Revisiting urban economics in light of data

Marc Barthelemy

CEA, Institut de Physique Théorique, Saclay, France

EHESS, Centre d'Analyse et de Mathématiques sociales, Paris, France

marc.barthelemy@cea.fr
www.quanturb.com

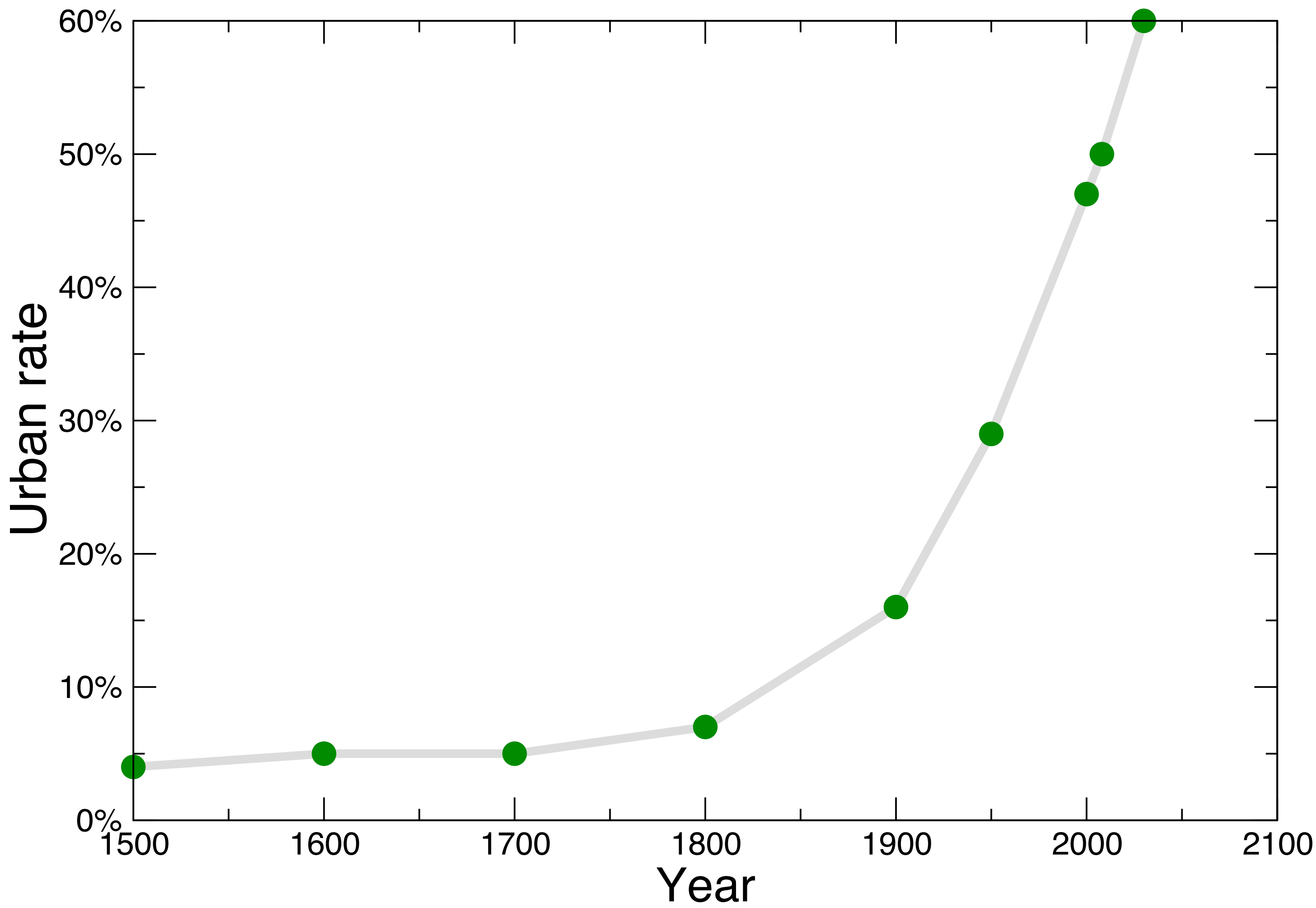
Outline

- Urban science: state of the art
- The polycentric structure of data
  - Mobile phone data
  - Patterns of commuting
  - Measuring hotspots
- Understanding polycentrism and scaling exponents
  - The edge-city model (Krugman)
  - classical model: Fujita-Ogawa
  - Revisiting the Fujita-Ogawa model
  - Computing scaling exponents

Outline

- Understanding mobility
  - Mobility: gravity law
  - The radiation model
- Relation between commuting distance and income
  - Empirical results
  - Testing the McCall model of job search
  - The 'closest opportunity' model
- Infrastructure
  - Multilayer networks
  - Time evolution of the road network

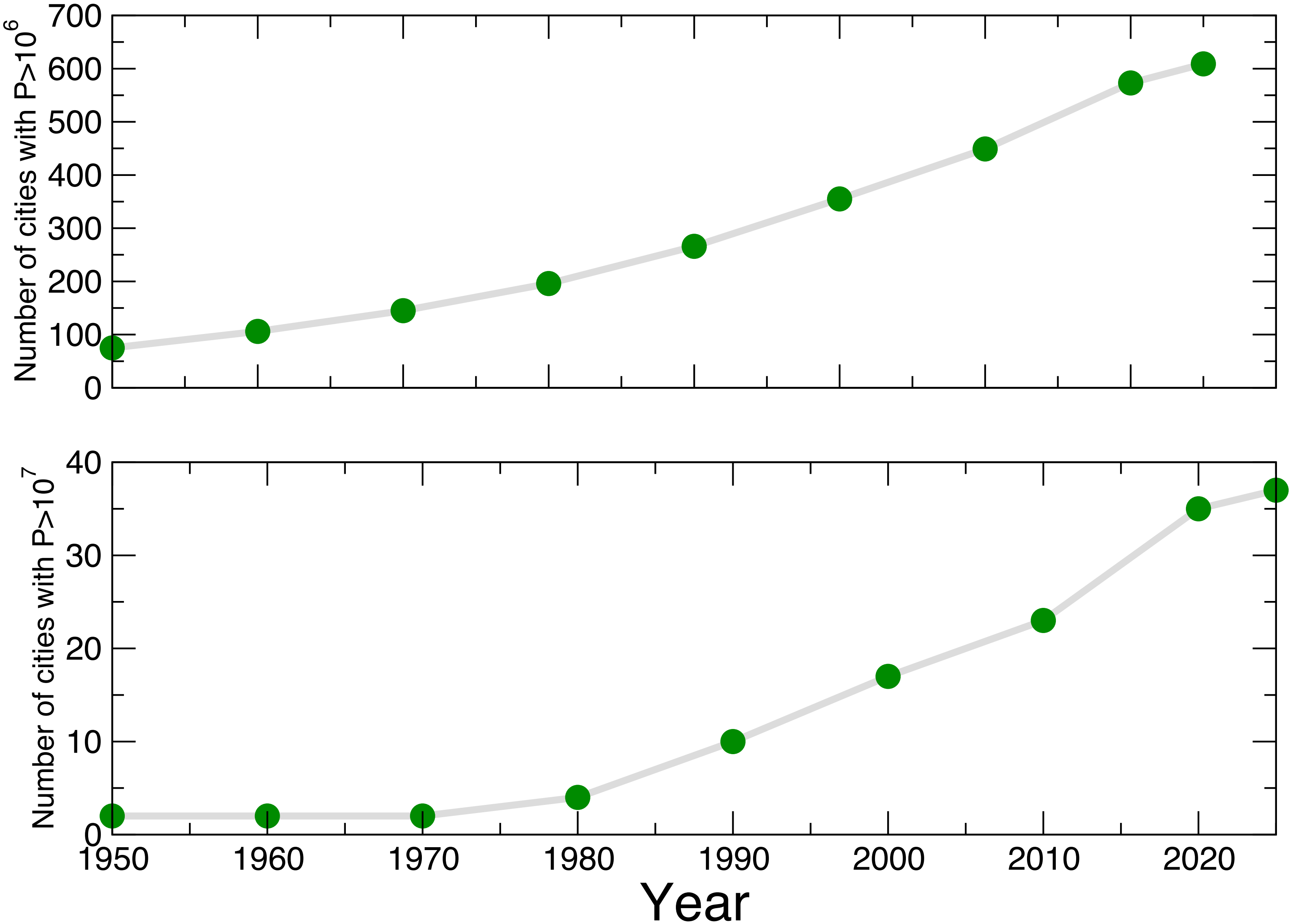
Importance of cities: urbanization rate



Projection: in 2050: 70% of the world population lives in cities

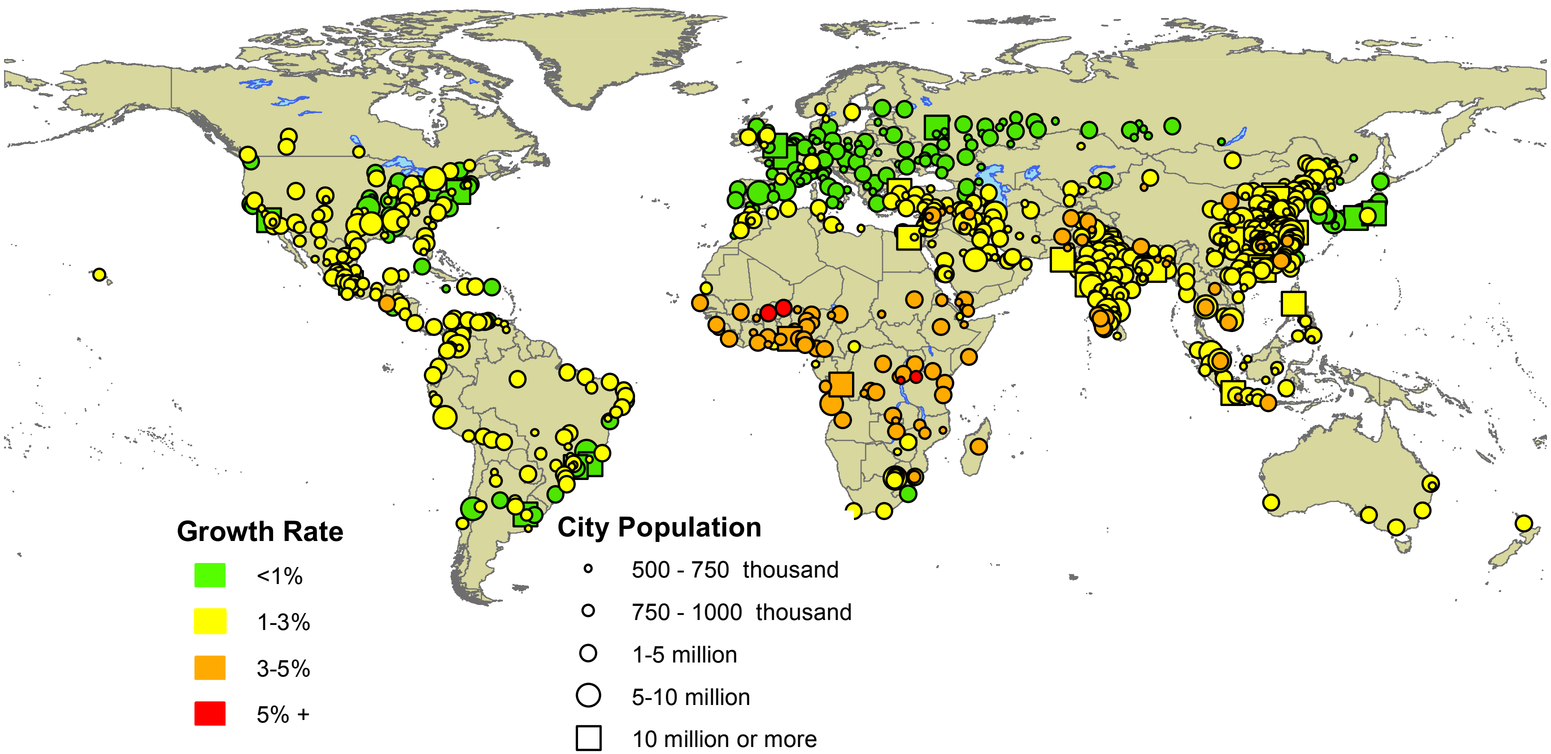
Data from: HYDE historical database

Importance of cities: megacities

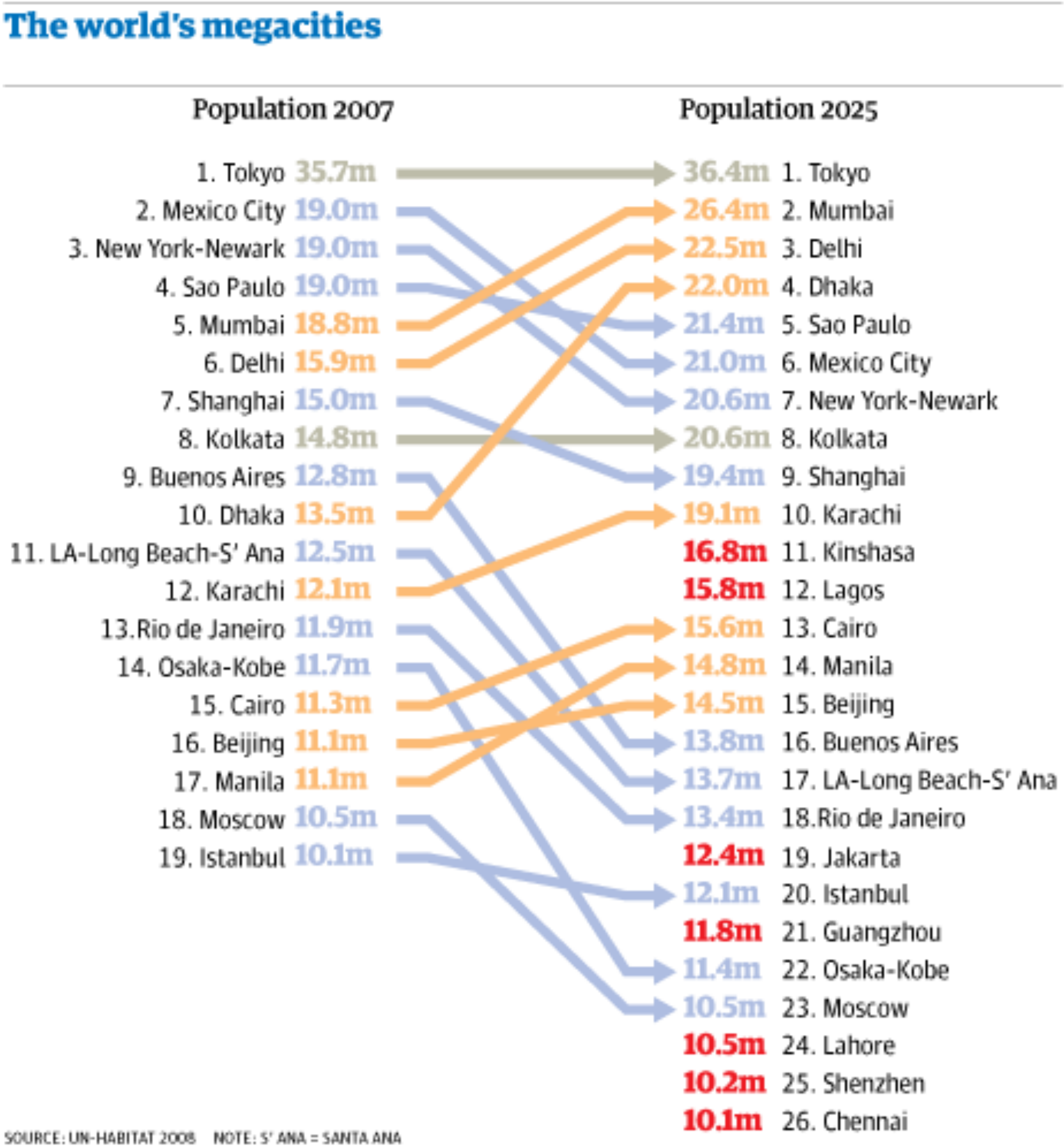


Data from www.geohive.com (UN data)

