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

Non Peer-reviewed author version

ECTORS, Wim; KOCHAN, Bruno; JANSSENS, Davy; BELLEMANS, Tom & WETS, Geert (2020) Zipf’s power law in activity schedules and the effect of aggregation. In: FUTURE GENERATION COMPUTER SYSTEMS-THE INTERNATIONAL JOURNAL OF ESCIENCE, p. 1014-1025.

DOI: 10.1016/j.future.2018.04.095
Handle: http://hdl.handle.net/1942/26622

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]

[a] *UHasselt - Hasselt University, Transportation Research Institute (IMOB), Agoralaan, 3590 Diepenbeek, Belgium*

Abstract

People's behavior depends on extremely complex, multidimensional processes. This poses challenges when trying to model their behavior. In the transportation modeling community, great effort is spent to model the activity schedules of people. Remarkably however, the frequency of occurrence of day-long activity schedules obeys a ubiquitous power law distribution, commonly referred to as Zipf's law. Previous research established the universal nature of this distribution and proposed potential application areas. However, these application areas require additional information about the distribution's properties. To stress-test this universal power law, this paper discusses the role of aggregation within the phenomenon of Zipf's law in activity schedules. Aggregation is analyzed in three dimensions: activity type encoding, aggregation over time and the aggregation of individual data. Five data sets are used: the household travel survey from the USA (2009) and from GBR (2009-2014), two six-week travel surveys (DEU MobiDrive 1999 and CHE Thurgau 2003) and a donated 450-day data set from one individual. To analyze the effect of aggregation in the first dimension, five different activity encoding aggregation levels were created, each aggregating the activity types somewhat differently. In the second dimension, the distribution of schedules is compared over multiple years and over the days of the week. Finally, in the third dimension, the analysis moves from study area-wide aggregated data

*Corresponding author. Tel.: +32-11-269114.
Email address: wim.ectors@uhasselt.be (Wim Ectors)

Preprint submitted to Future Generation Computer Systems *May 8, 2018*

to subsets of the data, and finally to individual (longitudinal) data.

Keywords: Zipf, power law, activity schedule, 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.

As discussed by Ectors et al. [1], the activity type exhibits a *universal* Zipf power law in study area-wide aggregated data. They suggested two practical uses of this universal distribution: (i) as an additional, necessary condition in a model's validation and (ii) as a possible way of extending mobility models which are based on universal mobility laws. Such models (e.g. TimeGeo [2] or DITRAS [3]) are predominantly based on universal mobility laws and less on disaggregated data, however they typically lack an integration of the activity type in their predicted mobility patterns. They have few tunable parameters.

For these applications, it is necessary to understand the extent of the universal distribution. Therefore, this paper attempts to stress-test the observed distribution by investigating the effects of aggregation in several dimensions. In this context, aggregation refers to the process of combining individual records into one dataset. In other words, the phenomenon of a universal law is investigated on different scales (e.g. from large to small spatial or temporal scales).

Aggregation is analyzed in three dimensions: (i) the activity type encoding, (ii) aggregation over time and (iii) the aggregation of individual records. For example, in the third dimension the analysis moves from highly aggregated data (e.g. data belonging to multiple individuals of a given study area) to subsets of the data (e.g. based on demographic properties) and finally to longitudinal data for single individuals. By systematically testing the limits of the observed law, modelers, researchers and practitioners receive confidence in its extensibility.

Chen et al. [4] list unresolved issues with respect to transportation planning applications. They mention how the ecological fallacy, i.e. the situation where a conclusion about individual behavior is drawn from data about aggregate behavior [5], is a contemporary issue. This paper partially addresses this issue in the context of the phenomenon of Zipf's law in activity schedules by investigating the properties of the distribution down to the lowest level of aggregation, that is the individual level. Due to the power law's nature, a *longitudinal* data set containing observed activity schedules is required.

In the remainder of this paper, first a literature review offers background information on Zipf's law and other universal distributions within transportation sciences. Subsequently, the data and basic methodology for estimating a power law fit are detailed, after which the effect of aggregation is analyzed in three dimensions: activity type encoding, aggregation over time and aggregation of individual data. A discussion section discusses some limitations of the research, and it interprets the analysis results with respect to the two suggested practical use cases of this distribution. The conclusion section finalizes this paper.

2. Literature review

The scale-free distribution which was observed in day-long activity schedules obeys a power law distribution [1]. The same distribution has been observed in diverse natural and social processes. It is often referred to as Zipf's law. The observation and analysis of power laws has a rich history. In 1913, Auerbach discovered that city sizes follow a power law. Estroup described in 1916 that a power law distribution governs the frequency of words, but it was not until after Zipf, an American linguist, published his work in 1949 that such power law distributions were called after him. Zipf investigated and popularized the distribution. It was revealed that the same power law distribution holds for a large number of events in various domains, extending from sizes of earthquakes, people's annual income, solar flares, to even the number of citations received on papers [6, 7, 8].

The power law distribution belongs to the family of heavy-tailed distributions, meaning that the distribution goes to zero more slowly than an exponential function (that is, they have a heavier tail than exponential distributions). Most commonly the discrete rank-size interpretation as Zipf's law is 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.:

$$\phi(r_i) = \frac{\phi(r_1)}{r_i} \tag{1}$$

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

Zipf's law and other power laws can also be linked to Pareto distributions, which take the more formal form of

$$P(X > x) = \begin{cases} \left(\frac{x_m}{x}\right)^{\alpha} & \text{if } x \geq x_m \\ 1 & \text{if } x < x_m \end{cases} \tag{2}$$

where $x_m > 0$ the minimum possible value of X, and $\alpha > 0$. Interpreting this equation for city size S yields

$$P(S > s) = \left(\frac{s}{s_{\min}}\right)^{-\alpha} = as^{-\alpha} \text{ for } s \geq s_{\min} \tag{3}$$

which states that the probability for a city to have a size greater than s decreases as $1/s$ if $\alpha = 1$ under Zipf's law. In this equation, a is a scaling factor [9]. The exponent α in this cumulative density function yields an exponent value of $\alpha + 1$ in the corresponding probability density function (PDF). The above equations also illustrate the scale-free nature of these distributions. Zipf's law (or sometimes called the zeta distribution) can be considered as a discrete version of a Pareto or power law distribution.

No conclusive proof exist against the existence of a natural power law mechanism, nor does a general agreement exist on the origin of the widespread manifestation of Zipf's law. The fact that many observations appear to share the same exponent value desires a universal mechanism which explains this distribution. However, most researchers agree that *several* mechanisms may lead to the observed power law distributions [8]. Examples of such mechanisms can be found in literature [7, 8, 9, 10, 11, 12, 13, 14, 15, 16] as this is not the focus of the current paper.

Still, some research argues against Zipf's apparent universality. In a large-scale study, 73 cities from across the world were analyzed for conformity with Zipf's law. The analysis showed that a Zipf's power law had to be rejected in more cities than expected [17]. A meta study based on 515 estimates from 29 studies on city size distributions found that the power law exponent is actually closer to 1.1, being statistically different from Zipf's value of 1.0 [18].

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 frequency [19], as well as in location visiting duration [20] and travel time in taxi travel [21]. Power law distributions were also observed in bus transport networks [22] and in airport networks [23]. Some researchers also used these universal distributions in their experiments [24, 25]. More recently, evidence for a universal Zipf power law in activity schedules was given [1].

