Daily travel behavior: Lessons from a week-long survey for the extraction of human mobility motifs related information*

Christian M. Schneider
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
schnechr@mit.edu

Christian Rudloff
Austrian Institute of Technology
Giefinggasse 2
1210 Vienna, Austria
christian.rudloff@ait.ac.at

Dietmar Bauer
Austrian Institute of Technology
Giefinggasse 2
1210 Vienna, Austria
dietmar.bauer@ait.ac.at

Marta C. González
Massachusetts Institute of Technology
77 Massachusetts Avenue
Cambridge, MA 02139
martag@mit.edu

ABSTRACT

Multi-agent models for simulating the mobility behavior of the urban population are gaining momentum due to increasing computing power. Such models pose high demands in terms of input data in order to be reliably able to match real world behavior. To run the models a synthetic population mirroring typical mobility demand needs to be generated based on real world observations. Traditionally this is done using travel diary surveys, which are costly (and hence have relatively low sample size) and focus mainly on trip choice rather than on activities for an entire day. Thus in this setting the generation of synthetic populations either relies on resampling identical activity chains or on imposing independence of various trips occurring during the day. Both assumptions are not realistic.

Using Call Detail Records (CDRs) it has been found that individual daily movement uses only a small number of movement patterns. These patterns, termed motifs, appear stably in many different cities, as has been shown for both CDR data as well as travel diaries.

In this paper the relation between these motifs and other mobility related quantities like the distribution of travel distances and times as well as mode choice is investigated. Additionally transition probabilities both for motifs (relevant for multi-day simulations) and mode transitions are discussed.

The main finding is that while some of the characteristics seem to be unrelated to motifs, others such as mode choice exhibit strong correlations which could improve the provision of synthetic populations for multi-agent models.

Thus the results in this paper are seen as one step further towards the creation of realistic (with respect to mobility behavior) synthetic populations for multi-agent models in order to analyze the performance of multi-modal transportation systems or disease spreading in urban areas.

Keywords: human mobility; multi-agent models; mobility demand modeling; motifs.

ACM Classification Codes:
I.6.5: SIMULATION AND MODELING: Model Development
J.2: Computer Applications: PHYSICAL SCIENCES AND ENGINEERING;

*Acknowledgments: This research was partially funded by the EMPORA project (financed by the Austrian KLIEN). Travel survey data has been made available from Verband Region Stuttgart which is gratefully acknowledged.
Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org. UrbComp13, August 11-14, 2013, Chicago, Illinois, USA. Copyright ©2013 ACM 978-1-4503-2331-4/13/08...$ 15.00

1. INTRODUCTION

Models of daily mobility are used in a number of applications including urban transport planning and monitoring [8] or disease spreading [7]. These models are typically based on multi-agent simulation tools such as MATSim [6] requiring as inputs the daily mobility profiles of the modeled population. In this respect the sequence of places visited and the means of transport used for the trips between visited places are the most important ones among various other aspects of

the daily travel behavior.
Traditionally the travel behavior of a sample from the population is collected for one or more days through travel diaries by household travel surveys (see e.g. Kitamura *et al.* [5] or Ortuzar *et al.* [8]). The resulting travel diaries are projected on the entire population to estimate the total travel behavior. This method is both time consuming and expensive as it incorporates time intensive interviews to ensure correct information. Moreover, accuracy of the corresponding findings is not guaranteed as surveying mostly aims at obtaining information on single trips while multi-agent models center around activities and trip chains.

To collect information about changes in daily behavior, travel surveys of up to six week have been collected (see e.g. Axhausen *et al.* [1]). While this data is more reliable with respect to recurring phenomena, new problems like missing data through survey fatigue emerge. Another approach to improve data accuracy in travel surveys is using GPS enabled devices to collect data (for an overview about travel survey techniques see Stopher *et al.* [11], for recent advances in this respect see [3]). However, improving the methodology of data collection does not avoid the problem of estimating the behavior of the entire population from a relatively small amount of data.