Importance of cities

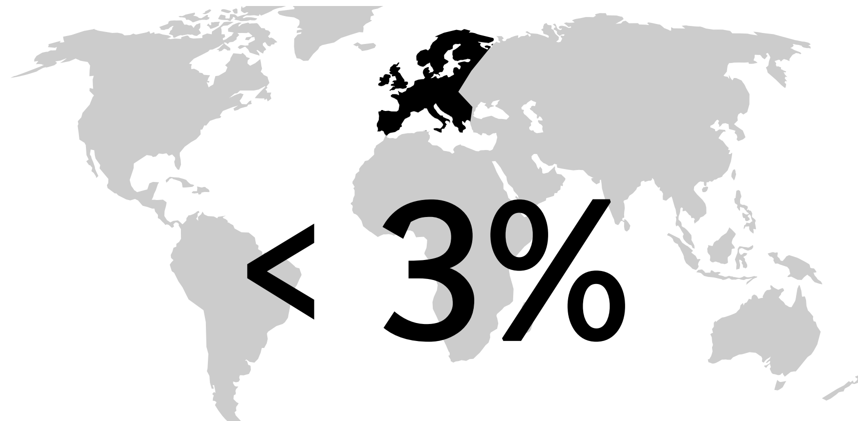


Heterogeneous distribution of growth rates


The world's megacities
Population 2007
Population 2025
1. Tokyo 35.7m
2. Mexico City 19.0m
3. New York-Newark 19.0m
4. Sao Paulo 19.0m
5. Mumbai 18.8m
6. Delhi 15.9m
7. Shanghai 15.0m
8. Kolkata 14.8m
9. Buenos Aires 12.8m
10. Dhaka 13.5m
11. LA-Long Beach-S' Ana 12.5m
12. Karachi 12.1m
13.Rio de Janeiro 11.9m
14. Osaka-Kobe 11.7m
15. Cairo 11.3m
16. Beijing 11.1m
17. Manila 11.1m
18. Moscow 10.5m
19. Istanbul 10.1m
36.4m 1. Tokyo
26.4m 2. Mumbai
22.5m 3. Delhi
22.0m 4. Dhaka
21.4m 5. Sao Paulo
21.0m 6. Mexico City
20.6m 7. New York-Newark
20.6m 8. Kolkata
19.4m 9. Shanghai
19.1m 10. Karachi
16.8m 11. Kinshasa
15.8m 12. Lagos
15.6m 13. Cairo
14.8m 14. Manila
14.5m 15. Beijing
13.8m 16. Buenos Aires
13.7m 17. LA-Long Beach-S' Ana
13.4m 18.Rio de Janeiro
12.4m 19. Jakarta
12.1m 20. Istanbul
11.8m 21. Guangzhou
11.4m 22. Osaka-Kobe
10.5m 23. Moscow
10.5m 24. Lahore
10.2m 25. Shenzhen
10.1m 26. Chennai
SOURCE: UN-HABITAT 2008 NOTE: S' ANA = SANTA ANA