Activity schedules are often discussed in transportation-related literature, especially within the context of activity-based modeling. Activity-based models are a class of state-of-the-art models that attempt to predict the demand for transportation (in an agent-based fashion) as a derived demand from the desire to participate in activities. Many such models typically build activity schedules for a synthetic population in a sequential fashion: mandatory activities are predicted first, after which non-mandatory (household maintenance) and discretionary (flexible) activities are predicted in succession [26]. In a similar sequential approach, Rinzivillo et al. [27] proposed an Activity-Based Cascading (ABC) classification strategy to enrich mobility data that misses activity purpose information. Interestingly, compared to many traditional activity type inference approaches which annotate each movement independently, the ABC approach takes the context of full activity schedules into account, yielding superior classification accuracy. This approach is especially efficacious given the highly imbalanced activity type distribution which needs to be predicted. The activity schedule distribution in this paper obeys a power law distribution, which may also be considered severely imbalanced. Such a cascaded classification approach might therefore be very useful when utilizing the universal activity type distribution within mobility models which are based on universal mobility laws (see section 1).

3. Data description

This research employs five data sets: (i) a Household Travel Survey (HTS) from the US, the USA National Household Travel Survey (NHTS) 2009 data set [28], (ii) the National Travel Survey (NTS) 2009-2014 from GBR [29], (iii) a six-week travel survey from Germany, DEU MobiDrive 1999 [30], (iv) a Swiss six-week travel survey CHE Thurgau 2003 [31] and (v) a 450-day set of trip data which was donated by one individual from Flanders, Belgium. The 450-day data was collected using the Moves smartphone application [32] combined with manual verification and trip purpose enrichment between June 26, 2016 and September 18, 2017. There were 435 days with out-of-home activities. The OVG HTS [33] activity encoding was used (10 classes). Table 1 tabulates the different datasets with their characteristics. It indicates which distinct aspects made each dataset suitable for an analysis in the indicated aggregation dimension in the following sections.

Out-of-home activity schedules are constructed out of trip purpose information from these data sets. Trip purposes are concatenated into a sequence which represents a schedule with the main out-of-home activities. From the NHTS 2009 data set 257,586 schedules could be extracted ($\pm$83,000 distinct schedules). From GBR NTS 2009-2014, 551,234 schedules were extracted. The DEU Mobidrive 1999 and CHE Thurgau 2003 data sets yield, respectively, 13,244 and 8,522 schedules.

Table 1: Characteristics of the datasets

Name	Origin	Total number of extracted out-of-home activity schedules	Number of activity type classes	Survey period per individual [days]	Survey period
USA NHTS 2009	United States	257,586 (iii)	37 (i)	1	03/2008 - 05/2009
GBR NTS 2009-2014	United Kingdom	551,234	23	7	2009-2014 (ii)
DEU MobiDrive 1999	Germany	13,244	22	42 (iii)	05/1999 - 12/1999
CHE Thurgau 2003	Switzerland	8,522	25	42 (iii)	08/2003 - 12/2003
450-day trip data	Belgium; donated	435	11	450 (iii)	06/2016 - 09/2017

Note: Distinct feature motivating the analysis in aggregation dimension:
(i) Activity type encoding, (ii) Time, (iii) Individual data

4. Description of the estimation procedure

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 [8, 34, 35]. 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. [34] proposed a method based on maximum likelihood estimation (MLE) combined with the Kolmogorov-Smirnov (KS) goodness-of-fit (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 $x \to 0$, resulting in an infinite area under the distribution. The cutoff parameter depicts the fact that few data sets 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" [36] was developed to automate the MLE + KS estimation process. The $x_{\min}$ parameter is optimized by means of the KS statistic. The package also 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 [34].

The PDF of a power law distribution takes the form of

$$p(x) = Cx^{-\alpha} \tag{4}$$

where C a constant and α the exponent of the power law. When fitting a Zipf's power law on non-numeric data such as words in a text (see Equation 1) or in this research activity schedules, one fits a power law on rank-ordered (frequency) distributions of the data. Doing so, one will estimate the parameters in $f(n) = C'n^{-\tau}$ where n the (relative) frequency in the rank-ordered distribution and τ the so-called Zipf exponent. For a given data set, the two exponents α and τ are related by Equation 5 [37, 38].

$$\alpha = 1 + \frac{1}{\tau} \tag{5}$$

The estimates in this paper from the PoweRlaw package are those based on Equation 4, that is the estimates tabulated are $\hat{\alpha}$. Zipf's exponent $\tau = 1$ yields an expected $\alpha = 2.0$ according to Equation 5, in order to confirm Zipf's law in activity schedules.

5. Aggregation in activity type encoding

This analysis is aimed at investigating the effects of a transformed activity type variable on the activity type schedule distribution. The activity type variable may be transformed in to a new aggregation level by grouping some of the activity types in to a new class.

Other research [39] found that there is a large effect from the choice of activity type classes on the activity type classification accuracy in the context of activity type inference in e.g. GPS data. In that research, the activity type variable was optimized, as it was demonstrated how an inappropriate choice of the classes could be used to artificially increase classification accuracy. Although the context is different, the effects of different encoding aggregation levels on the distribution's shape needs to be evaluated.

5.1. Encoding aggregation levels

To analyze the effect of aggregation in the first dimension, the activity type encoding, different activity type encoding aggregation levels were created for the

USA NHTS 2009 data set as this data set contains one of the richest activity type (travel motive) variables. Starting from the original 37 activity types, denoted here as Level 0, four more sets of encodings were proposed, each aggregating (or *grouping*) the activity type classes 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 grouping of the second digits was conducted according to common-practice and targeted at reducing the number of distinct activity types, yet not as strongly as for Level 2. 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 HTS [33] to each NHTS category. This re-coding strategy was used as it results in the same number of activity type classes as in Level 2b (see further), yet it is made up out of different classes. This way, one can evaluate whether the choice of activity type classes has an influence, independent of the number of classes. 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 [40].

These five activity encoding schemes were used to construct day-long activity schedules for the individuals in the NHTS data set. Table 6 in Appendix tabulates the different encoding levels side-by-side.

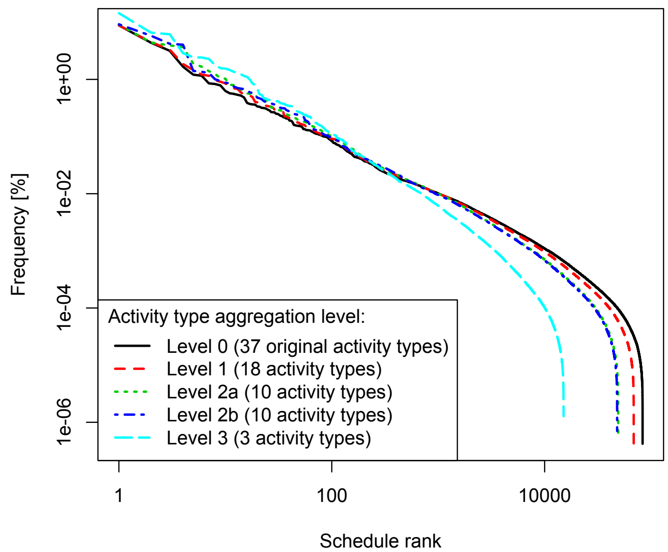


Figure 1: Activity schedule distribution in the USA NHTS 2009 data set based on five different activity encoding aggregation schemes.

5.2. Encoding effects on the distribution

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 data set. 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 2 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.