In recent years a new data source became available through the collection of mobile phone data from call details records (CDR). Usually CDRs contain information about the calling time, its duration and the serving cell phone tower (calls include voice and text messages). Such data arises naturally as phones need to be localized for enabling communication with the network. The resulting traces of movement provide another data source for inferring mobility behavior. Contrary to travel diaries this data collection method imposes no burden on the surveyed persons. As mobile phones are ubiquitous, the sample size is significantly larger, typically ranging in several tenth of the population depending on market shares of the corresponding provider.

At the same time the amount of information collected from an individual depends crucially on their phone usage. Frequent users provide ample information, persons that only use their phone seldom leave less detailed traces with the potential of missing trips altogether. Furthermore, the management of the mobile phone network and the construction of the network pose difficulties in matching observed movement in the mobile phone data sets to real movement. Examples for this are load balancing and hierarchies in cell coverage (see Ray and Schrom-Feiertag [9] for examples).

Therefore, it is not obvious that mobile phone related data can be used in order to extract information on daily mobility of the entire population. Consequently, much research was recently directed towards exploring the potential of this new data source. First steps of extracting mobility behavior from CDR data were already taken in Gonzalez *et al.* [4] where human travel patterns are studied, in Caceres *et al.* [2] where origin-destination data is extracted or in Toole *et al.* [12] where land use was inferred from CDR records. Wang *et al.* [13] provide details on subsample selection and accuracy of commuting flow estimation based on mobile phone data to generate origin-destination matrices.

In Schneider *et al.* [10] a first step of comparing daily mobility behavior (in contrast to trip based approaches) extracted from mobile phone data with data from travel surveys was conducted. It was shown that the number of visited places per day and the order of these visits can be extracted from mobile phone data resulting in so called *human mobility motifs*. Motifs are networks consisting of nodes and directed links, representing the locations where individuals perform activities and the trips between these locations, respectively. While millions of possible combinations exist to form daily networks, surprisingly only a limited number of them are of statistical significance. These 17 daily networks are called motifs, all with less than six activity locations categorized by four rules (see below for details).

The motifs appear not only in mobile phone data, but also in daily travel diaries from different cities in different countries. In this work, we investigate the motifs in more detail based on survey data in two cities:

i) We connect the motifs with travel distance, travel time and travel mode.

ii) We show that motifs created with the same rule but different locations show similar behaviour.

iii) We verify the previously unsupported claim that an individual has a characteristic motif over several days with our survey data.

Based on [10] we show that daily travel behavior as summarized by some indicators is strikingly similar for different cities, in our case for Chicago (USA) and the Stuttgart area (Germany). However, the behavior shows differences that can be explained by the difference in population densities, unemployment rate and urbanisation of the two study regions.

All these findings are important for multi-agent simulators to build realistic models for human travel activity in urban areas. They are a step further on the path to foster the combined use of CDR and survey data in order to provide reliable, accurate input data for multi-agent simulations used in the prediction of urban transportation system performance and disease spreading.

The organization of this paper is as follows: In section 2 of the paper we first describe the available data sources. Afterwards the frequency of occurrence of the different motifs, travel distances and travel times is compared for the two surveys (sub-section 2.2). In sub-section 2.3 the stability of daily motifs is tested for the week long Stuttgart survey data. In section 3 the occurrence of different modes and mode combinations is studied for the Stuttgart survey data. We conclude the paper with a summary of the results and an outlook to future research.

2. COMPARISON OF MOTIF OCCURRENCE AND CHARACTERISTICS FOR DIFFERENT SURVEYS

In [10] it was shown that daily mobility patterns extracted from two travel surveys (Chicago and Paris) as well as from mobile phone records from Paris gave very similar distributions of motifs. As these data sources all come from very similar densely populated metropolitan regions one of the questions arising is if the same holds for motif distributions in less densely populated areas. For this comparison data from a travel survey conducted in the Stuttgart Metropolitan Region is used. This region comprises the city of Stuttgart (population 600,000) as well as neighboring cities and towns. While the regions are comparative in area, the population density of the Stuttgart area is with 343/km^2 sparser than the Chicago area with 509/km^2.