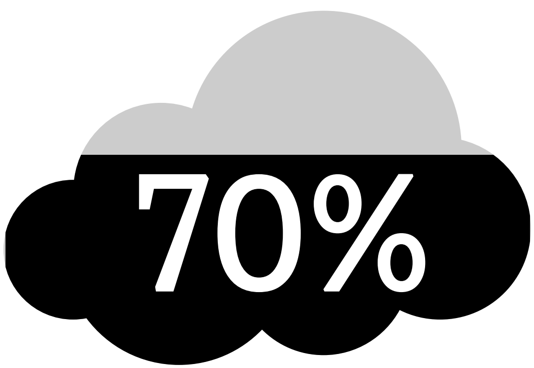
Cities are about concentration

Urbanized area



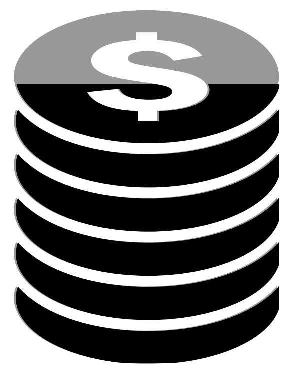
NASA 2001

CO2 emissions



ONU-HABITAT 2011

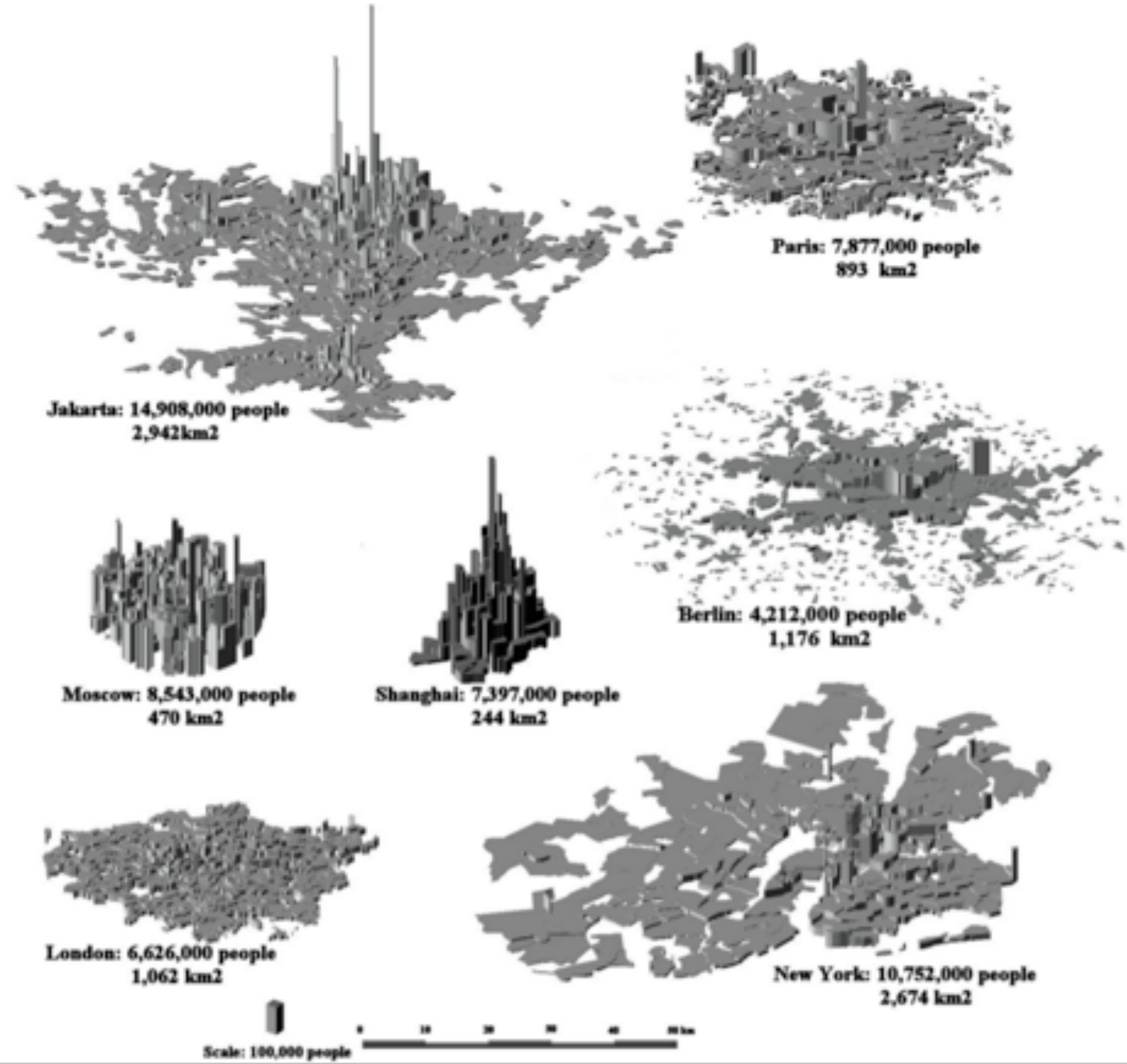
GDP



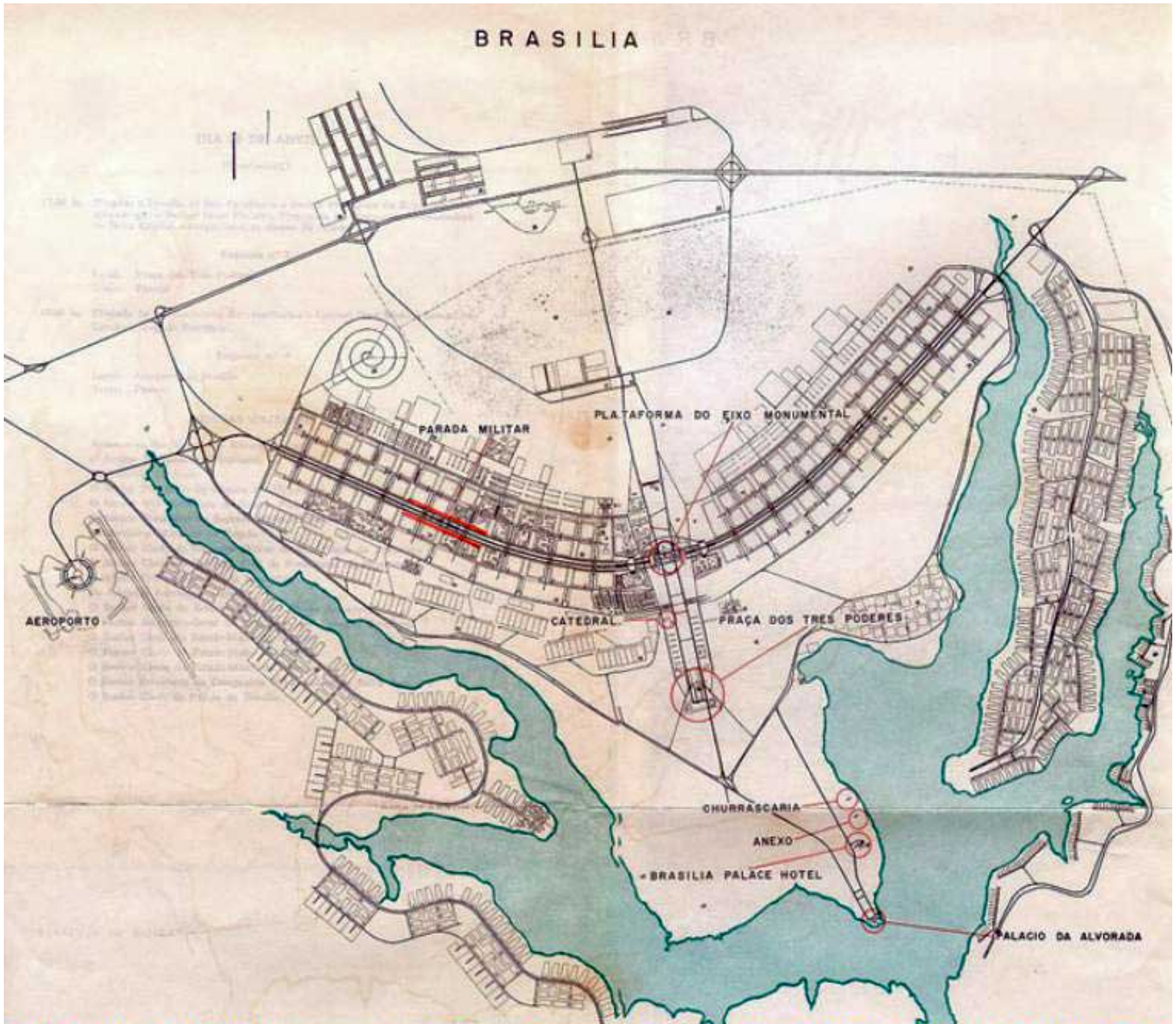
80%

ONU 2011

Finding the logic behind the apparent diversity of cities

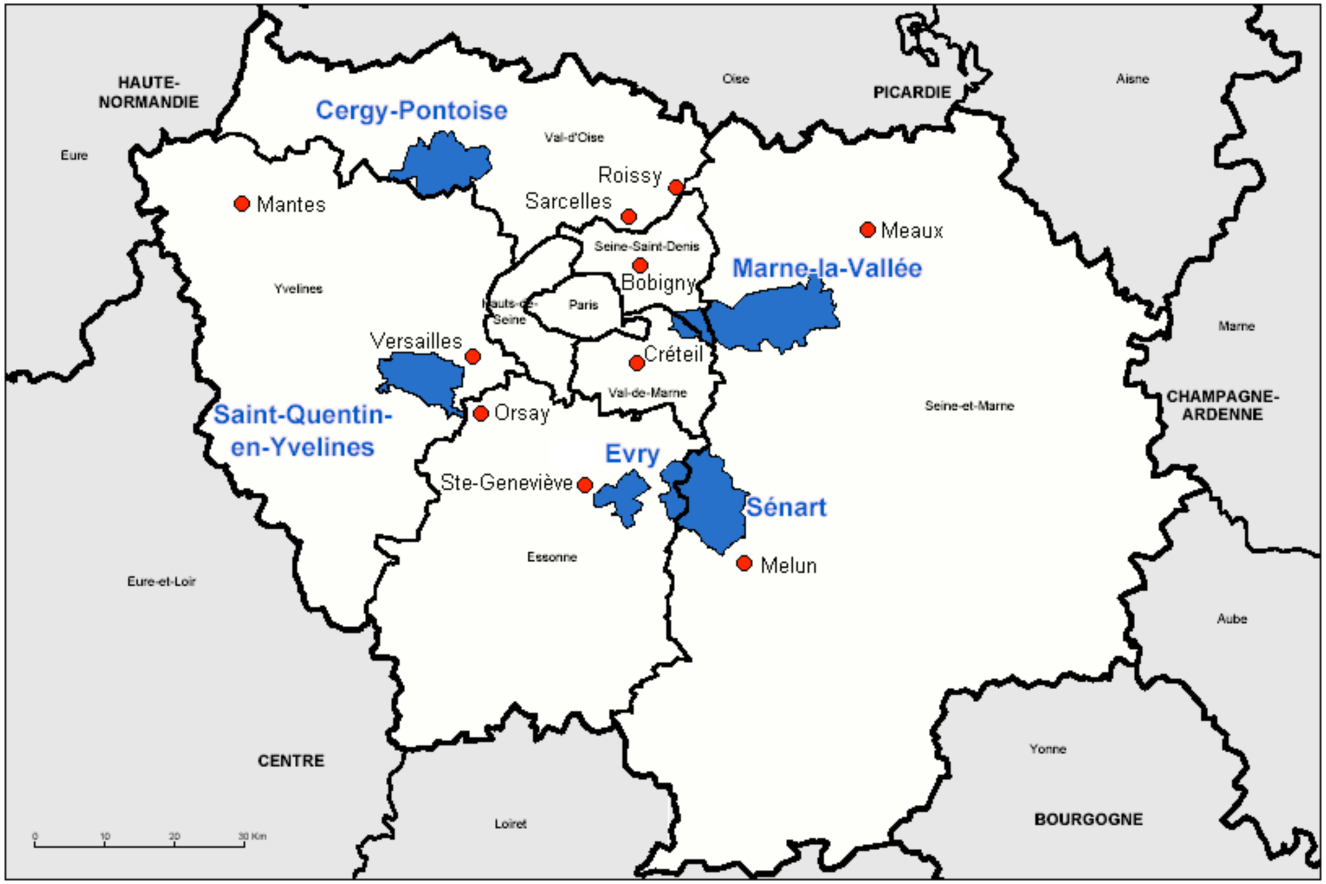


Many 'theories' in urbanism…and nevertheless



Brasilia (1960)
- Thought for cars
- Plan too rigid

Many 'theories' in urbanism…and nevertheless



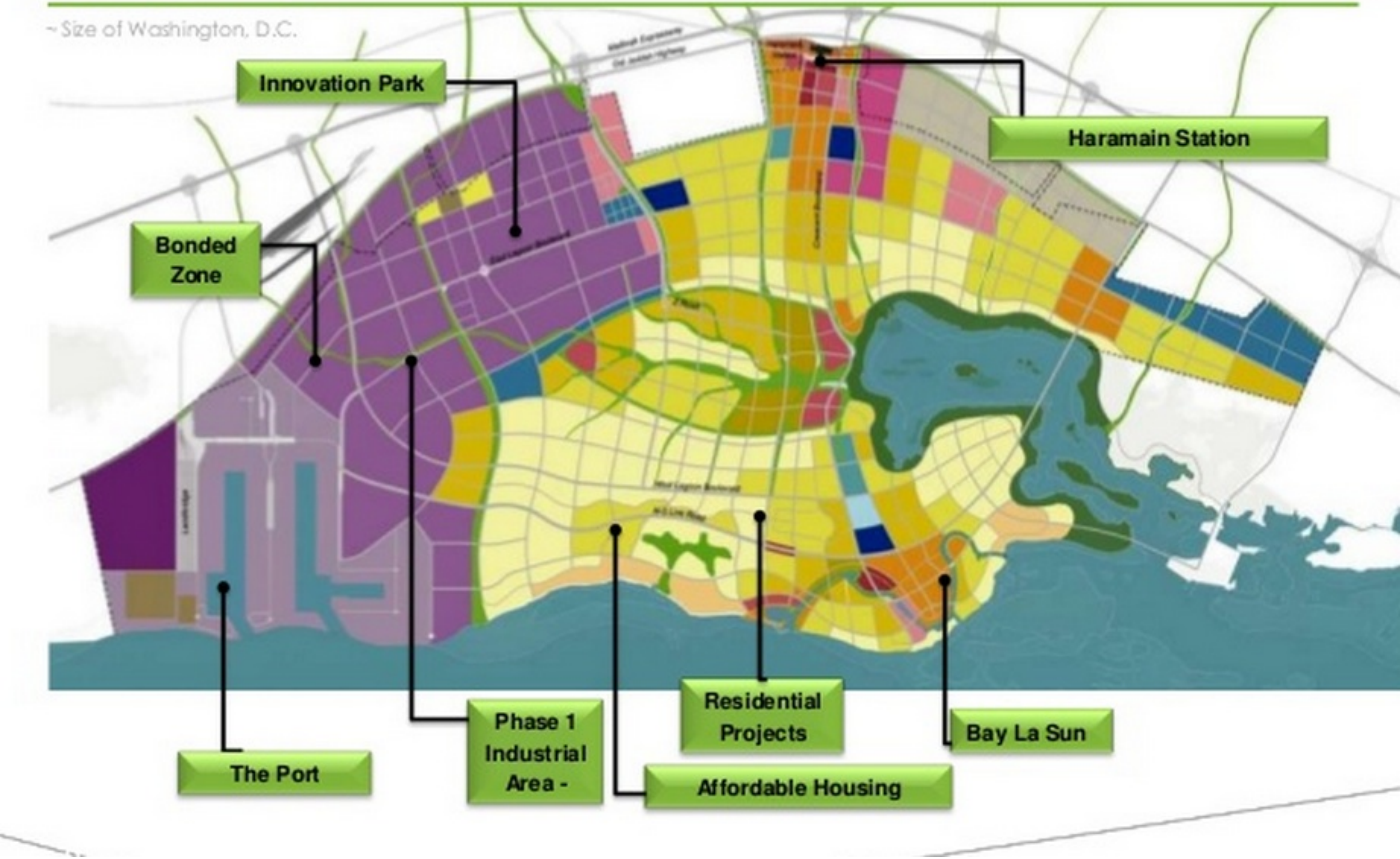
"Villes nouvelles" (1960): pendular movement…enhanced !

Many ‘theories’ in urbanism…and nevertheless

Urban sprawl: high environment, social and economical cost

KAEC is the Largest Integrated Economic City at Over 168km² With 64km of Waterfront

~ Size of Washington, D.C.



Many 'theories' of urbanism but nevertheless, we observe a large number of problems !

- Social and economical problems (spatial income segregation, crime, accessibility, …)

- Traffic problems; pollution

- Sustainability of these structures ?

- We need a robust theoretical guide for urbanism
- Necessity to understand these phenomena and to achieve a 'science of cities' (or 'quantitative urbanism') validated by data (in particular, for large-scale projects)

Long term goal: 'quantitative urbanism'

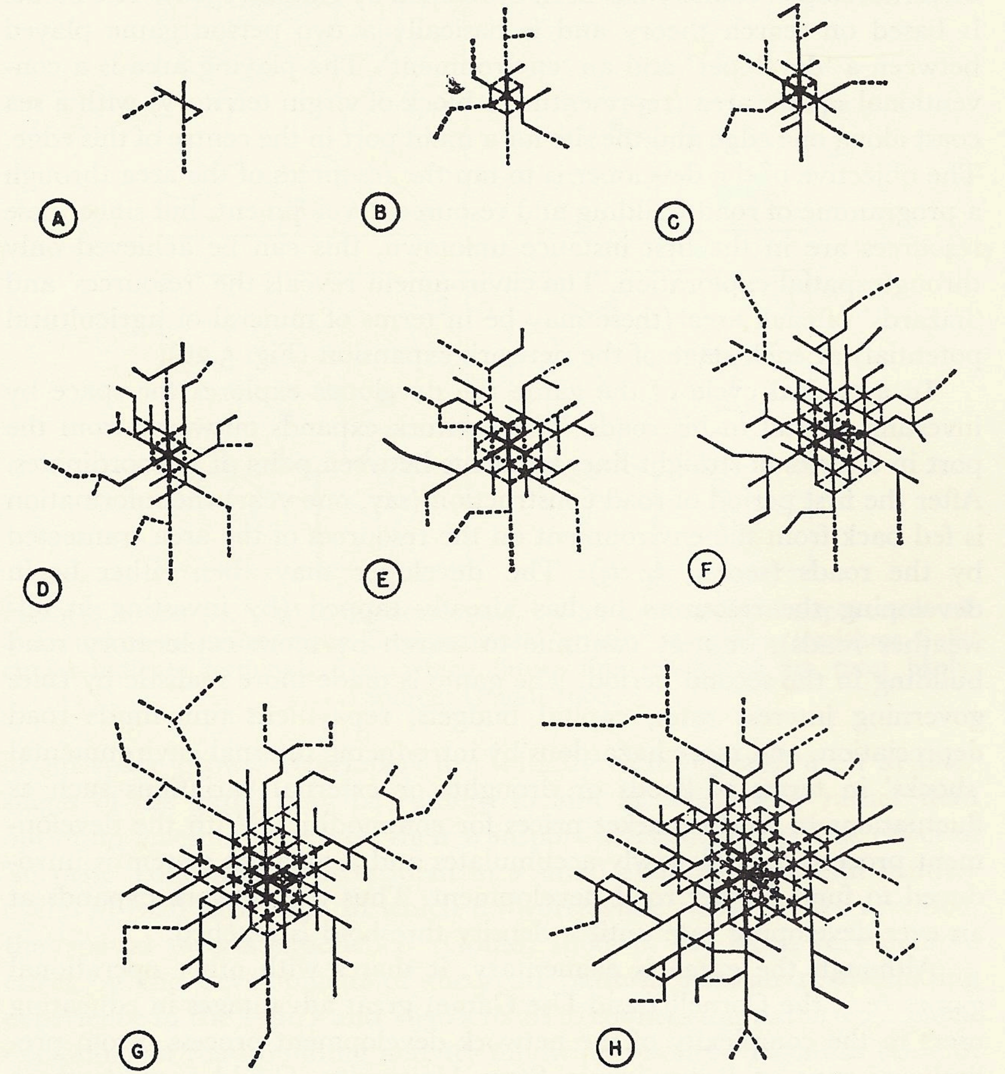


Fig. 5.25. Simulation of road network around a central node by Monte Carlo procedures. Source: Morrill, 1965.

- Long standing, interdisciplinary effort !
  Quantitative geography (1960s)
- Morrill (1965)
  Stochastic model of road network evolution
- Historical approaches
  - Cellular-automaton
  - Percolation, DLA
  - Urban economics

Science and cities: state of the art

Number of parameters

Urban economics (physics also):

Very abstract models, empirical tests ?

Applicability of the model ?

Complex simulations (LUTI models):

Validity ? Large (random) perturbation ?

Minimal model: the smallest number of parameters and the largest number of verified predictions

Loop: theory-empirical data

Machine learning: still a black box…

(Urban) Economics and real-world systems

- **Equilibrium**: systems such as cities are not in equilibrium (existence of many time scales)

- **Utility**: Choice usually impacts the analytical form of functions… You cannot measure an utility but usually the outcome of a theory

- **Decision process**: rationality is not driving all our decisions. A large number of factors, and large fluctuations among individuals…

- **Urban systems**: Monocentric assumption, homogeneous infrastructure, complete social graph, etc.

Towards a (new) science of cities

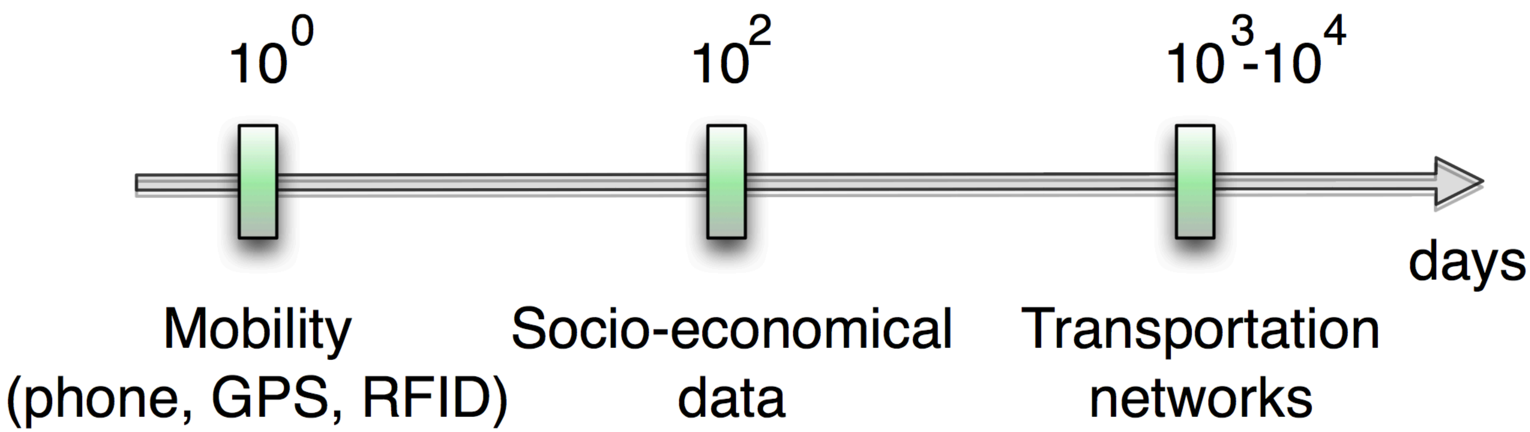
- "The physicist point of view":
  - 1. Get the data
  - 2. Extract useful information; ask (interesting) questions
  - 3. Propose a hierarchy of mechanisms
  - 4. Propose a model, extract predictions
  - 5. Compare predictions with data if not ok go back to 3 and 4

Note 1: the model should be with the smallest number of parameters and able to reproduce a large number of empirical facts

Note 2: Modeling the city, or an aspect of the city ?

