Urban Mobility Scaling: Lessons from „Little Data“

Galen Wilkerson*, Ramin Khalili**, **Stefan Schmid****

* TU Berlin, Germany
** TU Berlin & T-Labs, Berlin, Germany

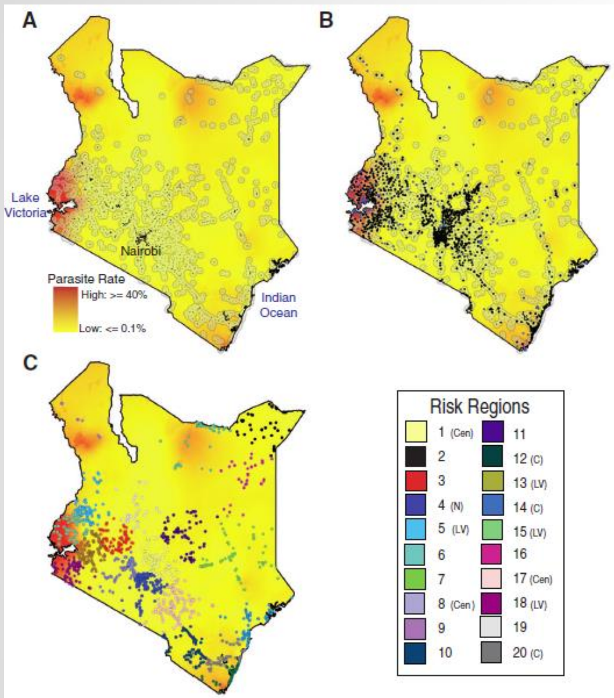
Urban Mobility Scaling: Lessons from „Little Data“

Galen Wilkerson*, Ramin Khalili**, **Stefan Schmid****

* TU Berlin, Germany
** TU Berlin & T-Labs, Berlin, Germany

Understanding Human Mobility

Model disease propagation (e.g., Malaria)

