Zipf's law in activity schedules

Wim Ectors, Bruno Kochan, Davy Janssens, Tom Bellemans & Geert Wets

hEART 2016

universiteit hasselt | imob
TRANSPORTATION RESEARCH INSTITUTE

Table of contents

- Introduction
  - Zipf's law?
- Methodology
- Aims
- Results
- Discussion and conclusion

Introduction

- Traffic demand models
  - Extremely complex, multidimensional processes
  - AB-models
    - (non-)mandatory activity schedulers
    - TOD models
    - Re-schedulers
    - Interaction with others
    - …

- Yet, day-long schedules obey a simple distribution:

Zipf's law *(Zipf, 1949)*

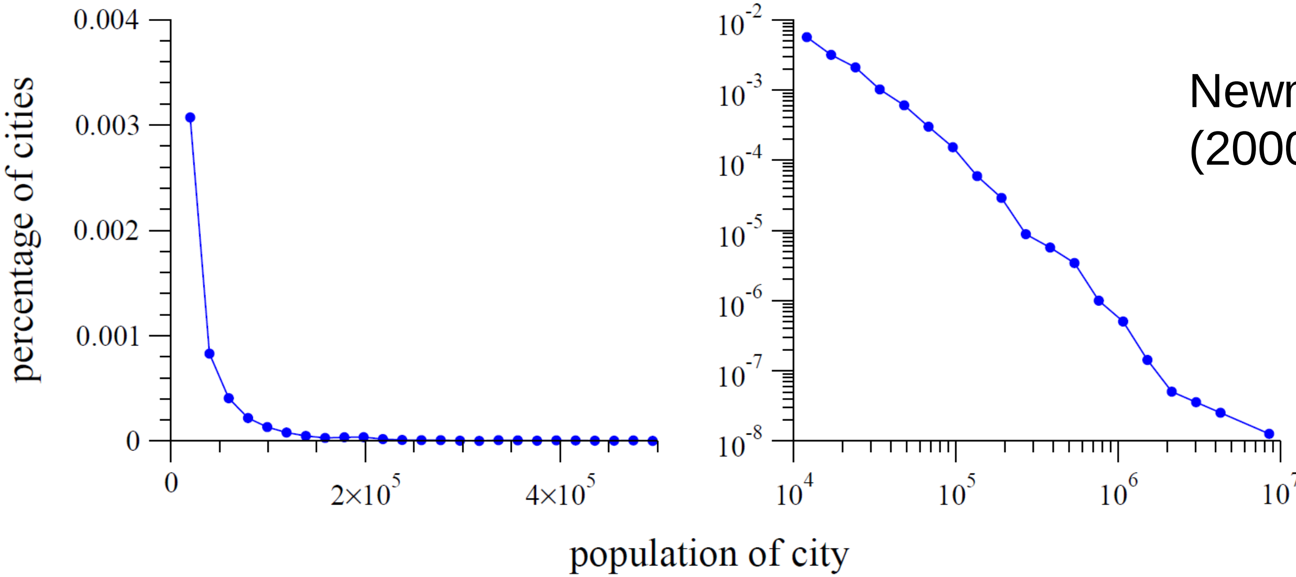
Introduction – Zipf's law

- Power-law distribution
  - Zipf: α=1

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

- Rank-size interpretation

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



Newman, 2005
(2000 US census)

universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE

Introduction – Zipf's law

- Observed in many different fields:
  - Word frequency
  - City size
  - Earthquake magnitude
  - Annual company income
  - Solar flares
  - Number of citations of papers
  - *Fujiwara, 2004; Furusawa and Kaneko, 2003; Maillart et al., 2008; Newman, 2005; Okuyama et al., 1999*
- General applicability still sometimes contested:
  - City size (*Soo, 2005)*

universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE

Introduction – Zipf's law

- In domain of transportation
  - Power-law like distributions in
    - Displacement distance
    - Gyration radius
    - Location visiting frequency
    - Location visiting duration
    - Bus transport networks

Displacement distance; Gyration radius; Location visiting frequency	*González et al. 2008*
Location visiting duration	*Brockmann et al. 2006*
Bus transport networks	*Yang et al. 2014*