Towards a (new) science of cities

- Game changer ? Always more data about cities !
- Different scales, different phenomena



OD matrix;
spatial structure
of cities (polycenters)

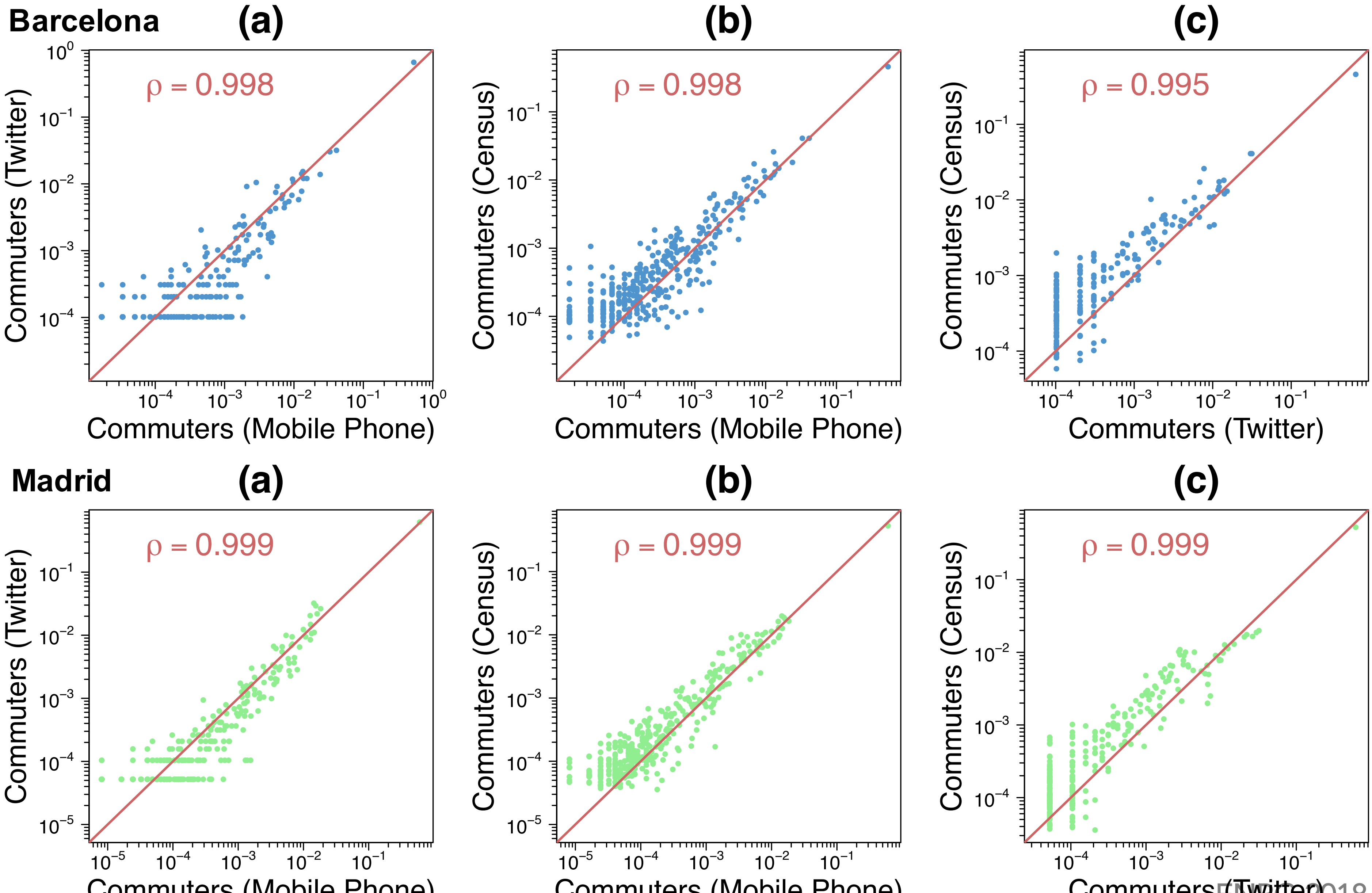
Scaling; Effect
of income
(segregation,
commuting)

Evolution of infrastructure
networks
(roads and transportation)

New data

- New datasets (mobile phone, GPS, RFIDs, etc) with detailed (real-time) origin-destination matrix, allow to:
  - Answer old questions: statistics of OD matrices; choice of trips; relation with socio-economical factors (income).
  - Ask new ones: information about the city structure (coarse-graining the OD matrix); relation mobility-social network; familiar strangers effect, etc.

Cross-checking different sources of mobility informat
(LeNormand et al. 2014, *arxiv.org:1404.0333*)



1. Mesoscale information from mobile phone data

How can we extract useful information from new data ?

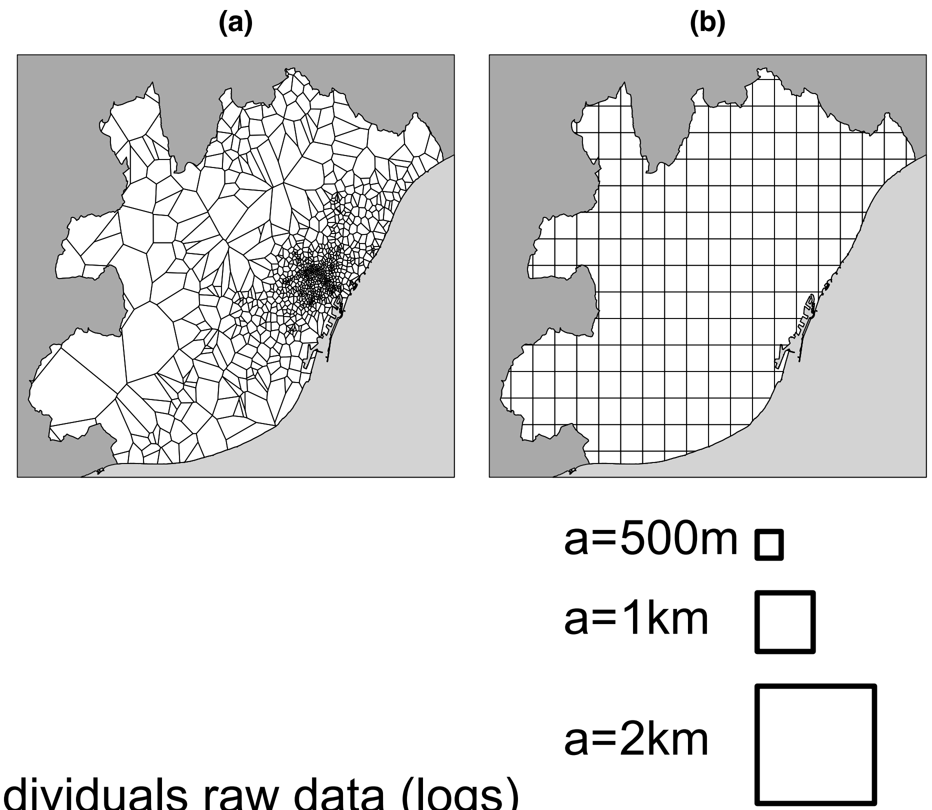
Mobile phone data

- Aggregated
  - For all antennas, at any time: number of mobile phones
  - No mobility information

- Individual data logs
  - Allow to ‘track’ individuals
  - Contains the OD matrix

```
id-caller|id-receiver|date|duration|op-caller|op-receiver|zone-start|zone-end|
9590|9899|01/10/2015 04:29:05|136|Other|Telefonica|A|A
9590|9899|01/10/2015 04:33:18|88|Other|Telefonica|A|A
9590|9899|01/10/2015 04:59:06|21|Other|Telefonica|A|C
9001|9899|01/10/2015 06:33:30|33|Other|Telefonica|B|D
9086|9875|01/10/2015 02:05:51|58|Other|Telefonica|C|C
…
```

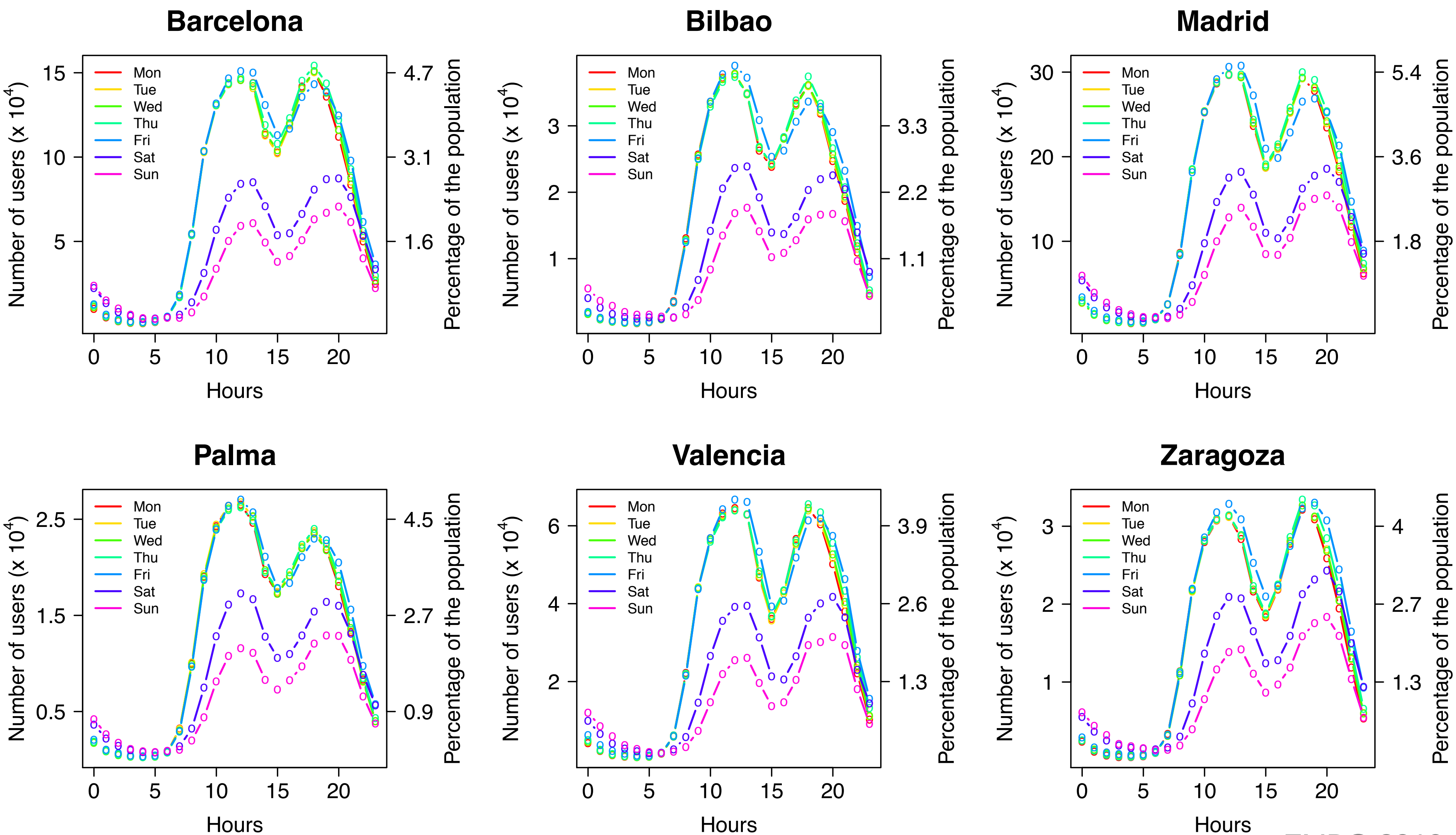
Data : two months of usage data from mobile phones in (31) Spanish urban areas