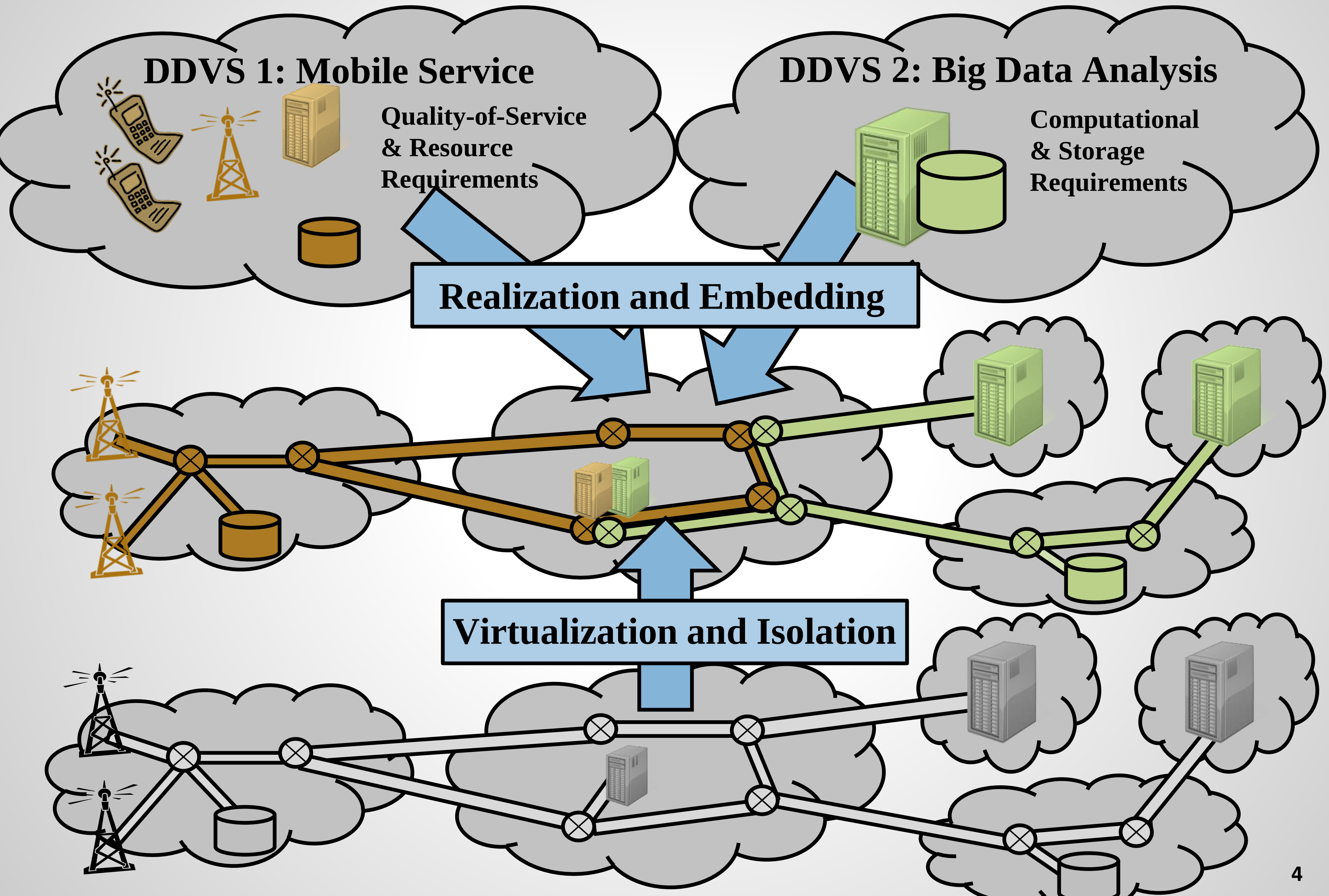


Reduce energy / CO2 consumption

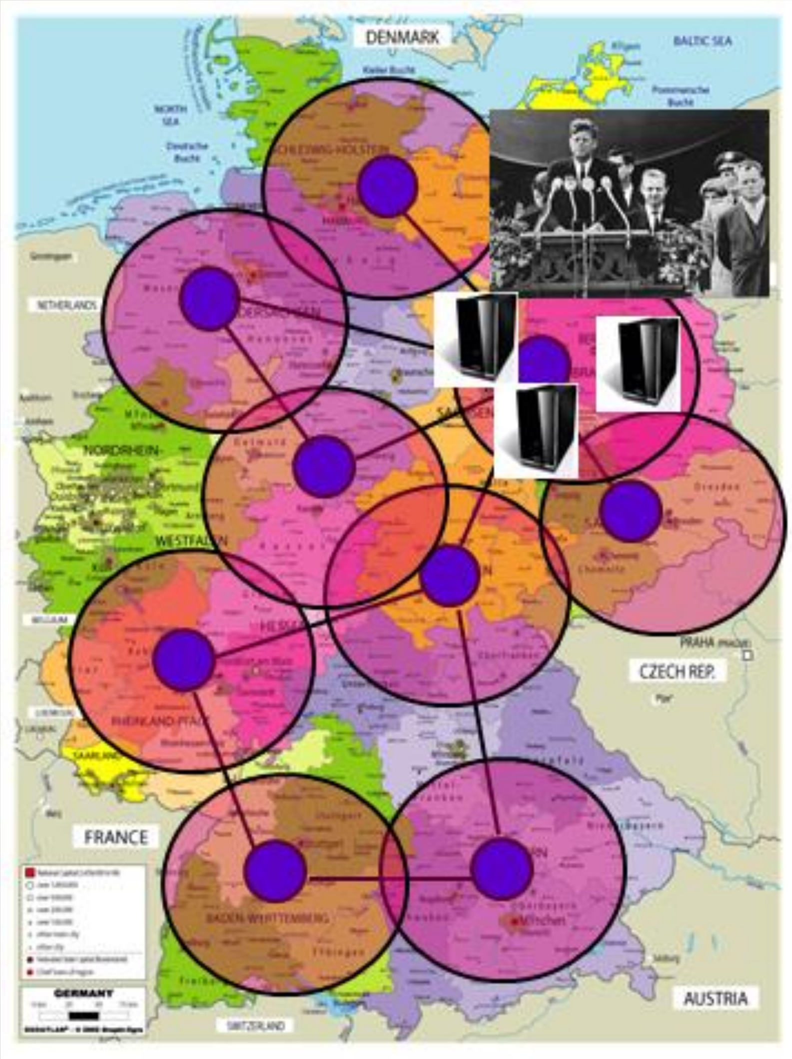
Avoid escape panic

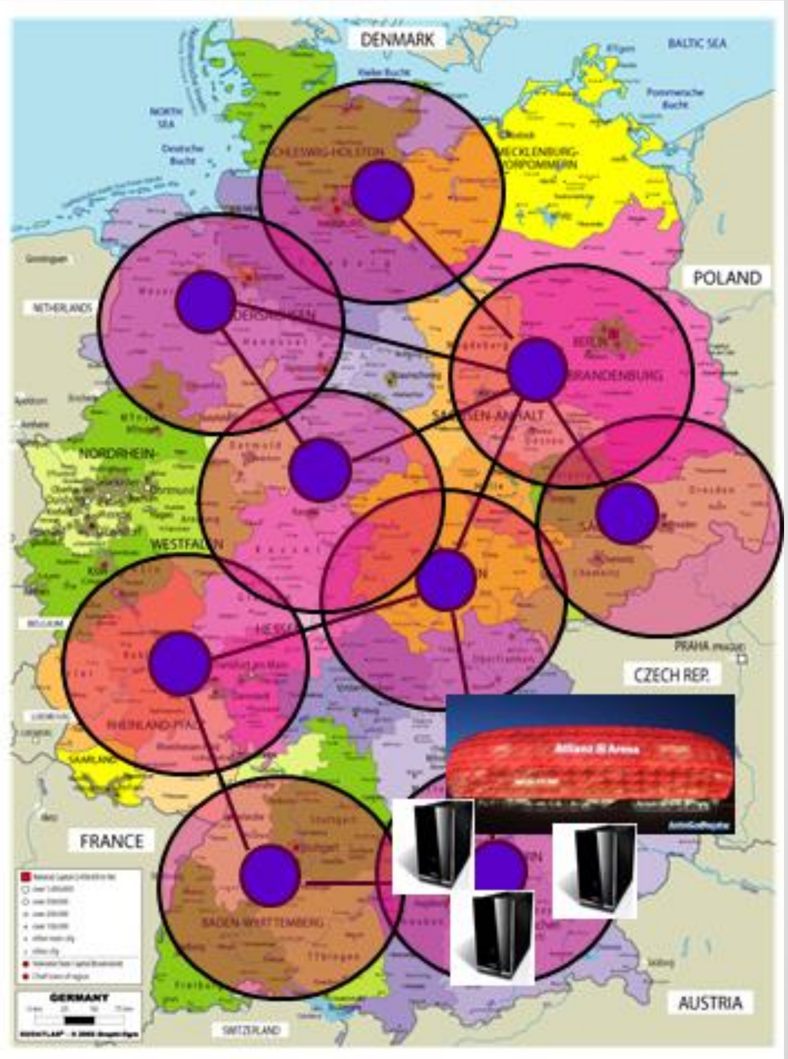
Infrastructure planning

Our Motivation: Network Virtualization



Our Motivation: Network Virtualization





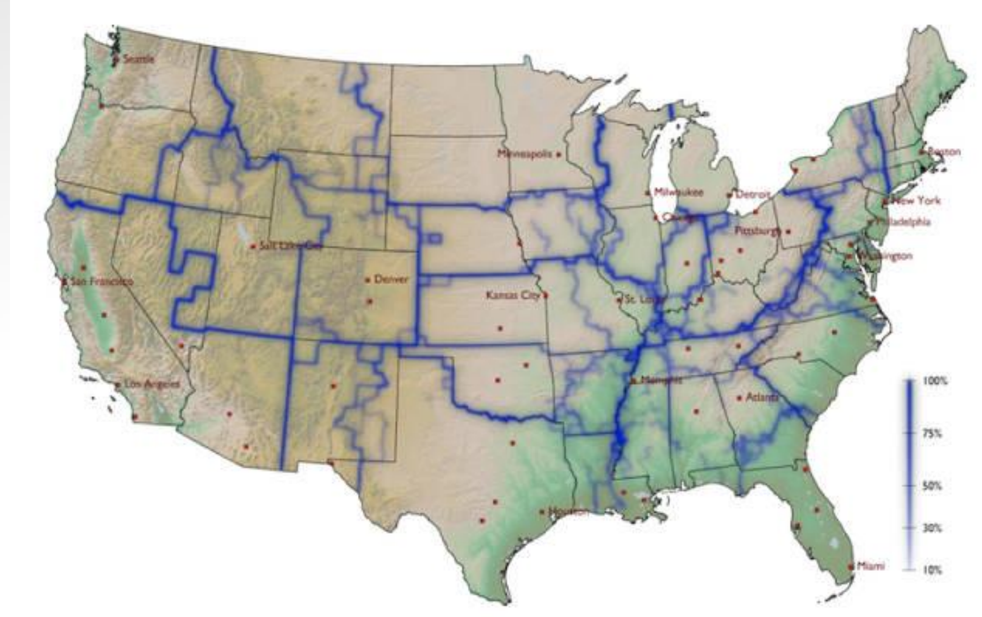
“Big Data” Studies

- Today’s data often checkin based (no “trajetory”)
  - wheresgeorge.com
  - Mobile call-data records (CDRs)
  - Geo-tagged social media (Foursquare, Twitter, ...)

Brockmann’s “dollar states” (unlikely to cross)

- Automatic data collection: often aggregated, no categories (meta-data)
- Based on “incidental sampling”
- Resolution often limited
- Euclidean distances only

“Big Data” Studies



Brockmann’s “dollar states” (unlikely to cross)

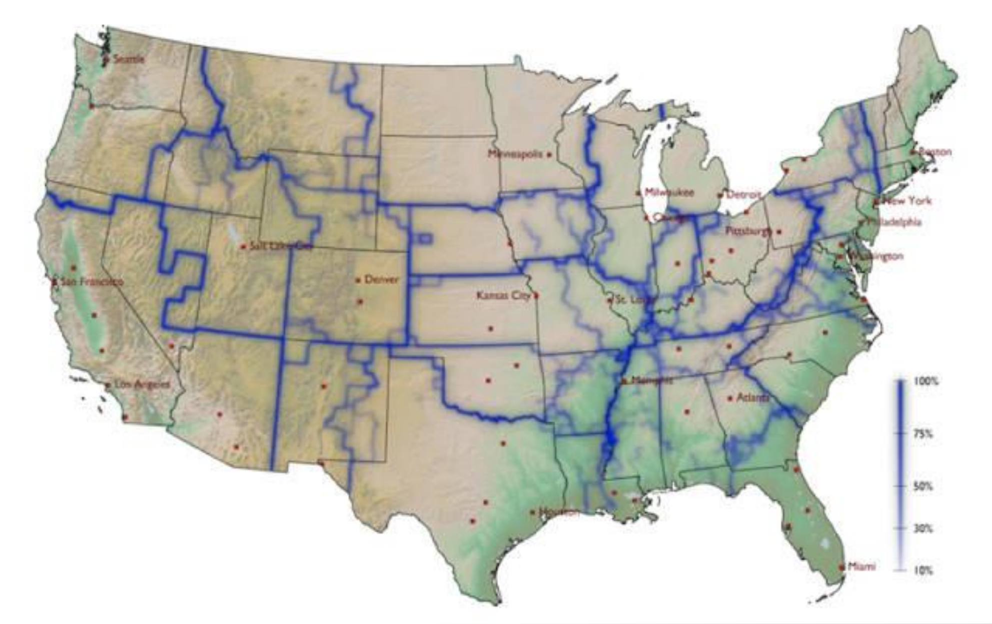
- ❑ Today’s data often checkin based (no “trajetory”)
  - ❑ wheresgeorge.com
  - ❑ Mobile call-data records (CDRs)
  - ❑ Geo-tagged social media (Foursquare, Twitter, ...)

- ❑ Automatic data collection: often aggregated, no categories (meta-data)
- ❑ Based on “incidental sampling”
- ❑ Resolution often limited
- ❑ Euclidean distances only

Treating samples as i.i.d. is imprudent!

“Big Data” Studies

- Today’s data often checkin based (no “trajetory”)
  - wheresgeorge.com
  - Mobile call-data records (CDRs)
  - Geo-tagged social media (Foursquare, Twitter, ...)



Brockmann’s “dollar states” (unlikely to cross)

- Automatic data collection: often aggregated, no categories (meta-data)
- Based on “incidental sampling”
- Resolution often limited
- Euclidean distances only

Especially problematic to study urban mobility!

Treating samples as i.i.d. is imprudent!

Evergreen: Power-law Trip Length Distribution

- Distribution of trip lengths l
- Scaling exponent: polynomial trip length distribution

$$P(l)=C\,l^{-\alpha}$$

- Where's George data: $\alpha = 1.59$
- Foursquare study: $\alpha = 1.50$

Different α imply significant differences in number of trips of a certain length!