  - This study:
    - Day-long activity pattern (=schedule) frequency

Methodology

- Zipf's law in this study:
  - Schedule frequency
    1. Compose schedules from HTS data
    2. One-way frequency table
    3. Sort & add rank column
    4. Plot rank vs schedule frequency

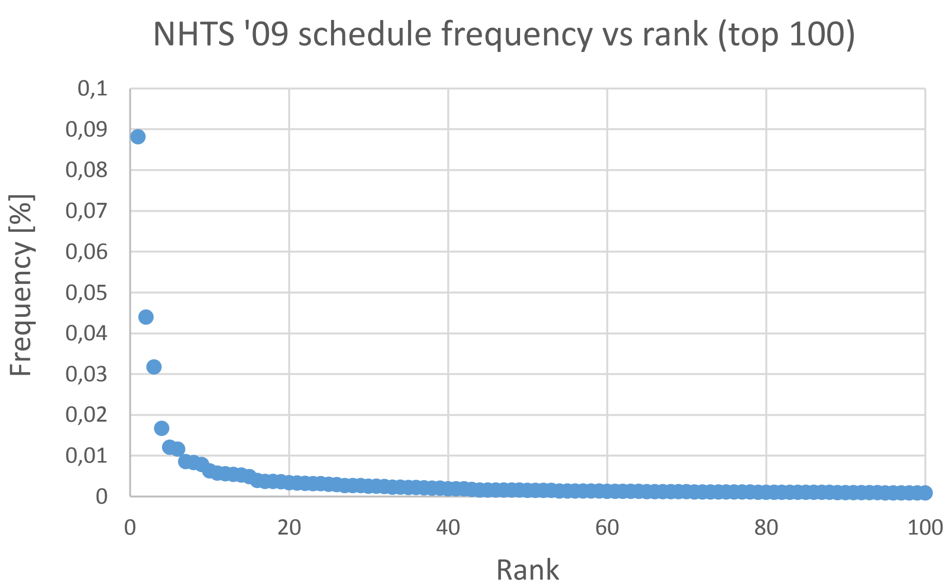
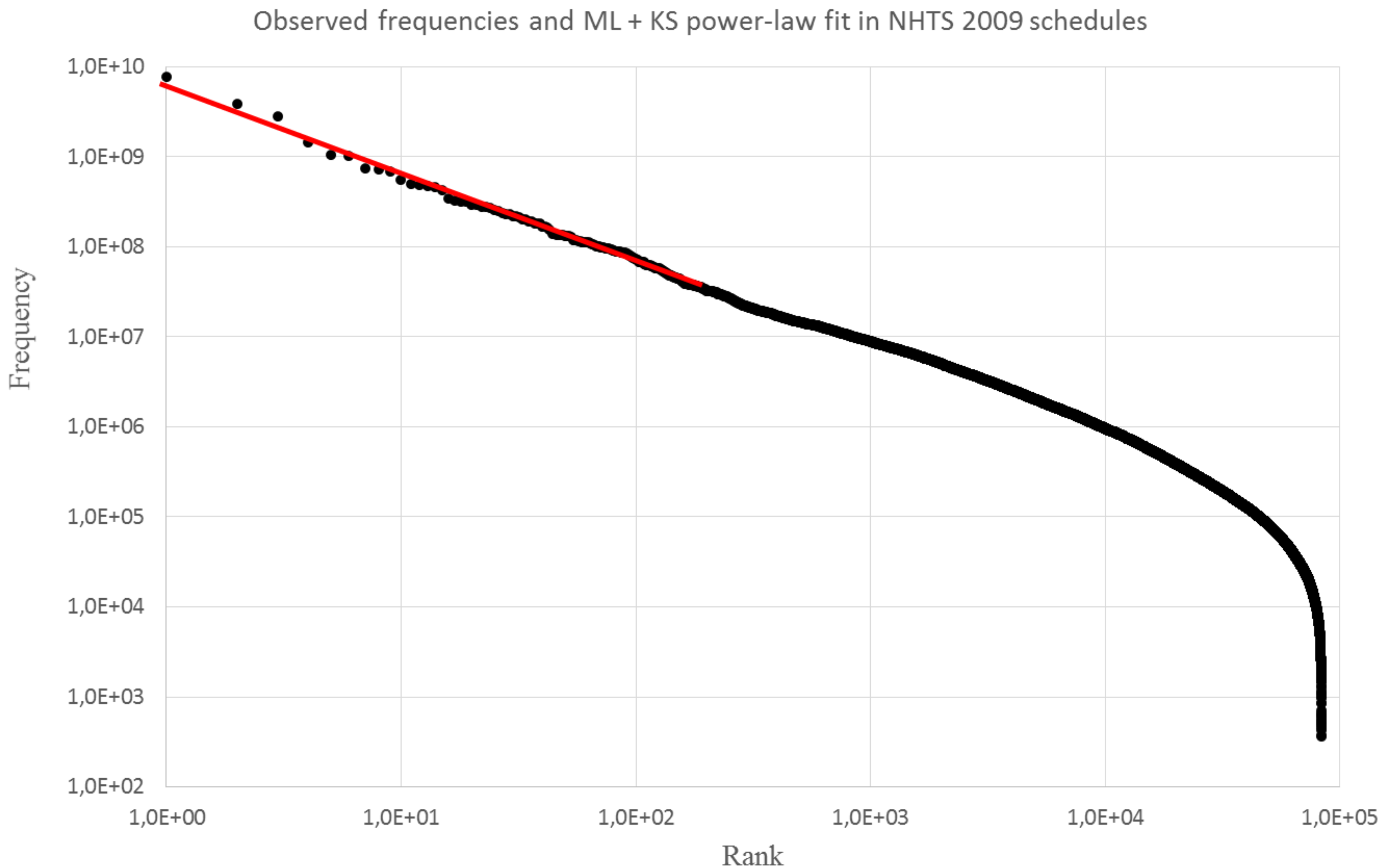


