An Exploration of human mobility motifs in the California component of the 2017 National Household Travel survey

Rongxiang Su and Konstadinos G. Goulias
Department of Geography and GeoTrans Laboratory
University of California Santa Barbara

In travel behavior and travel demand forecasting developing taxonomies of daily activity-travel patterns is the first step in the design of simulation models and in comparative studies across contexts and cultures. Often these daily patterns are summaries of daily trips and activities correlated with sociodemographic groups (Garikapati et al., 2016; Timmermans et al., 2003). These daily summaries, however, do not capture the interdependence among episodic activities, locations of activities, trips in between activity locations, and modes used. Explicit correlation among the daily choices people make in building their schedules has been studied using history dependence models (Kitamura et al., 1997), correlation patterns among activities and destinations (Axhausen et al., 2002), scheduling model systems based on heuristics and rules (Arentze & Timmermans, 2000; Miller & Roorda, 2003), and building econometric model systems designed for insertion in microsimulators (Bhat et al., 2012; Goulias et al., 2011). All these methods explore some facets of the spatio-temporal organization of activities and travel by taking slices of time and space to build model components. In one recent application (McBride, Davis, & Goulias 2019), daily activity and travel patterns are viewed in the entire daily pattern as sequences of episodes that happen at specific places during a day. This more holistic approach classifies each minute in a "state" and has the needed information to examine transitions among states while accounting for the amount of time allocated to each state. This holistic approach with minute-by-minute daily schedules enables richer analyses such as distinguishing between persons that have fragmented schedules from persons that follow very simple patterns in time allocation. It also allows to build taxonomies of typical and atypical schedules (McBride, Davis, & Goulias 2019). However, a thorny issue in all these activity-travel models is location choice for activity participation and the relationship among the locations visited by individuals as they execute their schedules. One way to explore interdependence among locations is to use a metaphor from the way a cell converts DNA to RNA in recurring patterns that are called network motifs (Alon, 2007; Shoval & Alon, 2010). The idea of motifs was transferred to human mobility patterns by Schneider et al. (Schneider et al., 2013) to study recurring patterns in daily travel behavior, verify the degree of spatial and temporal pattern regularity, and test the existence of simple universal rules underlying human movement (Song et al. 2010; González, Hidalgo, & Barabási 2008). To this end, (Schneider et al., 2013) describe 17 unique motifs that are able to capture 90% of daily mobility patterns in samples of data tested that include surveys and GPS traces. Extending the motif idea further, Cao et al. (2019) using GPS data proposed two new concepts of the location-based motif and the activity-based motif showing that a relatively small number of location-based motifs and activity-based motifs can describe 99.9% of human movement patterns.

These data analysis methods are informative and provide the base of our study here. However, passive data collection and analysis such as GPS data, fail to explain who are the persons that follow a certain type of mobility pattern in a day and the time scheduling of these motifs. In this paper we rectify this. First, we derive a set of human mobility motif representations using the 2017 California component of the National Household Travel Survey (NHTS) which contains

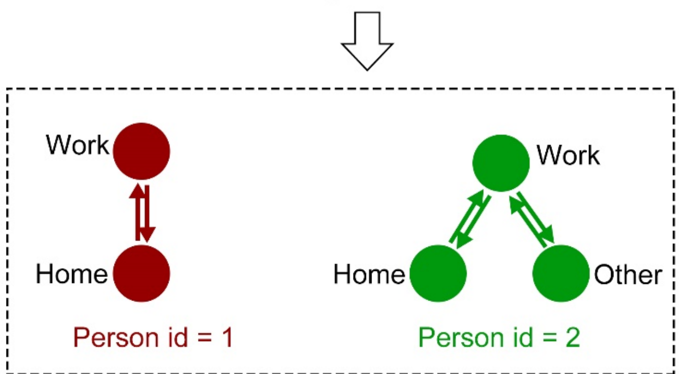
43,871 individual daily travel diaries. We then correlate these motifs with daily time allocation in different types of activities and with person characteristics. We conclude the paper with next steps.

Data used and network motifs

Motif is a directed network, in which nodes represent visited locations and directed edges represent trips between locations. In the California-NHTS data, we have a trip diary recording trips for each individual in a day. We also have data about the person and household characteristics for each survey participant. In order to construct motif representations, we need to use the origin and destination locations as nodes and connect nodes with a directed edge if there is a trip between them. An example is shown in Figure 1. The first person is a typical commuter traveling only between the home and work locations. Two nodes represent home and work locations respectively and two directed edges between the two nodes represent the trips from home to work and work to home. The second person in Figure 1, has a similar commute but also visits another destination labeled other. Therefore, this person has a motif with three nodes representing home, work, and other location and four directed edges (the trips).