(aka. "universality classes")

Little data, but “oho”!

- Data from «Mobility in Germany 2008» survey
  - February 2009 - March 2009
  - Specifically collected to study human mobility

- Conventional data: by phone, online and mail surveys
- Methodology: Maximum likelihood and Kolmogorov-Smirnov

- Largest household survey in Germany (apart microcensus)
  - 25,922 households; 60,713 individualys
  - ~ 200,000 trips: without overnight stay
  - ~ 25,000 travels: with overnight stay

- No interpolation: actual reported distances

Little data, but “oho”!

- Data from «Mobility in Germany 2008» survey
  - February 2009 - March 2009

Many categories: mode (car, bike, ...), time of day, purpose (work, shopping), duration, etc.

Granularity down to 100m!

  - 25,922 households; 60,713 individualys
  - ~ 200,000 trips: without overnight stay
  - ~ 25,000 travels: with overnight stay

- No interpolation: actual reported distances

Little data, but "oho"!

- Data from «Mobility in Germany 2008» survey
  - February 2009 - March 2009

Many categories: mode (car, bike, ...), time of day, purpose (work, shopping), duration, etc.

Granularity down to 100m!

Good to study urban mobility!

The urban scale is hardly understood: Check-in data too coarse-grained to study trip lengths (Euclidean interpolation inaccurate), many alternative modes, etc.

- ...mirnov
- ...s)
  - 2...
  - ~...
  - ~...
- No i...

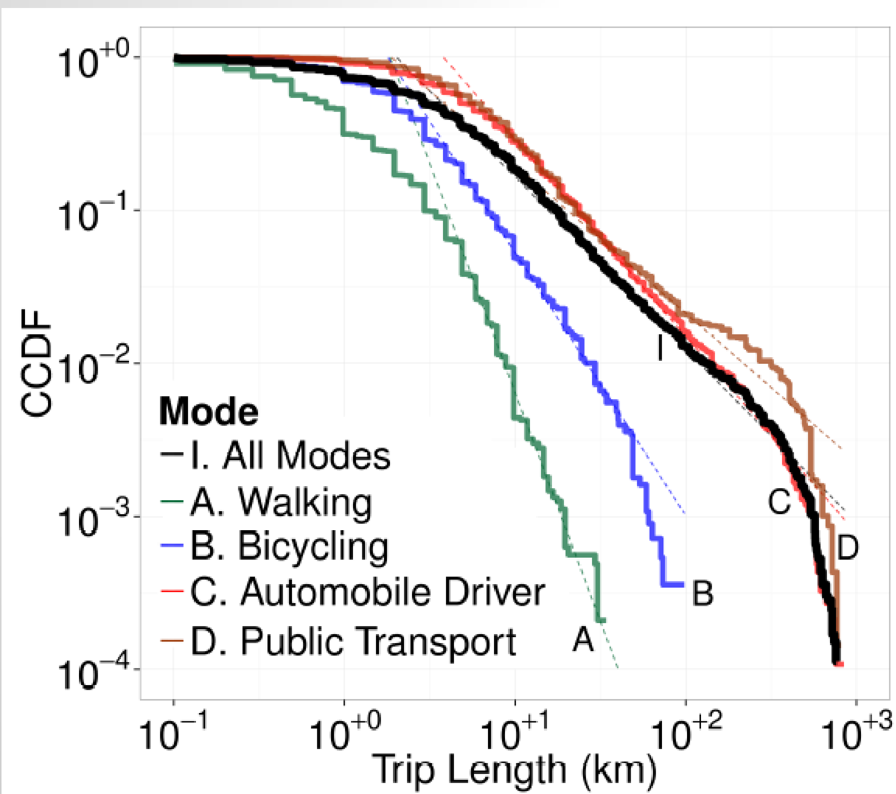
Overview of Results

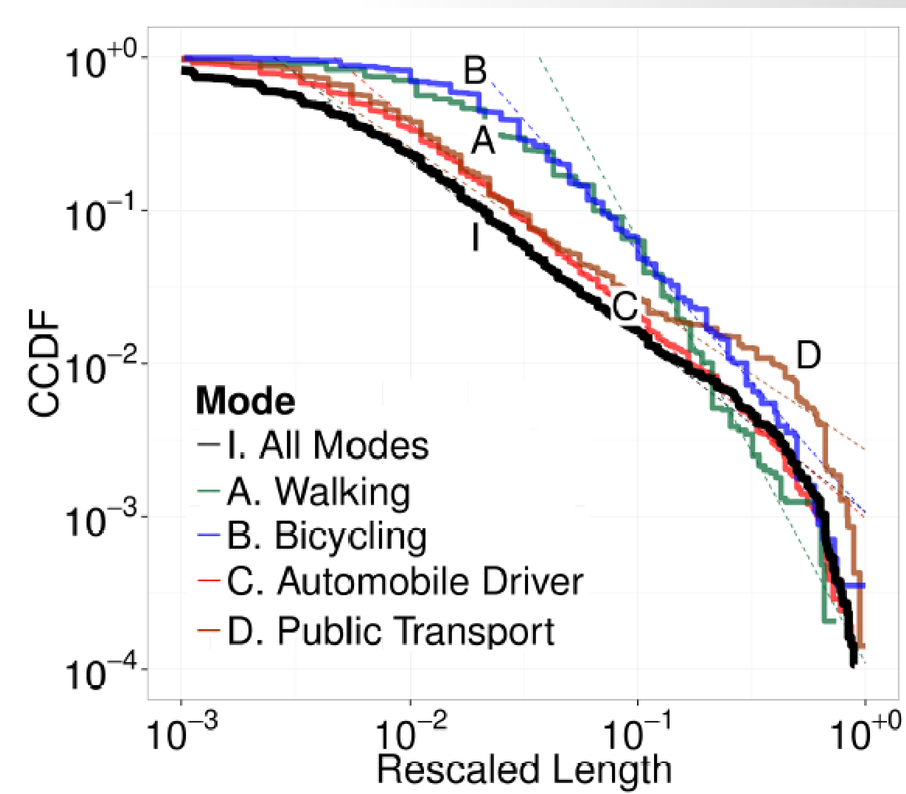
- At urban scale (~10km), mode differences are evident
- While at an aggregated level, we get similar alpha values as in big data studies, they depend on mode
  - Aggregation inaccurate: tails merge with heads!
- Interestingly, trip length does not depend on population of city
- Other factors matter: purpose, time, ...

