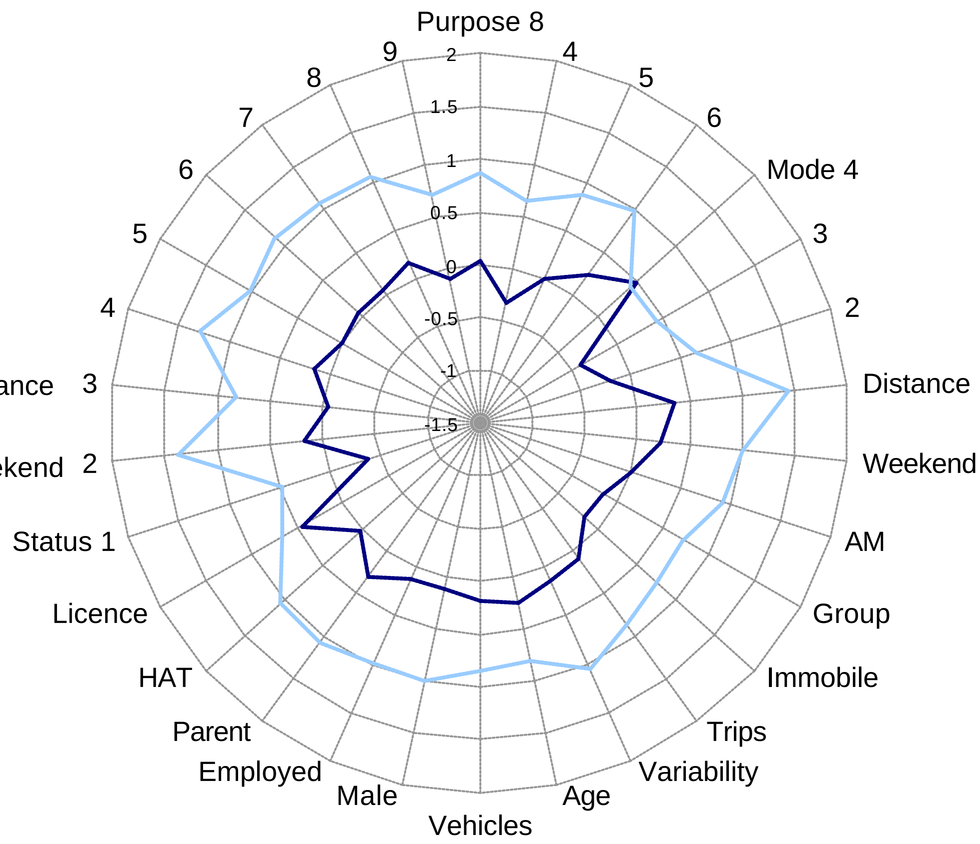
- about 15200 days in Mobidrive; 115 million comparisons
- 170 comparisons: 1 Minute
- total computing time for comparing all days: 15 months

- Initial compromise: one random (week)day per person
- SQA used for inter-personal comparisons

