



The 8th International Conference on Ambient Systems, Networks and Technologies
(ANT 2017)

Zipf's power law in activity schedules and the effect of aggregation

Wim Ectors^{a,*}, Bruno Kochan^a, Davy Janssens^a, Tom Bellemans^a, Geert Wets^a

^aHasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium

Abstract

Modeling people's behavior in e.g. travel demand models is an extremely complex, multidimensional process. However, the frequency of occurrence of day-long activity schedules obeys a ubiquitous power law distribution, commonly referred to as Zipf's law.¹ This paper discusses the role of aggregation within the phenomenon of Zipf's law in activity schedules. Aggregation is analyzed in two dimensions: activity type encoding and the aggregation of individual data in the dataset. This research employs four datasets: the household travel survey (HTS) NHTS 2009, two six-week travel surveys (MobiDrive 1999 and Thurgau 2003) and a 24-week set of trip data which was donated by one individual. Maximum-likelihood estimation (MLE) and the Kolmogorov-Smirnov (KS) goodness-of-fit (GOF) statistic are used in the "PoweRlaw" R package to reliably fit a power law. To analyze the effect of aggregation in the first dimension, the activity type encoding, five different activity encoding aggregation levels were created in the NHTS 2009 dataset, each aggregating the activity types somewhat differently. To analyze aggregation in the second dimension, the analysis moves from study area-wide aggregated data to subsets of the data, and finally to individual (longitudinal) data.

1877-0509 © 2017 The Authors. Published by Elsevier B.V.
Peer-review under responsibility of the Conference Program Chairs.

Keywords: Zipf; power law; activity schedules; data aggregation; activity type classes

1. Introduction

The transportation research community invests heavily in understanding travel behavior. Modeling people's behavior in travel demand models is an extremely complex, multidimensional process. However, as demonstrated by Ectors et al.¹, the frequency of occurrence of day-long activity schedules obeys a remarkably simple, scale-free distribution.

This distribution has been observed in many natural and social processes and obeys a power law. It is commonly referred to as Zipf's law. Auerbach discovered in 1913 that city size is governed by such a power law. The American linguist Zipf described a power law distribution in word frequency in 1949 (although it had first been noticed by Estroup in 1916). Zipf famously investigated this distribution more in detail, revealing that the same power law distribution holds for a large number of events in different domains, ranging from sizes of earthquakes, people's annual income, solar flares, to the number of citations received on papers.^{2,3,4}

* Corresponding author. Tel.: +32-11-269114.
E-mail address: wim.ectors@uhasselt.be

The rank-size interpretation of Zipf's law is most commonly mentioned (Equation 1). For example, within the context of city sizes, the size of a city at rank r_i scales with a factor $1/r_i$ relative to the size of the largest city. The second largest city is half the first city's size, the third largest one-third its size etc.

$$f(r_i) = \frac{f(r_1)}{r_i} \quad (1)$$

where f represents *frequency* and r the *rank*. In other words, the size of a city is inversely proportional to its rank.

In its more formal form, Zipf's law states that the probability for a city to have a size greater than S decreases as $1/S$:

$$P(\tilde{S} > S) = aS^{-\zeta} \quad (2)$$

where \tilde{S} is the size of a particular city, ζ denotes the exponent (≈ 1 for Zipf's law), a a scaling factor.⁵

To the authors' best knowledge, no conclusive proof exists rejecting the existence of a natural power law mechanism, nor does a general agreement exist on the origin of Zipf's law's ubiquitousness. Additionally, many of the observations seem to share the same exponent, which desires a universal mechanism. Yet, most researchers agree that *several* mechanisms may lead to the observed power law distributions.⁴

Still, some research objects against Zipf's apparent universality. In a large-scale study, 73 cities from across the world were analyzed for conformity with Zipf's law. Zipf's law was rejected in more cities than expected⁶. A meta study including 515 estimates from 29 studies concluded that the power law exponent is statistically different from Zipf's value of 1.0, actually being closer to 1.1.⁷

Zipf's law has not been mentioned often within the domain of transportation sciences. Still, power law-like distributions have been proven in displacement distance, gyration radius and location visiting frequency⁸, as well as in location visiting duration⁹ and travel time in taxi travel¹⁰. Power law distributions were also observed in bus transport networks¹¹ and in airport networks¹². Some researchers also used these universal distributions in their experiments.^{13,14} More recently, evidence for a universal Zipf power law in activity schedules was given.¹

This paper discusses the role of aggregation within the phenomenon of Zipf's law in activity schedules. In the remainder of this paper, first the data and basic methodology for estimating a power law fit are detailed, after which the effect of aggregation is analyzed in two dimensions: activity type encoding and dataset aggregation level. The conclusion section finalizes this paper.

