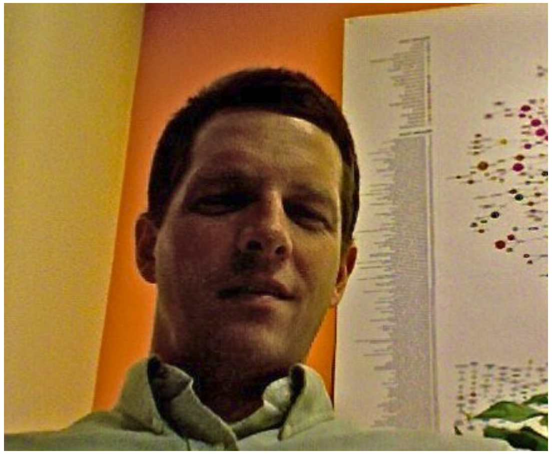
Urban Mobility Scaling: Lessons from 'Little Data':

Developing a Science of Cities



Galen J. Wilkerson
Technische Universität - Berlin
gjwilkerson@gmail.com

Motivation

Science of cities:

- Understanding major factors, statistical laws of how people choose to live together.
- Especially, scaling and salient parameters and their relationships:
  - Scaling – scale-invariant system phenomenon
  - E.g. Population, area, density patterns, energy availability, transportation mode share (Verkehrsmittelanteil), cultural parameters
- Sustainability

Motivation

In order to develop science of cities, Big (Mobility) Data is very interesting and useful!

Here we

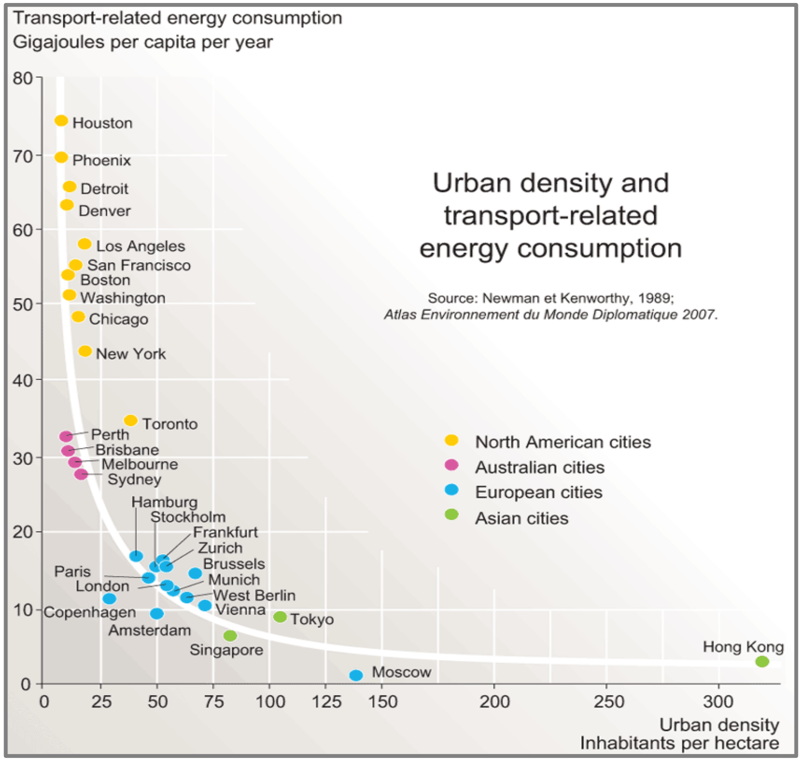
- elucidate challenges of these new data sources
- give some compelling preliminary results from conventional data.

Overview

- Background:
  - Urban Scaling, Complex Systems, Big (Mobility) Data
- Our preliminary findings
  - Categories matter!
    - Human-powered modes are different!
  - Urban Scale (~10km) is revealing
  - Mobility scaling confirms previous results
  - Distance vs. intervening opportunity mechanisms, some results
  - Time matters – aggregation is dangerous
- Future work

Background

- Large-scale urban phenomena
  - Newman and Kenworthy (1980) (urban planning, environmental science)
    - Per capita energy vs. density



[Newman, Kenworthy]

  - Michael Batty et al. (many publications since 1970s) (urban planning/geography)
    - spatial distribution of populations (fractal) and related topics (spatial entropy), also scaling.

[M. Batty]

Roads colored by connectivity

Background

- Urban Scaling
  - Luis Bettencourt, Dirk Helbing, Geoffrey West (2007) (statistical physics)
    - “allometric scaling” of urban phenomena with city population

$$Y(t) = Y_0 N(t)^\beta$$

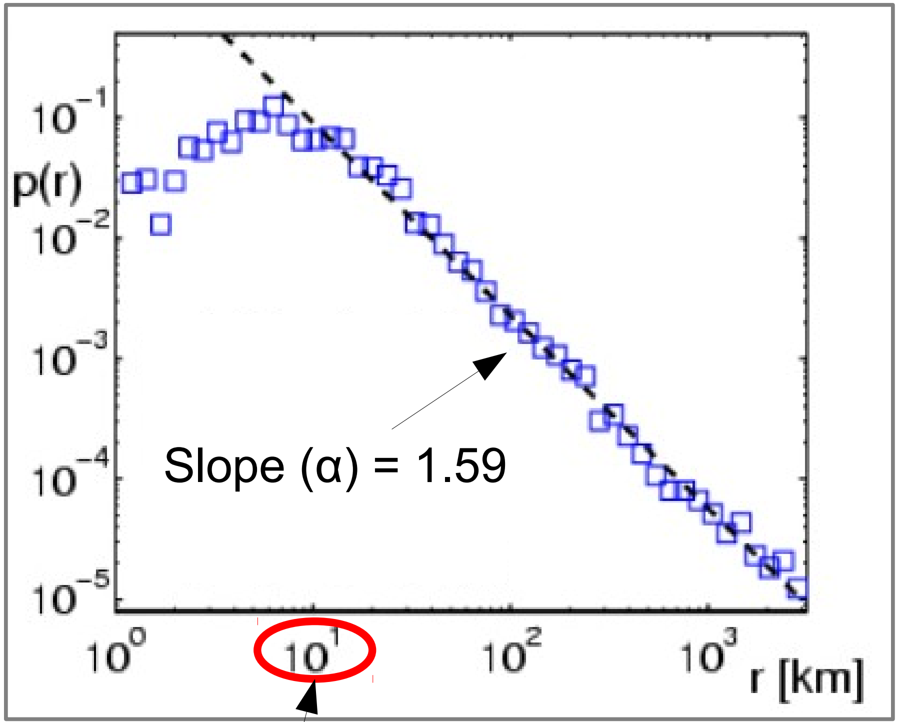
      - Similar to non-linear scaling of metabolism with mass of animal (mouse vs. whale)
      - Compare patents to gasoline sales
  - Sante Fe Institute Sustainability, ETHZ (Helbing et al.) FuturICT