The Mode Share

- Non-motorized modes have higher exponents ("less powerlaw")
- At intra-urban scale, a non-negligible number of non-motorized trips
- Note: Variance and mean not defined for small exponents

Mode	Count	α
I. All Modes	52973	2.13
A. Walk	14303	3.99
B. Bicycle	5581	2.72
C. Auto. Driver	18484	2.29
D. Public Trans.	6944	1.97
Auto. Passenger	7658	2.00





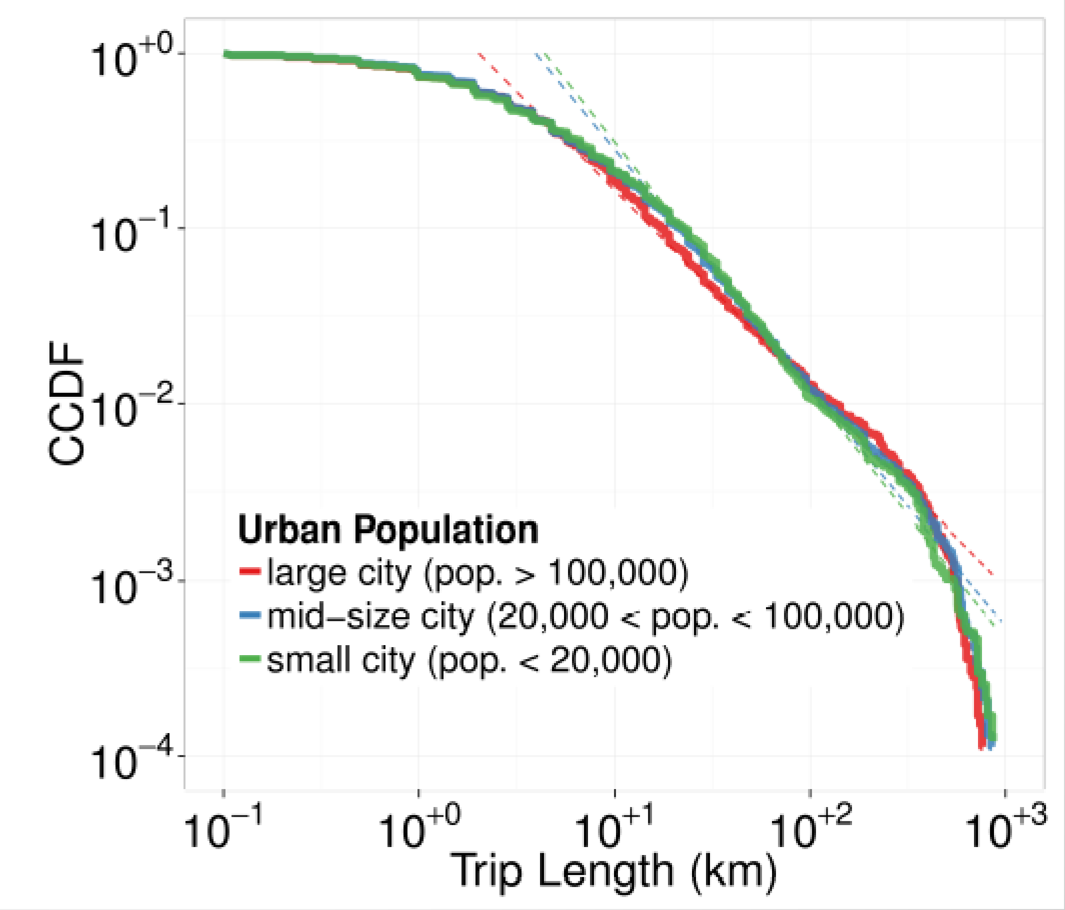
Population

- Population size only has a minor influence on trip length distribution

- Confirms previous study

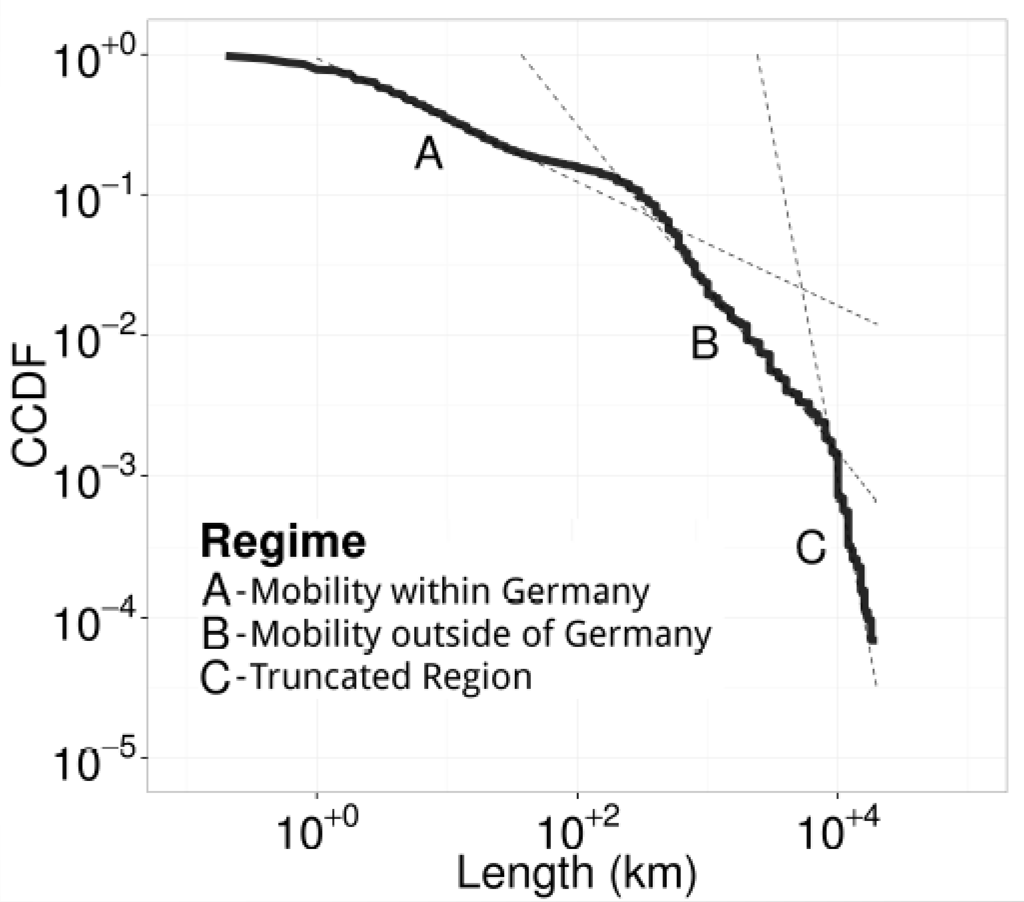