Table 1: Top-10 schedules NHTS '09

Schedule	Freq. [%]	Rank
1-11-1	8,82	1
1-21-1	4,39	2
1-41-1	3,17	3
1-53-1	1,67	4
1-51-1	1,21	5
1-22-1	1,16	6
1-30-1	0,85	7
1-82-1	0,83	8
1-41-41-1	0,78	9
1-54-1	0,63	10

Methodology



Methodology

- ML fit
- Kolmogorov-Smirnov cutoff criterion
  - Power-law for values $> x_{min}$
- *R* package: poweRlaw (*Gillespie 2015)*

Aims

1. To provide evidence for a rank-frequency Zipf's law in activity schedule frequencies
2. To test the law's dependency on the aggregation level of activity types
3. To test the law's validity for each day of the week

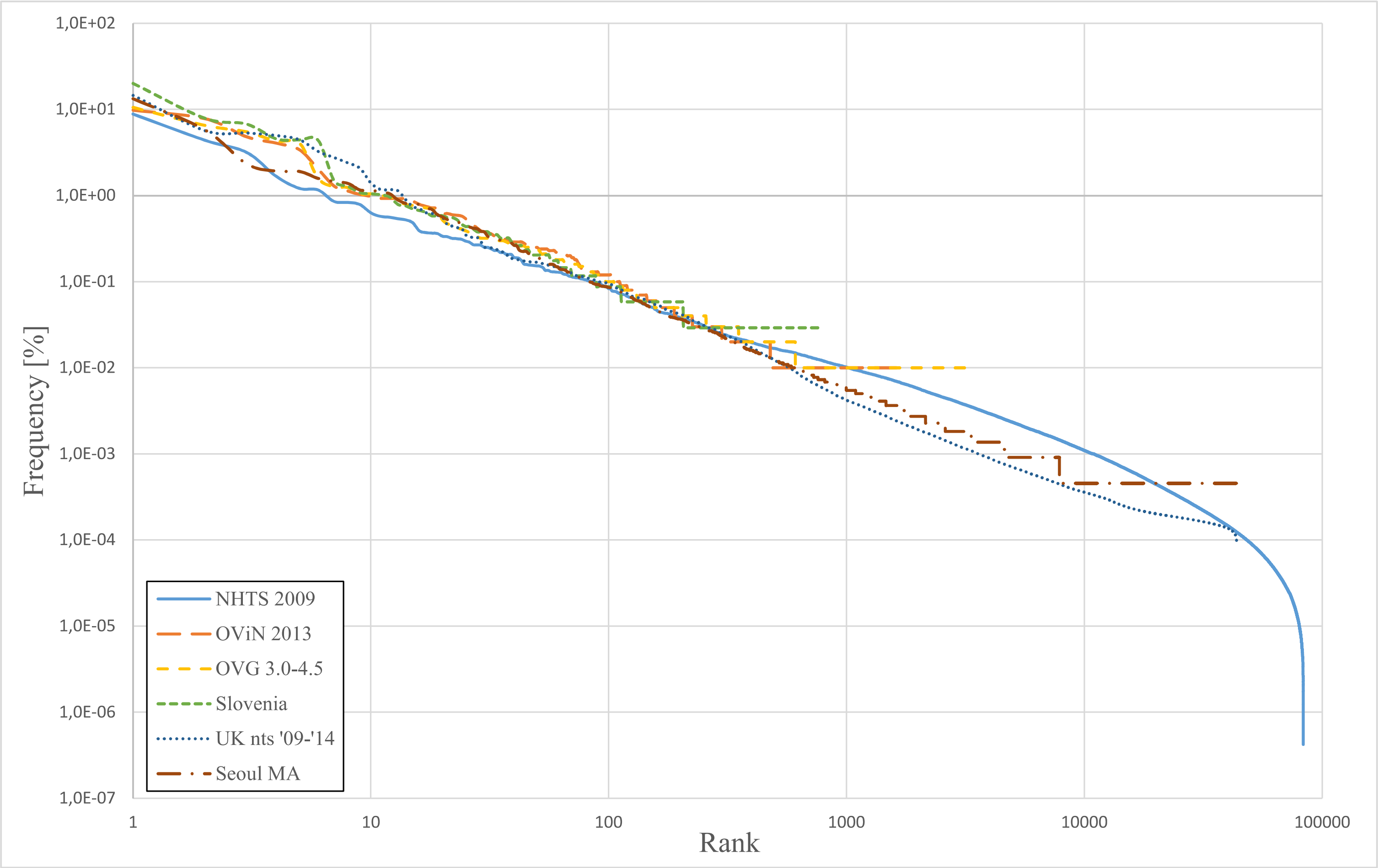
Results

1. Providing evidence for a rank-frequency Zipf's law in activity schedule frequencies

Dataset	poweRlaw estimations (ML + Kolmogorov-Smirnov)				Bootstrapping uncertainty evaluation (5000 (*2000) simulations)		
	α	Xmin	Cum. Pct discarded	n_tail	AM	SD	P-value
NHTS 2009*	2.003	36809977	55%	181	2.006	0.070	0.255
OViN 2013	1.885	1325	20%	721	1.877	0.058	0.005
OViN 2013 (OVG encoding)	1.830	2378	22%	406	1.862	0.052	0.336
OVG 3.0-4.5	1.947	2	23%	529	1.953	0.051	0.149
Ljubljana (Slovenia)	1.995	9	31%	43	2.028	0.208	0.449
UK nts 2009-2014*	1.862	4,071	9%	5042	1.869	0.117	0?
Seoul MA 2010*	1.859	26	32%	520	1.869	0.048	0.156