Table 2: Estimation results from the *R* package poweRlaw for activity schedule distributions.

		poweRlaw estimations (MLE + KS)			Bootstrapping uncertainty evaluation		
Data set	Aggregation or subset	$\hat{\alpha}$	$\hat{x}_{\min}$	Cum. pct rejected	AM($\hat{\alpha}$)	SD($\hat{\alpha}$)	P-value
USA NHTS 2009	Level 0 (37 original act. 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
USA NHTS 2009	Employed	2.344	83702196	64%	2.357	0.213	0.368
USA NHTS 2009	Unemployed	2.089	35493956	63%	2.102	0.134	0.907
USA NHTS 2009	Using public transport	2.497	5614144	43%	2.308	0.270	0.947
USA NHTS 2009	Not using public transport	1.997	36809978	55%	2.000	0.070	0.258
GBR NTS 2009	All data aggregated	1.802	4.213	14%	1.837	0.066	0.002
GBR NTS 2010	All data aggregated	1.802	4.649	15%	1.837	0.061	0.004
GBR NTS 2011	All data aggregated	1.832	3.642	13%	1.852	0.054	0
GBR NTS 2012	All data aggregated	1.908	24.271	21%	1.88	0.085	0.197
GBR NTS 2013	All data aggregated	1.803	4.06	13%	1.829	0.052	0.008
GBR NTS 2014	All data aggregated	1.852	12.918	18%	1.842	0.066	0.126
GBR NTS 2009-2014	All data aggregated	1.862	4.071	9%	1.869	0.117	0
USA NHTS 2009	All data aggregated	2.003	36809977	55%	2.006	0.070	0.255
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	2.625	1	0%	2.299	0.296	0.169

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

6. Aggregation over time

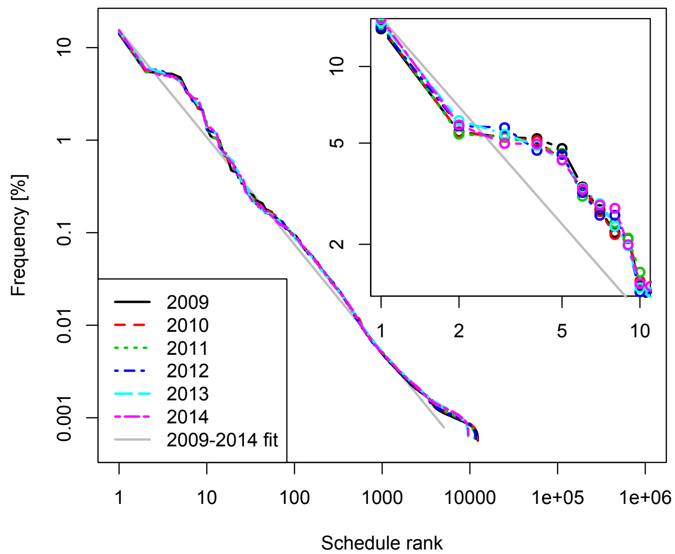
This section discusses the distribution of activity schedules over time. Aggregation over time is interpreted here as the analysis over different temporal resolutions. Possible seasonality effects or other long-term variations are investigated. To this end, the GBR NTS 2009-2014 and USA NHTS 2009 data were employed. Figure 2 illustrates the data.

6.1. Long-term variations

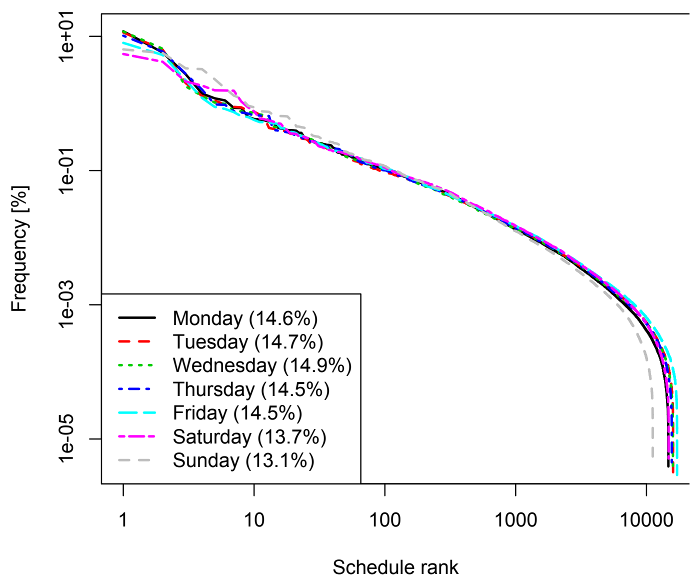
As can be seen in Figure 2a, the distribution of activity schedules appears to be extremely stable over a time period of several years. The estimates in Table 2 confirm this observation. The estimated power law exponent varies between 1.802 and 1.908 ($\bar{\alpha} = 1.833$) when using MLE and the KS cutoff criterion. The mean exponent value in the bootstrapping procedure ranges only between 1.829 and 1.880 ($\bar{\alpha} = 1.846$).

When aggregating the NTS data of these six years into a single data set, a power law exponent of 1.862 is estimated through MLE + KS and 1.869 by means of the bootstrapping procedure. These values are conform the previous analysis, providing evidence that data may be aggregated over time without significantly influencing the distribution, thanks to the apparent stability over time. This is an important finding as it enables the analysis of activity schedule distributions through aggregation of the data over time in study areas of which the HTS sample sizes are rather small.

One has to remark that the bootstrapping GoF test reveals that the distribution in the GBR NTS data is not a clean power law. Only in 2 cases the null hypothesis (of a power law distribution being appropriate) is not rejected at a significance level of 10%. However, no other typical alternate distribution such as the exponential, log-normal or truncated power law distribution seems more appropriate than a power law. The rejection of the null hypothesis in this particular data is most likely caused by the distinctive behavior in the high-frequency region. This might be caused by the survey design or by other factors. Still, a power law seems (visually) appropriate.



(a) Different years in the GBR NTS 2009-2014 data, with fitted power law (ML +KS) of the combined data



(b) Different days of the week in the USA NHTS 2009 data

Figure 2: Analyzing activity schedule distribution over different time spans

6.2. Day-of-the-week variations

The time dimension was also analyzed on a smaller temporal resolution, i.e. for the days of the week. For this, the USA NHTS 2009 was used. Figure 2b shows the distribution of activity schedules for different days of the week. Visually, their distributions are nearly identical. Table 2 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 data set. This suggests that some schedules may be more typical for a particular day of the week, yielding higher frequencies for the top-ranked schedules on that particular day. 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. Still, this seems not to have an 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.

6.3. Multiple observation windows in individual longitudinal data

Lastly, the aggregation of longitudinal individual data over time will be discussed in more detail in subsection 7.3 in order to investigate the buildup and evolution of the power law distribution.

7. Aggregation of individual data

The fact that Zipf's law seems valid on aggregated schedules for a whole study area was established [1]. 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 gender, employment status and public transport (PT) usage were taken from the USA NHTS 2009 data set and a power law distribution was

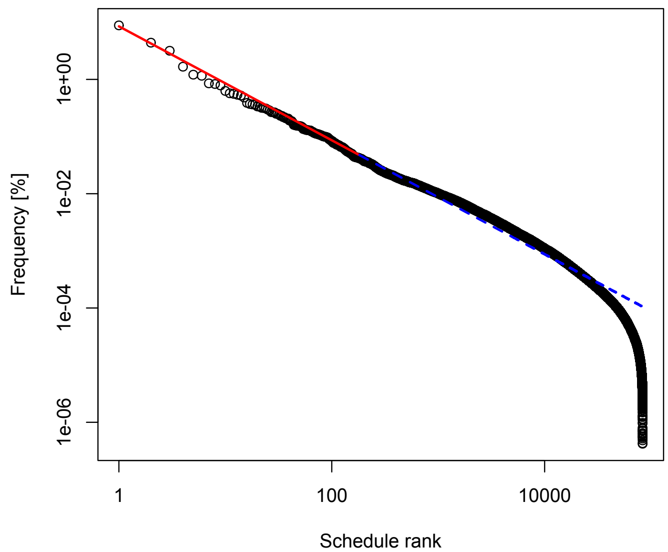


Figure 3: Activity schedule distribution in the USA NHTS 2009 data set. The red full line represents the fitted power law (according to the MLE + KS), the dotted blue line is the extrapolation of this fit.