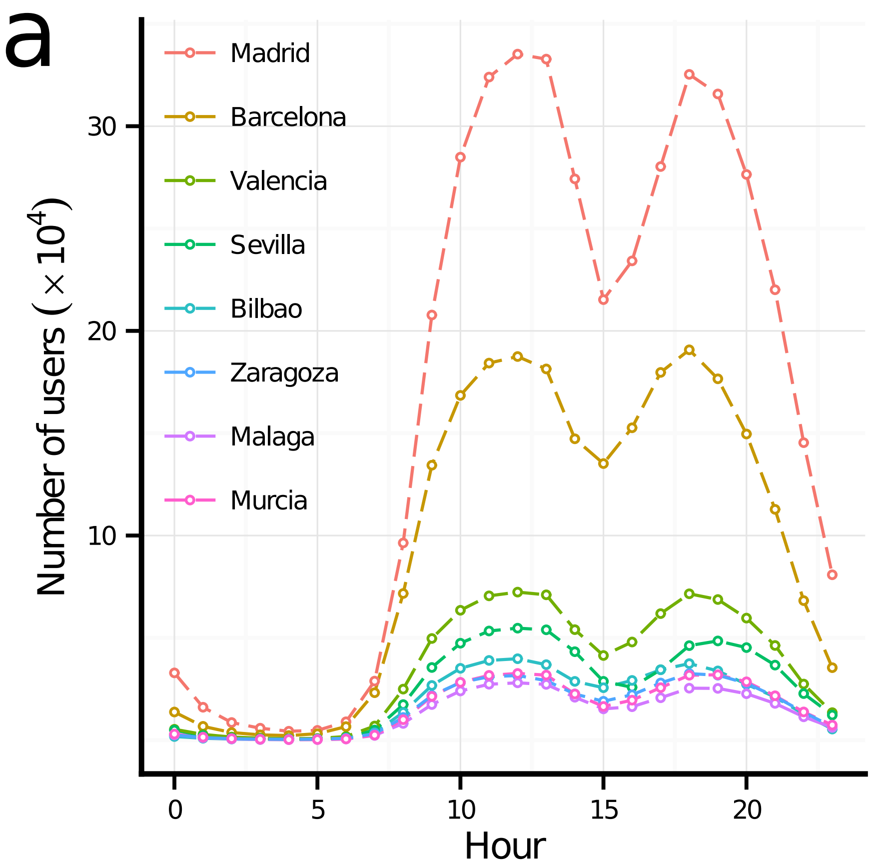
Individuals raw data (logs)

```
id-caller|id-receiver|date|duration|op-caller|op-receiver|zone-start|zone-end|
9590|9899|01/10/2015  04:29:05|136|Other|Telefonica|A|A
9590|9899|01/10/2015  04:33:18|88|Other|Telefonica|A|A
9590|9899|01/10/2015  04:59:06|21|Other|Telefonica|A|C
9001|9899|01/10/2015  06:33:30|33|Other|Telefonica|B|D
9086|9875|01/10/2015  02:05:51|58|Other|Telefonica|C|C
...
```

Mobile phone activity vs. time



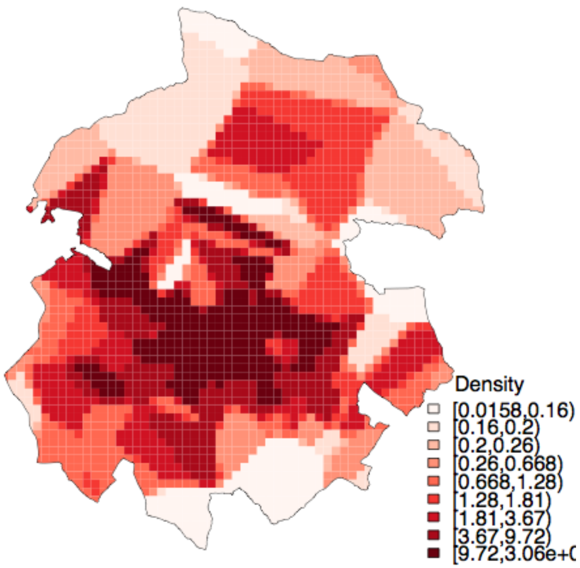
A common urban rhythm



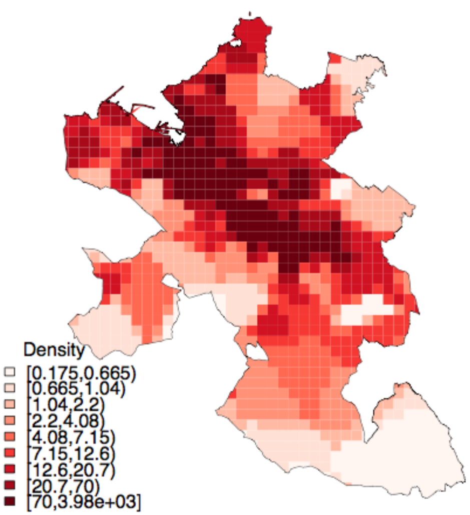
Space time varying density

- State of the art: how to represent a time and space varying density
  - 1. A first approach with “Hotspots” or “Activity centers”: an important concept in urban studies and spatial economics
  - 2. Weighted quantities (Venables index)

Zaragoza urban area



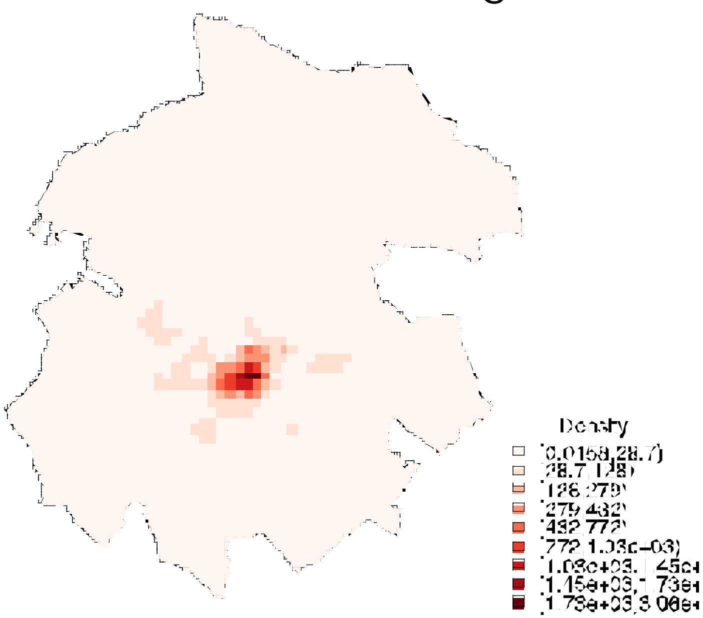
Bilbao urban area



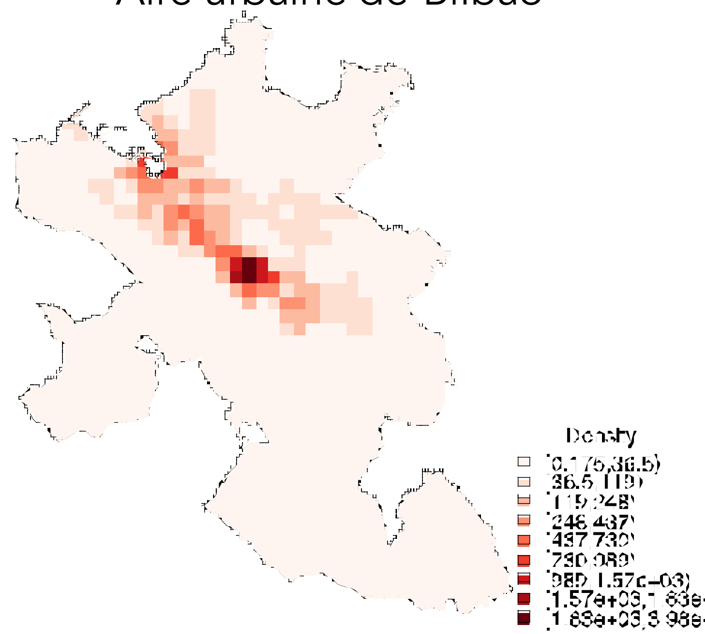
Hotspots: local maxima of density

City structure (mono- vs. polycentric)

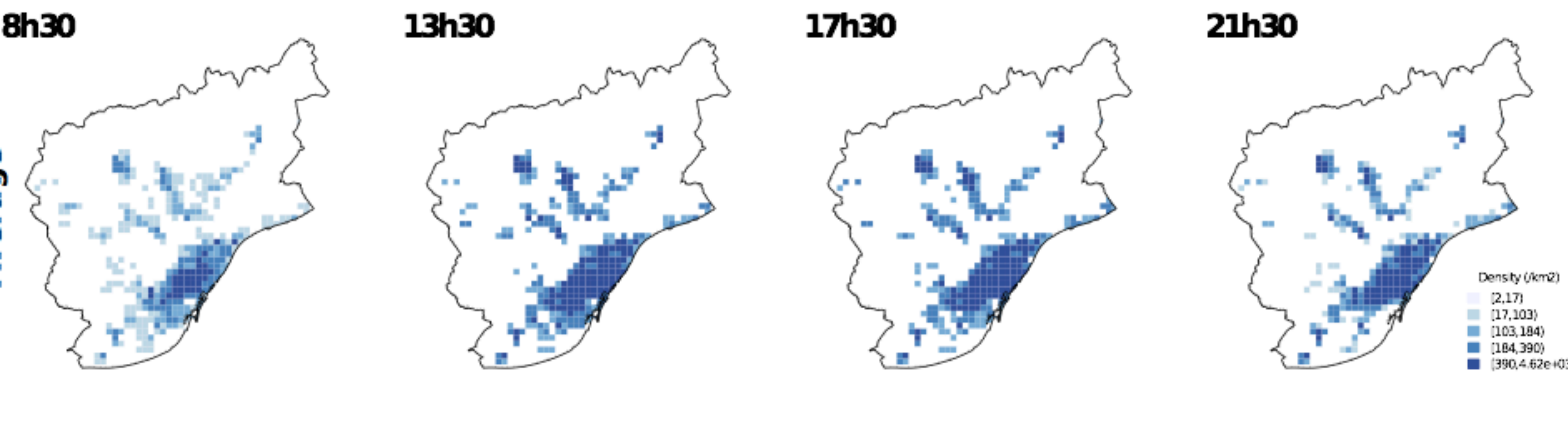
Aire urbaine de Zaragoza



Aire urbaine de Bilbao



Hotspots location
Example : the urban area of Barcelona



Hotspot identification

- State of the art
  - No clear method
  - Density larger than a given threshold is a hotspot
  - Problem of the threshold choice ?

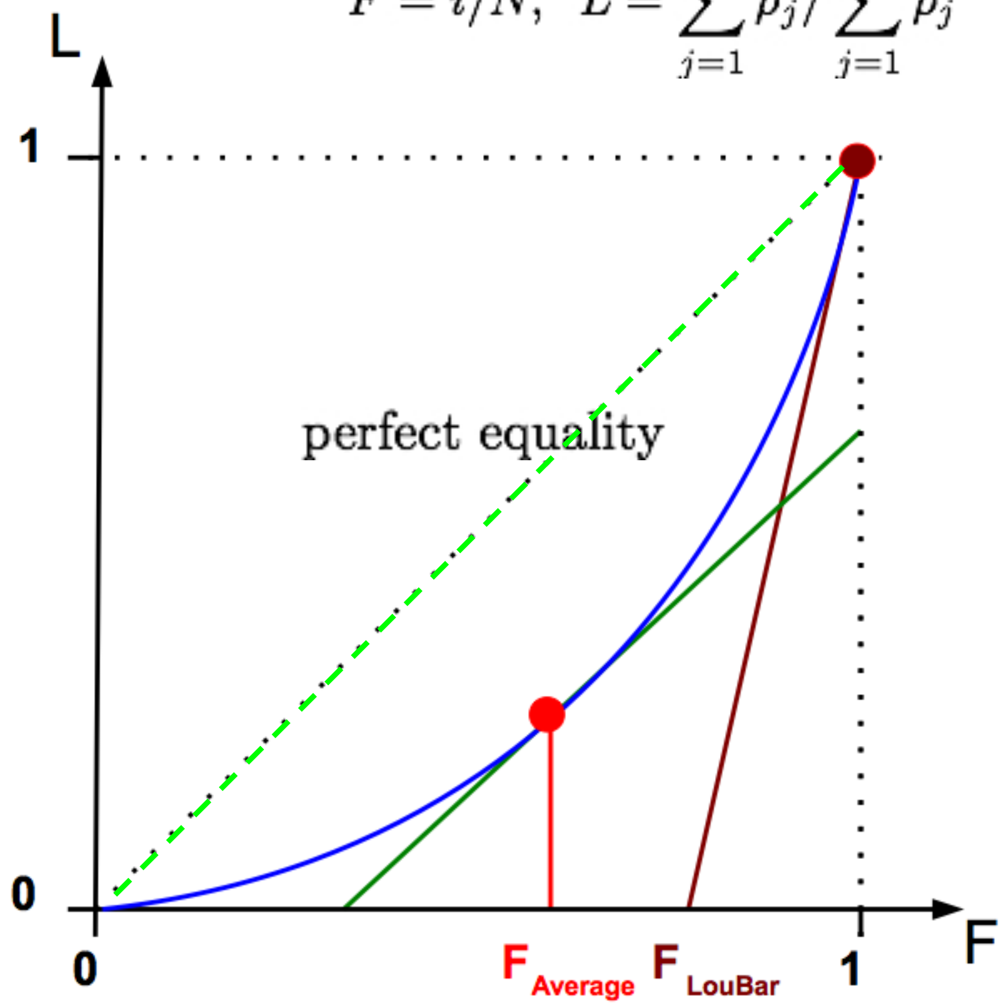
$$i \text{ hotspot } \Leftrightarrow \rho_i > \delta$$

Louail, et al, Sci. Rep. 2014

Hotspot identification

- A simple approach
  - Discussion on the Lorentz curve
  - Identify a lower and upper threshold

$$F = i/N, \quad L = \sum_{j=1}^{i} \rho_j / \sum_{j=1}^{N} \rho_j$$



$$F_{LouBar} = 1 - \frac{\bar{\rho}}{\rho_{max}}$$

Louail, et al, Sci. Rep. 2014

Numbers of hotspots vs. population size of the city

We can now count the hotspots:

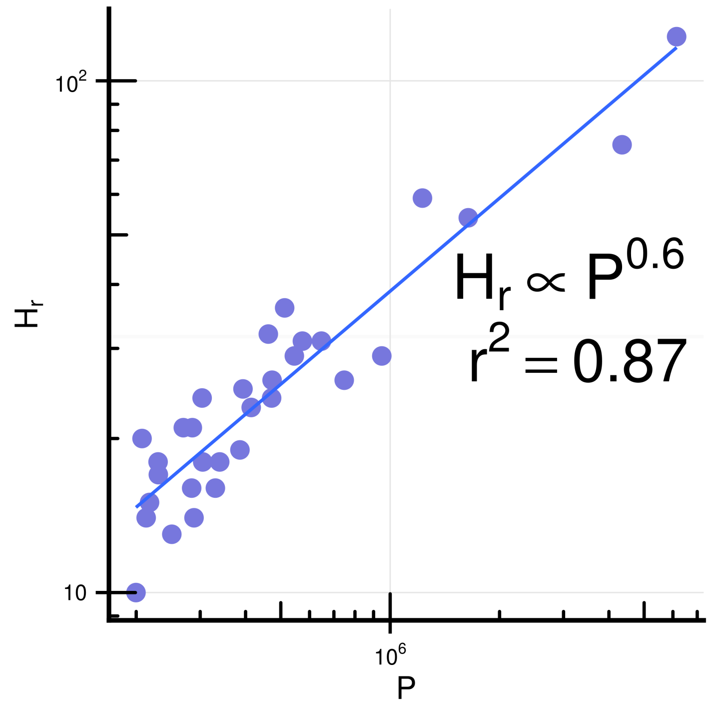
- residential hotspots
- activity hotspots