Results

1. Providing evidence for a rank-frequency Zipf's law in activity schedule frequencies



universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE

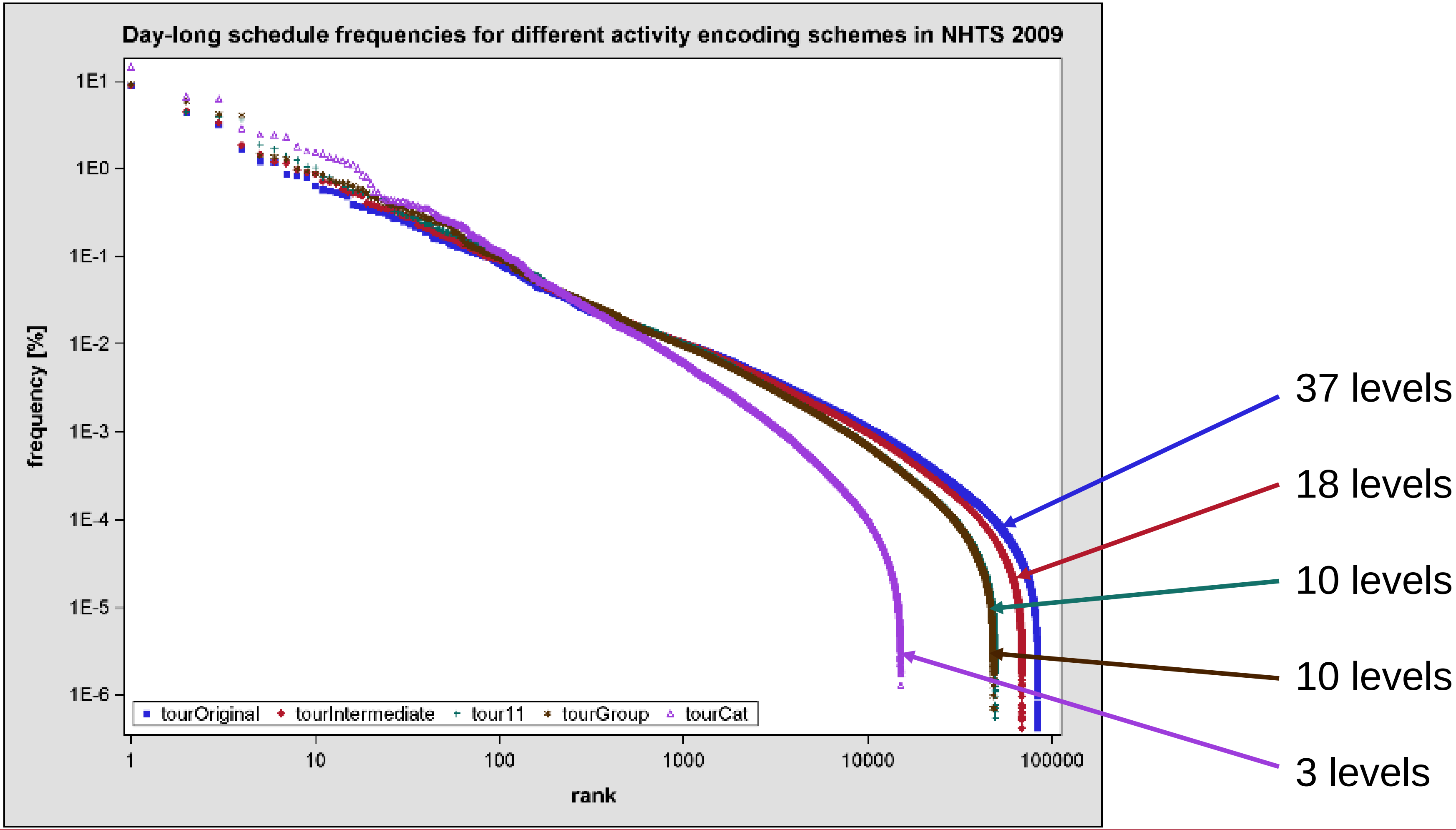
Results

2. Testing the law's dependency on the aggregation level of activity types

Dataset	# levels [encoding]	poweRlaw estimations (ML + Kolmogorov-Smirnov)				Bootstrapping uncertainty evaluation (5000 (*2000) simulations)		
		α	Xmin	Cum. Pct discarded	n_tail	AM	SD	P-value
NHTS 2009*	37 ["Original"]	2.003	36809977	55%	181	2.006	0.070	0.255
	18 ["Intermediate"]	1.967	36837451	50%	193	1.972	0.065	0.166
	10 ["OVG"]	1.934	46135634	43%	171	1.939	0.065	0.998
	10 ["Group"]	1.892	60781076	45%	123	1.899	0.071	0.741
	3 ["Cat"]	1.890	109512566	28%	90	1.891	0.084	0.835