2. Description data and estimation procedure

This research employs four datasets: 1) a HTS from the US, the USA NHTS 2009¹⁵ dataset, 2) a six-week travel survey from Germany, DEU MobiDrive 1999¹⁶, 3) a Swiss six-week travel survey CHE Thurgau 2003¹⁷ and 4) a 24-week set of trip data which was donated by one individual. The 24-week data was collected using the Moves smartphone application¹⁸ combined with manual verification and trip purpose enrichment. There were 163 days with out-of-home activities. The OVG HTS¹⁹ activity encoding was used (10 classes).

Out-of-home activity schedules are constructed out of trip purpose information from these datasets. Trip purposes are concatenated into a sequence which represents a schedule with the main out-of-home activities. From the NHTS 2009 dataset 257,586 schedules could be extracted (83,000 distinct schedules). The Mobidrive 1999 and Thurgau 2003 datasets yield, respectively, 13,244 and 8,522 schedules.

In order to evaluate the role of aggregation, first the methodology of fitting a power law distribution to the data needs to be defined. Often, a linear regression (using least-squares) is fitted to log transformed variables, yet this method is flawed.^{4,20,21} The slope estimate may exhibit systematic, large errors. Additionally, the traditional R^2 cannot be used as evidence for a power law distribution. Clauset et al.²⁰ proposed a method based on a MLE fitting approach with the KS GOF as a cutoff criterion. Some cutoff x_{\min} is needed since the power law probability distribution $p(x) = Cx^{-\alpha}$ with $\alpha \geq 1$ diverges for $\alpha \rightarrow 0$, resulting in an infinite area under the distribution. Under Zipf's law, an exponent value of 2.0 (or $\alpha - 1 = 1.0$ for the cumulative distribution) is expected. The cutoff parameter depicts the fact that few datasets follow a power law distribution across their entire range; in most cases a certain fraction (e.g. the low frequency area) deviates from the power law distribution. The R package called "PoweRlaw"²² was developed to automate the MLE + KS estimation process. The x_{\min} parameter is optimized by means of the KS statistic. The package also

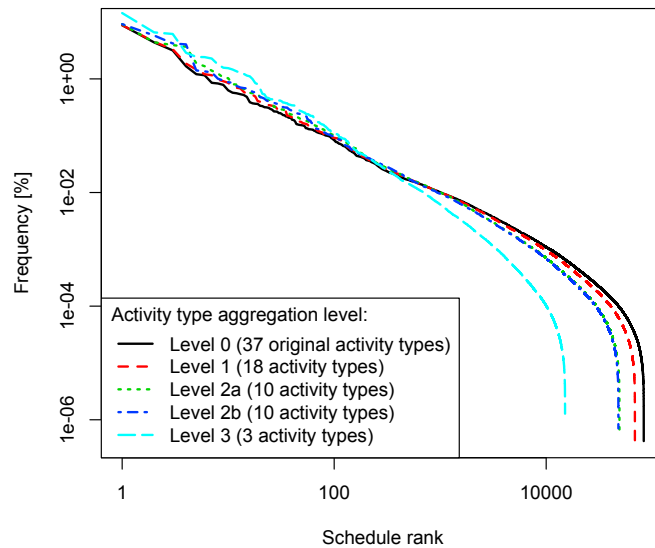


Fig. 1: Activity schedule distribution in the USA NHTS 2009 dataset based on five different activity encoding aggregation schemes.

supports bootstrapping procedures to evaluate parameter estimation uncertainty and to perform a hypothesis test with null hypothesis that a power law distribution is appropriate. A 10% significance level is recommended in this test.²²

3. Aggregation in activity type encoding

To analyze the effect of aggregation in the first dimension, the activity type encoding, different activity type encoding aggregation levels were created in the USA NHTS 2009 dataset. Starting from the original 37 activity types, denoted here as Level 0, four more sets of encodings were proposed, each aggregating some activity types or grouping them somewhat differently. The approach corresponds to constructing an encoding tree and pruning the branches to increase the aggregation level.