Examples: Clusters for person-attribute matrix

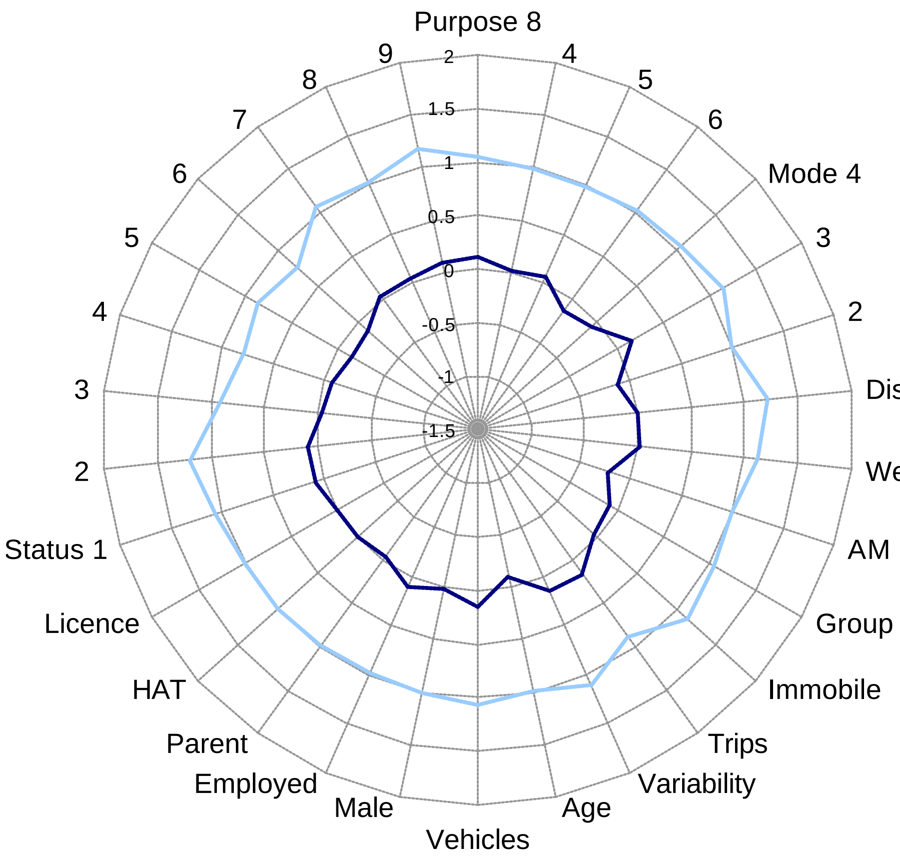


Cluster 3

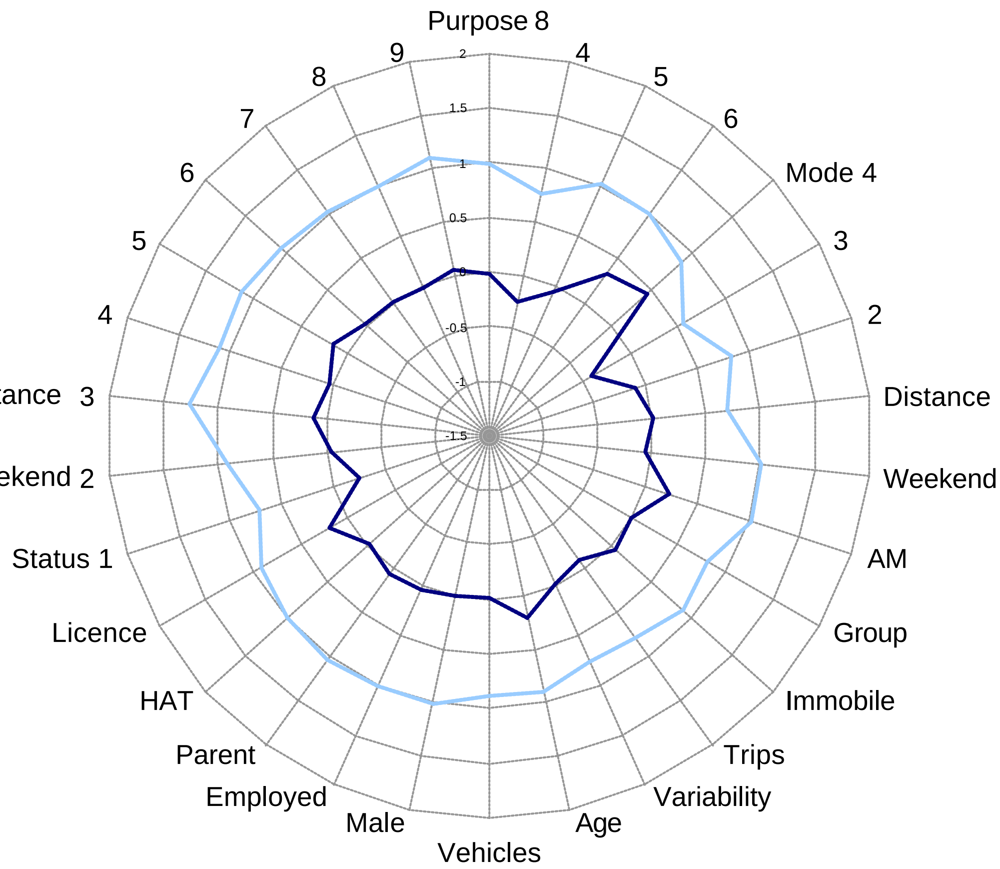


Cluster 4

Examples: Clusters for random days matrix

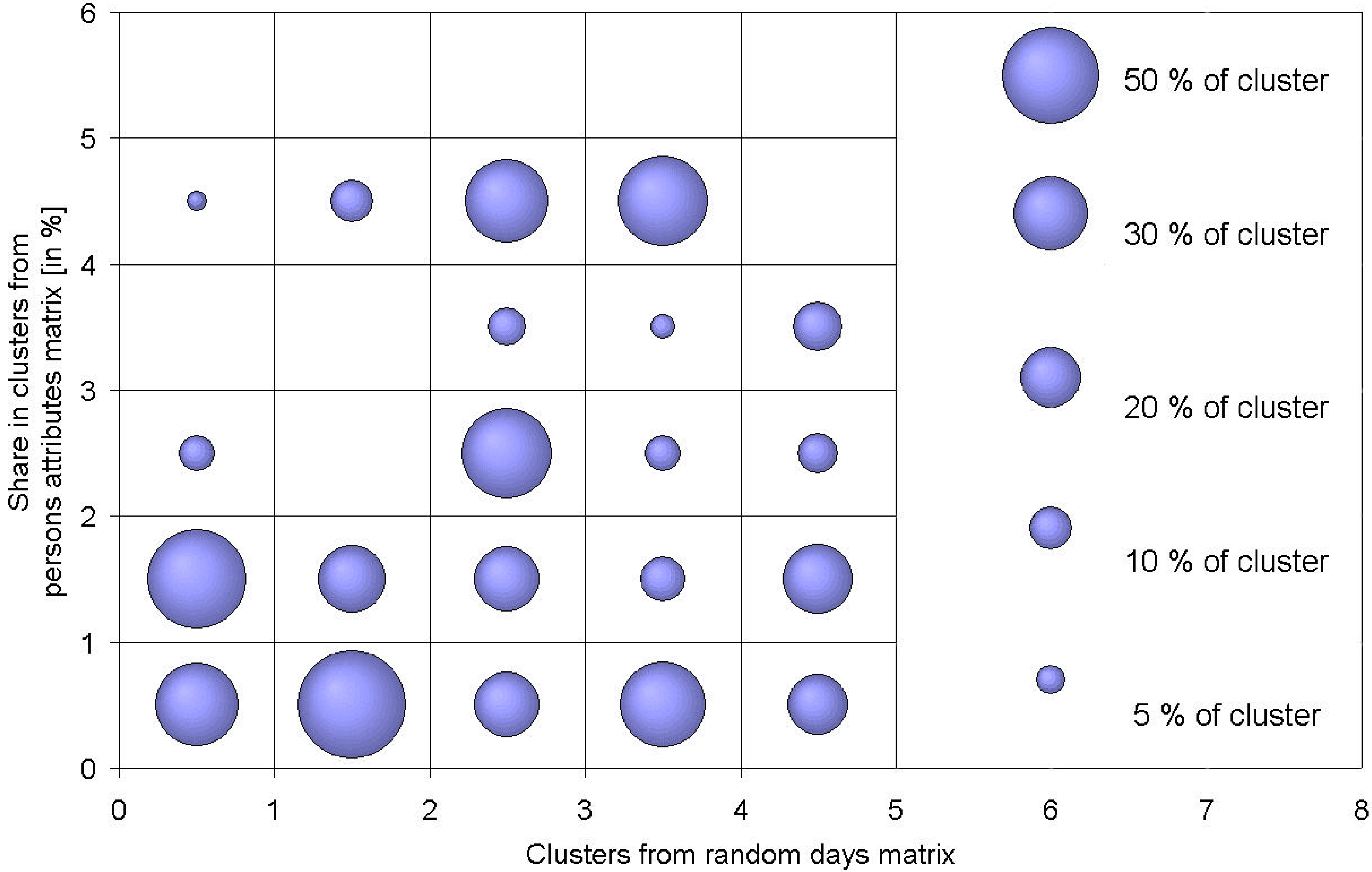


Cluster 1

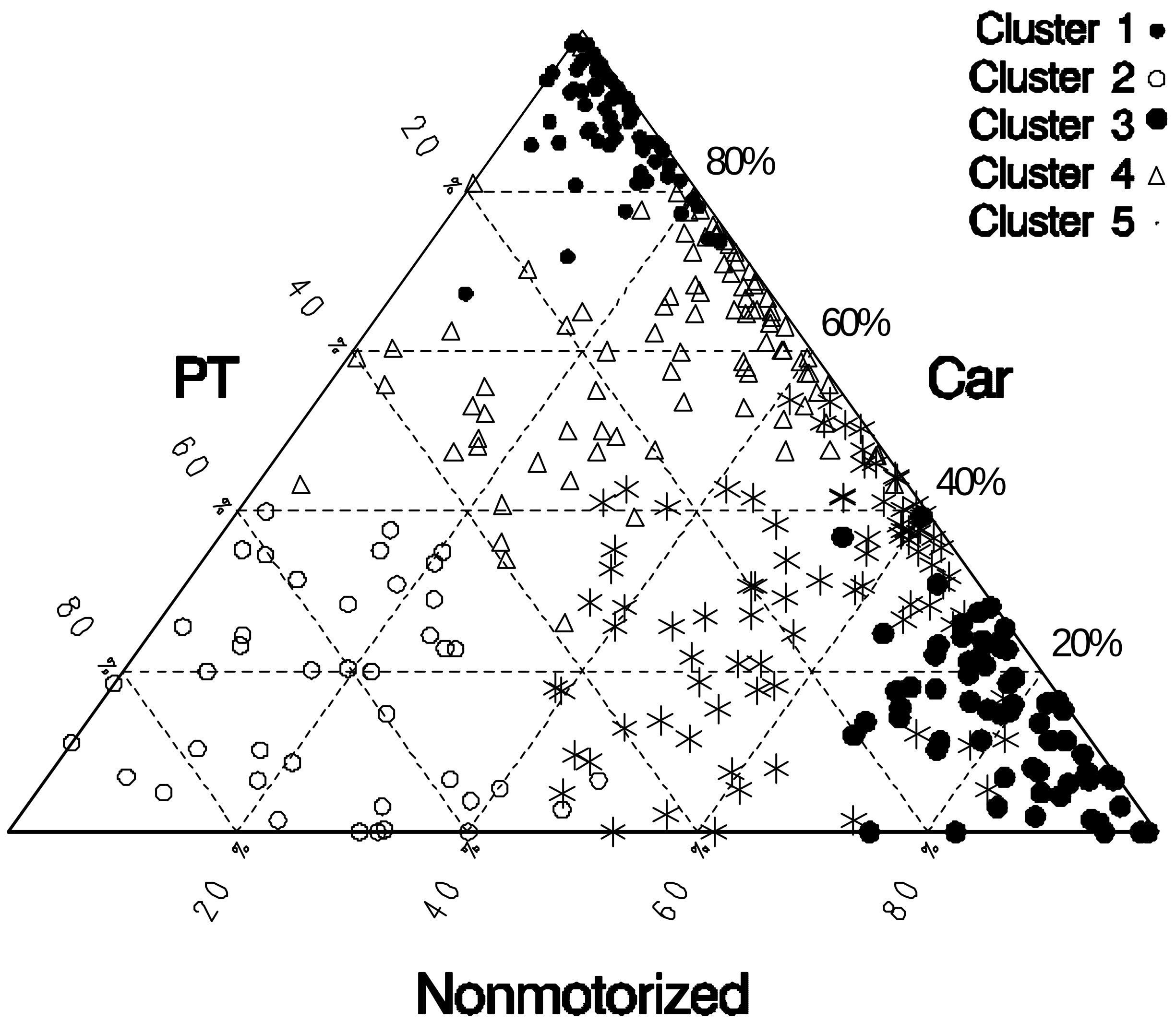


Cluster 3

Cross classification



Person attribute matrix: Share of modes



Description of the person-attribute matrix clusters

Cluster 1: “Working men”

++	distance per trip, share of working trips, male persons, employed person, cars per household, morning trips, car trips
+	immobile days, parents
o	number trips/day, intrapersonal variabilty
-	shopping trips

Cluster 2: “Stable behaviour”

++	school trips, leisure trips, pupils, young persons, public transport
+	employed persons
--	car trips, intrapersonal variability, number of trips per day, shopping trips

Description of the person-attribute matrix clusters

Cluster 3: “Local Cluster”

++	school trips, pupils, young persons, unmotorised trips, retirees
+	share of immobile days, women, trips in morning
	employed persons, parents, trips at weekend
--	distance per trips

Cluster 4: “Active families”

++	parents, trips per day, intrapersonal variability
+	employed persons, average distance per trip, car trips
--	immobile days

Description of the person-attribute matrix clusters

Cluster 5: “Average cluster”

- \+ unmotorised trips
- o employed persons, age, parents, different trip purposes, number of trips/day intrapersonal variability
- \- average distance

Summary

Classification based on a comparison of person attributes

- 5 cluster solution
- Good differentiation in terms of travel characteristics
- Reasonable differences for the sociodemographic characteristics

Classification based on a comparison of one random day with multidimensional sequence alignment :

- 5 cluster solution does not give different clusters in terms of sociodemographics
- Additional information from order of activities

Outlook

Further research: Sequence alignment

- Check for more than one random day
- Check robustness of the approach
- Check other classification methods

Further research: Travel behaviour

- Relevance for transport policy