2.1 Available Data

The Stuttgart travel survey contains information of 13731 persons living in 5581 households for one week during the period from 14.9.2009 until 26.10.2010. A total of 275898 trips (average of 20 trips per person or roughly 3 per day) are recorded and for each trip the typical trip characteristics (start time, end time, starting point, end point, duration, distance, purpose, primary mode) are collected. The socio-demographic variables of each individual are also included in the data set. Locations are geocoded in terms of longitude and latitude values, which are rounded to two decimal places effectively anonymizing the data.

The Chicago activity survey took place in 2007 containing data of 32366 persons living in 10552 households and the number of recorded activities is 159856. The data is structured in a slightly different fashion having an entry for each activity performed on the two days collected per person. The information for each activity are a starting and end time, a location, the duration of the trip from the location of the last activity, the duration of the activity, distance from the last activity, the purpose of the stop and the mode used to reach the activity. The first activity of the day always starts at 3 am and ends at 3 am on the following day. Up to two days of data are collected per person. Different locations are assigned unique ids and are geocoded using the longitude and latitude of the census tract they are in. Like the Stuttgart data the Chicago data contains different socio-demographic information about the persons and households.

2.2 Detected types of motifs

In a first step for Chicago and Stuttgart the distribution of occurrence of the various motifs in the sampled population is calculated and the results are shown in Figure 1. It can be seen that although the overall behavior is similar, the motif distribution in Stuttgart and Chicago shows differences.

There is a comparatively higher number of motifs with a small number of nodes. In particular the motifs with six stops (not shown in the figure) do not appear as often as in the Chicago data. However, the order in which the different motifs appear is similar in both cases. In Figure 2 it can be seen that there is a larger amount of longer daily distances in Stuttgart than in Chicago. This might lead to a tendency to avoid extra stops.

In [10] the most often occurring motif have been classified using four construction rules. The rules were defined in the following way: Based on the definition of a central node C, let $T(n)$ denote a cyclical tour with n locations starting at C and ending there. Then a construction rule is a set of

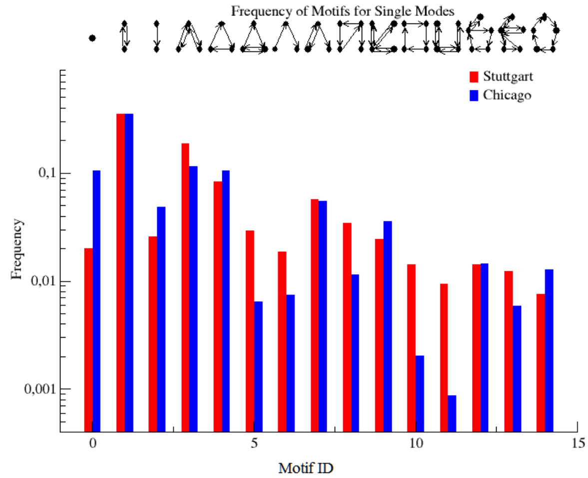

Figure 1: Occurrence frequency of motifs for different cities.

cycles in the motif starting at C. Using this notation the predominant four construction rules for the most frequent motifs are I) T(1), T(n-2), II) T(n-1), III) T(2), T(n-3), IV) T(1), T(1), T(n-3). That is construction rule I) consists of a long tour plus a short tour visiting an additional node not contained in the long tour, rule II) only of one cyclical tour with no other nodes visited. Rule III) has one long and one short tour without joint locations, rule IV) finally features two back and forth trips and a long cycle. The most frequent motifs fall into these four construction rules.

If one disaggregates the traveled distance according to the various motifs and calculates the distributions of traveled distances for given motif with distances rescaled by dividing by the mean motif travel distance, then Figure 3 shows the similarities between Chicago and Stuttgart for motifs following construction rule II, i.e. round trips for $n = 1$ up to $n = 4$ locations different than the center C.

Looking at the travel times of motifs in Figure 4, it can be seen that for motifs constructed using the same rule the distributions are not only very similar for the two cites but for all motifs with the same construction rule. While for the travel distances rescaling is necessary, for the travel times it is not. This might be due to the difference in the study areas. While in the Stuttgart area, more trips might use overland highways, the Chicago area is densely populated and most traffic is on roads with strict speed restrictions. People seem to travel very similar times, but due to higher speeds, shorter distances are traveled.