The first digit of the original Level 0 encoding corresponds to a higher-level group, while the second digit specifies the activity type in more detail. This is exploited to construct other encoding schemes. The level 1 encoding was constructed by retaining the first digit and subsequently grouping some of the second digits. This moderate aggregation halved the number of activity types from 37 to only 18 distinct categories. The Level 2a encoding was formed by allocating the most appropriate category from the OVG HTS¹⁹ (which is well-known to the authors) to each NHTS category. Only ten distinct activity type categories remain. The Level 2b encoding provides the same level of aggregation (ten distinct categories), but is simply based on the first digit of the original USA NHTS 2009 encoding. The final activity encoding scheme, Level 3, offers the highest level of aggregation into only three distinct classes. For this scheme the original activity types were identified as either being of 'Mandatory', 'Maintenance' or 'Discretionary' nature. These five activity encoding schemes were used to construct day-long activity schedules for the individuals in the NHTS dataset.

The distributions of the resulting sets of schedules are illustrated in Figure 1. One observes that the power law regime (the linear trend on a log-log plot) breaks down relatively quickly only in case of the most severe aggregation of Level 3; for the other cases it seems valid for the majority of observations. In general, the more aggregation is applied to the activity types, the less Zipf's law seems to hold across the whole dataset. The effect seems in practice only significant at extreme levels of aggregation. Figure 1 also shows how the sets of schedules based on Level 2a and Level 2b (both ten distinct activity types) are nearly indistinguishable, although their activity coding is different in some instances. Table 1 lists the power law estimates from the MLE + KS estimation procedure. With increasing activity type aggregation also the deviation from the theoretical Zipf's exponent increases. Still, a power law distribution remains appropriate. The bootstrapping estimates are consistent with those based on the singular MLE + KS procedure.

4. Aggregation of individual data

The fact that Zipf's law seems valid on aggregated schedules for a whole study area was already established. It is however interesting to explore the limits of Zipf's law when using less aggregated data. This section will analyze this effect, moving from study area-wide aggregated data to individual longitudinal data. First a power law distribution is fitted to fully aggregated data. Subsequently, subsets based on the day of the week (DOW) and gender were taken from the USA NHTS 2009 dataset and a power law distribution was fitted to these subsets. Next, the six-week travel surveys DEU Mobidrive 1999 and CHE Thurgau 2003 allow to consider individual schedules, representing the least amount of aggregation possible. Finally, a 24-week trip history belonging to one person tests the validity of Zipf's law (for this particular individual) in longitudinal data.

4.1. Aggregation to study area level

Figure 2 illustrates the remarkable power law in activity schedules for a complete study area based on a single-day HTS. A nearly identical distribution is found for the DEU Mobidrive 1999 and CHE Thurgau 2003 datasets when each recorded day is treated independently and subsequently aggregated. Table 1 lists the estimates for these experiments. All three datasets have exponent values very close to Zipf's value of 2.0. It appears that aggregated schedules from multiple individuals will consistently exhibit a power law distribution, also analysed in more detail in Ectors et al.¹

4.2. Subsets of a study area

The USA NHTS 2009 was used to analyze subsets as it is a significantly large dataset. This avoids incorrectly rejecting a power law distribution due to insufficient data. As illustrated in Figure 3, subsets were generated based on DOW and gender. Visually, their distributions are nearly identical. Table 1 lists the power law fit estimates. Again, estimates close to Zipf's law's value of 2.0 were found. They appear consistently slightly higher than the estimate for the full dataset. This suggests that some schedules may be more typical for a particular subset of the data, yielding higher frequencies for the top-ranked schedules in that subset of the data. Each subset does not (necessarily) have the same schedule at each rank. The effect is however small since e.g. there are only small differences in the distributions of weekdays and weekends (where a different travel behavior is expected). There are however fewer distinct schedules on Sundays. This has however no effect on the power law exponent estimate because of the x_{\min} cutoff value. Additionally, none rejects the null hypothesis of a power law distribution being an appropriate distribution

Table 1: Estimation results from the R package `powerLaw` for activity schedule distributions.