Household Id	Person Id	Origin	Destination	Origin type	Destination type
30000328	1	10000	100	Home	Work
30000328	1	100	10000	Work	Home
30000328	2	10000	100	Home	Work
30000328	2	100	20001	Work	Other
30000328	2	20001	100	Other	Work
30000328	2	100	10000	Work	Home

Original data



Motif representations

Figure 1. Example of construct a motif from trip data

Analysis of the daily patterns in California-NHTS shows 1046 distinct types of individual-level network motifs. Figure 2 shows 14 unique motifs from 35,523 persons capturing approximately 82.18% of the person days in this sample. Figure 3 enumerates each of the 14 most popular motifs, the structure of each motif, and its percentage. Motif 2 (with two locations visited in a day) accounts for the highest percentage. Motif 2 is followed by 12,867 people which is about 29.76% of the survey sample. This is consistent with the location-based motif in Schneider et al. (2013) and Cao et al. (2019).

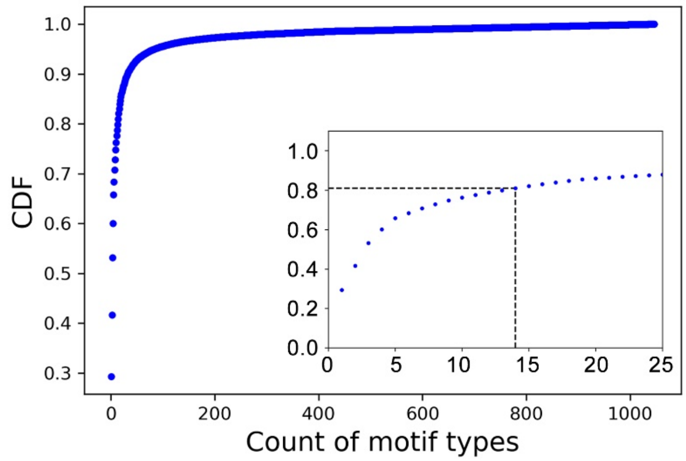


Figure 2. Cumulative distribution function for count of individual motif types

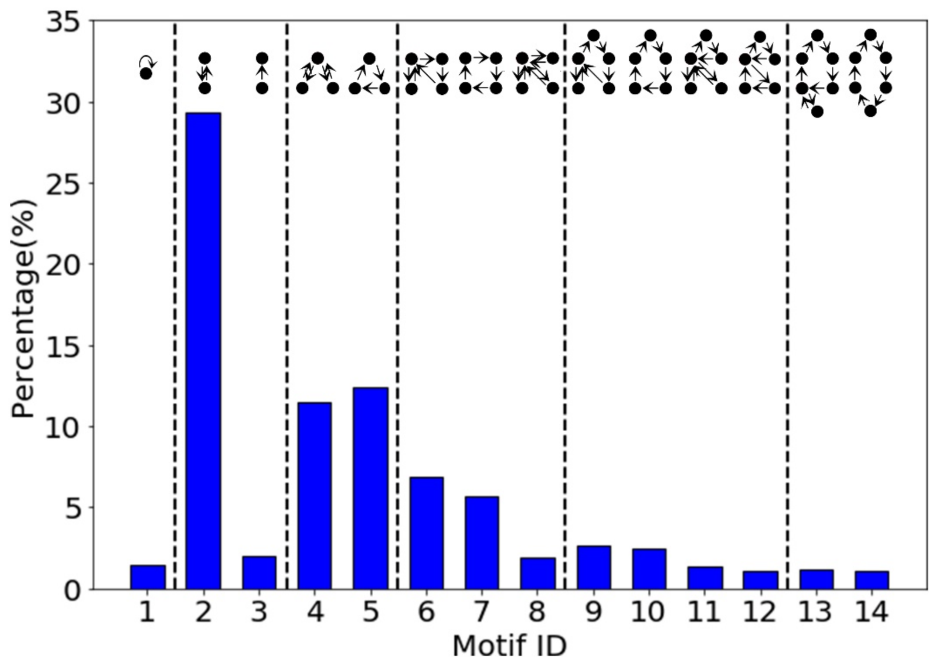


Figure 3. Percentage of individual motif patterns

Composition of the motifs

For each of these motifs we compute the average characteristics of the person's daily travel behavior and the average characteristics of the persons within each of these motifs. Table 1 provides the list of the 14 motifs plus one category with other motifs not enumerated here. For each motif it also provides the within motifs average characteristics of time allocation by the 43871 survey respondents. The column complexity contains the values of an indicator that combines the diversity of time allocated in activities and travel and the number of activities and trips a person makes. This summary indicator captures daily activity-travel patterns for each individual in a succinct mathematical way (McBride et al., 2020). Table 1 shows that complexity of daily schedules increases with the number of combinations of nodes and edges in the motifs as expected. In general, time travelling also increases from simple motifs to more complex motifs. Table 1 includes another synoptic indicator of daily travel called Travel Time Ratio (Dijst & Vidakovic, 2000), defined as the total travel time in a day divided by the sum of the total time in activities outside the home plus the total travel time in a day (TTR). TTR is very high for the first motif because some persons went for a very long trip with both origin and destination home. The rest of motifs show similar TTRs. Table 2 shows the relationship between each motif and the within motif