Results

2. Testing the law's dependency on the aggregation level of activity types



universiteit hasselt | imob
TRANSPORTATION RESEARCH INSTITUTE

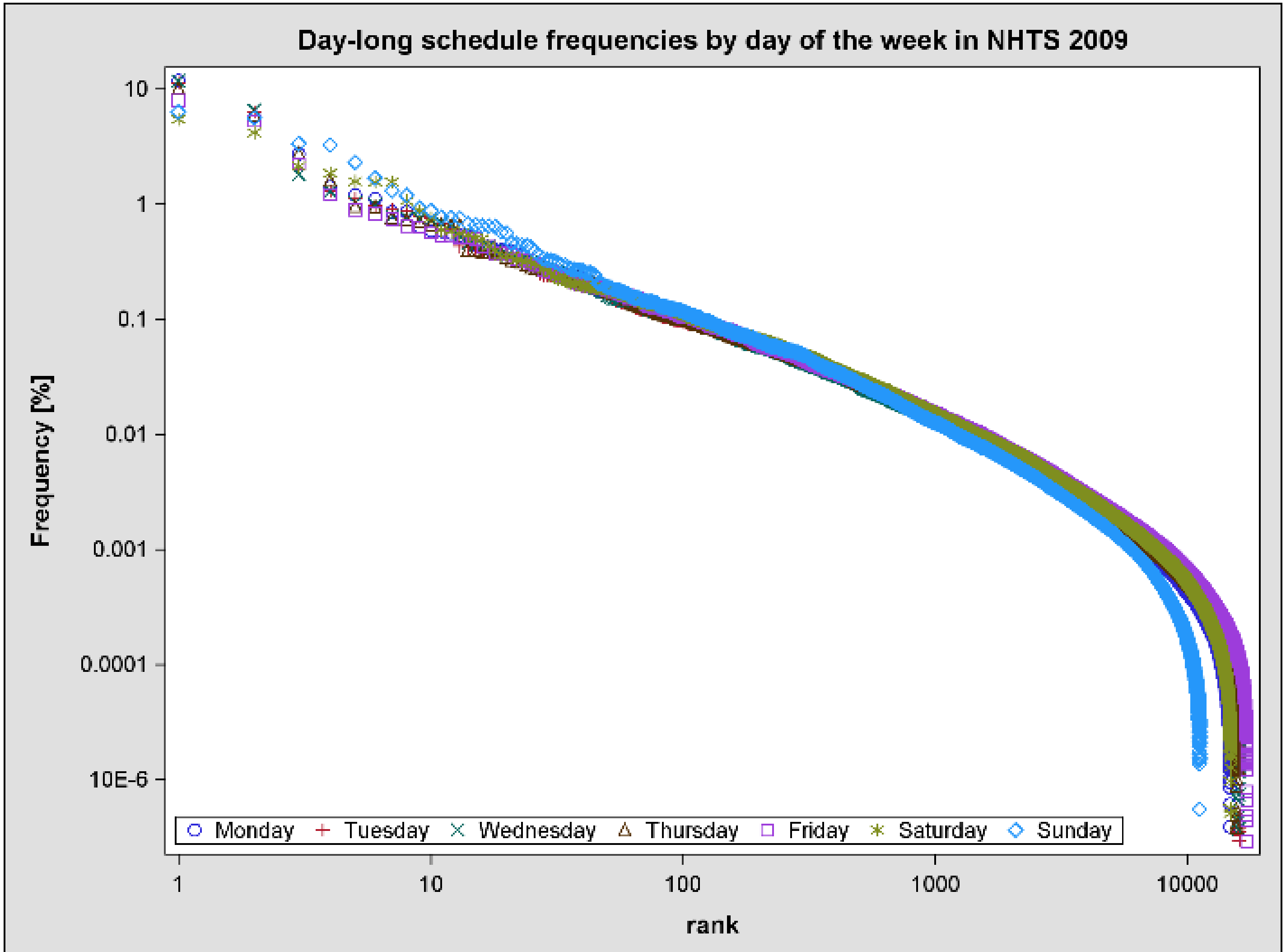
Results

3. Testing the law's validity for each day of the week

Dataset	Subset	poweRlaw estimations (ML + Kolmogorov-Smirnov)				Bootstrapping uncertainty evaluation (5000 (*2000) simulations)		
		α	Xmin	Cum. Pct discarded	n_tail	AM	SD	P-value
NHTS 2009*	All data	2.003	36809977	55%	181	2.006	0.070	0.255
NHTS 2009	Monday	2.290	46616705	67%	22	2.270	0.359	0.831
	Tuesday	2.161	35581917	67%	26	2.182	0.236	0.820
	Wednesday	2.152	45646004	68%	20	2.172	0.267	0.679
	Thursday	2.088	48120314	71%	17	2.140	0.282	0.221
	Friday	2.279	34509610	72%	28	2.284	0.250	0.901
	Saturday	2.182	61045896	76%	15	2.176	0.288	0.134
	Sunday	2.091	52160661	66%	21	2.060	0.200	0.982

Results

3. Testing the law's validity for each day of the week



universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE

Discussion and conclusion

- A power-law distribution (Zipf) appears to be
  - valid for most schedules and to be study area independent
  - largely independent of activity type encoding aggregation
  - valid on different days of the week
- Further research on
  - Evidence general applicability (≠ study areas)
  - Explore the extent of validity (aggregated → disaggregated data)
  - Investigate origin

References

- Brockmann, D, L Hufnagel, and T Geisel. 2006. “The Scaling Laws of Human Travel.” *Nature* 439 (7075): 462–465. doi:10.1038/nature04292.
- Fujiwara, Yoshi. 2004. “Zipf Law in Firms Bankruptcy.” *Physica A: Statistical Mechanics and Its Applications* 337 (1-2): 219–230. doi:10.1016/j.physa.2004.01.037.
- Furusawa, Chikara, and Kunihiko Kaneko. 2003. “Zipf’s Law in Gene Expression.” *Physical Review Letters* 90 (8): 1–11. doi:10.1103/PhysRevLett.90.088102.