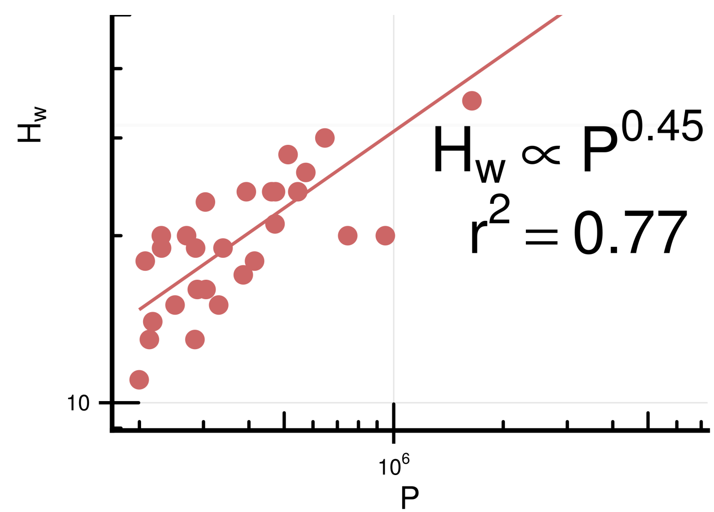
Numbers of hotspots vs. population size of the city

Exponent value is remarkably smaller for work/school/daily activity hotspots

→ in Spanish urban areas, the number of activity places grows slower than the number of major residential places.

Sublinear in both cases !!!

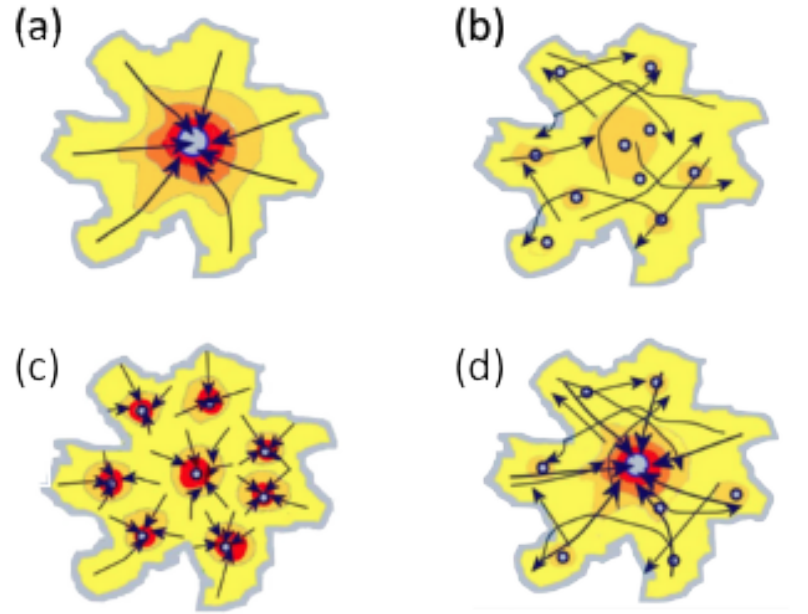




Typology of mobility patterns (journey to work trips)

Motivation:
Compare the spatial structure of mobility patterns in many cities

Question:
How to build a quantitative typology of cities based on the spatial structure of the mobility patterns ?

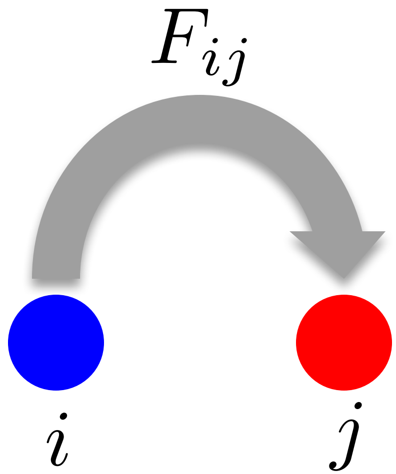


(Bertaud & Malpezzi 2003)

How to compare OD commuting matrices of different cities?

- The OD matrix is a large and complicated object

- Difficult to compare different cities !
  - Different sizes
  - Potentially different spatial resolutions



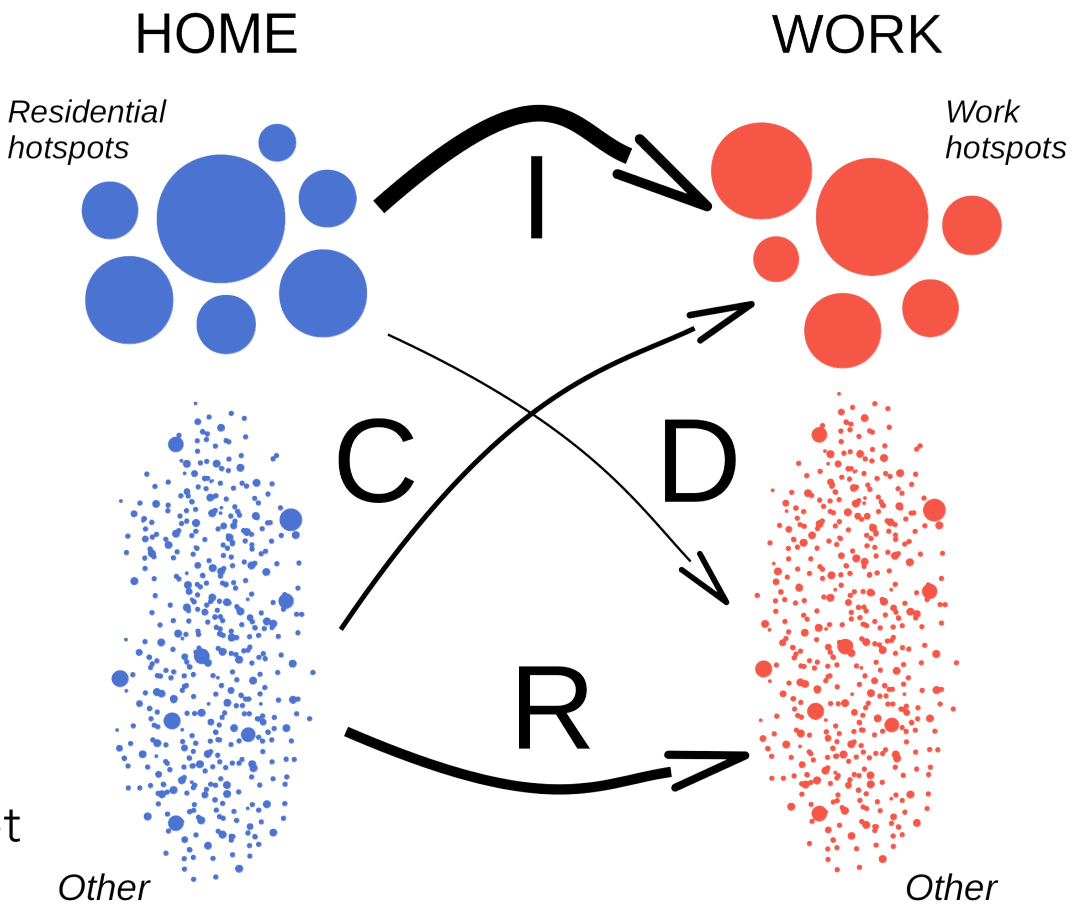
- We need a **simpler**, **clearer** picture: coarse-grained information

How to compare OD commuting matrices of different cities?

1. Determine Residential and work hotspots (Louail et al, 2014)

2. Separate 4 categories of flows: I, C, D, R

Integrated: Hotspot->Hotspot
Convergent: Non hotspot->hotspot
Divergent: Hotspot->non hotspot
Random: non hotspot->non hotspot



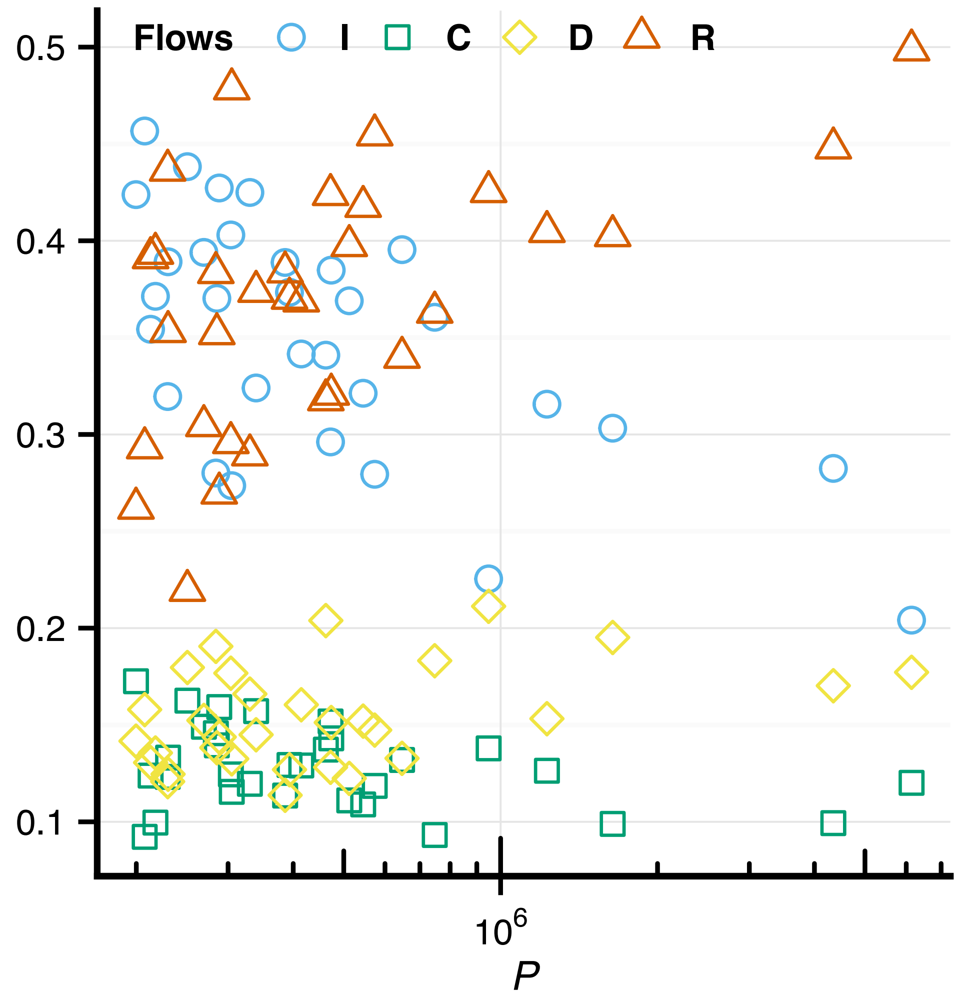
Louail, et al, Nature Comms 2015

Structure of flows versus population
(30 largest urban areas in Spain)

The importance of **I**ntegrated flows decreases when population size increases, in favor of an increase of "**R**andom" flows

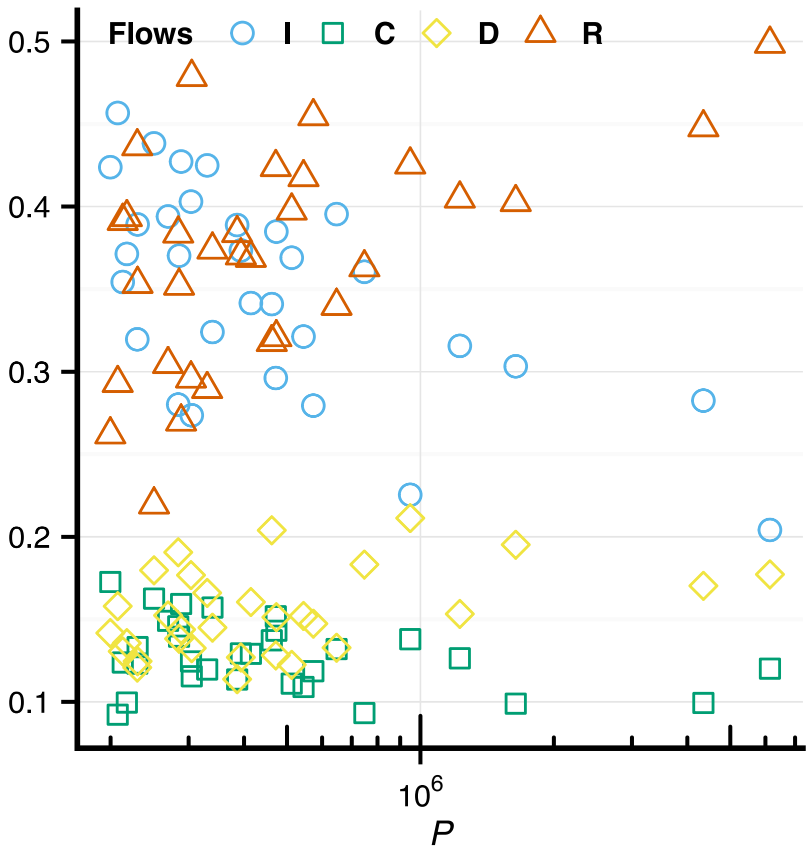
Weights of **D**ivergent and **C**onvergent flows are constant