Y	β
New patents	1.27
Inventors	1.25
Private R&D employment	1.34
”Supercreative” employment	1.15
R&D establishments	1.19
R&D employment	1.26
Total wages	1.12
Total bank deposits	1.08
GDP	1.15
GDP	1.26
GDP	1.13
Total electrical consumption	1.07
New AIDS cases	1.23
Serious crimes	1.16
Total housing	1.00
Total employment	1.01
Household electrical consumption	1.00
Household electrical consumption	1.05
Household water consumption	1.01
Gasoline stations	0.77
Gasoline sales	0.79
Length of electrical cables	0.87
Road surface	0.83

[PNAS 2007]

Background

- Mobility Scaling
  - Dirk Brockmann et al. (2006)
    (statistical physics – *now at HU Berlin!*)
  - pioneer of 'Big' mobility check-in data for scaling
    (wheresgeorge.com)
    - Movement data as 'side effect' of experiment
    - Low-dimensional
  - Extension of animal-foraging literature
    - Monkeys, fish movements
  - Dollar-bill movement to estimate human mobility, compared to diffusion/random walk (displacement over time)
  - Heavy-tailed long-distance displacements! (i.e. long-distance trips occur more than we might expect, movement of money is *super-diffusive*): scaling exponent α = 1.59

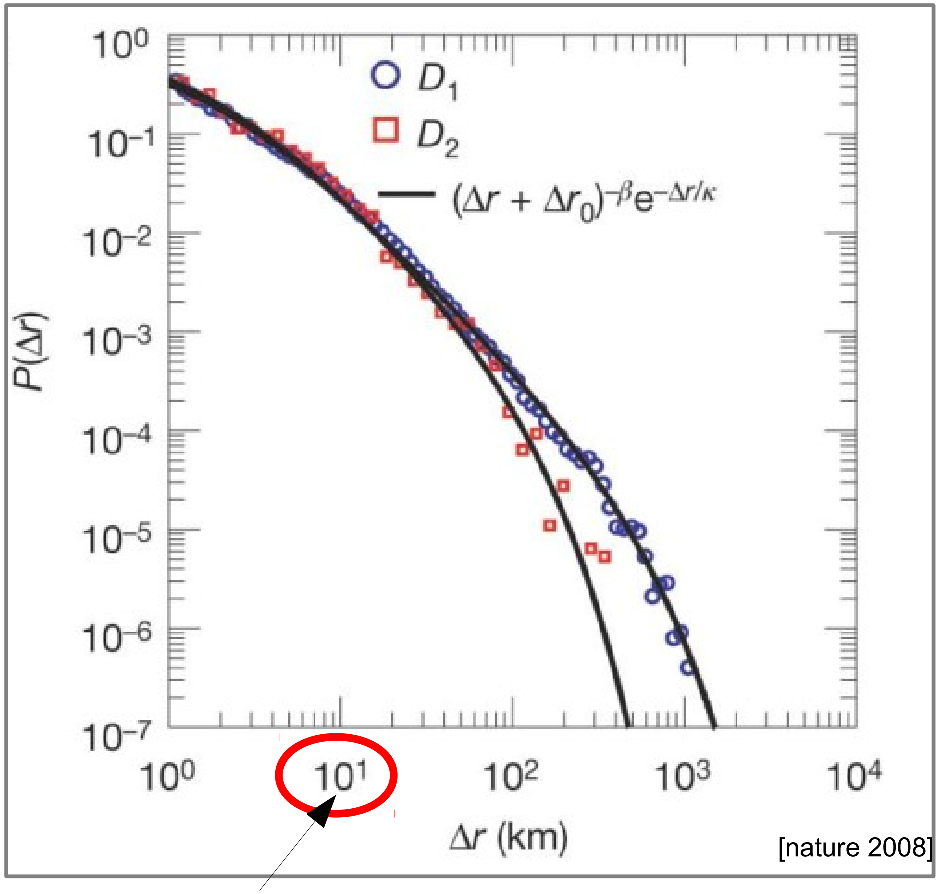


[nature 2006]