Dataset	Aggregation or subset	powerLaw estimations (MLE + KS)			Bootstrapping uncertainty evaluation		
		α	x_{\min}	Cum. pct rejected	AM(α)	SD(α)	P-value
USA NHTS 2009	Level 0 (37 original activity types)	2.003	36809977	55%	2.006	0.070	0.255
USA NHTS 2009	Level 1 (18 activity types)	1.967	36837451	50%	1.972	0.065	0.166
USA NHTS 2009	Level 2a (10 activity types)	1.934	46135634	43%	1.939	0.065	0.998
USA NHTS 2009	Level 2b (10 activity types)	1.892	60781076	45%	1.899	0.071	0.741
USA NHTS 2009	Level 3 (3 activity types)	1.890	109512566	28%	1.891	0.084	0.835
USA NHTS 2009	Monday*	2.290	46616705	67%	2.270	0.359	0.831
USA NHTS 2009	Tuesday*	2.161	35581917	67%	2.182	0.236	0.820
USA NHTS 2009	Wednesday*	2.152	45646004	68%	2.172	0.267	0.679
USA NHTS 2009	Thursday*	2.088	48120314	71%	2.140	0.282	0.221
USA NHTS 2009	Friday*	2.279	34509610	72%	2.284	0.250	0.901
USA NHTS 2009	Saturday*	2.182	61045896	76%	2.176	0.288	0.134
USA NHTS 2009	Sunday*	2.091	52160661	66%	2.060	0.200	0.982
USA NHTS 2009	Women	2.104	37421218	61%	2.115	0.114	0.551
USA NHTS 2009	Men	2.157	36416801	58%	2.165	0.116	0.783
DEU Mobidrive 1999	All data aggregated	2.053	23	52%	2.002	0.133	0.714
CHE Thurgau 2003	All data aggregated	1.929	16	49%	2.009	0.113	0.317
Donated	Schedules from an individual	2.454	4	59%	2.629	0.708	0.689

Note: the different scales of x_{\min} are caused by different weight variables.

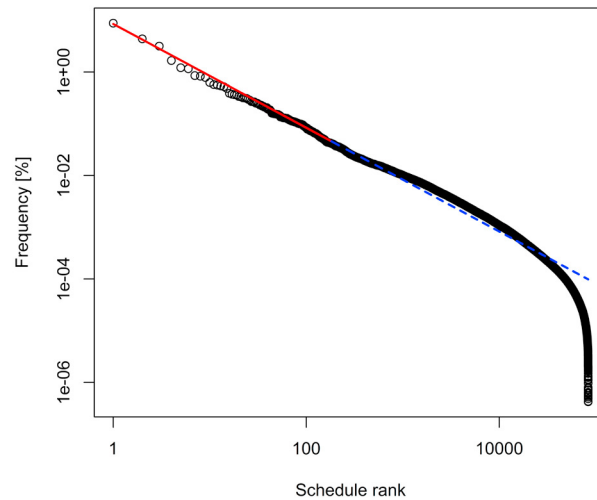


Fig. 2: Activity schedule distribution in the USA NHTS 2009 dataset. The red full line represents the fitted power law (according to the MLE + KS), the dotted blue line is the extrapolation of this fit.