I and R alone seem enough to characterize cities

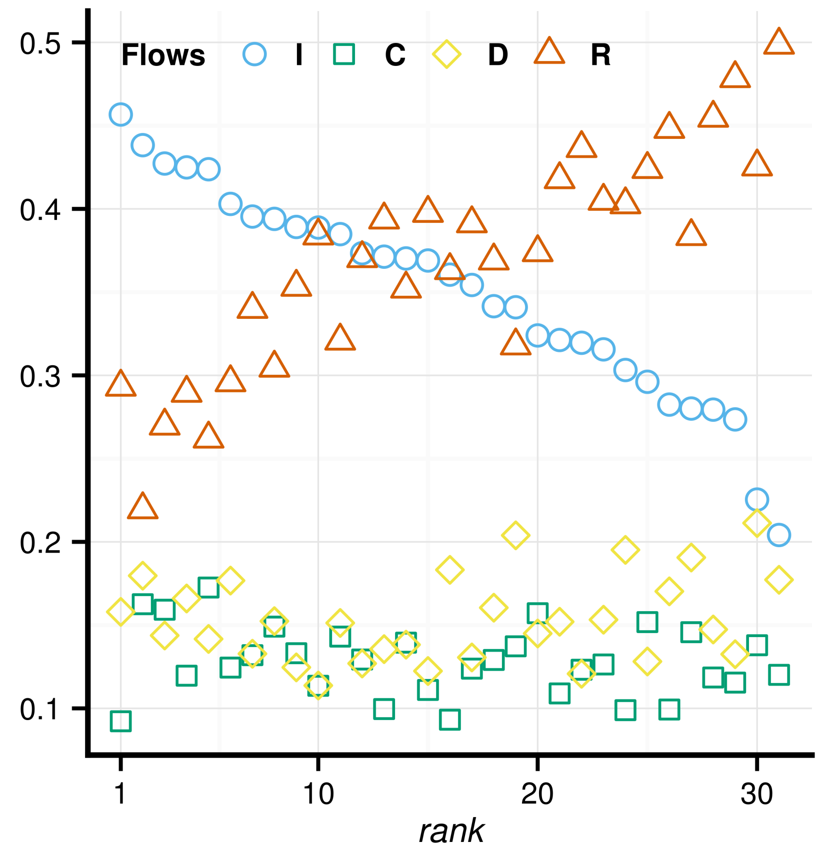


Louail, et al, Nature Comms 2015

ICDR vs. Population size



ICDR ranked by decreasing I



The importance of **I**ntegrated flows seems to decrease when population size increases, in favor of an increase of "**R**andom" flows

Weights of **D**ivergent and **C**onvergent flows seem constant whatever the city size

I and **R** alone seem sufficient to classify cities

Structure des flots (Espagne)

R	27%	36%	41%	46%
I	43%	37%	31%	25%

Vient des possibilité plus grandes dans les grandes villes de se deplacer (?)

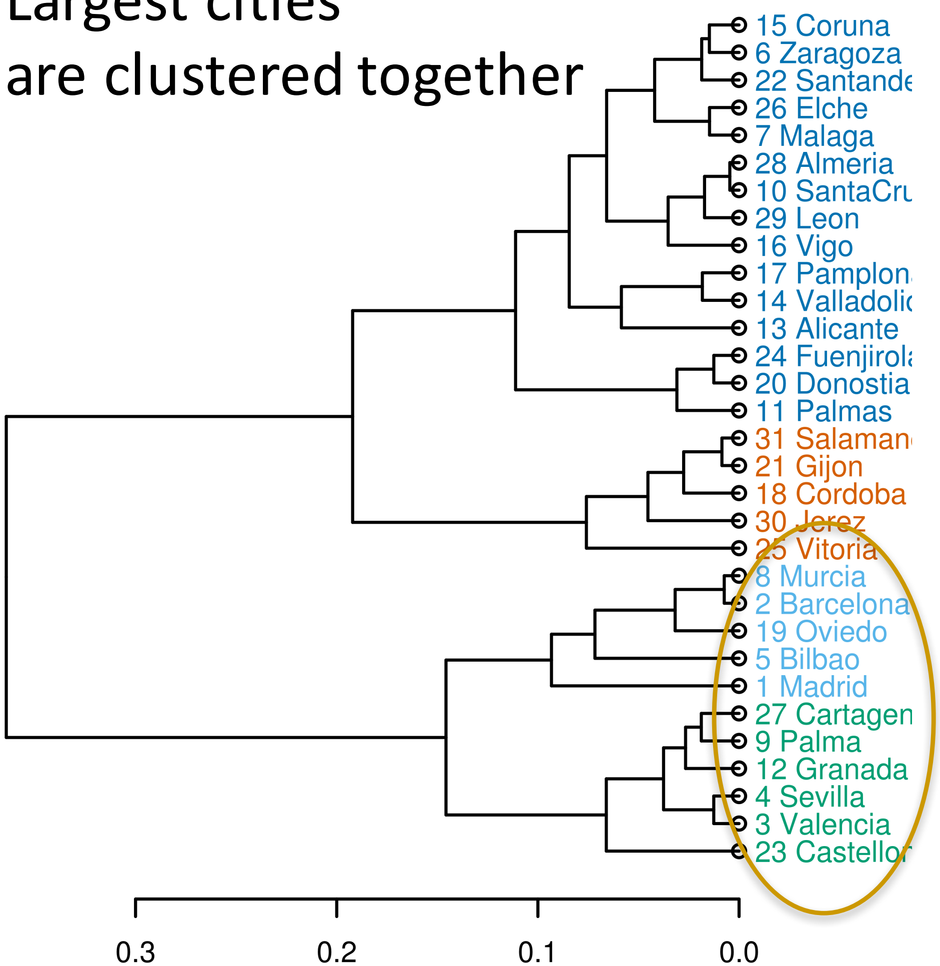
Structure spatiale "délocalisée" des grandes villes

Hierarchical clustering of cities based on their I, C, D, R values

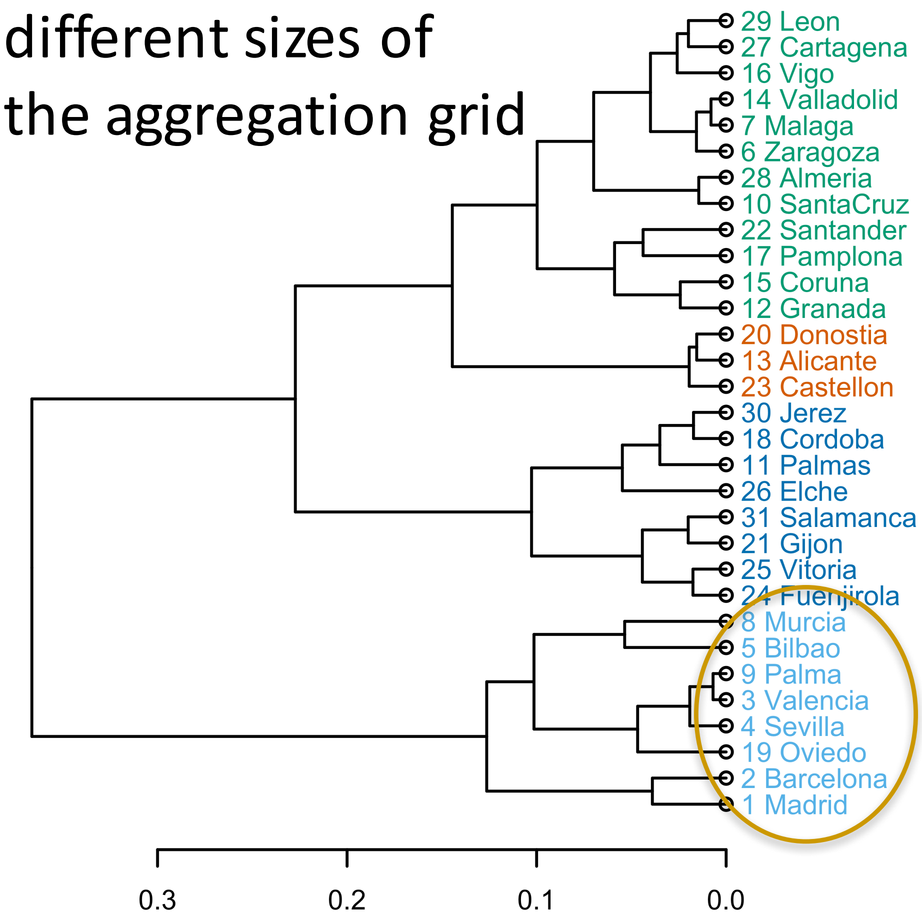
1km * 1km grid

2km * 2km grid

Largest cities are clustered together



Robust with different sizes of the aggregation grid



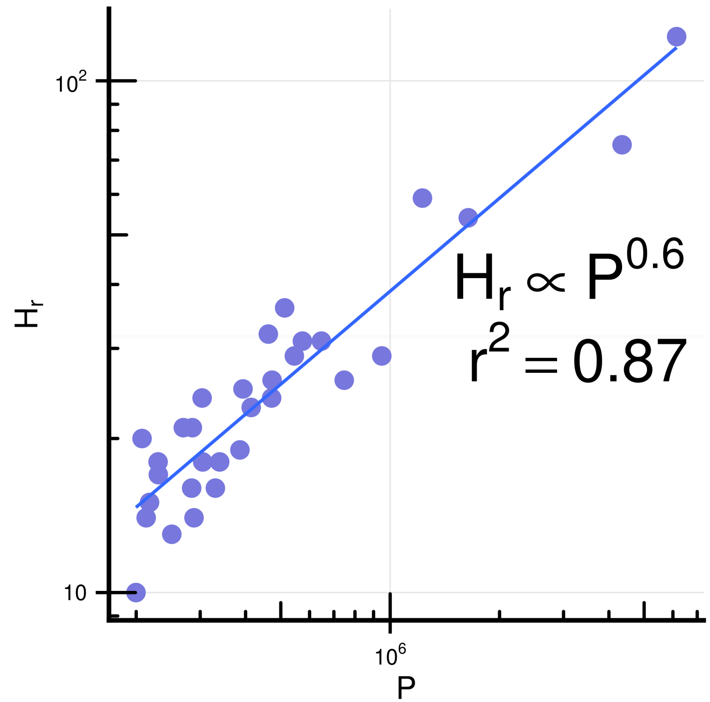
II. Understanding the polycentric structure

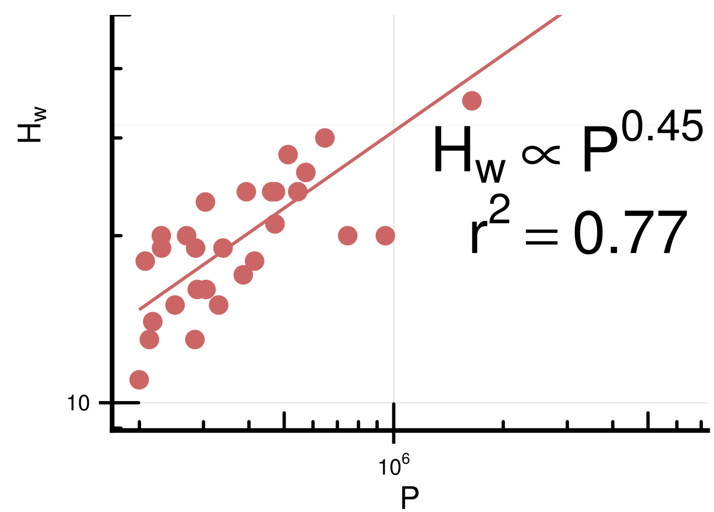
Numbers of hotspots vs. population size of the city

Exponent value is remarkably smaller for work/school/daily activity hotspots

→ in Spanish urban areas, the number of activity places grows slower than the number of major residential places.

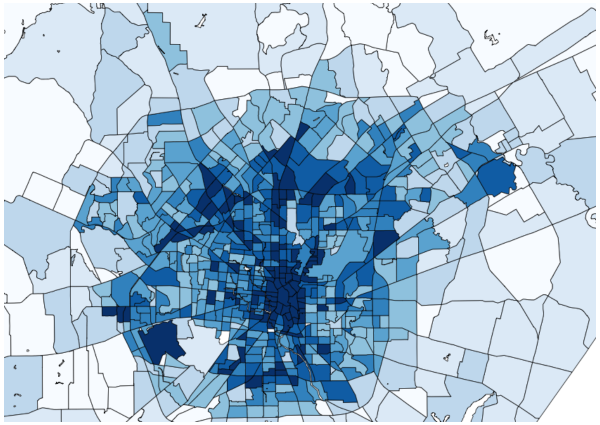
Sublinear in both cases !!!