urban scale

Background

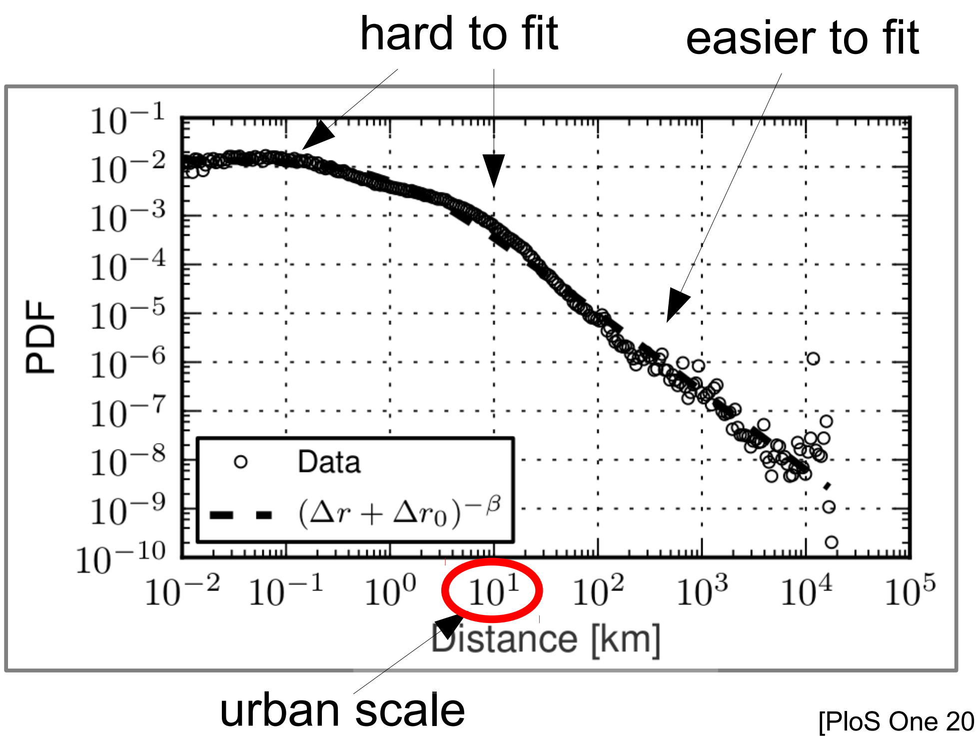
- Mobility Scaling
  - Gonzalez, Hidalgo, Barabasi (2008) (statistical physics)
  - 'Big' mobility data (phone calls: D1 = 16,264,308 displacements, constant locations (cell towers) D2 = 10,407 displacements)
  - More complex function due to fitting truncation
  - Perhaps challenging to characterize urban-scale mobility due to
    - sampling method (for D1, only during phone calls)
    - spatial resolution: Cell-phone tower resolution (3km)
  - Found scaling exponent of 1.75 (here called β)



urban scale

Background

- Mobility Scaling
  - Noulas et al. (2012)
    - Foursquare across planet
    - For long trips, measured scaling exponent of 1.50
    - Claims difficult to characterize mobility at urban scale from big data and distance mechanism alone
    - Intervening opportunity mechanisms seem to determine urban mobility patterns (rather than purely distance-based mechanisms)
      - e.g. we go to ~nearest grocery store



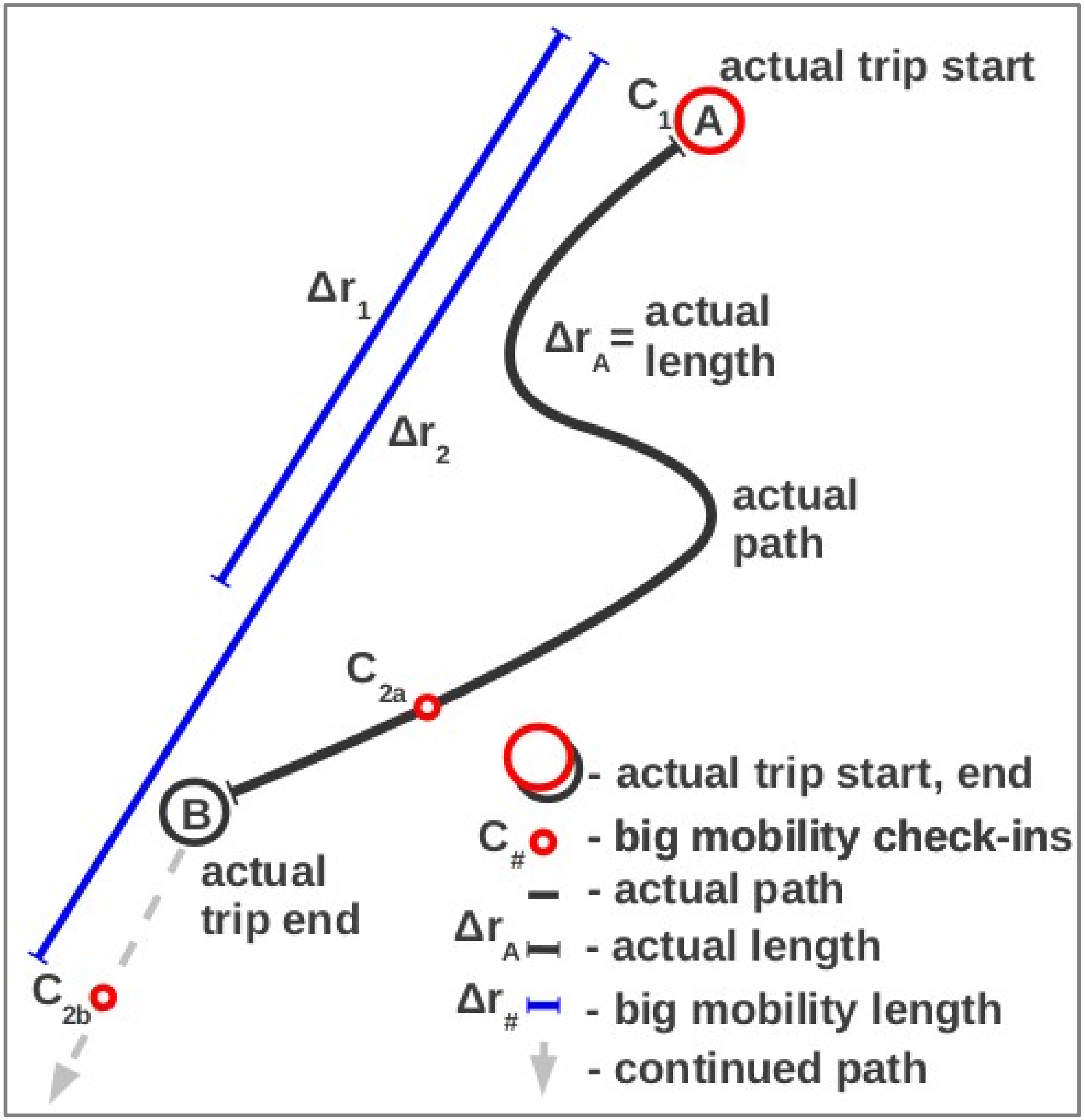
[PloS One 2012]

Background

- Mobility Scaling

  'check-ins' problem (especially at smaller urban scale!)