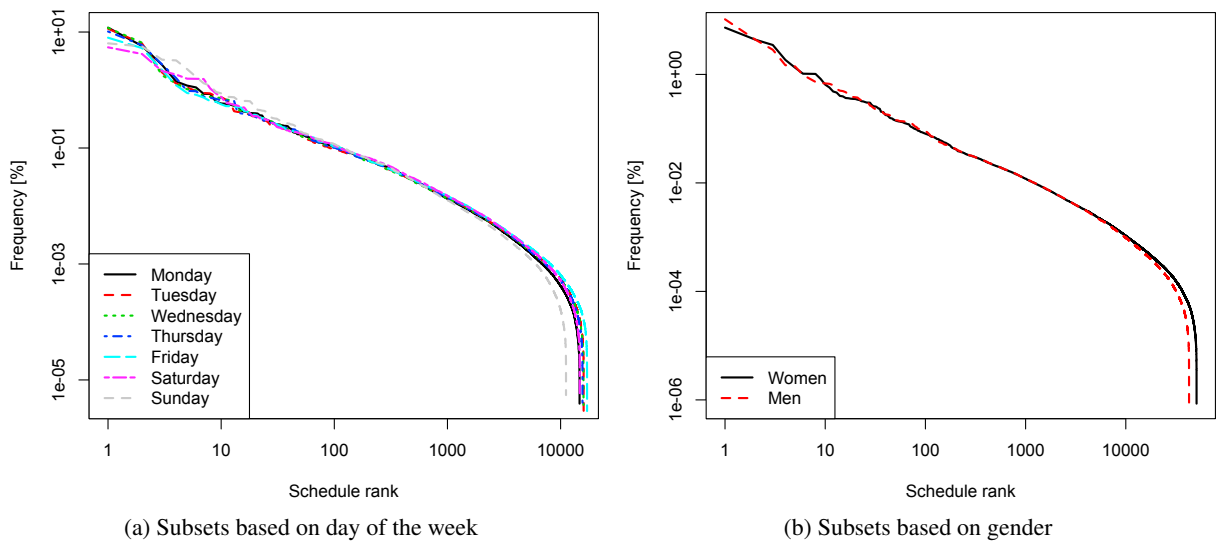


Fig. 3: Activity schedule distribution in subsets of the USA NHTS 2009 dataset (using the original activity encoding).

for the subsets. It appears that subsets of the data will also exhibit a power law distribution, possibly with slightly deviating exponent values and different schedules at similar ranks, provided that the subsets are not made too small.

4.3. The individual level

A much tougher question is whether Zipf's law is valid for activity schedules from each individual separately, similarly to other universally distributed quantities like displacement distance, location visiting frequency etc.⁸ The schedules are not fully independent in this case, but belong to one individual. To analyze this question, three datasets are used: two six-week travel surveys (DEU Mobidrive 1999 and CHE Thurgau 2003) and the donated 24-week trip dataset from one individual.

To analyze the six-week travel surveys, a variable (present in the original dataset) with 10 trip purpose classes is used instead of the 23 classes originally in the survey. As the data is limited (six weeks) this will ensure the highest possible frequencies for each schedule, so a power law might be discovered in 'only' six weeks of data. As discussed

before, this choice should not negatively influence the estimation results. Some individuals in the data have very few days when trips were made, resulting in bad fits and outlier-like exponent estimates. These 'outliers' were removed according to a threshold of minimum number of schedules (days). This threshold was put at 21 schedules, which is half the theoretically maximal number of schedules (6 weeks \times 7 schedules per week = 42 schedules). The DEU MobiDrive 1999 data set contains 361 individuals. After filtering out some outlier-like individuals (with less than half of the schedules reported), 352 individuals remained.

After generating frequency tables for each individual, very low frequencies are observed. At schedule ranks greater than 2 they are certainly lower than 5. A simple Chi-square GOF test is therefore not possible, as the assumption of expected frequencies greater than 5 is violated. The KS GOF test seemed most appropriate. In SAS, a distribution can be tested against a predefined distribution based on this statistic. The desired test could be achieved by imposing the Beta distribution with some fixed parameters, leading to a power law distribution. The null hypothesis H_0 is that a power law distribution with specified α is a good fit.

If one α is imposed for all individuals, 12% (43 out of 352) have a distribution which is not significantly different from a power law distribution, based on a significance level of 5%. Similar results are obtained using the CHE Thurgau 2003 dataset: 4% of the individuals (9 out of 230) have a distribution that is not significantly different from a power law distribution. Curiously, when α is allowed to vary across the individuals, *more* cases reject H_0 . These results are not supporting the theory that Zipf's law is also valid for individuals. However, as can be seen in Figure 4a, the cases where the H_0 of a good fit is rejected seem to be not fully developed, having a large horizontal tail at the end of the distribution.