By scaling the distance distribution in relation to the mean distance of motifs for the different motif construction rules, the same motifs exhibit very similar distance distributions (Figure 3). Focusing on the travel times of motifs, it can be seen that all motifs constructed using the same rule the distributions are very similar independent of the city.

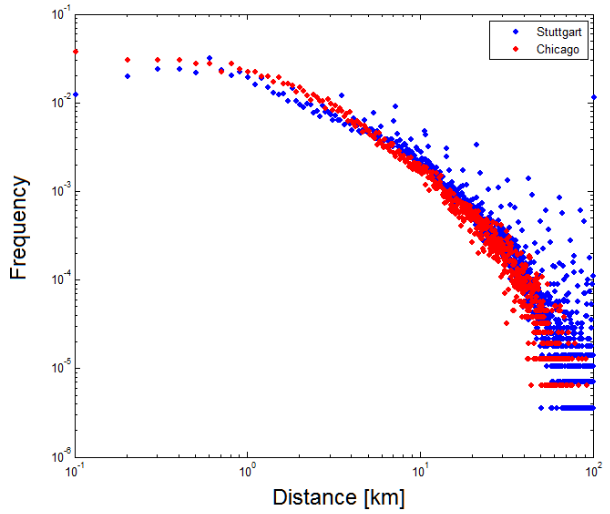

Figure 2: Difference in travel distances between daily trip routines in Chicago (red) and Stuttgart (blue).

2.3 Weekly chains of motifs and the usability of mobile phone data for motif generation

In [10] the co-occurrence of motifs over several weekdays was determined using mobile phone data from Paris. Due to the non-availability of suitable multi-day survey data sets for Chicago and Paris, a verification of this finding was not possible. Thus it is an open question whether the proposed relations found in the mobile phone data occur in travel diaries or the relation is only an artifact of the data collection methodology using mobile phone data.

As the Stuttgart travel survey covers multiple days for each person, the correlation matrix of [10] can be reproduced for travel diary data. The result are shown in Figure 5. This matrix shows qualitatively similar pattern as the matrix shown in Figure 4 of [10].

As an example this shows that persons switch between motif 6 (round trip with four locations) and 10 (round trip with five locations) often as well as between 7 (back and forth to three separate locations) and 11 (two times back and forth and one cycle with two locations). In contrast a switch between 6 and 7 is not likely. In the first two cases only one new location is added to the general pattern rather than changing the pattern altogether.

Figure 6 provides information on the corresponding transition matrix. The transition matrix contains as row i the distribution of motifs for the following day given that today motif i was chosen. Thus different to co-occurrence the transitions also take the temporal ordering into account. In Figure 6 the transition matrix is compared to the overall motif distribution: The image shows color coded the ratio between the conditional distributions and the overall distribution such that a value greater than one indicates higher than expected follow up by a motif and a value lower than one less likely following of the row motif by the column motif.

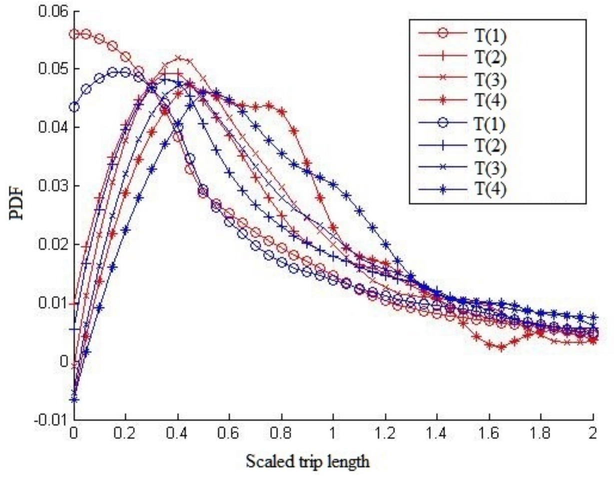

Figure 3: Scaled and smoothed travel distance distributions for motifs constructors according to rule II) in Chicago (red) and Stuttgart (blue).