  - Assumes linear displacements
  - Assumes check-ins occur at trip start and end
  - Sampling rate variability (e.g. walking vs. driving)



[G.Wilkerson 2014]

Background

- Characteristics of 'Big' mobility data
  - Big data 'check-ins': calls, tweets, posts, where's george submissions
    - No survey design - side-effect of other activity
    - Low-dimensional (little known about trips or displacements)
    - Sometimes low-density (you need a lot of it)
      - Aggregated in space and time → *mean-field approximations:* ”If we get enough data, things average out” (do they?)
    - → **'Big' is not only defining feature**

  - Allows large-scale system perspective- Exciting/enticing/revolutionary (be careful!)
    - A “problem looking for a set of tools” → ~ statistical physics
      - random process underlying phenomenon. (e.g. rare events outliers of normal dist. or tail of 'heavy-tail' dist.?)
    - Humans are not particles or molecules
      - Integrate statistical physics perspective with urban & transportation planning German DLR, US RITA (DOT), European Union, IPCC, Think Tanks, etc.

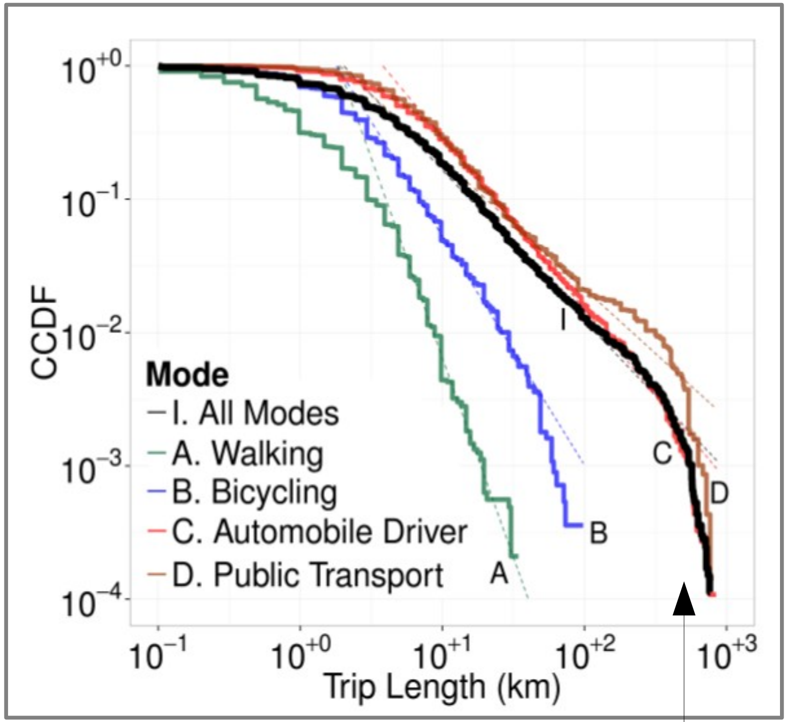
Our preliminary findings

- Our Data
  - Mobilität in Deutschland 2008 (MID 2008)
  - Conventional ('little' – not really) transportation survey
    - 25,922 households, 60,713 individuals, **193,290 trips and 36,182 travel** (w/overnight stay) events
    - Each trip has ~140 columns (dimensions) of information (weather, purpose, transportation mode, gender, age, etc.)
      - Allows investigation of many relationships
    - *Intentional Survey* designed to be large, balanced, and have resolution down to 100m
    - Our data describes actual trip lengths,
      - not linear displacements,
      - not 'check-ins' - all trips on survey day

- Methods
  - We just look at relative scaling of trip lengths, without fitting (for now).

Our preliminary findings

- Categories matter!
  - Trips are high-dimensional (mode, purpose, gender, weather, etc.)
    - **Trip lengths should be sensitive to mode!**
    - Scaling exponents → different universality classes for modes (e.g. walking, vs. driving)



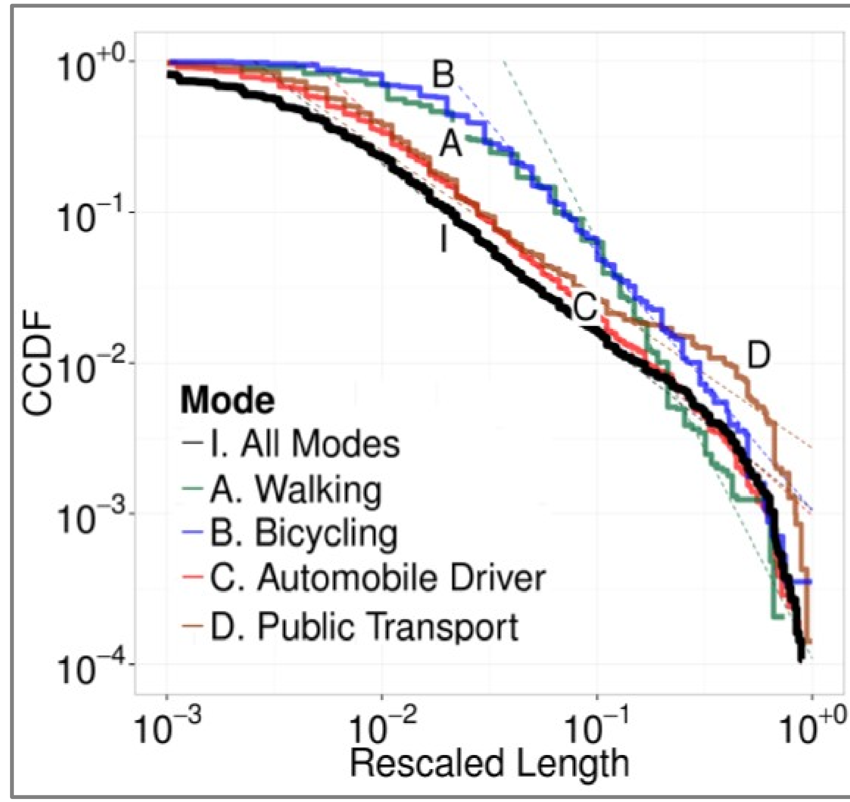
Truncation at 'diameter' of Germany (~$10^{2.83}$ km)