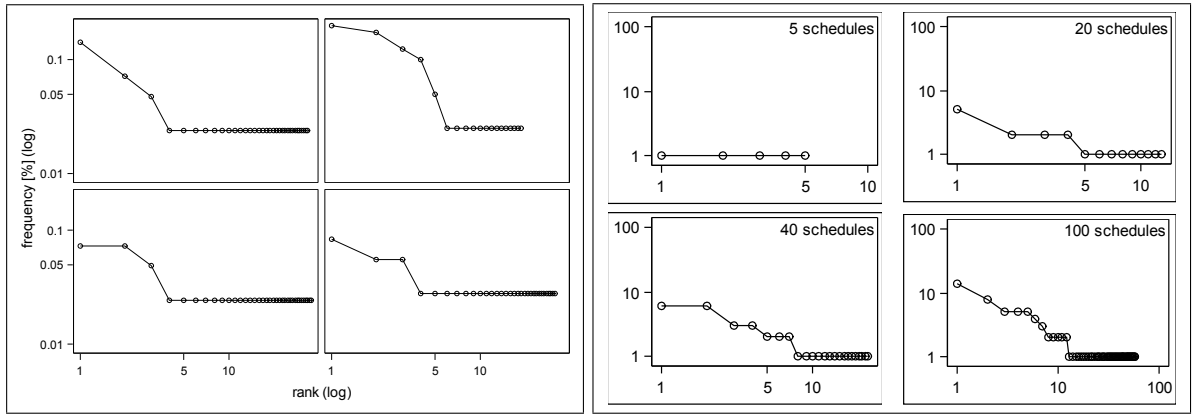
A simulation was build to reveal how a power law distribution may be formed. The activity schedule frequency distribution of the DEU Mobidrive 1999 data was plotted in increasing fractions of the data (after randomization). Some examples are given in Figure 4b. One observes a rather flat distribution at first which then, over time, starts to grow into a power law distribution starting from the left-hand side. The flat tail of the distribution reduces and gradually moves to the right bottom side of the chart. This illustrates the fact that sufficient data is needed to obtain sufficiently large schedule frequencies which exhibit a power law distribution.

It appears that the individuals with a good power law fit have a quite advanced evolution of their power law distribution, whilst the individuals without a good fit seem still at the transition phase in the evolutionary process (still having a long flat tail) as visible in Figure 4a. At small sample sizes, the power law distribution simply cannot be accurately determined. In literature, a sample size of $n \gtrsim 50$ is proposed as a rule of thumb.²⁰ The mean sample size for the Mobidrive individuals is $37.625 < 50$ (this is after excluding outliers). Therefore, more than six weeks of data are needed to consistently obtain power law distributions, allowing infrequent schedules the chance to occur at sufficient numbers. The exact sample size most likely differs for each individual. Additionally, a person's schedules might not be independent which could increase the need for sufficient data (e.g. there is a higher probability to have another home-work-home schedule after a *home-work-home* schedule than a *home-shopping-home* schedule usually taking place during the weekend). Future research will try to correlate the stage of evolution to person characteristics.

Significantly more data than six weeks of trip data (incl. trip purpose) may be needed in order to verify the above theory. To the author's best knowledge, such data does not exist for a large group of individuals. However, a 24-week dataset of trip data was donated by a punctual user of the Moves smartphone application¹⁸. This data exhibits a clear power law, as illustrated in Figure 5. The results from running the *powerLaw* algorithms on this data are tabulated in Table 1. The estimated exponent is greater than estimated for other datasets. This could be a consequence of a still-evolving distribution, or perhaps the exact exponent value depends on person characteristics such as the intensity of activity participation, age or employment. The null hypothesis of a good fit cannot be rejected.

5. Conclusion

Modeling people's behavior in e.g. travel demand models is an extremely complex, multidimensional process. However, the frequency of occurrence of day-long activity schedules obeys a ubiquitous power law distribution, commonly referred to as Zipf's law.¹ This paper discussed the role of aggregation within the phenomenon of Zipf's law in activity schedules. Aggregation was analyzed in two dimensions: activity type encoding and the aggregation of individual data in the dataset. Maximum likelihood estimation and the Kolmogorov-Smirnov goodness-of-fit statistic were used to correctly fit a power law to the data. The R package "*powerLaw*" was used to this end. Bootstrap-



(a) Some random examples of individual distributions rejecting the null hypothesis of a power law being a good fit in DEU Mobidrive 1999

(b) Simulation of the power law formation process based on increasing samples from the randomized DEU Mobidrive 1999 dataset.

Fig. 4: Illustrations regarding the potential effect of sample size in power law fitting.