- Also relatively independence of city area: but Pearon values low

Urban Population	Count	α
small ($< 20k$)	23433	2.41
medium ($20k$-$100k$)	53038	2.35
large ($> 100k$)	53011	2.13



Regimes

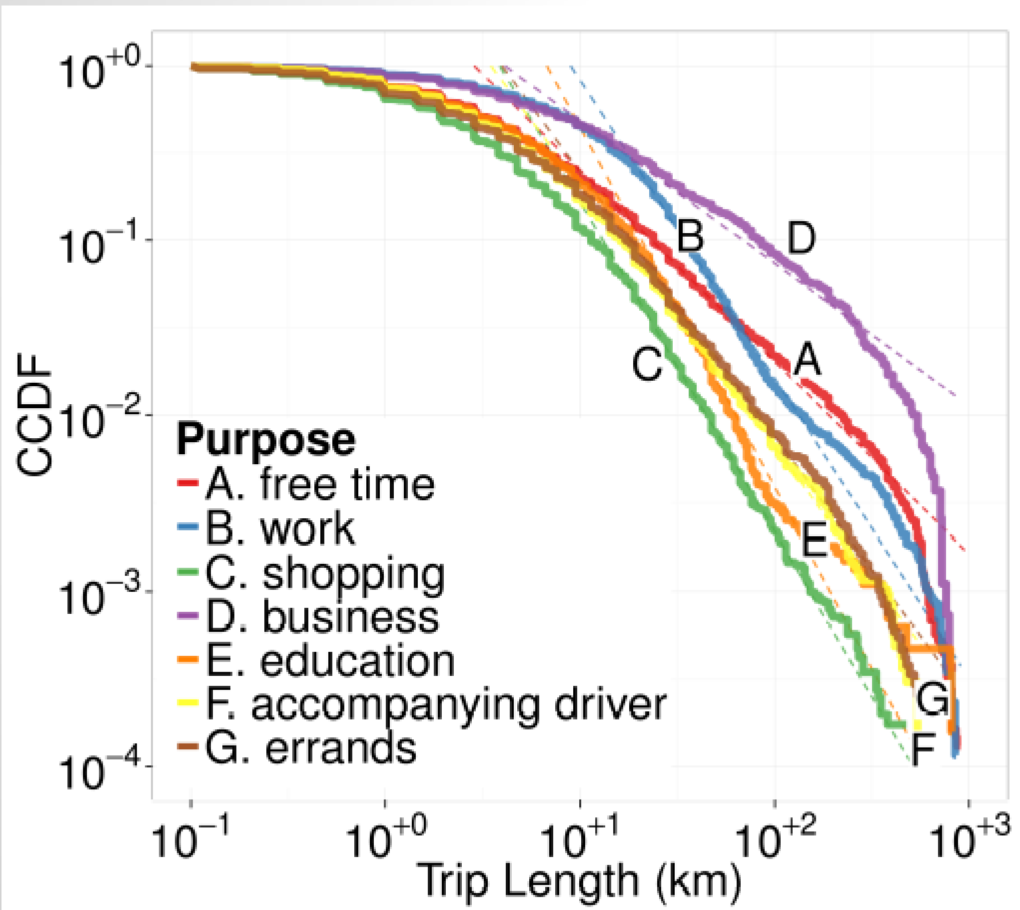
- Overall three regimes: within Germany, outside Germany, truncated
- Trips within Germany yield similar exponents as found in *Where's George* data
- Travels have larger exponents, but biased (origin must be Germany)



Purpose Matters

Trip lengths depend on purpose and "intervening opportunities" / facilities: supports intuitive model by Noulas