Mode	Count	α	ℓ_0 (km)	$\bar{\ell}$ (km)	σ^2
I. All Modes	52973	2.13	29.40	9.99	1313.79
A. Walk	14303	3.99	6.37	1.37	3.77
B. Bicycle	5581	2.72	6.37	3.47	30.06
C. Auto. Driver	18484	2.29	39.90	13.06	1331.84
D. Public Trans.	6944	1.97	27.98	16.34	2875.92
Auto. Passenger	7658	2.00	24.32	17.69	2949.11

[G.Wilkerson 2014]

Our preliminary findings

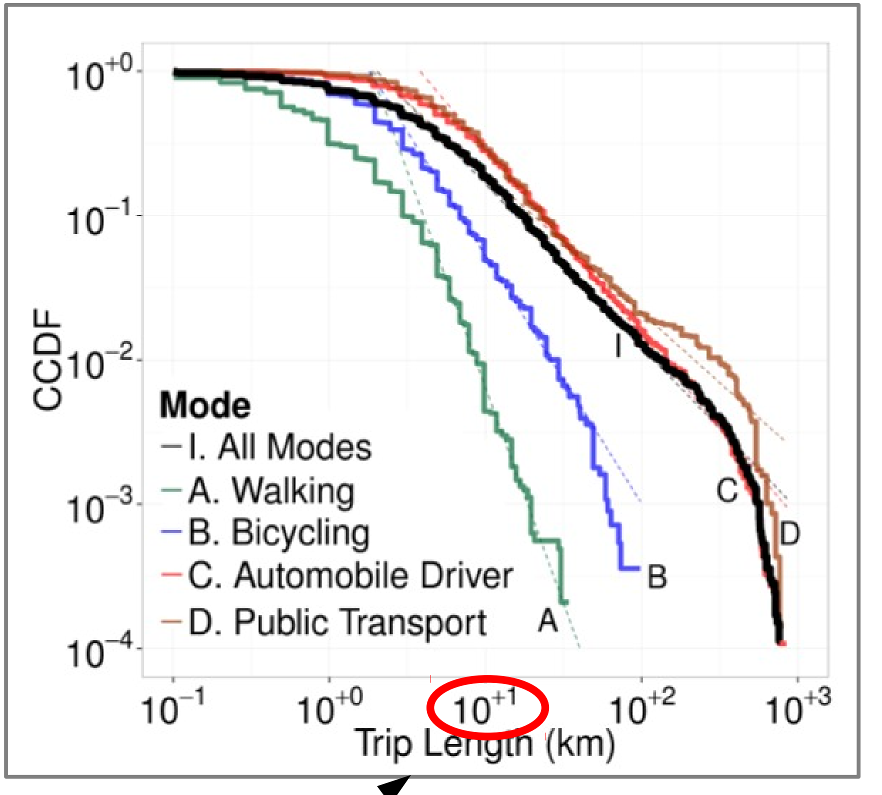
- Categories matter!
  - Trip lengths normalized by fraction of maximum are revealing!
  - Walking, bicycling may have different underlying process vs. fossil-fuel modes! (perhaps exponential? - i.e. almost no rare events – seem not heavy- tailed)
  - May agree with very recent research on human mobility in cities: “Unraveling the origin of exponential law in intra-urban human mobility”, Scientific Reports 3, 2013



[G.Wilkerson 2014]

Our preliminary findings

- Urban scale
  - In Germany, cities over 100,000 pop. have avg. 'diameter' of 14.89km (≈ $10^{1.17}$ km)
  - Contrary to Noulas et al., mobility scaling exponents can be distinguished (and probably statistically fit) well within the urban scale!
  - Previous papers average together *tail* of walking, bicycling with *head* of motorized modes!
  - Arguably, urban trips are complicated, not in straight line, influenced by many factors (including street geometry) vs. straight flights

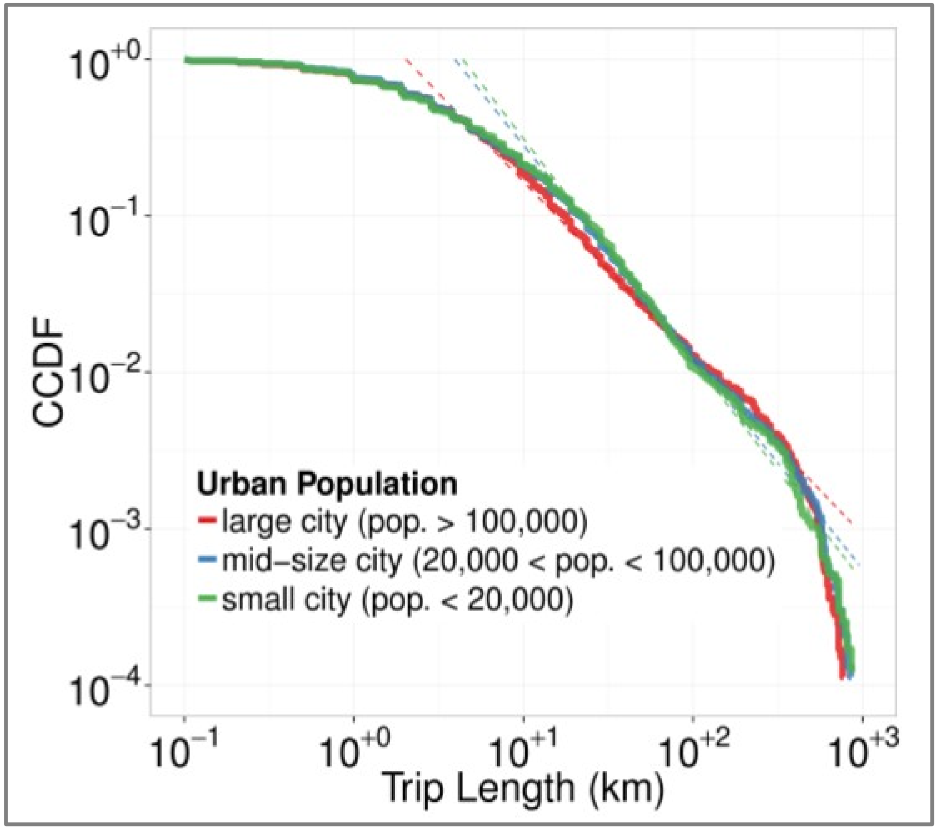


urban scale
(≈ $10^{1.17}$ km)