person characteristics. We see a few major trends. Students tend to be concentrated in motifs 2, 3, 4, 5 that are also the motifs with on average higher number of minutes at school per day. It also shows that people with a driver's license are more likely to also have motifs with more locations and trips. Retired persons are spread throughout the different motifs except motif 1 that has a low percentage.

Table 1 Relationship of Motif with Time Based Behavioral Indicators (including size and sample percentage of Motif membership)

Motif	Complexity	Time* at School	Time* at Work	Time* Outside Home	Time* Traveling	Travel Time Ratio	Time at Home Ratio	Persons in Sample	Percent in Sample
1	0.012	7.2	15.9	244.1	78.5	0.802	0.830	644	1.5%
2	0.023	63.1	184.2	404.2	58.5	0.253	0.719	12867	29.3%
3	0.018	34.7	274.4	871.3	140.3	0.257	0.394	899	2.0%
4	0.038	50.2	182.8	461.7	87.2	0.257	0.679	5029	11.5%
5	0.030	42.5	156.0	418.0	77.8	0.265	0.710	5420	12.4%
6	0.044	29.7	158.1	469.3	102.4	0.271	0.674	3038	6.9%
7	0.036	27.7	118.3	418.5	96.3	0.297	0.709	2494	5.7%
8	0.048	26.6	97.6	433.1	103.0	0.298	0.699	858	2.0%
9	0.049	28.1	135.4	480.9	115.3	0.286	0.666	1144	2.6%
10	0.042	20.2	106.7	425.1	106.1	0.299	0.705	1065	2.4%
11	0.053	16.4	91.5	447.9	121.6	0.314	0.689	598	1.4%
12	0.050	32.3	96.5	478.3	124.2	0.297	0.668	474	1.1%
13	0.054	23.1	101.6	479.1	126.8	0.300	0.667	501	1.1%
14	0.047	14.6	76.0	437.8	135.4	0.337	0.696	492	1.1%
All Other	0.055	19.7	136.4	611.6	185.8	0.341	0.575	8348	19.0%

*Time is measured in minutes per day

Note: The background color in a gradient is according to the data in each column.

Table 2 Motifs and Respondents Characteristics (Percent of persons within each motif group)

Motif	Woman	Student	Has Driver's License	Homeworker	Full time Employed	Part time Employed	Retired
1	51.71%	4.97%	81.68%	11.49%	15.06%	7.92%	18.17%
2	49.13%	16.13%	77.70%	6.28%	37.96%	10.86%	44.23%
3	47.27%	12.90%	80.76%	7.23%	45.27%	9.45%	49.94%
4	49.06%	12.89%	83.69%	7.52%	41.54%	12.37%	47.98%
5	53.32%	13.06%	81.88%	7.05%	36.86%	11.42%	43.34%
6	53.55%	9.28%	87.13%	7.54%	42.00%	12.38%	49.21%
7	57.26%	9.86%	83.72%	7.02%	33.04%	11.35%	39.45%
8	49.77%	9.56%	88.11%	10.96%	35.90%	12.24%	41.96%
9	54.02%	9.09%	87.67%	8.92%	37.50%	13.72%	46.59%
10	58.22%	8.92%	85.26%	6.67%	29.48%	11.36%	35.31%
11	48.66%	6.69%	90.80%	12.71%	35.95%	16.56%	43.48%
12	56.96%	12.66%	85.02%	9.07%	33.33%	13.08%	40.72%
13	54.49%	11.38%	86.23%	11.38%	36.53%	14.57%	44.11%
14	57.32%	6.71%	89.02%	9.15%	28.25%	11.79%	33.33%
All other	54.53%	7.75%	89.81%	9.80%	41.58%	13.31%	48.16%

Note: The background color in a gradient is according to the data in each column.