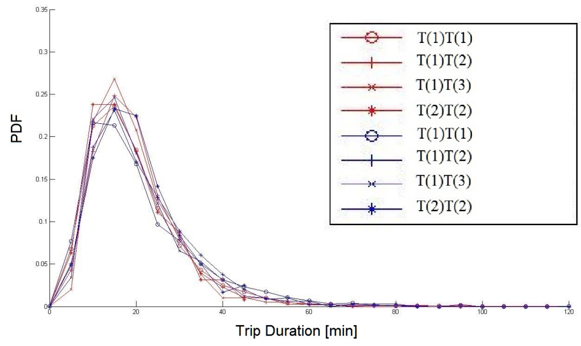

Figure 4: Travel time distributions for motifs given as a histogram with 15 minute bins in Chicago (red) and Stuttgart (blue).

Clearly persons showing one motif on one day are very likely to follow the same routine on the other days. This is especially pronounced for motifs 4 (one round trip to two locations other than the center C and one back and forth to one of them) and 8 (back and forth between three locations) while transitions between 6 and 10 and 7 and 11 respectively are occurring relatively frequently. Furthermore for each motif there is only a limited set of alternative motifs that are more likely to occur conditional on the previous day compared to overall. In other words the conditional distributions differ significantly from the unconditional ones.

3. MOTIFS AND MODE CHOICES

Besides the motifs also the modes of transport used for the trips are of interest. In the following the interplay between mode choice and motifs is studied for the Stuttgart survey.

Figures 7 and 8 show the frequency of the most common motifs in the Stuttgart survey for people using either only

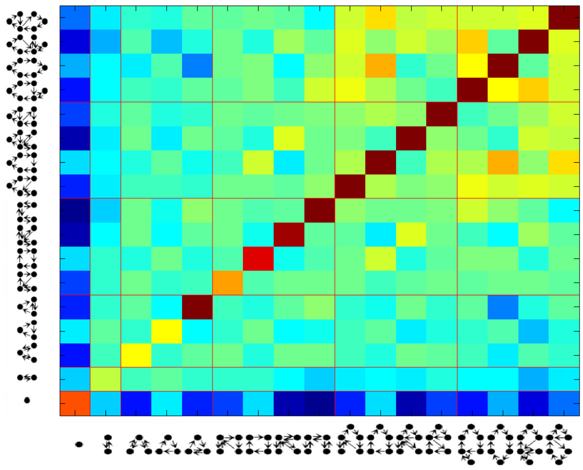

Figure 5: Co-occurrence of motifs (calculated as in equation [3] of [10]) over observation period in Stuttgart data set. Color coded according to MATLABs HSV colormap: blue refers to low values, red to large values.

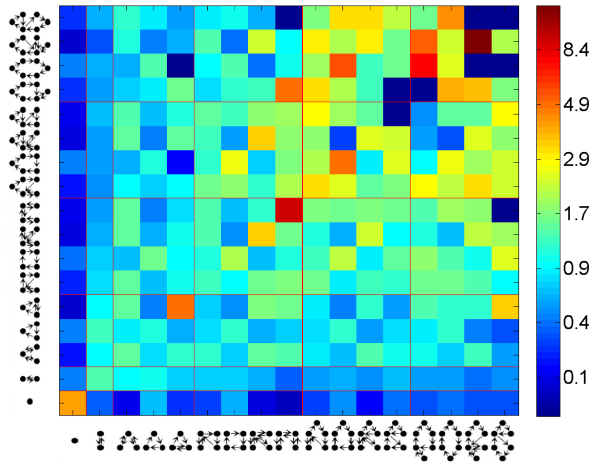

Figure 6: Relative transition matrix of motifs over weekdays in the observation period in Stuttgart data set: Each row corresponds to the conditional distribution for the next day of the motif corresponding to the row relative to the overall frequencies.

one mode for their daily tours or two modes, respectively. Motifs with only one mode used back-and-forth movement starting at the central node occurs with higher frequency than overall expected. For longer cycles this is only the case for car and car passenger trips. For motifs returning to the central node more often, the frequency is lower. This suggests, that mode changes do not occur within cycles very frequently but happen to a large extent at the central node. Furthermore, in particular on foot (less pronounced for bicycles), motifs with a small number of edges occur with much higher frequency.