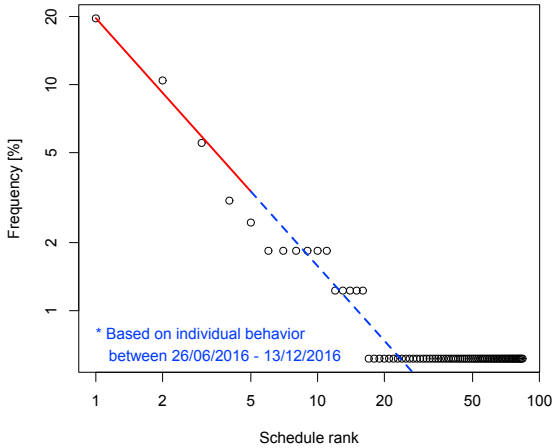


Fig. 5: Activity schedule distribution in the donated 24-week trip dataset. The red full line represents the fitted power law (according to the MLE + KS), the dotted blue line is the extrapolation of this fit.

ping procedures were used to evaluate parameter estimation uncertainty and to perform a hypothesis test with null hypothesis that a power law distribution is appropriate. This research worked with four datasets: the NHTS 2009, two six-week travel surveys (MobiDrive 1999 and Thurgau 2003) and a 24-week set of trip data which was donated by one individual. The latter was collected using the Moves smartphone application¹⁸ combined with manual verification and trip purpose enrichment.

To analyze the effect of aggregation in the first dimension, the activity type encoding, five different activity encoding aggregation levels were created in the NHTS 2009 dataset, each aggregating the activity types somewhat differently. The approach corresponds to constructing an encoding tree and pruning the branches to increase the aggregation level. Except for extreme levels of activity type aggregation, the effect on the power law distribution is negligible and one could state that Zipf’s law in activity schedules is not significantly influenced by activity type encoding aggregation.

To analyze aggregation in the second dimension, the analysis moved from study area-wide aggregated data to subsets of the data, and finally individual (longitudinal) data. A power law distribution was fitted to fully aggregated NHTS 2009 data. Subsequently, power laws were fitted to subsets based on the day of the week and gender. No considerable effect of subsetting the data was observed, provided that the the subset is sufficiently large. The two six-week travel surveys (Mobidrive 1999 and Thurgau 2003) allowed to analyze individual schedules. This analysis, in which power law distributions were fitted to each individual’s data, did not support Zipf’s law. However, subsequent

simulation and literature suggested that this is a consequence of insufficient data, i.e. the distributions seem underdeveloped even though they are based on six weeks of data. Finally, a 24-week trip history belonging to one person tested the validity of Zipf's law (for this particular individual) in longitudinal data. A good fit was found and the null hypothesis of a power law distribution being appropriate could not be rejected. The estimated exponent is slightly larger than expected under Zipf's law, but this could result from individual variation or a still-developing distribution.

Future research will try to correlate the stage of evolution of a power law activity schedule distribution to person characteristics, as well as modeling the mechanism that leads to Zipf's power law in activity schedules. Additionally, more tests could be done on simulated longitudinal data or on activity schedules inferred from GPS trajectories.

Acknowledgments

The authors would like to thank prof. dr. Kay Axhausen for providing the Mobidrive 1999¹⁶ and Thurgau 2003¹⁷ datasets. They are also thankful to the U.S. Department of Transportation, Federal Highway Administration, for making the NHTS 2009¹⁵ data freely available. The authors thank the donor of the 24-week of individual trip data.