- Gillespie, Colin S. 2015. “Fitting Heavy Tailed Distributions: The poweRlaw Package.” *Journal of Statistical Software* 64 (2): 1–16. http://www.jstatsoft.org/v64/i02.
- González, Marta Cecilia, César Augusto Hidalgo, and Albert-László Barabási. 2008. “Understanding Individual Human Mobility Patterns.” *Nature* 453 (7196): 779–782. doi:10.1038/nature07850.
- Maillart, T, D Sornette, S Spaeth, and G. von Krogh. 2008. “Empirical Tests of Zipf’s Law Mechanism in Open Source Linux Distribution.” *Physical Review Letters* 101 (21): 218701. doi:10.1103/PhysRevLett.101.218701.
- Newman, Mej. 2005. “Power Laws, Pareto Distributions and Zipf’s Law.” *Contemporary Physics* 46: 323–351. doi:10.1080/00107510500052444.
- Okuyama, K, M. Takayasu, and H Takayasu. 1999. “Zipf’s Law in Income Distribution of Companies.” *Physica A: Statistical Mechanics and Its Applications* 269 (1): 125–131. doi:10.1016/S0378-4371(99)00086-2.
- Soo, Kwok Tong. 2005. “Zipf’s Law for Cities: A Cross-Country Investigation.” *Regional Science and Urban Economics* 35 (3): 239–263. doi:10.1016/j.regsciurbeco.2004.04.004.
- Yang, Xu Hua, Guang Chen, Sheng Yong Chen, Wan Liang Wang, and Lei Wang. 2014. “Study on Some Bus Transport Networks in China with Considering Spatial Characteristics.” *Transportation Research Part A: Policy and Practice* 69. Elsevier Ltd: 1–10. doi:10.1016/j.tra.2014.08.004.
- Zipf, George Kingsley. 1949. *Human Behaviour and the Principle of Least Effort*. Reading: Addison-Wesley.

universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE

Thank you

Wim Ectors*
Bruno Kochan
Davy Janssens
Tom Bellemans
Geert Wets

**wim.ectors@uhasselt.be*

universiteit hasselt | imob TRANSPORTATION RESEARCH INSTITUTE