The findings described above are confirmed in Figure 8 where the relative frequencies of the motifs for people using two modes on a day are given. One can see that motifs with one or more returns to the central node are clearly more frequent than in the entire data. It is also interesting to notice that driving a car is not a mode that is combined often with other modes as the combinations presented in the graph are the most frequent combinations in the data. This suggests that car drivers have a very small willingness to switch to a different mode. An exception to the mode changes at a central location are those where a longer cycle is split into walking/public transport and car passenger/public transport. It seems that travelers often perform one or more stages of a public transport motif by foot or are either driven or picked up somewhere and use public transport for the rest of their journeys.

Above the distribution of motifs for a given mode combinations are discussed. But also the other conditional distributions (i.e. conditioning on the motif rather than on the mode used) show connections between motif and mode choices: Inspecting the conditional distribution of different combinations of modes used conditional on the chosen motif reveals slight deviations from randomness. Figure 9 provides a plot of the distributions conditional on motif chosen relative to the overall distribution. A number of interesting facts can be seen: For the simplest motif (two locations, going back and forth) relatively often only one mode is used or one trip is made as car passenger, the other using public transport. Tours with four or more locations are more often than average made using three or more modes and less than average using only one mode with the exception of driving a car.

Beside the overall mode usage also the switching between different modes is of interest for daily tours where several modes are used. In Table 1 the transfer probabilities between different modes are presented for all observations. One can see that staying on the same mode is the preferred choice overall with a chance of at least 70% of continuing the next trip in the mode of the last trip. Car drivers are most predictable remaining in their cars at even 93% of the occasions.

Table 1: Transfer probabilities between different modes (Bike, Walk, Car Driver (D), Car Passanger (P), Public Transport (PT)).

from , to	Bike	Walk	D	P	PT
Bike	0.7890	0.0485	0.0735	0.0567	0.0322
Walk	0.0147	0.7083	0.0887	0.0913	0.0963
D	0.0049	0.0295	0.9310	0.0239	0.0104
P	0.0066	0.0667	0.0657	0.7975	0.0625
PT	0.0130	0.0968	0.0528	0.0910	0.7444

Within tours beginning and ending at the same location this is even more pronounced. In Table 2 this can be seen for two examples, taking only observations arising in motifs $T(1)$ and $T(2)$ respectively, i.e. the round trip with two and three different locations. In particular, it can be seen that the tendency to change mode when a car is driven or a bike

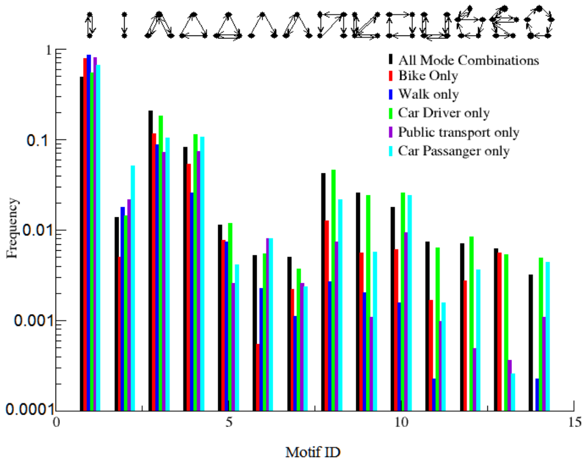

Figure 7: Frequency of motifs for days with a single chosen mode.