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 450-day trip history belonging to one person tests the validity of Zipf's law (for this particular individual) in longitudinal data.

7.1. Aggregation to study area level

Figure 3 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 GBR NTS 2009-2014, DEU Mobidrive 1999 and CHE Thurgau 2003 data sets when each recorded day is treated independently and subsequently aggregated. Table 2 lists the estimates for these experiments. All three data sets 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 analyzed in more detail in Ectors et al [1].

7.2. Subsets of a study area

The USA NHTS 2009 was used to analyze subsets. It is a significantly large data set, which avoids incorrectly rejecting a power law distribution due to insufficient data. Furthermore, each time only 2 subsets were created in this experiment. As illustrated in Figure 4, subsets were generated based on gender, employment status and the use of PT. These subsets were chosen because they might yield different transportation behavior (and we could observe different distributions for the subsets). Additionally, the subsets go from approximately equal sizes in case of gender, to a highly unbalanced ratio for the use of PT. Visually, their distributions are nearly identical. Table 2 lists the power law fit estimates. Again, estimates close to Zipf's law's value of 2.0 were found. The estimated values of subsets differ slightly. 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. Each subset does not (necessarily) have the same schedule at each rank. Additionally, all subsets have a p-value > 0.10, so the null hypothesis of a power law being an appropriate distribution cannot be rejected. 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.

Figure 4a illustrates that the activity schedule distribution of men and women are nearly equal. Table 3 lists the top 10 schedules for both groups. One observes how the highest-ranked schedules are the same for men and women, though occurring at slightly different frequencies. In general (and without surprise), the simplest schedules involving only one out-of-home activity occur with the highest frequency. Onward from rank five, differences between men and women begin to manifest. Most of the differences seem to confirm stereotypical presuppositions, e.g. men work out or do sports more often, women most often do the grocery shopping.

As illustrated in Figure 4b, there are some differences in the higher-frequency range between employed and unemployed persons. However, in general the distribution seems to obey a power law. Table 4 lists the top 10 schedules according to employment status. For workers, the schedule with one work activity clearly dominates whilst for unemployed persons the schedule with a single shopping activity has the highest frequency. For workers, this schedule appears at a much lower frequency than for unemployed persons, yet workers appear to frequently schedule their shopping activities after their work, partially compensating for the lower 'H - buy goods - H' schedule frequency. The disproportionately high frequency of 'H - work - H' for workers is most likely causing the power law exponent estimate to inflate; it has a value of $\hat{\alpha} = 2.344$ whilst for the group of unemployed persons the MLE + KS procedure yielded an estimate of $\hat{\alpha} = 2.089$. In general, all schedules with only one out-of-home activity occur at least 50% less frequent for workers than for unemployed persons. This shows the impact of work activities and the resulting need for chaining activities. It confirms the well-known practice of considering 'work' as a mandatory activity in activity-based models, being predicted with priority over other (non-mandatory) activity types [41]. Although the ranks of corresponding schedules differ, still the distribution is a power law.

Finally, Figure 4c shows the schedule distribution for subsets based on PT usage. The subset of PT users represents only 4.4% of the weighted USA NHTS 2009 sample. A person is a member of this subset if he or she used a PT mode on the surveyed travel day. Despite the small subset, it still displays a power law distribution. However, the power law regime breaks down sooner compared to the non-PT users since the number of distinct schedules is also much smaller. With decreasing subset size, the power law regime will become smaller to the point where insufficient observations are present to reliably observe a power law. Table 5 lists the top 10 of schedules according to PT usage. Somewhat surprisingly, no large differences can be observed. Out of the two mode categories, the non-PT mode is often considered most flexible as almost every destination can be reached from door to door. Remarkably, this does not yield more complex activity schedules: in fact the subgroup using PT has slightly more schedules with greater than one out-of-home activity in its top 10, compared to the subgroup not using PT. This suggests that the PT group tends to chain more activities

Table 3: Top 10 schedules for men and women in NHTS 2009. ‘H’ is short for ‘home’

	Men		Women	
Rank	Schedule	Frequency [%]	Schedule	Frequency [%]
1	H - work - H	10.382	H - work - H	7.288
2	H - education - H	4.401	H - education - H	4.385
3	H - buying goods - H	2.847	H - buying goods - H	3.488
4	H - visit friends/relatives - H	1.480	H - visit friends/relatives - H	1.854
5	H - gym/exercise/play sports - H	1.404	H - religious activity - H	1.333
6	H - religious activity - H	0.980	H - medical/dental services - H	1.031
7	H - get/eat meal - H	0.834	H - buy goods - buy goods - H	1.026
8	H - work - get/eat meal - return to work - H	0.728	H - gym/exercise/play sports - H	1.021
9	H - work - H - gym/exercise/play sports - H	0.693	H - get/eat meal - H	0.827
10	H - go out/hang out - H	0.683	H - work - buy goods - H	0.649

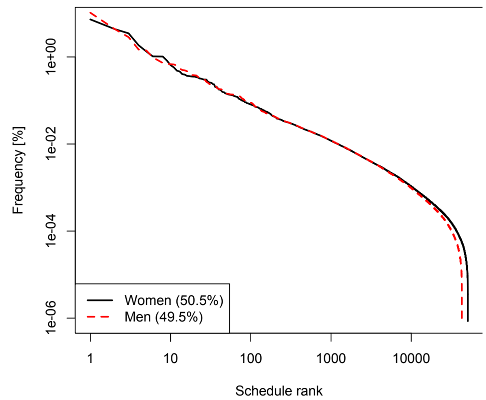
Table 4: Top 10 schedules for employed and unemployed people in NHTS 2009. ‘H’ is short for ‘home’

	Employed		Unemployed	
Rank	Schedule	Frequency [%]	Schedule	Frequency [%]
1	H - work - H	15.317	H - buy goods - H	6.002
2	H - buy goods - H	2.086	H - education - H	2.902
3	H - work - buy goods - H	0.992	H - visit friends/relatives - H	2.801
4	H - work - H - buy goods - H	0.942	H - medical/dental services - H	2.412
5	H - work - get/eat meal - return to work - H	0.938	H - gym/exercise/play sports - H	2.007
6	H - work - H - gym/exercise/play sports - H	0.911	H - religious activity - H	1.720
7	H - visit friends/relatives - H	0.872	H - buy goods - buy goods - H	1.595
8	H - religious activity - H	0.716	H - work - get/eat meal - H	1.408
9	H - gym/exercise/play sports - H	0.651	H - go out/hang out - H	0.833
10	H - work - get/eat meal - H	0.600	H - shopping/errands (other) - H	0.768

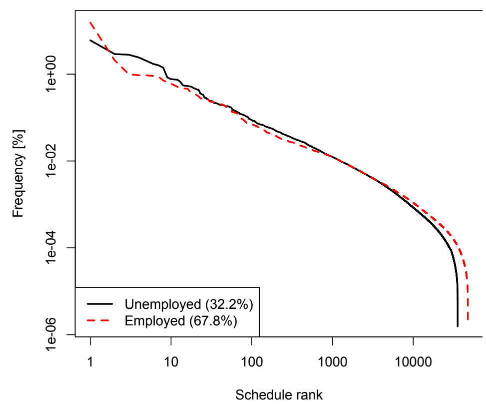
than users of other modes. Another difference is for example the schedule ‘H - religious activity - H’ which occurs at a much lower frequency (0.326%) for PT users than for others, yet the combination of a religious activity with eating out and/or shopping afterwards does occur at a higher frequency for PT users compared to others.