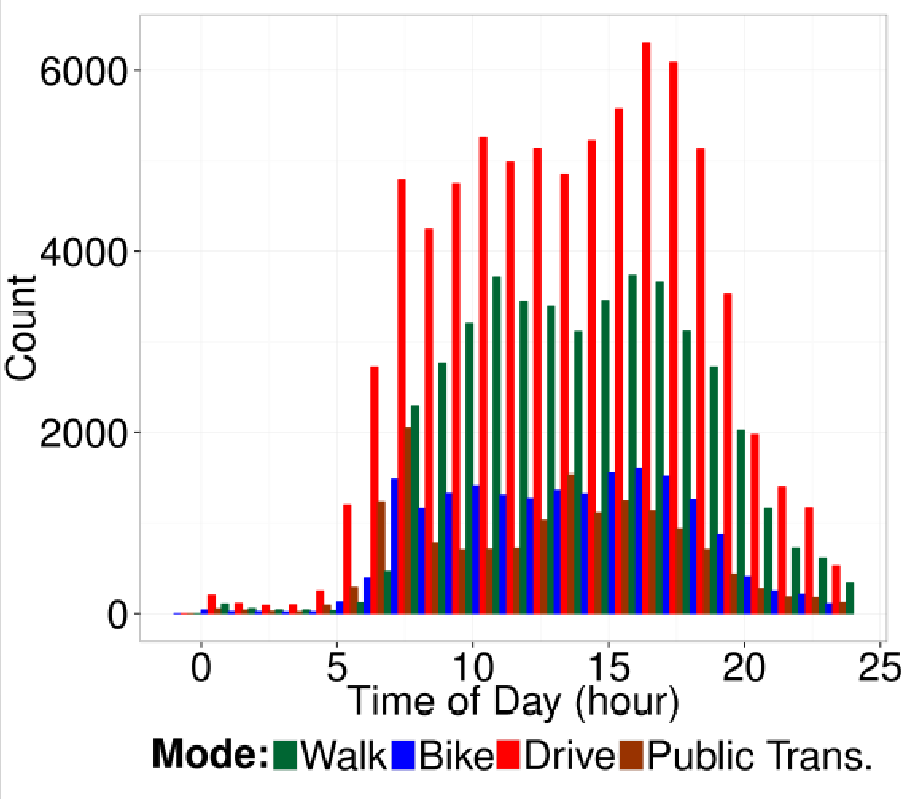
Purpose	Count	α
education	12704	3.06
shopping	40322	2.88
work	25808	2.71
errands	23716	2.51
accompanying driver	16447	2.50
free time	61152	2.10
business	2706	1.82

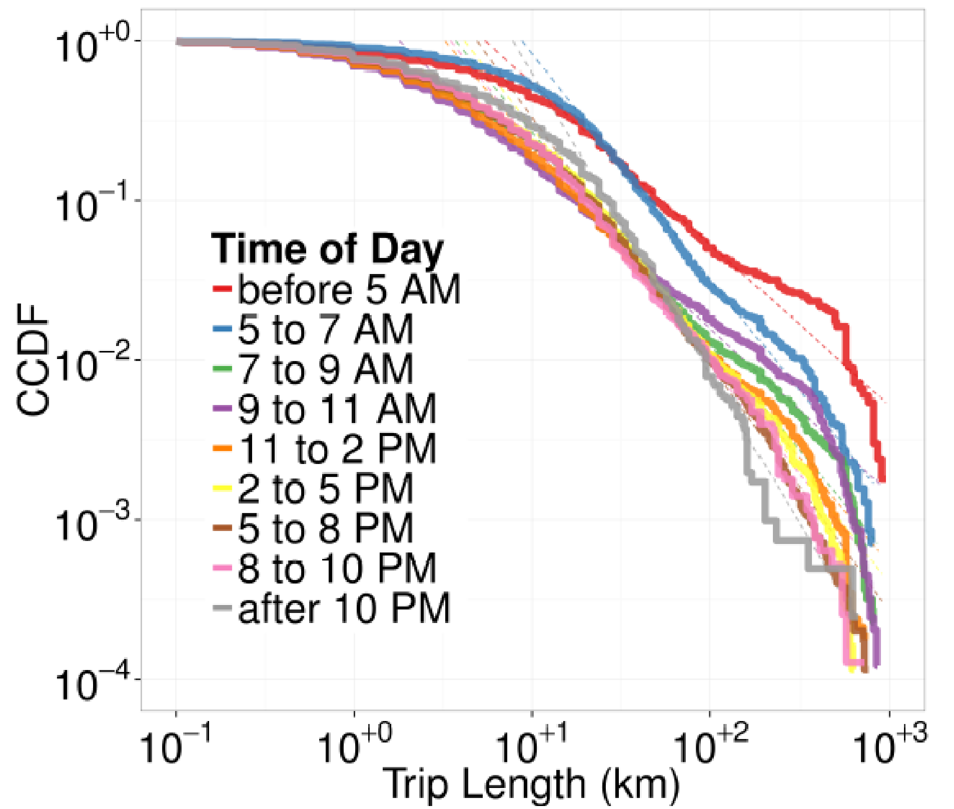


E.g., shopping trips shorter than business trips.

Time Matters (1)