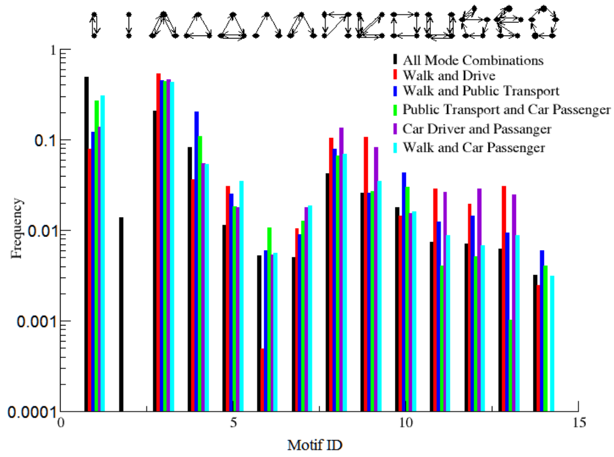

Figure 8: Frequency of motifs for days where two different modes are used.

ridden is very small within a tour, both for one intermediate goal and two intermediate goal. On the contrary, the introduction of the second intermediate goal drastically reduces transition probability from walk to walk from almost 90% to only 37% with a huge increase of public transport usage (rising from 2% to 43%).

Table 2: Transfer probabilities for two different motifs of length two and three.

T(1)	Bike	Walk	D	P	PT
Bike	0.9409	0.0186	0.0132	0.0195	0.0078
Walk	0.0081	0.8979	0.0148	0.0544	0.0242
D	0.0016	0.0079	0.9790	0.0087	0.0207
P	0.0033	0.0679	0.0301	0.8199	0.0777
PT	0.0020	0.0151	0.0058	0.0.385	0.9377
T(2)	Bike	Walk	D	P	PT
Bike	0.8697	0.0252	0.0252	0.0504	0.0294
Walk	0.0066	0.3693	0.0854	0.1130	0.4258
D	0.0006	0.0119	0.9668	0.0123	0.0078
P	0	0.0499	0.0438	0.8380	0.0657
PT	0.056	0.1671	0.0238	0.0873	0.7135

Finally as expected at the central location, the transition probabilities for mode changes are much higher compared to the general mode change probabilities (See Table 3). As the car is the dominant mode in the survey, the changes into driving a car are more frequent (conditional frequencies of more than 20% for all modes) than for the other modes. But also for biking the probabilities to change from bike to another means of transport at the central location is more than twice as likely compared to the overall frequency.

4. CONCLUSIONS AND FUTURE WORK

Previous results demonstrate that motifs represent basic movement patterns that consistently have been found in data sets for a number of regions in different countries. Therefore they should be the core building block for multi-agent mobility models.

Table 3: Transfer probabilities between different modes at the central location

from , to	Bike	Walk	D	P	PT
Bike	0.3772	0.1134	0.2439	0.1850	0.0804
Walk	0.0404	0.3890	0.2604	0.2132	0.0963
D	0.0202	0.0966	0.7796	0.0753	0.0280
P	0.0379	0.1543	0.2067	0.5207	0.0804
PT	0.0379	0.1825	0.2063	0.2428	0.3150

While the most important motifs have been found in many different data sets their relative importance differs depending on the characteristics of the considered region. The corresponding percentages can be obtained from CDRs building a good data base in this respect.

However, for movement simulation also other parameters are of interest. The results presented in this paper show that there are a number of interesting relations between motifs, travel distances, travel times and mode choice.

The results in this paper demonstrate that while motifs and the distribution of travel times appear to be unrelated, there are correlations between mode choice and motif choice as well as between mode transitions and position of the trip within the motif network structure. Also the motif chosen by individuals on any given day is not random, but persons are more likely to choose the same motif on another day instead of changing it. However, if they switch their motif, some motif combinations are more likely to occur than others.

Some of the results contained in this paper have only been obtained for the survey data from Stuttgart so far. Verification that the findings are stable also for other data sources such as travel diaries in Chicago and Paris are left for future research.