What is the relationship between motif patterns and people's social demographic characteristics?

We explore further in Table 3 the relationship between motifs and social demographic characteristics using a multinomial logit model (MNL). Since the percentages for some motifs are very low, we merge motifs into five groups based on the number of nodes. Group 1 is motif 1 with only one node. Group 2 includes motifs 2 and 3 with two nodes. Group 3 includes motifs 4 and 5 with three nodes. Group 4 includes motifs 6, 7 and 8 with four nodes. And all the other motifs are classified as Group 5. In this MNL, the dependent variable is motif group with the five group categories. We used as explanatory variables all the variables in Tables 1 and 2, age groups, and day of the week and include in Table 3 only the significantly different than zero coefficients. The model shows that all the younger age groups are more likely to use daily pattern groups 2 to 5 but not the single node motif when compared to the 75 and older group. The coefficients (and the implied log odds) in this MNL show that younger respondents tend to use more complex motifs. This is as expected because younger people are more mobile than older people. Considering that, for motif group 2, minutes in School has the highest positive significant coefficient and the weekdays are all positive and with a highest coefficients compared with other groups. This indicates that young students tend to have very simple daily patterns (Home-School and back) in their typical going to school days. However, not all students follow this motif and as Table 2 shows a substantial number of students have by far more complex daily motifs. In addition, a full time worker is most likely to have group 2's motifs among all the other groups when compared to part time workers. Minutes at work also has the highest positive coefficient for group 2 showing that people spending more time at work they are more likely to have group 2 motifs with only two visited locations in a day. Homeworkers (persons that work from home) – telecommuters are more likely to have group 1's motif which means that they prefer not to visit other places. Table 2 confirms this with a substantial number of telecommuters making loop trips (origin and destination is the same place) but some of them have much more complex daily motifs (motifs 8, 11 and higher). Retired people are most likely to have group 2's motif, but they also very likely to have more complex motifs. Women and persons with a driver's license are more likely to follow more complex motifs, which contain more visited locations, and this is consistent with other research on women travel (McBride et al., 2020). Working days of the week are all showing a positive significant coefficient for group 2 with highest size of coefficient compared with the other groups of motifs indicating this is a typical working day pattern (already captured by the fulltime and student indicators and discussed above). The lowest value of coefficient of group 2's travel time ratio is indicative of a motif that has more time in activities and less time traveling (Table 1 shows this is the motif 2 with the lowest travel time) when compared to other motifs. This also indicates relative proximity of the activity location (work, school, and other) from home.

Table 3. Estimated parameters of multinomial logit model

	Dependent variable			
	Group 2	Group 3	Group 4	Group 5
(Intercept)	13.733***	6.842***	2.622***	-1.733***
Age15 and lower	2.499***	2.360***	2.158***	2.509***
Age16to17	2.362***	1.597***	0.966***	1.109***
Age18to25	2.269***	1.789***	1.478***	1.347***
Age26to45	1.848***	1.639***	1.427***	1.360***
Age46to65	0.833***	0.631***	0.535***	0.423***
Age66to75	0.297***	0.186***	0.187***	0.139***
Woman	0.125***	0.171***	0.300***	0.340***
Has Driver's License	0.560***	0.895***	0.883***	1.162***
Homeworker	-0.694***	-0.769***	-1.108***	-1.250***
Fulltime Employed	0.298***	0.138***	0.114***	0.021
Retired	0.781***	0.422***	0.310***	0.143***
Monday	0.539***	0.258***	0.183***	0.140***
Tuesday	0.302***	0.138***	-0.082**	-0.124***
Wednesday	0.816***	0.555***	0.293***	0.287***
Thursday	0.945***	0.688***	0.534***	0.570***
Friday	0.540***	0.373***	0.283***	0.306***
Complexity	-236.345***	25.852***	138.389***	221.105***
Minutes at School	0.003***	0.0002	-0.002***	-0.004***
Minutes at Work	0.004***	0.001***	-0.0001	-0.001**
Minutes outside Home	-0.007***	-0.007***	-0.007***	-0.006***
Minutes Travel	0.019***	0.013***	0.010***	0.011***
Travel Time Ratio	-13.415***	-10.459***	-9.081***	-7.129***
Akaike Inf. Crit.	72,092.280	72,092.280	72,092.280	72,092.280
Note: *p<0.1; **p<0.05; ***p<0.01				