7.3. The individual level

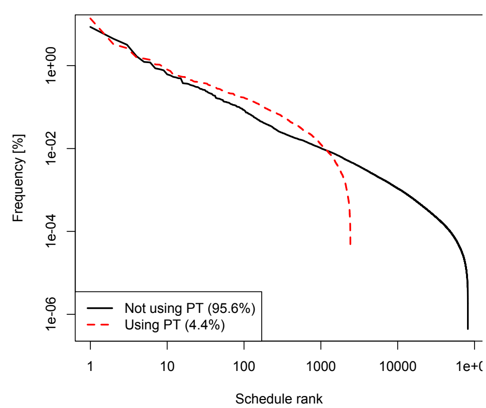
It is a challenge to recognize whether Zipf’s law is valid for activity schedules from each individual separately, similarly to other universally distributed quantities like displacement distance, location visiting frequency etc. [19]. The



(a) Subsets based on gender



(b) Subsets based on employment status



(c) Subsets based on PT usage

Figure 4: Activity schedule distribution in subsets of the USA NHTS 2009 data set (using the original activity encoding).

Table 5: Top 10 schedules according to PT usage on the the travel day in NHTS 2009. 'H' is short for 'home'

	Using PT		Not Using PT	
Rank	Schedule	Frequency [%]	Schedule	Frequency [%]
1	H - work - H	13.765	H - work - H	8.591
2	H - education - H	3.287	H - education - H	4.445
3	H - buy goods - H	2.642	H - buy goods - H	3.196
4	H - medical/dental services - H	1.578	H - visit friends/relatives - H	1.683
5	H - work - get/eat meal - return to work - H	1.478	H - gym/exercise/play sports - H	1.230
6	H - visit friends/relatives - H	1.384	H - religious activity - H	1.197
7	H - work - H - buy goods - H	1.069	H - work - get/eat meal - H	0.862
8	H - buy goods - buy goods - H	1.055	H - medical/dental services - H	0.819
9	H - go out/hang out - H	0.913	H - buy goods - buy goods - H	0.772
10	H - gym/exercise/play sports - H	0.805	H - go out/hang out - H	0.615

schedules are not fully independent in this case, but belong to one individual. To analyze this question, three data sets were used: two six-week travel surveys (DEU Mobidrive 1999 and CHE Thurgau 2003) and the donated 450-day trip data set from one individual.

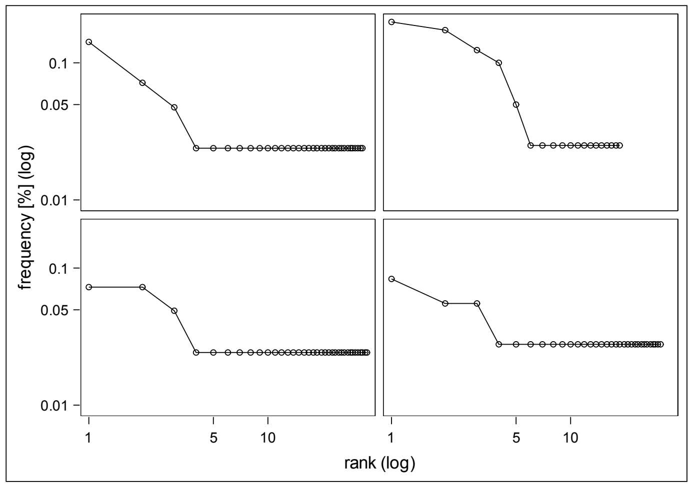
To analyze the six-week travel surveys, a variable (present in the original data set) with 10 trip purpose classes is used instead of the 23 classes originally in the survey. As the data is limited (six weeks) this will ensure the highest possible frequencies for each schedule, so a power law might be discovered in 'only' six weeks of data. As discussed in section 5, this choice should not negatively influence the estimation results. Some individuals in the data have very few days within which 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 was used instead. Each observed distribution was tested against a predefined distribution in SAS based on this statistic. The null hypothesis H_0 is that a power law distribution with specified α is a good fit.

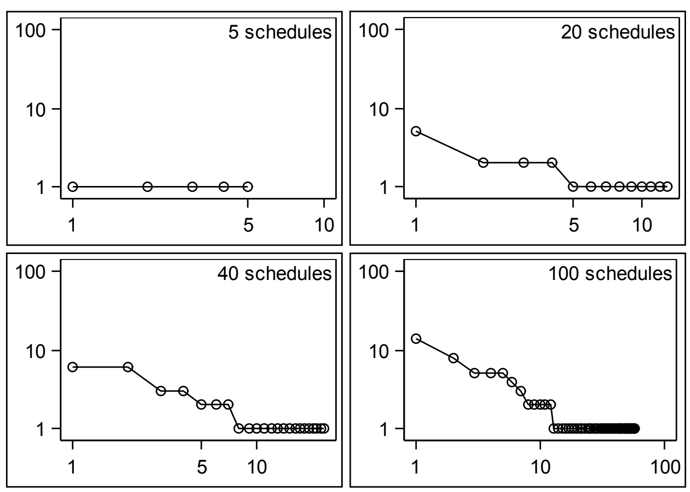
If one α is imposed for all individuals, 12% (43 out of 352) have a distribution which is not significantly different from a power law distribution, based on a significance level of 5%. Similar results are obtained using the CHE Thurgau 2003 data set: 4% of the individuals (9 out of 230) have a distribution 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 do not support the theory that Zipf's law is also valid for individuals. However, as can be seen in Figure 5a, 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 5b. 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 long flat tails) as visible in Figure 5a. At small sample sizes, the power law distribution simply cannot be accurately determined. In literature, a minimum sample size of $n \gtrsim 50$ is proposed as a rule of thumb to reliably fit a power law [34]. The mean sample size for the Mobidrive individuals is $37.625 < 50$ (this is even after excluding outliers). Therefore, more than six weeks of data



(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 data set.

Figure 5: Illustrations regarding the potential effect of sample size in power law fitting.

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 450-day data set of trip data was donated by a punctual user of the Moves smartphone application [32]. This data exhibits a clear power law, as illustrated in Figure 6. Two power law fits were included, the first based on the mean exponent value $\bar{\alpha}$ from the bootstraps, the second based on a separate MLE step without excluding any of the data. The difference between both fits illustrates how the distribution might still be evolving. Infrequent schedules did not have the chance to occur at a sufficient frequency to guarantee a power law regime over a large range. Like in the other distributions (e.g. Figure 4), the power law regime breaks down at low frequencies. In previous figures, individual weights were used to calculate frequencies which resulted in a smooth curve, whilst in Figure 6 data of a single user is plotted without the use of weights. Therefore, discrete plateaus of schedules occurring at the same frequency are visible.

The results from running the poweRlaw algorithms on this data are tabulated in Table 2. The estimated exponent is greater than estimated for other data sets, although the bootstrapping results yield $\bar{\alpha} = 2.299$ which is not an extreme value. Remarkably is also that the KS criterion does not exclude any data ($x_{\mathrm{min}} = 1$, the minimum frequency in this data). A higher than expected $\hat{\alpha}$ could also 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

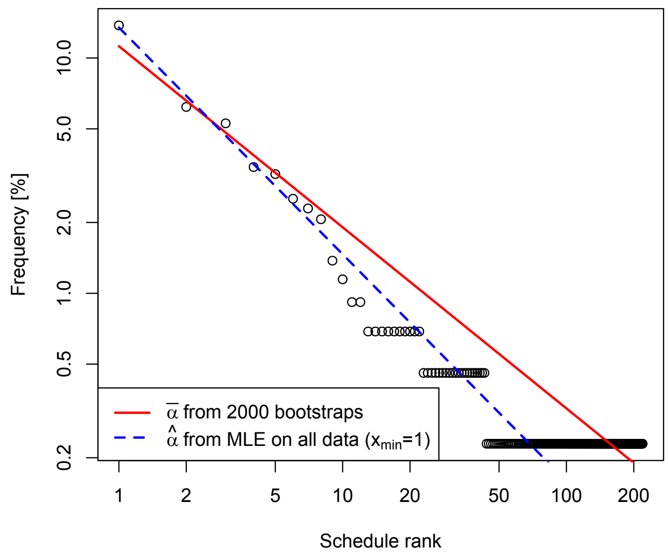


Figure 6: Activity schedule distribution of all donated 450 days of annotated Moves trip data of one individual

rejected.