References

1. Ectors, W., Kochan, B., Janssens, D., Bellemans, T., Wets, G.. Zipf's law in activity schedules. In: *hEART 2016: The 5th symposium arranged by European Association for Research in Transportation, Delft, The Netherlands, 14-16 September 2016*. 2016, URL: <http://hdl.handle.net/1942/22707>.
2. Zipf, G.K.. *Human Behaviour and the Principle of Least Effort*. Reading: Addison-Wesley; 1949.
3. Ki Baek, S., Bernhardtsson, S., Minnhagen, P.. Zipf's law unzipped. *New Journal of Physics* 2011;**13**(4):043004. URL: <http://stacks.iop.org/1367-2630/13/i=4/a=043004>. doi:10.1088/1367-2630/13/4/043004. arXiv:1104.1789.
4. Newman, M.. Power Laws, Pareto Distributions and Zipf's Law. *Contemporary physics* 2005;**46**:323–351. URL: <http://www.tandfonline.com/doi/abs/10.1080/00107510500052444>. doi:10.1080/00107510500052444. arXiv:0412004v3.
5. Gabaix, X.. Zipf's Law for Cities: an Explanation. *Quarterly Journal of Economics* 1999;**114**(August):739–767.
6. Soo, K.T.. Zipf's Law for cities: A cross-country investigation. *Regional Science and Urban Economics* 2005;**35**(3):239–263. doi:10.1016/j.regsciurbeco.2004.04.004.
7. Nitsch, V.. Zipf zipped. *Journal of Urban Economics* 2005;**57**(1):86–100. doi:10.1016/j.jue.2004.09.002.
8. González, M.C., Hidalgo, C.A., Barabási, A.L.. Understanding individual human mobility patterns. *Nature* 2008;**453**(7196):779–782. URL: <http://www.ncbi.nlm.nih.gov/pubmed/18528393>. doi:10.1038/nature07850. arXiv:0806.1256v1.
9. Brockmann, D., Hufnagel, L., Geisel, T.. The scaling laws of human travel. *Nature* 2006;**439**(7075):462–5. URL: <http://dx.doi.org/10.1038/nature04292>. doi:10.1038/nature04292. arXiv:0605511.
10. Zheng, Z., Rasouli, S., Timmermans, H.. Two-regime Pattern in Human Mobility: Evidence from GPS Taxi Trajectory Data. *Geographical Analysis* 2016;**48**(2):157–175. doi:10.1111/gean.12087.
11. Yang, X.H., Chen, G., Chen, S.Y., Wang, W.L., Wang, L.. Study on some bus transport networks in China with considering spatial characteristics. *Transportation Research Part A: Policy and Practice* 2014;**69**:1–10. URL: <http://dx.doi.org/10.1016/j.tra.2014.08.004>. doi:10.1016/j.tra.2014.08.004.
12. Paleari, S., Redondi, R., Malighetti, P.. A comparative study of airport connectivity in China, Europe and US: Which network provides the best service to passengers? *Transportation Research Part E: Logistics and Transportation Review* 2010;**46**(2):198–210. URL: <http://dx.doi.org/10.1016/j.tre.2009.08.003>. doi:10.1016/j.tre.2009.08.003.
13. Guidotti, R., Trasarti, R., Nanni, M.. TOSCA: Two-Steps Clustering Algorithm for Personal Locations Detection. *Proceedings of the 23rd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 2015;(January). doi:10.1145/2820783.2820818.
14. Song, C., Koren, T., Wang, P., Barabasi, A.L.. Modelling the scaling properties of human mobility. *Nature Physics* 2010;**6**(10):1–6. URL: <http://dx.doi.org/10.1038/nphys1760>. doi:10.1038/NPHYS1760. arXiv:1010.0436.
15. U.S. Department of Transportation, Federal Highway Administration, . 2009 National Household Travel Survey. 2009. URL: <http://nhts.orl.gov>.
16. Chalasani, V., Axhausen, K.W.. Mobidrive: A six week travel diary. 2004. URL: <https://www.ethz.ch/content/dam/ethz/special-interest/baug/ivt/ivt-dam/vpl/tsms/tsms2.pdf>.
17. Loechl, M.. Stability of Travel Behaviour: Thurgau 2003. 2005. URL: <http://archiv.ivt.ethz.ch/vpl/publications/tsms/tsms16.pdf>.
18. ProtoGeo Oy, . Moves - Activity Diary for iPhone and Android. 2016. URL: <https://moves-app.com/>.
19. Janssens, D., Declercq, K., Wets, G.. Onderzoek Verplaatsingsgedrag Vlaanderen 4.5 (2012-2013). Tech. Rep.; Hasselt University, Transportation Research Institute (IMOB); 2014. URL: <http://www.mobielvlaanderen.be/pdf/ovg45/ovg45-analyse-globaal.pdf>.
20. Clauset, A., Shalizi, C.R., Newman, M.E.J.. Power-Law Distributions in Empirical Data. *SIAM Review* 2009;**51**(4):661. URL: <http://link.aip.org/link/SIREAD/v51/i4/p661/s1/&Agg=doi>. doi:10.1137/070710111. arXiv:0706.1062v2.
21. Urzúa, C.M.. Testing for Zipf's law: A common pitfall. *Economics Letters* 2011;**112**(3):254–255. URL: <http://dx.doi.org/10.1016/j.econlet.2011.05.049>. doi:10.1016/j.econlet.2011.05.049.
22. Gillespie, C.S.. Fitting Heavy Tailed Distributions: The powerLaw Package. *Journal of Statistical Software* 2015;**64**(2):1–16. URL: <http://www.jstatsoft.org/v64/i02>.