In conclusion, we find that 14 unique motifs can capture more than 82.18% of the 2017 NHTS-California sample data. The results of cross tabulations between motifs and daily time allocation and travel behavior and an MNL model tell us that a substantial number of students and fulltime workers are more likely to be choosing a simple two node motif (Home-Work and back, Home-School and back). However, women, telecommuters, and drivers are more likely to visit more than two places in a day. Younger people are also more likely to visit more places than older people. We also find that retired people actually do not just stay at home, instead, they are more likely to visit more than two places in a day. In this exploratory analysis we find that motifs are an efficient way to analyze diary data and correlate well with daily summary indicators and social and demographic characteristics of survey respondents. Next steps include: a) pinpoint the relationship between motifs and other household characteristics; b) study the correlation between motifs, activities, and modes used; c) understand the relationships among household members motif choices; d) compare motifs across different spatial contexts; and e) test motifs as predictive tools.

References

Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461.

Arentze, T., & Timmermans, H. (2000). *Albatross: A learning based transportation oriented simulation system*. Citeseer.

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

Bhat, C. R., Goulias, K. G., Pendyala, R. M., Paleti, R., Sidharthan, R., Schmitt, L., & Hu, H. (2012). A household-level activity pattern generation model for the simulator of activities, greenhouse emissions, networks, and travel (SimAGENT) system in Southern California. *91st Annual Meeting of the Transportation Research Board, Washington, DC*.

Cao, J., Li, Q., Tu, W., & Wang, F. (2019). Characterizing preferred motif choices and distance impacts. *PLOS ONE*, *14*(4), e0215242. https://doi.org/10.1371/journal.pone.0215242

Dijst, M., & Vidakovic, V. (2000). Travel time ratio: The key factor of spatial reach. *Transportation*, *27*(2), 179–199.

Garikapati, V. M., Pendyala, R. M., Morris, E. A., Mokhtarian, P. L., & McDonald, N. (2016). Activity patterns, time use, and travel of millennials: A generation in transition? *Transport Reviews*, *36*(5), 558–584.

González, M. C., Hidalgo, C. A., & Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, *453*(7196), 779–782. https://doi.org/10.1038/nature06958

Goulias, K. G., Bhat, C. R., Pendyala, R. M., Chen, Y., Paleti, R., Konduri, K. C., Huang, G., & Hu, H.-H. (2011). Simulator of activities, greenhouse emissions, networks, and travel (SimAGENT) in Southern California: Design, implementation, preliminary findings, and integration plans. *2011 IEEE Forum on Integrated and Sustainable Transportation Systems*, 164–169.

Kitamura, R., Chen, C., & Pendyala, R. M. (1997). Generation of synthetic daily activity-travel patterns. *Transportation Research Record*, *1607*(1), 154–162.

McBride, Elizabeth C, Davis, A. W., & Goulias, K. G. (2019). Fragmentation in Daily Schedule of Activities using Activity Sequences. *Transportation Research Record: Journal of the Transportation Research Board*, *2673*(4), 844–854. https://doi.org/10.1177/0361198119837501

McBride, Elizabeth C, Davis, A. W., & Goulias, K. G. (2020). Sequence analysis of place-travel fragmentation in California. In *Mapping the Travel Behavior Genome* (pp. 371–398). Elsevier.

Miller, E. J., & Roorda, M. J. (2003). Prototype model of household activity-travel scheduling. *Transportation Research Record*, *1831*(1), 114–121.

Schneider, C. M., Belik, V., Couronné, T., Smoreda, Z., & González, M. C. (2013). Unravelling daily human mobility motifs. *Journal of The Royal Society Interface*, *10*(84), 20130246. https://doi.org/10.1098/rsif.2013.0246

Shoval, O., & Alon, U. (2010). SnapShot: Network motifs. *Cell*, *143*(2), 326–326.

Song, C., Koren, T., Wang, P., & Barabási, A.-L. (2010). Modelling the scaling properties of human mobility. *Nature Physics*, *6*(10), 818–823. https://doi.org/10.1038/nphys1760

Timmermans, H., van der Waerden, P., Alves, M., Polak, J., Ellis, S., Harvey, A. S., Kurose, S., & Zandee, R. (2003). Spatial context and the complexity of daily travel patterns: An international comparison. *Journal of Transport Geography*, *11*(1), 37–46.