[G.Wilkerson 2014]

Our preliminary findings

- Allometric scaling of trip length according to population
  - Possible but not conclusive



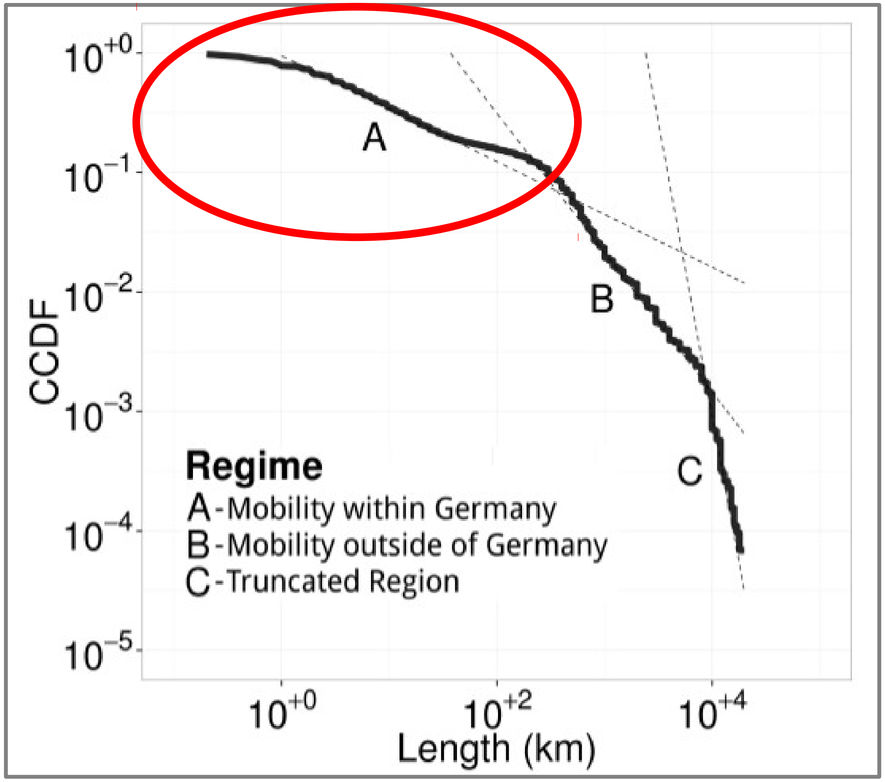
Urban Population	Count	α	ℓ_0 (km)	$\bar{\ell}$ (km)	σ^2
small ($< 20k$)	23433	2.41	43.32	10.52	1202.28
medium ($20k$-$100k$)	53038	2.35	30.38	10.62	1329.72
large ($> 100k$)	53011	2.13	29.40	9.99	1312.92

[G.Wilkerson 2014]

Our preliminary findings

- Mobility scaling results seems to agree with previous studies when trips and travel taken together (daily & overnight mobility)
  - We have α = 1.44, previous values ranged from α = 1.50 to 1.75

- Universality classes: within 'diameter' of Germany (~$10^{2.83}$ km) and distance of Thailand (~$10^{3.94}$ km)

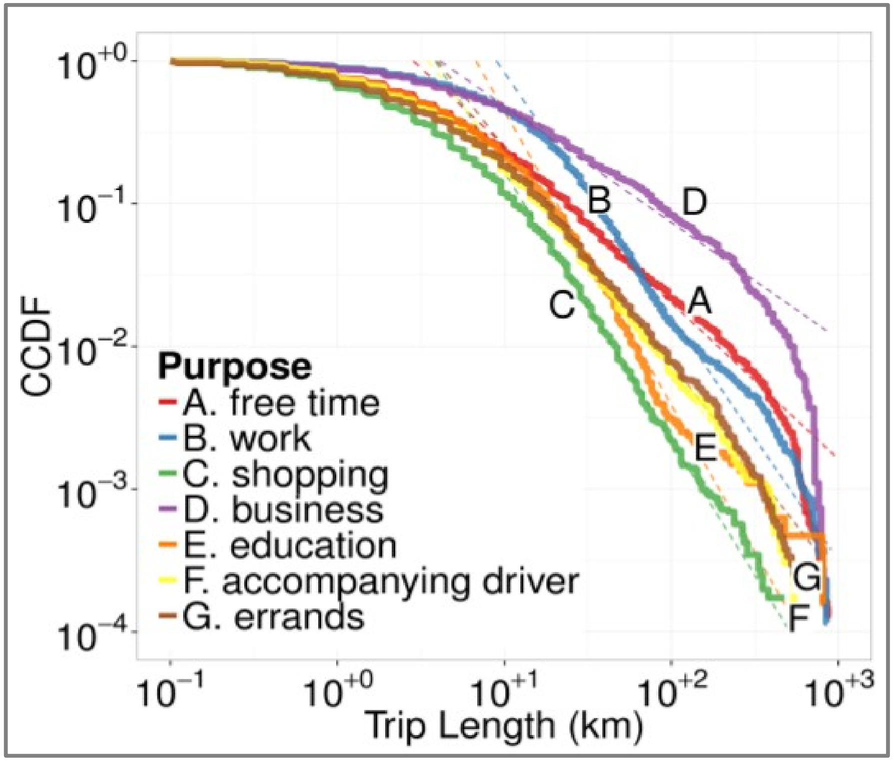


Regime	Count	α	ℓ_0 (km)	$\bar{\ell}$ (km)	σ^2
A	209,045	1.44	1.81	48.97	14,727.00
B	8,055	2.17	816.00	1,670.36	2,172,741.49
C	380	5.91	11,000.00	11,312.92	7,047,781.70