The buildup and evolution of the power law distribution may be investigated by analyzing the staged aggregation of individual data over time. Figure 7 shows the observed schedule distribution of one individual after certain periods of time since the recording started. For this also the 450-day data was used. One observes how after a few weeks an unmistakable power law becomes visible. Over time, more distinct schedules occur and the relative frequencies of all schedules decrease (the power law distribution seems to move down). This process seems to saturate at some point, up to the point where no new schedules are made. This saturation point has most likely not yet been reached after 450 days of observations. Still, the evolution of the distribution seems to slow down considerably when comparing the change between 1 month & 6 months ($\Delta t = 5$ months) and 6 months & 1.23 years ($\Delta t \approx 8.75$ months)(law of diminishing returns). Throughout its evolution, the distribution appears to maintain a relatively constant slope on this log-log plot, which is analyzed quantitatively next.

Figure 8 illustrates the evolution of the power law exponent $\hat{\alpha}$ on a continuous

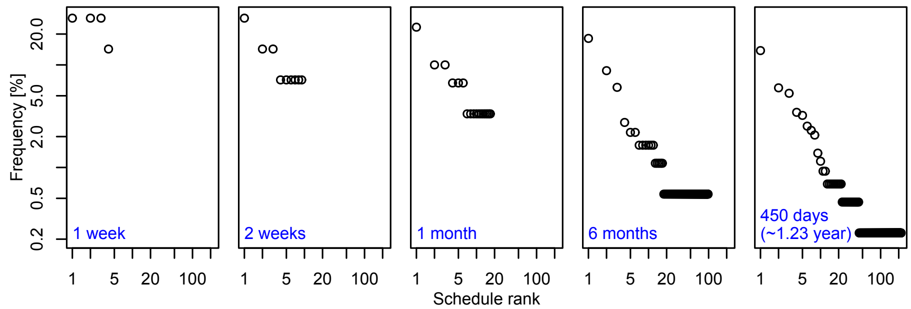


Figure 7: Growth and evolution of the distribution of schedules in the donated 450 days of individual data

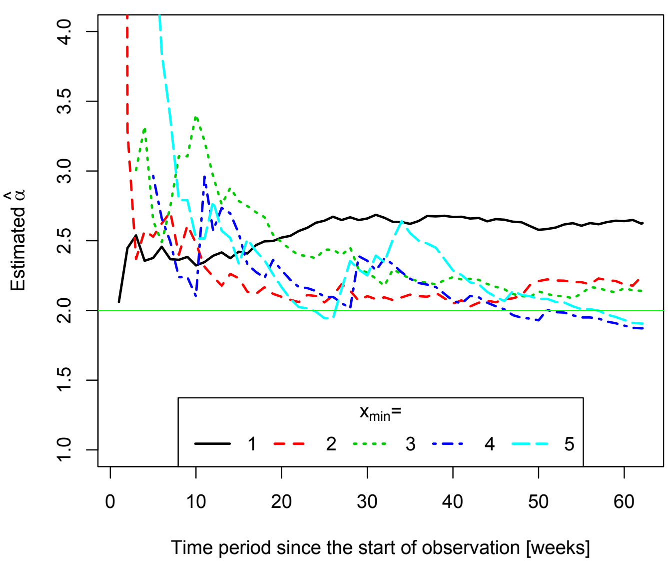


Figure 8: Estimated $\hat{\alpha}$ in function of the considered time period, for different values of $x_{\min}$, based on the schedules in the donated 450 days of individual data

scale. The evolution is plotted for several values of $x_{\min}$ so that the underlying and uncontrolled effect of the KS criterion for a cutoff value $x_{\min}$ is excluded. In this data, an $x_{\min} > 5$ is not expected. The frequency count of 1 is represented by the bottom plateaus in Figure 7, that of 2 the second to last plateau etc. One observes in Figure 8 how the evolution of $\hat{\alpha}$ with $x_{\min} = 1$ (bottom plateau is included in the estimation of $\hat{\alpha}$) is clearly different from the other ones (where $x_{\min} > 1$ and the bottom plateau (or more) is excluded). For values of $x_{\min} > 1$, $\hat{\alpha}$ seems to evolve to a value close to the expected value of 2.0. For time periods smaller than ten to fifteen weeks, the power law estimates are highly unstable. This explains why in the two six-week travel surveys (DEU Mobidrive 1999 and CHE Thurgau 2003) no consistent power law distributions could be found. For these time periods a power law distribution cannot be fitted reliably to day-long schedule data due to insufficient observations.

8. Discussion

This research worked with five data sets as discussed in section 3 and Table 1. The 450-day data was collected using the Moves smartphone application combined with manual verification and trip purpose enrichment. Unfortunately, only longitudinal data from one person could be obtained. Investigating the distribution of activity schedules in longitudinal individual data is challenging as such data is very scarce (and perhaps nonexistent for a large numbers of people). There are many challenges in collecting such data; most likely the main difficulty is to ensure participant commitment throughout a very long time period since the user has to consistently keep track of his or her activities. We have attempted to work with this issue by including two six-week household travel survey data sets in addition to the donated schedules from a single user. To the authors' knowledge, the two six-week data sets are the largest available travel survey data sets for a considerate number of people.

Another approach to deal with the scarceness of longitudinal data is to start from a large amount of mobility data (e.g. GPS traces with stop detection) and then infer the activity type by means of a classification approach [27, 39]. Another approach would be to use the predicted schedules from an activity-based model. There seem to be some issues or challenges with both approaches. A conceptual issue is that *manipulated* (having uncertainty) or completely *synthetic* data would be employed in the process of evidencing limitations in the observed law. Additionally, using the output from an activity-based model is characterized by challenges since model validation remains a challenge in this domain, and validating such model output is actually the intended application (circular reasoning). Therefore, this invalidates such an approach. The analysis of individual activity schedules remains however an interesting and challenging topic which will be addressed in detail in future research.

Previous research [39] found that there is a large effect from the choice of activity type classes on the activity type classification accuracy in the context of activity type inference in e.g. GPS data. This effect allowed to artificially increase the classification accuracy. Satisfyingly, no such effect is present here since moderate changes to the activity type encoding do not significantly affect the distribution's shape. In most modeling situations, the activity type classes may be chosen practically without limitations. The optimization of the activity type variable will not affect the universal law property of the data.

As mentioned in section 1, previous research [1] suggested two practical uses of the universal activity schedule distribution: (i) as an additional, necessary condition in a model's validation and (ii) as a possible way of extending mobility models which are based on universal mobility laws, but which typically lack an integration of the activity type. This paper attempts to stress-test the observed distribution by investigating the effects of aggregation in several dimensions in order to understand the extent of the universal distribution. By systematically testing the limits of the observed law, modelers, researchers and practitioners receive confidence in its extensibility. Relevant findings with respect to these applications include:

- From a practical point of view it is necessary to always consider a sufficiently large number of schedules in order to (visually) reproduce a power law distribution.