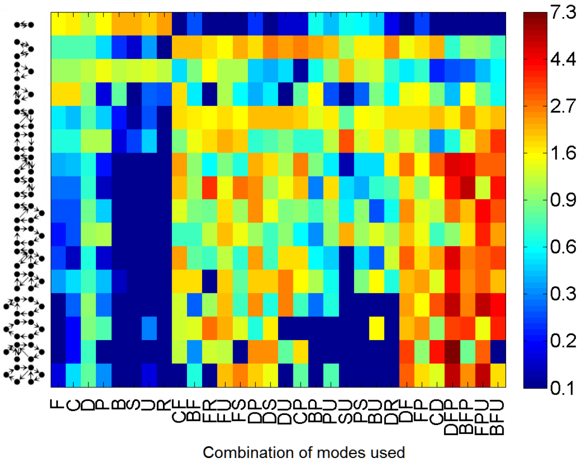

Figure 9: Relative frequency of motifs for trip chains with different mode combinations in Stuttgart. Modes: foot travel (F), cycle (C), car as a driver (D), car as a passenger (P), commuter-rail (S), subway (U), rail (R), bus (B).

Nevertheless the results presented in the paper suggest that these relations should be included in any simulation modeling the movement of persons over the course of one or several days. We have argued here that the data underlying such simulations should be combined from diverse data sources such as CDRs and travel diaries. The optimal combination of these data for obtaining realistic synthetic populations in the multi-agent setting is an open research question. Based on current knowledge methods to obtain the relative frequency of the most likely motifs from CDR data allows for a relatively low cost extraction of data for a large subsample of the population while the relation to mode choice and mode transitions could be obtained from travel diaries with smaller sample sizes as they do not need to be projected to obtain travel demand for the whole population but only relate to mode choice.

This paper is seen as the first step into this direction. However, the question how much can be gained in terms of realism by using the combination of travel diaries and CDRs as opposed to only using travel diaries or only using CDRs remains still as an open research question.

5. REFERENCES

[1] Axhausen, K., Zimmermann, A., Schönfelder, S., Rindsfüser, G., Haupt, T., Observing the rhythms of daily life: A six-week travel diary, Transportation, 29 (2), 95–124 (2002).

[2] Caceres, N., Widberg, J., Benitez, F., Deriving origin destination data from a mobile phone network, Intelligent Transportation Systems, IET, 1(1), 15–26, (2007).

[3] Cottrill, C. D., Pereira, F. C., Zhao, F., Dias, I. F., Lim, H. B., Ben-Akiva, M. E., Zegras, P. C., Future Mobility Survey: Experience in Developing a Smart-Phone-Based Travel Survey in Singapore, Proceedings of the Annual Meeting of the TRB, Washington, paper no. 13-4849 (2013).

[4] Gonzalez, M., Hidalgo, C., Barabási, A., Understanding individual human mobility patterns, Nature, 453, 779–782, (2008).

[5] Kitamura, R., Chen, C, Prndayala, R., Narayanan, R., Micro-Simulation of daily activity-travel patterns for travel demand forecasting, Transportation, 27, 25–51 (2000).

[6] MATSim: http://www.matsim.org/

[7] Nicolaides, C., Cueto-Felgueroso, L., González, M.C., Juanes, R., A Metric Influential Spreading during Contagion Dynamics through the Air Transportation Network, PLoS ONE 7(7): e40961 (2012).

[8] Ortuzar, J., Willumnsen, L., Modelling Transport, Fourth Edition, Wiley, 2011.

[9] Ray, M., Schrom-Feiertag, H., Cell-based Finding and Classification of Prominent Places of Mobile Phone Users, Proceedings of the 4th International Symposium on Location Based Services & TeleCartography (2007)

[10] Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., González, M. C., Unraveling Daily Human Mobility Motifs, Journal of the Royal Society Interface 10 20130246 (2013)

[11] Stopher, P., Greaves, S., Household travel surveys: Where are we going?, Transportation Research Part A: Policy and Practice, 41 (5), 367–381 (2007)

[12] Toole, J., Ulm, M., Bauer, D., González, M. C., Inferring land use from mobile phone activity, proceedings of the International Workshop on Urban Computation (UrbComp2012), Beijing, 2012.

[13] Wang, P., Hunter, T., Bayen, A., Schechtner, K., González, M. C., Understanding Road Usage Patterns in Urban Areas, Nature Scientific Reports **2**, Article 1001, doi:10.1038/srep01001 (2012)