[G.Wilkerson 2014]

Our preliminary findings

- Purpose matters !
  - Education vs. business, vs. others
  - Purely distance vs. intervening opportunity mechanisms
    - Response of length to purpose



Purpose	Count	α	ℓ_0 (km)	$\bar{\ell}$ (km)	σ^2
education	12704	3.06	31.07	8.15	574.29
shopping	40322	2.88	35.15	5.19	196.73
work	25808	2.71	38.95	17.40	1654.51
errands	23716	2.51	45.13	8.06	593.81
accompanying driver	16447	2.50	32.30	7.74	476.70
free time	61152	2.10	30.38	13.55	2209.65
business	2706	1.82	12.35	36.58	8011.00

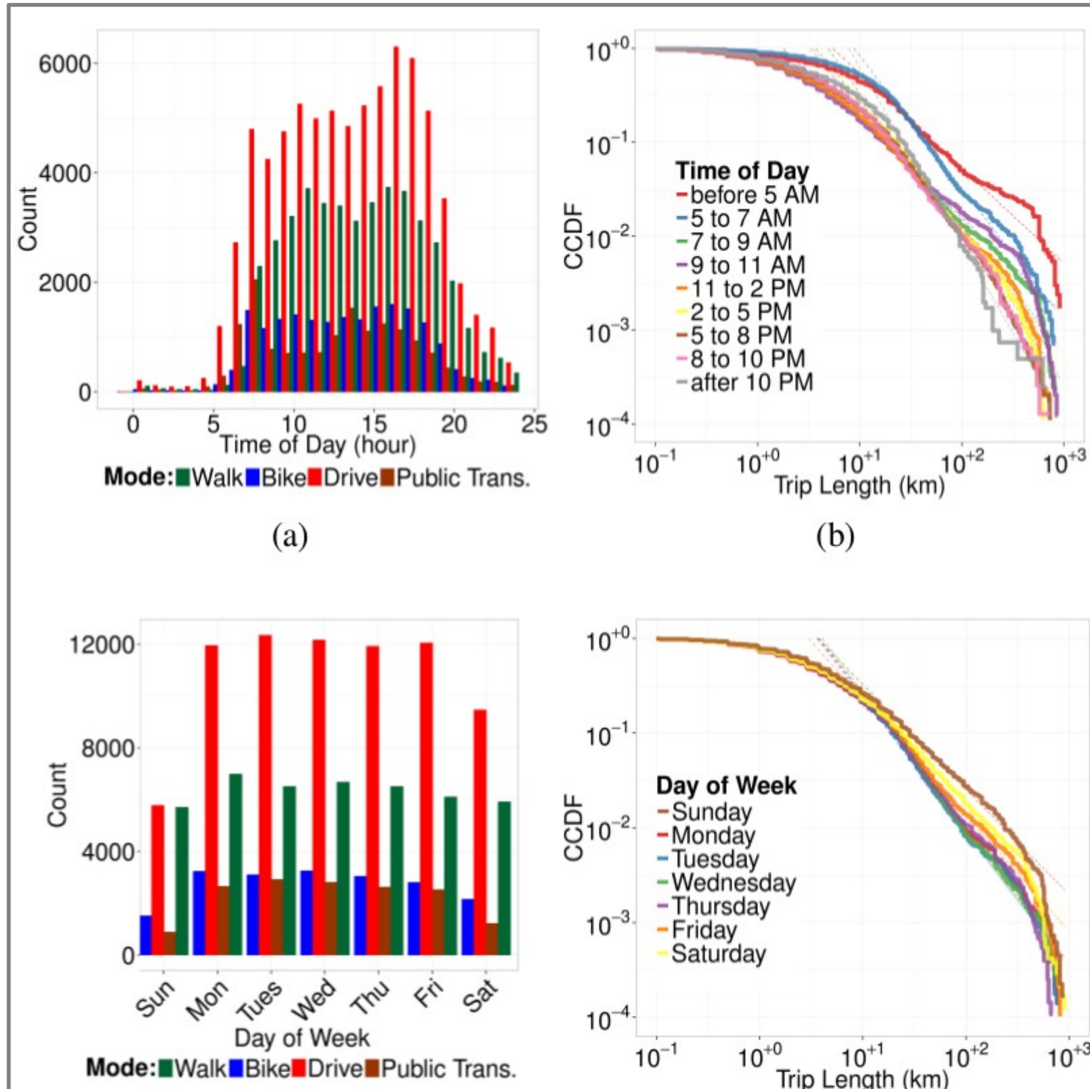
[G.Wilkerson 2014]

Our preliminary findings

- Time matters! We can't just aggregate.

Time of Day	Count	α	ℓ_0 (km)	$\bar{\ell}$ (km)	σ^2
before 5 AM	1670	2.01	25.27	32.92	10,123.99
5 to 7 AM	7026	2.41	20.58	23.71	3,268.39
7 to 9 AM	21991	2.33	16.15	11.29	1,717.64
9 to 11 AM	24511	2.03	10.45	11.31	2,159.07
11 to 2 PM	37693	2.32	31.36	9.49	1,046.26
2 to 5 PM	43375	2.43	51.30	10.34	868.00
5 to 8 PM	34742	2.55	31.36	9.68	705.90
8 to 10 PM	7819	2.39	34.30	9.58	684.20
after 10 PM	4060	2.89	30.40	10.94	550.62

Day of Week	Count	α	ℓ_0 (km)	$\bar{\ell}$ (km)	σ^2
Sunday	17768	2.11	32.34	15.84	2,652.07
Monday	28476	2.42	34.20	9.66	1,026.79
Tuesday	28449	2.42	38.81	9.47	919.62
Wednesday	28649	2.46	48.45	9.86	966.94
Thursday	27787	2.38	38.95	10.07	943.46
Friday	27878	2.22	43.23	11.46	1,507.96
Saturday	23880	2.23	32.30	12.60	1,789.28