- In most modeling situations, the activity type classes may be chosen almost without limitations. The optimization of the activity type variable (as in [39]) will not affect the universal law property of the data.
- A temporally consistent distribution across different modeling years should be observable.
- Models producing individual, multi-day (long-term) schedules can be calibrated or validated using this distribution (multi-day training data are scarce and preferably entirely used for training the model, making them inadmissible for subsequent validation).
- Different subsets of the population should also exhibit the universal distribution.
- Though subsets of the data may exhibit the same universal rank-distribution, they could have different activity schedules at each rank, affecting the way they could be assigned to a synthetic population. If the model distinguishes subsets of the population (or if it distinguishes between week- or weekend days), the model accuracy can be improved by using subset-specific distributions.

9. Conclusion

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, the frequency of occurrence of day-long activity schedules obeys a remarkably simple, ubiquitous and scale-free distribution commonly referred to as Zipf's law. This paper discussed the role of aggregation within the phenomenon of Zipf's law in activity schedules. Aggregation was analyzed in three dimensions: activity type encoding, aggregation over time and the aggregation of individual data, in which the analysis moved from study area-wide aggregated data to subsets of the data, and finally individual (longitudinal) data.

The analysis in three dimensions concludes that, 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. The distribution appears stable throughout time, looking at different temporal scales. No considerable effect of subsetting the data were observed, provided that the the subset is sufficiently large. The two six-week travel surveys allowed to analyze individual schedules, yet this analysis 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, the 450-day 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. After roughly ten to fifteen weeks of collecting individual data, a power law exponent could be determined with relative confidence.

Previous research [1] suggested two practical applications for the observed universal law: (i) as an additional component in a model's validation, and (ii) as an extra dimension in universal law-based transportation models. This work provides information about the limitations or surprising consistencies modelers might expect in their implementations. The analysis results were discussed with respect to these applications.

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 will be done on simulated longitudinal data (originating from activity-based models) or on activity schedules inferred from GPS trajectories combined with an accurate activity type annotation. Furthermore, concrete test with respect to the suggested applications will be conducted.

Acknowledgments

The authors would like to thank prof. dr. Kay Axhausen for providing the DEU Mobidrive 1999 [30] and CHE Thurgau 2003 [31] data sets. They are also thankful to the U.S. Department of Transportation, Federal Highway Administration, and the Department for Transport for making the NHTS 2009 [28], respectively the GBR NTS 2009-2014 [29] data freely available. The authors thank the donor of the 450-day of individual trip data.

References

[1] W. Ectors, B. Kochan, D. Janssens, T. Bellemans, G. Wets, Exploratory analysis of Zipf's universal power law in activity schedules, Transportation`doi:10.1007/s11116-018-9864-9`.
URL http://link.springer.com/10.1007/s11116-018-9864-9

[2] S. Jiang, Y. Yang, S. Gupta, D. Veneziano, S. Athavale, M. C. González, The TimeGeo modeling framework for urban motility without travel surveys, Proceedings of the National Academy of Sciences 113 (37) (2016) E5370–E5378. `doi:10.1073/pnas.1524261113`.
URL http://www.pnas.org/lookup/doi/10.1073/pnas.1524261113

[3] L. Pappalardo, F. Simini, Data-driven generation of spatio-temporal routines in human mobility, Springer US, 2017. `arXiv:1607.05952`, `doi:10.1007/s10618-017-0548-4`.
URL https://doi.org/10.1007/s10618-017-0548-4

[4] C. Chen, J. Ma, Y. Susilo, Y. Liu, M. Wang, The promises of big data and small data for travel behavior (aka human mobility) analysis, Transportation Research Part C: Emerging Technologies 68 (2016) 285–299. `doi:10.1016/j.trc.2016.04.005`.
URL http://dx.doi.org/10.1016/j.trc.2016.04.005

[5] D. A. Freedman, The Ecological Fallacy (2002).
URL http://www.stat.berkeley.edu/{~}census/ecofall.txt

[6] G. K. Zipf, Human Behaviour and the Principle of Least Effort, Addison-Wesley, Reading, 1949.

[7] S. Ki Baek, S. Bernhardsson, P. Minnhagen, Zipf's law unzipped, New Journal of Physics 13 (4) (2011) 043004. arXiv:1104.1789, doi:10.1088/1367-2630/13/4/043004.
URL http://stacks.iop.org/1367-2630/13/i=4/a=043004

[8] M. Newman, Power Laws, Pareto Distributions and Zipf's Law, Contemporary physics 46 (5) (2005) 323–351. arXiv:0412004v3, doi:10.1080/00107510500052444.
URL http://www.tandfonline.com/doi/abs/10.1080/00107510500052444

[9] X. Gabaix, Zipf ' S Law for Cities : an Explanation, Quarterly Journal of Economics 114 (August) (1999) 739–767.

[10] W.-C. Chen, On the Weak Form of Zipf's Law, Journal of Applied Probability 17 (3) (1980) 611–622. doi:10.2307/3212955.

[11] B. Corominas-Murtra, R. V. Solé, Universality of Zipf's law, Physical Review E 82 (1) (2010) 9. arXiv:1001.2733, doi:01110210.1103/PhysRevE.82.011102.

[12] Y. M. Ioannides, H. G. Overman, Zipf's law for cities: An empirical examintion, Regional Science and Urban Economics 33 (2) (2003) 127–137. doi:10.1016/S0166-0462(02)00006-6.

[13] B. Jiang, T. Jia, Zipf's Law for All the Natural Cities in the United States: A Geospatial Perspective, International Journal of Geographical Information Science (2010) 10arXiv:1006.0814, doi:10.1080/13658816.2010.510801.
URL http://arxiv.org/abs/1006.0814

[14] W. Li, Random Texts Exhibit Zipf's-Law-Like Word Frequency Distribution, IEEE Transactions on Information Theory 38 (6) (1992) 1842–1845. doi:10.1109/18.165464.

[15] M. Marsili, Y.-c. Zhang, Interacting Individuals Leading to Zipf ' s Law, Physical Review Letters 80 (12) (1998) 2741–2744. arXiv:9801289v1, doi:10.1103/PhysRevLett.80.2741.

[16] W. J. Reed, The Pareto, Zipf and other power laws, Economics Letters 74 (1) (2001) 15–19. doi:10.1016/S0165-1765(01)00524-9.

[17] K. T. Soo, Zipf's Law for cities: A cross-country investigation, Regional Science and Urban Economics 35 (3) (2005) 239–263. doi:10.1016/j.regsciurbeco.2004.04.004.

[18] V. Nitsch, Zipf zipped, Journal of Urban Economics 57 (1) (2005) 86–100. doi:10.1016/j.jue.2004.09.002.

[19] M. C. González, C. A. Hidalgo, A.-L. Barabási, Understanding individual human mobility patterns, Nature 453 (7196) (2008) 779–782. arXiv:0806.1256v1, doi:10.1038/nature07850.
URL http://www.ncbi.nlm.nih.gov/pubmed/18528393

[20] D. Brockmann, L. Hufnagel, T. Geisel, The scaling laws of human travel., Nature 439 (7075) (2006) 462–5. arXiv:0605511, doi:10.1038/nature04292.
URL http://dx.doi.org/10.1038/nature04292

[21] Z. Zheng, S. Rasouli, H. Timmermans, Two-regime Pattern in Human Mobility: Evidence from GPS Taxi Trajectory Data, Geographical Analysis 48 (2) (2016) 157–175. doi:10.1111/gean.12087.

[22] X. H. Yang, G. Chen, S. Y. Chen, W. L. Wang, L. Wang, Study on some bus transport networks in China with considering spatial characteristics, Transportation Research Part A: Policy and Practice 69 (2014) 1–10. doi:10.1016/j.tra.2014.08.004.
URL http://dx.doi.org/10.1016/j.tra.2014.08.004

[23] S. Paleari, R. Redondi, P. Malighetti, 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 46 (2) (2010) 198–210. doi:10.1016/j.tre.2009.08.003.
URL http://dx.doi.org/10.1016/j.tre.2009.08.003

[24] R. Guidotti, R. Trasarti, M. Nanni, TOSCA : TwO-Steps Clustering Algorithm for Personal Locations Detection, Proceedings of the 23nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (January). doi:10.1145/2820783.2820818.