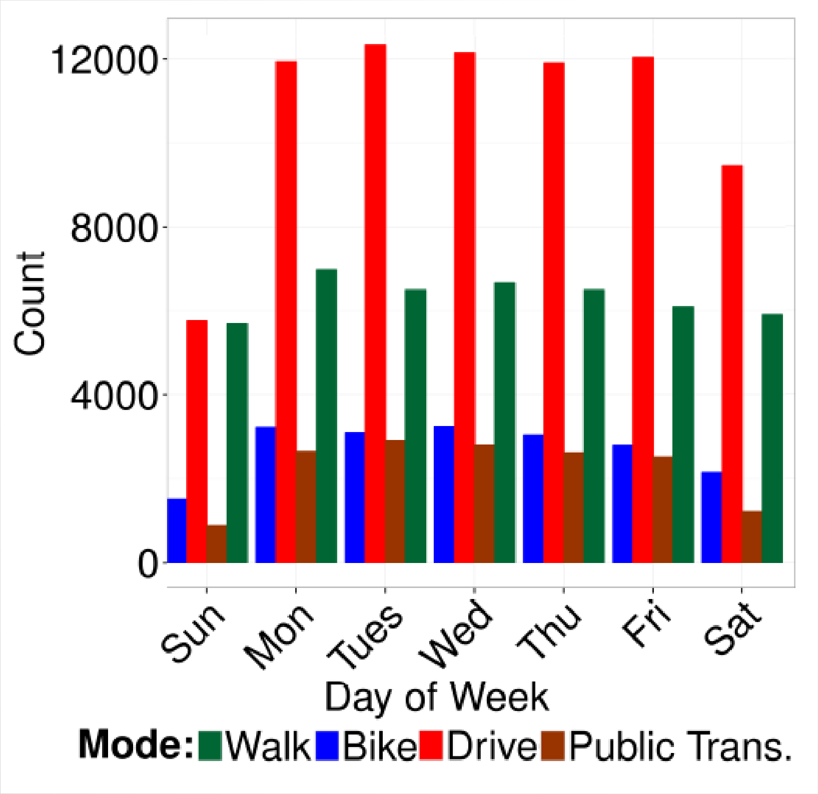
Time of day effects: trips between 5 AM and 7 AM significantly longer

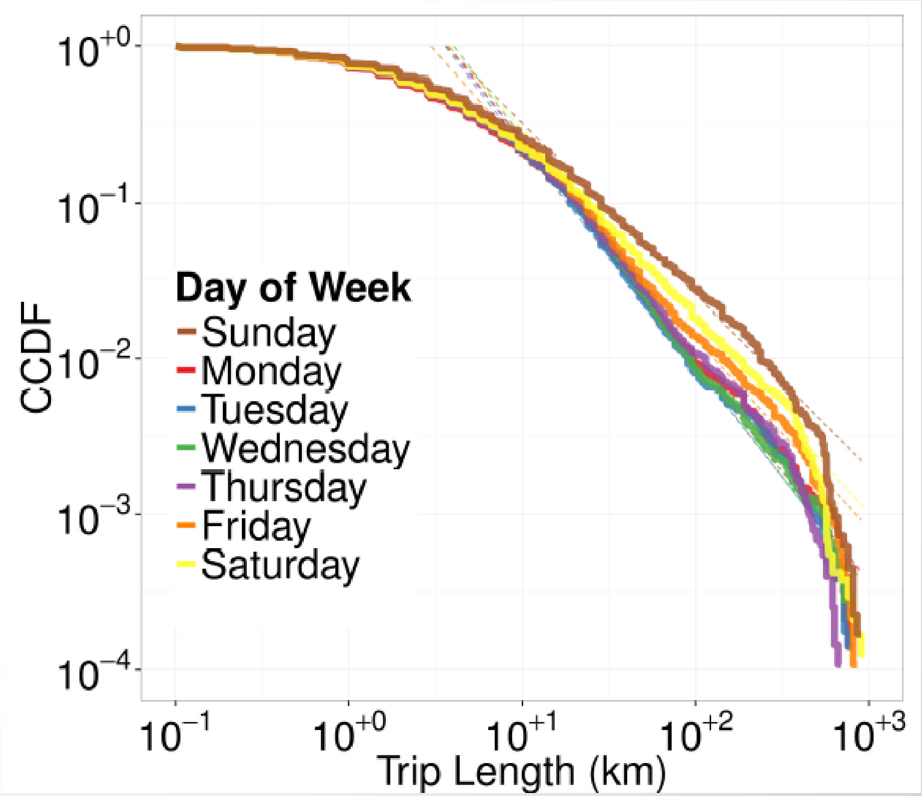




Time Matters (2)

Day of week effects: Sunday have a different trip frequency (lower), mode share (less cars), and lengths (shorter).





Conclusion

- Treating trip lenghts as i.i.d. introduces errors
- Aggregation inaccurate: tails merge with heads!
- Mode matters
  - At urban scale (~10km), many different modes
  - While at an aggregated level, we get similar alpha values as in big data studies, they depend on mode
- Time matters
  - Longer trips in the early morning
  - Different mode share on weekends
- Purpose matters
  - Shopping trips shorter than business trips

Open Questions

- An empirical first look

- Urban mobility, non-motorized modes: rather exponential?
  - **Brockmann**: Human travel displacements follow **power-law** (Dollar bill study)
  - **Gonzàlez**: European mobile phone users described as **truncated power-law**
  - Florence car drivers and mobility in London subway **unlikely power-law**

- Models, models, models
  - In large space, including long-distance, human mobility has **Lévy walk** characteristic and scale-invariant step lengths; also observed in anmial mobility
  - Noulas: mobility driven by **points-of-interest**, **not power-law, not Lévy**

Big Data Controversy


FINANCIAL TIMES

Home | UK | World | Companies | Markets | Global Economy | Lex | Comment | Management | Personal Finance

Arts | Magazine | Food & Drink | House & Home | Lunch with FT | Style | Books | Pursuits | Travel | Columnists | How To Spend It

March 28, 2014 11:38 am

Big data: are we making a big mistake?

By Tim Harford

Big data is a vague term for a massive phenomenon that has rapidly become an obsession with entrepreneurs, scientists, governments and the media

- Big data : “the end of theory”? (Wired 2008)
  - “With enough data, the numbers speak for themselves!”

- The first success: Google Flu Trends
  - Faster than Center for Disease Control

- Euphory lower today: Hard to distinguish between correlation and causality

“There are many small data problems that occur in big data. They don’t disappear because you’ve got lots of the stuff. They get worse.” (Spiegelhalter)

- Recent victim: Google Flu Trends

Thank you.


Deutsche Telekom Laboratories
Wir machen aus Ideen Zukunft.
TECHNISCHE UNIVERSITÄT