[G.Wilkerson 2014]

Summary

- Complex Systems, scaling approach may be useful to characterize most salient urban features
- Categories matter!
- Urban Scale (~10km) is revealing
- Human-powered vs Fossil fuel mobility is different (esp. in cities)
- Mobility scaling confirms previous results
- Distance vs. intervening opportunity mechanisms may be connected by purpose.
- Time matters – aggregation is dangerous

Future Work

- Connection to planning literature

- Careful statistical fitting of heavy tails (non-trivial) and understanding of underlying processes
  - Exponential vs. heavy tails (human-powered vs. fossil-fuel-powered)

    cf. Animal foraging and recent findings on urban-scale exponential trip lengths
  - Trip duration, velocity, energy (!!) distributions → city as energetic (metabolic) system
    - Allometric scaling

- 'Dimensionality reduction' (Clustering, Principle Component Analysis, ANOVA, etc.)
  (~140 parameters, which determine trip length, duration, etc.?)
  - Intervening-opportunity vs. purely distance-based arguments

- Comparison to ther data sources (carefully - big data)

Thanks! and References

gjwilkerson@gmail.com

Wilkerson, G., Khalili, R., and Schmid, S., “Urban Mobility Scaling: Lessons from 'Little' Data”, Netscicom 2014 (IEEE Infocom) http://arxiv.org/abs/1401.0207

General background on Complex Systems and Complex Networks:

Mitchell, Melanie. *Complexity: A guided tour.* Oxford University Press, 2009.

Publications by Mark Newman, Schlomo Havlin, Eugene Stanley, Per Bak, Laszlo Barabasi, and many others.

Urban scaling:

Bettencourt, Luís MA, et al. "Growth, innovation, scaling, and the pace of life in cities." Proceedings of the National Academy of Sciences 104.17 (2007): 7301-7306.

Batty, Michael. Cities and complexity: understanding cities with cellular automata, agent-based models, and fractals. The MIT press, 2007.

Muneepeerakul, R., & Qubbaj, M. R. (2012). The effect of scaling and connection on the sustainability of a socio-economic resource system. Ecological Economics, 77(C)

Mobility scaling:

Brockmann, Dirk, Lars Hufnagel, and Theo Geisel. "The scaling laws of human travel." Nature 439.7075 (2006): 462-465.

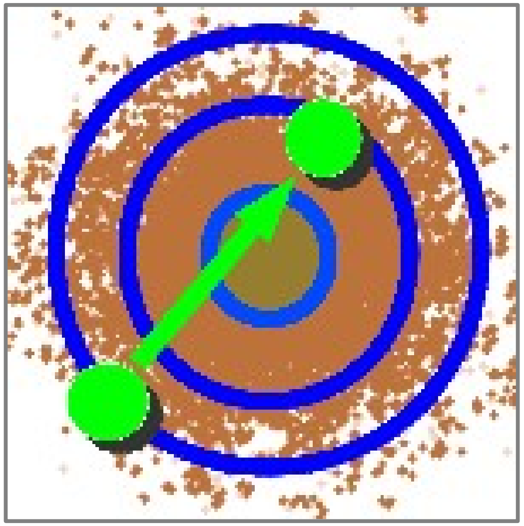
Gonzalez, Marta C., Cesar A. Hidalgo, and Albert-Laszlo Barabasi. "Understanding individual human mobility patterns." Nature 453.7196 (2008): 779-782.

Noulas, Anastasios, et al. "A tale of many cities: universal patterns in human urban mobility." PloS one 7.5 (2012): e37027.

Cho, Eunjoon, Seth A. Myers, and Jure Leskovec. "Friendship and mobility: user movement in location-based social networks." Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 2011.

Liang, Xiao, et al. "Unraveling the origin of exponential law in intra-urban human mobility.", Scientific reports 3 (2013).

Future Work

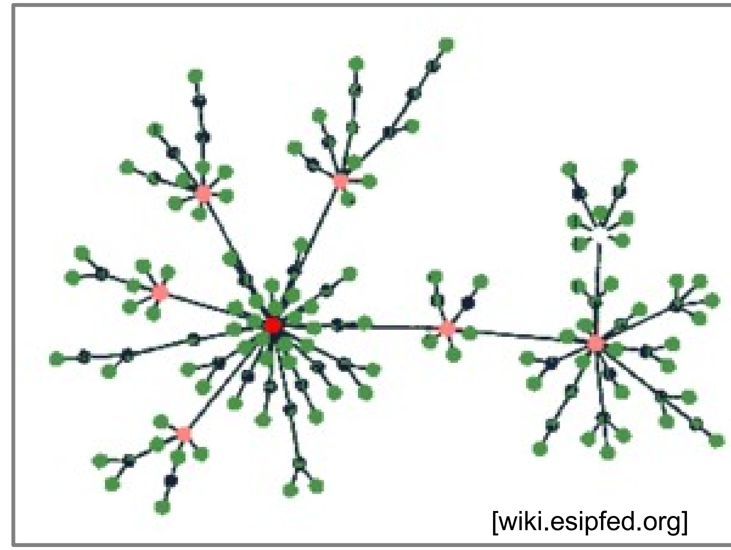


Purely geometric mechanisms

- Distance distribution between points A, B chosen randomly in a disc?

    … in a 2-D Gaussian?

Future Work



- Complex Networks
  - E.g. Infrastructure, social networks
  - Another view of systems: interactions between parts.
    - Also big - described statistically (e.g. number of connections per node)
    - ***__Many__*** examples of *real-world networks*:
      metabolic networks, co-author networks, airport connections, etc.
    - Shared with theoretical computer science (graphs)
    - Often number of neighbors (degree) follows a heavy-tail
    - Very large literature and recent work!

Future Work

- Infrastructure networks
  - Highways as 'small-world' connections vs. 'lattice' of city
  - Requires energy to build and use them (autos) ! (due to time constraints)
  - Relation to transportation mode share?