[25] C. Song, T. Koren, P. Wang, A.-L. Barabasi, Modelling the scaling properties of human mobility, Nature Physics 6 (10) (2010) 1–6. arXiv:1010.0436, doi:10.1038/NPHYS1760.
URL http://dx.doi.org/10.1038/nphys1760

[26] S. Rasouli, H. Timmermans, Activity-based models of travel demand: promises, progress and prospects, International Journal of Urban Sciences 18 (1) (2014) 31–60. doi:10.1080/12265934.2013.835118.
URL http://www.tandfonline.com/doi/abs/10.1080/12265934.2013.835118

[27] S. Rinzivillo, L. Gabrielli, M. Nanni, L. Pappalardo, D. Pedreschi, F. Giannotti, The purpose of motion: Learning activities from Individual Mobility Networks, DSAA 2014 - Proceedings of the 2014 IEEE International Conference on Data Science and Advanced Analytics (2014) 312–318doi:10.1109/DSAA.2014.7058090.

[28] U.S. Department of Transportation, Federal Highway Administration, 2009 National Household Travel Survey (2009).
URL http://nhts.ornl.gov

[29] Department for Transport, National Travel Survey, 2002-2014 [computer file]. 9th Edition. (2015). doi:10.5255/UKDA-SN-5340-5.

[30] V. Chalasani, K. W. Axhausen, 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

[31] M. Loechl, Stability of Travel Behaviour: Thurgau 2003 (2005).
URL http://archiv.ivt.ethz.ch/vpl/publications/tsms/tsms16.pdf

[32] ProtoGeo Oy, Moves - Activity Diary for iPhone and Android (2016).
URL https://moves-app.com/

[33] D. Janssens, K. Declercq, G. Wets, 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

[34] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-Law Distributions in Empirical Data, SIAM Review 51 (4) (2009) 661. arXiv:0706.1062v2, doi:10.1137/070710111.
URL http://link.aip.org/link/SIREAD/v51/i4/p661/s1{&}Agg=doi

[35] C. M. Urzúa, Testing for Zipf's law: A common pitfall, Economics Letters 112 (3) (2011) 254–255. doi:10.1016/j.econlet.2011.05.049.
URL http://dx.doi.org/10.1016/j.econlet.2011.05.049

[36] C. S. Gillespie, Fitting Heavy Tailed Distributions: The poweRlaw Package, Journal of Statistical Software 64 (2) (2015) 1–16.
URL http://www.jstatsoft.org/v64/i02

[37] L. A. Adamic, B. A. Huberman, Zipf's law and the Internet, Glottometrics 3 (2002) 143–150.

[38] R. Hanel, B. Corominas-Murtra, B. Liu, S. Thurner, Fitting power-laws in empirical data with estimators that work for all exponents, PLoS ONE 12 (2) (2017) 1–15. arXiv:1609.05357, doi:10.1371/journal.pone.0170920.

[39] W. Ectors, S. Reumers, W. D. Lee, K. Choi, B. Kochan, D. Janssens, T. Bellemans, G. Wets, Developing an optimised activity type annotation method based on classification accuracy and entropy indices, Transportmetrica A: Transport Science 13 (8) (2017) 742–766. doi:10.1080/23249935.2017.1331275.

[40] H. Contrino, N. McGuckin, Using NHTS to Estimate Activity Patterns, in: Presented at the 11th TRB National Transportation Planning Applications Conference, 2007.
URL http://www.trbappcon.org/2007conf/program.html

[41] B. M. Paul, P. Vovsha, J. E. Hicks, G. Vyas, V. Livshits, K. Jeon, Generation of Mandatory Activities and Formation of Mandatory Tours: Application to the Activity-Based Model for Phoenix, AZ, in: TRB 94th Annual Meeting Compendium of Papers, no. x 250, 2015, pp. 1–14.

Appendix

Table 6: Five activity type aggregation levels for the USA NHTS 2009 dataset

USA NHTS 2009 activity description	**Weighted freq.**	**Level 0**	**Level 1**	**Level 2a**	**Level 2b**	**Level 3**
Appropriate skip	*2016940865*	*-1*				
Refused	*56316311*	*-7*				
Don't know	*175030832*	*-8*				
Not ascertained	*27723241*	*-9*				
Home	1,34819E+11	1	1	1	Home	Mandatory
Work	215609	10	10	3	Work	Mandatory
Go to work	31062036426	11	11	3	Work	Mandatory
Return to work	5732878676	12	11	3	Work	Mandatory
Attend business meeting/trip	1066903014	13	12	3	Work	Mandatory
Other work related	7901898136	14	10	3	Work	Mandatory
School/religious activity	1132537921	20	20	11	School/Religious	Mandatory
Go to school as student	11830627020	21	21	6	School/Religious	Mandatory
Go to religious activity	6980876310	22	22	11	School/Religious	Discretionary
Go to library: school related	453575041	23	21	6	School/Religious	Discretionary
OS - Day care	828988699	24	21	8	School/Religious	Maintenance
Medical/dental services	6302927234	30	30	10	Medical/dental services	Maintenance
Shopping/errands	7097239018	40	40	4	Shopping/Errands	Maintenance
Buy goods: groceries/clothing/hardware store	44001480325	41	41	4	Shopping/Errands	Maintenance
Buy services: video rentals/dry cleaner/post office/car service/bank	11224064829	42	42	10	Shopping/Errands	Maintenance
Buy gas	6603091100	43	41	4	Shopping/Errands	Maintenance
Social/recreational	3779680002	50	50	9	Social/Recreational	Discretionary
Go to gym/exercise/play sports	13430438123	51	51	9	Social/Recreational	Discretionary
Rest or relaxation/vacation	3276538854	52	52	9	Social/Recreational	Discretionary
Visit friends/relatives	17562038581	53	53	5	Social/Recreational	Discretionary
Go out/hang out: entertainment/theater/sports event/go to bar	6838625710	54	52	9	Social/Recreational	Discretionary
Visit public place: historical site/museum/park/library	1852249711	55	52	9	Social/Recreational	Discretionary
Family personal business/obligations	4484117764	60	50	11	Family personal business/obligations	Discretionary
Use professional services: attorney/accountant	1109208170	61	42	10	Family personal business/obligations	Maintenance
Attend funeral/wedding	667934182	62	53	11	Family personal business/obligations	Discretionary
Use personal services: grooming/haircut/nails	1467981696	63	42	10	Family personal business/obligations	Discretionary
Pet care: walk the dog/vet visits	2939462521	64	52	7	Family personal business/obligations	Maintenance
Attend meeting: PTA/home owners association/local government	1609806545	65	53	9	Family personal business/obligations	Maintenance
Transport someone	309113327	70	70	8	Transport Someone	Mandatory
Pick up someone	11035542385	71	70	8	Transport Someone	Mandatory
Take and wait	1186149745	72	70	8	Transport Someone	Mandatory
Drop someone off	11961497342	73	70	8	Transport Someone	Mandatory
Meals	791727089	80	80	9	Meals	Discretionary
Social event	2485502724	81	52	9	Meals	Discretionary
Get/eat meal	20351291660	82	80	9	Meals	Maintenance
Coffee/ice cream/snacks	2976589359	83	80	9	Meals	Discretionary
Other reason	2592958077	97	90	11	Other	Discretionary
# of distinct valid activity type classes:		**37**	**18**	**10**	**10**	**3**