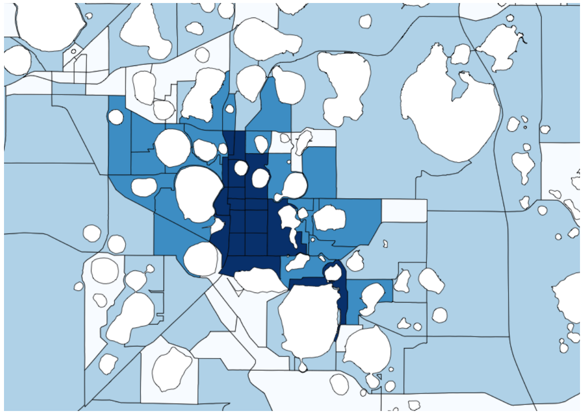


Polycentric structure

- Activity centers (# of employees per zip code, USA)



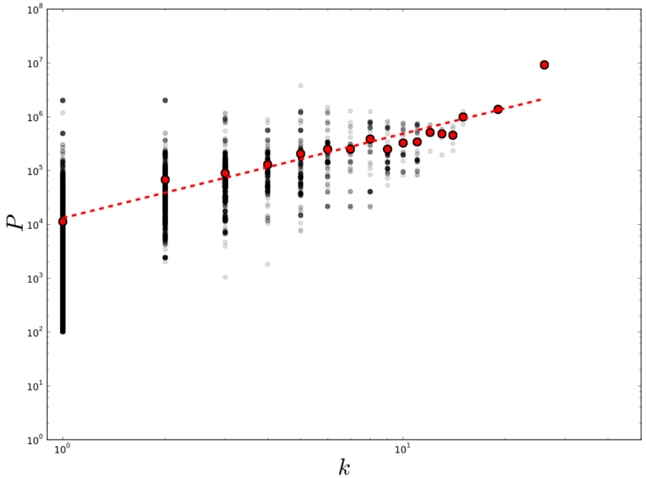
San Antonio (TX), USA



Winter Haven (FL), USA

- For each city, we can count the secondary centers (9000 cities US, 1994-2010)

$$k \sim P^{\beta} \quad \beta \simeq 0.64$$



Polycentric structure

- We have a polycentric structure, evolving with P
- We can count the number H of centers

$$H \sim P^{\beta} \quad \beta \approx 0.5 - 0.6$$

- Mobility is the key: we need to model how individuals choose their home and work place
- Problem largely studied in geography, and in spatial economics: Edge-city model (Krugman), Fujita-Ogawa model (1982): utility maximization
- Revisiting Fujita-Ogawa: predicting the value of β

Mobility is the key

- We have a polycentric structure, evolving with P

- We need to model how individuals move from home to work

  - Connected to the spatial structure of the city

  - Once known allows to compute all mobility/transport related quantities

Spatial economics: the edge city model (Krugman 1996)

- The important ingredient is the ‘market potential’

$$\Pi(x) = \int K(x-z)\rho(z)\mathrm{d}z$$

- Describes the spillovers due to the density in z
- Specifically

$$K(u) = A(u) - B(u)$$

- The average market potential is

$$\overline{\Pi} = \frac{1}{\Omega}\int \Pi(x)\rho(x)\mathrm{d}x$$

Spatial economics: the edge city model (Krugman 1996)

- The equation for the evolution of business density is

$$\frac{\mathrm{d}\rho}{\mathrm{d}t} = \gamma\left(\Pi(x) - \overline{\Pi}\right)$$

- Linearize around flat situation $\rho(x) = \rho_0 + \delta\rho(x)$

$$\delta\rho(k) \sim \mathrm{e}^{\gamma K(k) t}$$

- At least one maximum at k=k*; the number of hotspots is then:

$$H \sim \Omega k_*^2$$

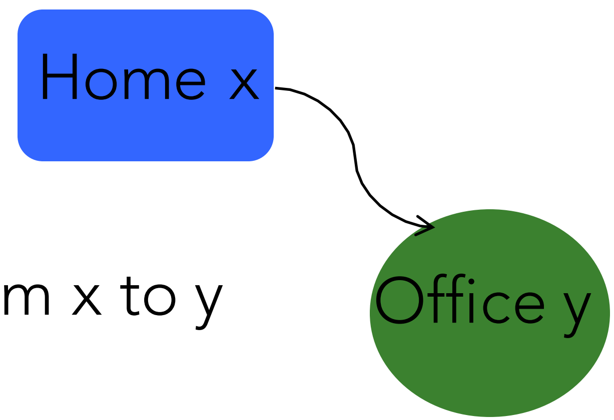
- Scaling with the population ?
- Link micromotives-macrobehavior ?

Spatial economics: Fujita-Ogawa (1982)

- A model for the spatial structure of cities: an agent will choose to live in *x* and work in *y* such that

$$Z_0(x, y) = W(y) - C_R(x) - C_T(x, y)$$

is maximum



- W(y) is the wage at y
- C_R(x) is the rent at x
- C_T(x,y) is the transportation cost from x to y [proportional to d(x,y)]

$$C_T(x, y) = td(x, y)$$

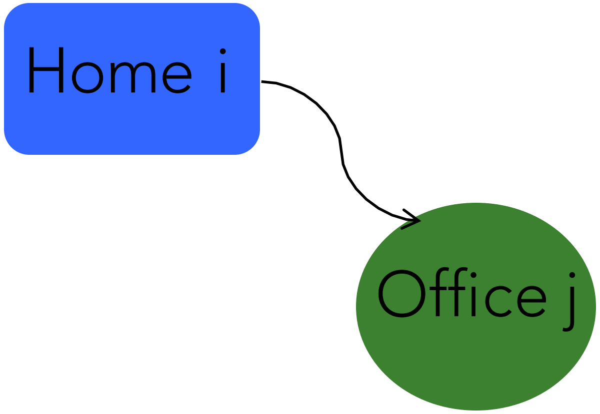
Spatial economics: Fujita-Ogawa (1982)

- And a similar equation for companies (maximum profit)

$$P(y) = \Pi(y) - C_R(y) - L(y)W(y)$$

- W(y) is the wage at y
- $C_R(y)$ is the rent at y
- L(y) number of workers ($N=ML_0$)
- **Π**(y) is the benefit to come to y:

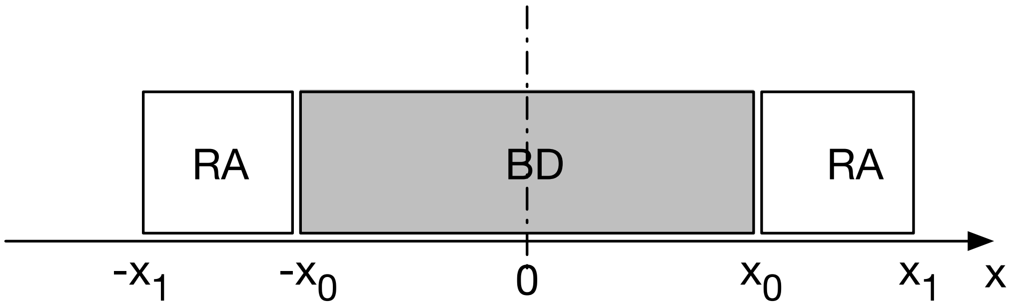
Agglomeration effect !



$$\Pi(y) = \int K(y-z)\rho(z)\mathrm{d}z$$

$$K(u) = k\mathrm{e}^{-\alpha|u|}$$

Spatial economics: Fujita-Ogawa (1982)



- Main result: monocentric configuration stable if

$$\frac{t}{k} \leq \alpha$$

  - t: transport cost
  - 1/α interaction distance between firms

- Effect of congestion: larger cost t

Spatial economics: Fujita-Ogawa (1982)

- This model is unable to predict the spatial structure and the number of activity centers....

- We have to simplify the problem !

Spatial economics: Fujita-Ogawa (1982)

- There are many problems with this model:
  - Not dynamical: optimization. We want an out-of-equilibrium model
  - No congestion (!) We want to include congestion (for car traffic)
  - No empirical test. Extract testable predictions (see the book: Spatial Economics, by Fujita, Krugman, Venables)

A physicist's variant of Fujita-Ogawa

- Assumptions and simplifications:
  - Assume that home is uniformly distributed (x): find a job !

$$Z_0(x,y) = W(y) - C_T(x,y)$$

  - We have now to discuss W and C_T

A physicist's variant of Fujita-Ogawa

- Assumptions and simplifications:

  - Add congestion (BPR function, t=cost/distance) and the generalized cost reads:

$$C_T(i,j) = t d_{ij} \left[1 + \left(\frac{T(j)}{c} \right)^{\mu} \right]$$

  - Wages: a typical physicist assumption (s: typical salary)

$$W(j) = s\eta_j$$

The 'attractivity' η is random (in [0,1]) (cf. Random Matrix Theory)

Note: length scales

- We have t: transportation cost per unit distance
- We have s: salary scale

$$\ell = s/t$$ (usually large: ~10^2-10^3kms)

A new length: effective commuting distance financially sustainable

=>No naive scaling…

$$\sout{L_{tot} \propto \sqrt{A}}$$

$$L_{tot} = \sqrt{A} F(\frac{\ell}{\sqrt{A}})$$

Summary: the model

- Every time step, add a new individual at a random i

- The individual will choose to work in *j* (among N_c possible centers) such that

$$Z(i,j) = \eta_j - \frac{d_{ij}}{\ell}\left[1 + \left(\frac{T(j)}{c}\right)^{\mu}\right]$$

is maximum

- $W(j)$ is the wage at j --> random

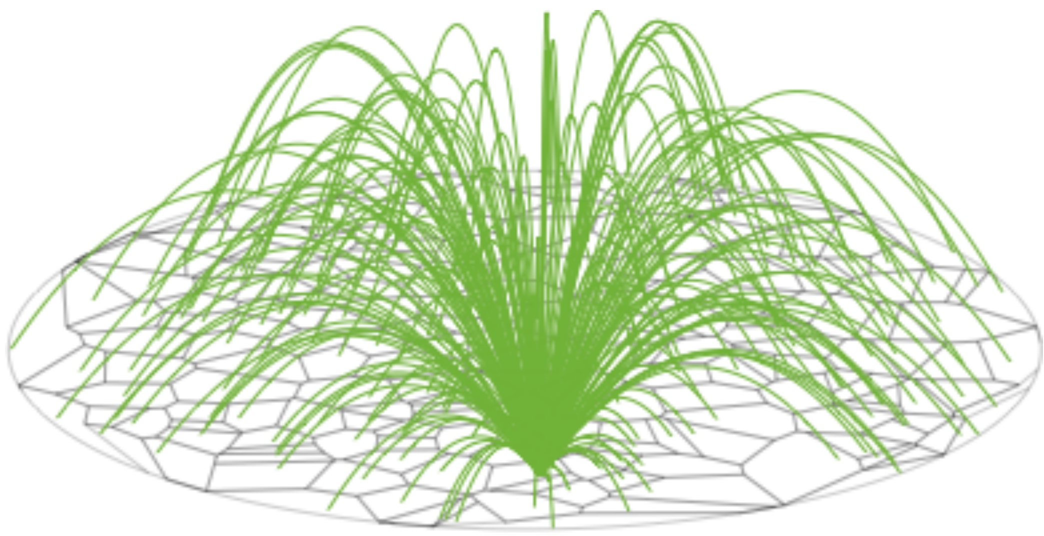
- $C_T(i,j)$ is the transportation cost from i to j: depends on the traffic from i to j --> congestion effects

Louf, MB, PRL 2013

Results

- Depending on the values of parameters, we see three type of mobility patterns:

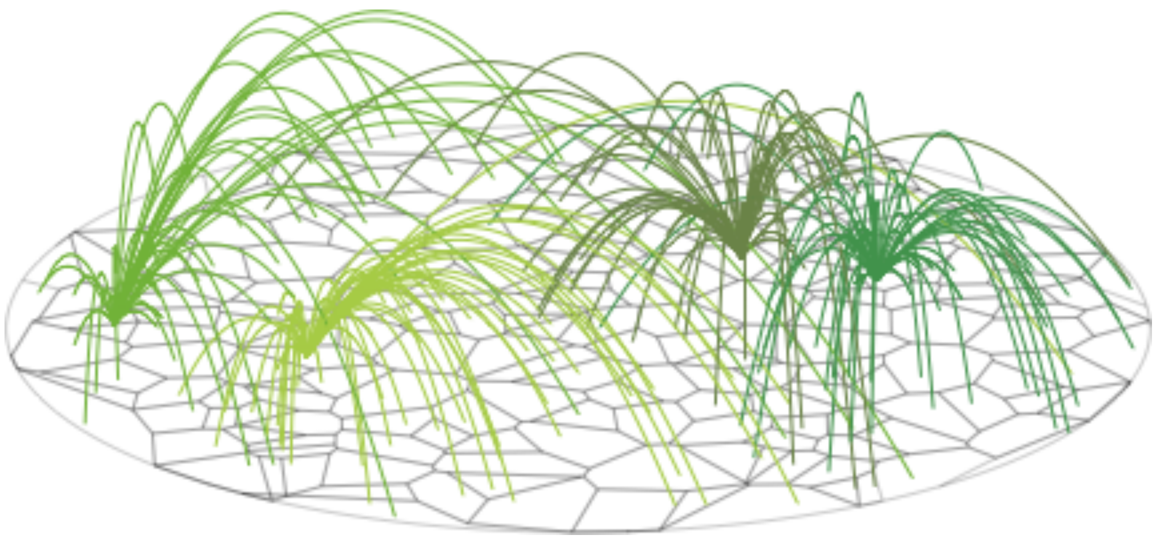
  1. Monocentric: one activity center



Results

- Depending on the values of parameters, we see three type of mobility patterns:

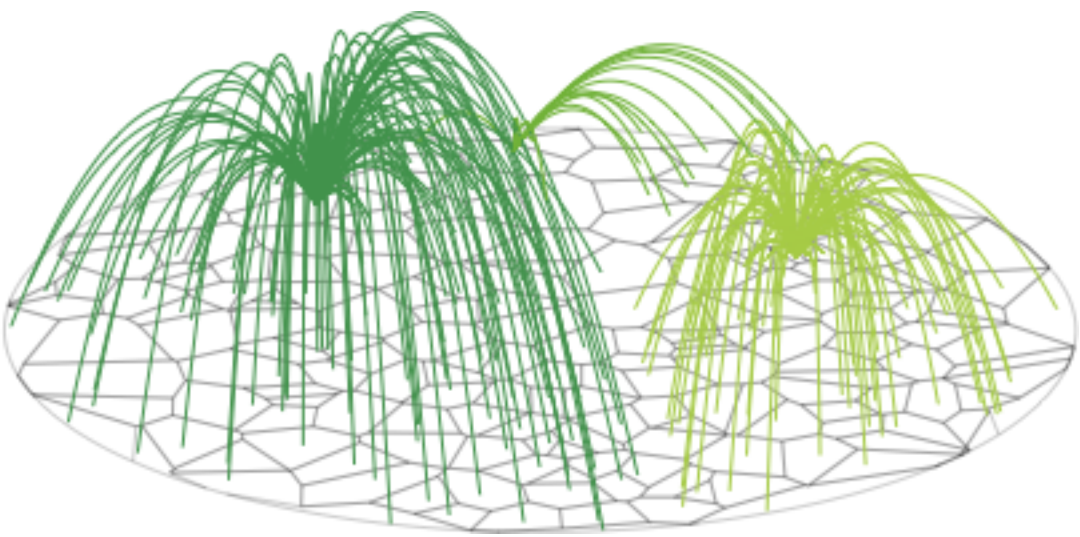
  2. Attractivity driven polycentrism: many activity centers, attractivity η dominates



Results

- Depending on the values of parameters, we see three type of mobility patterns:

  3. Spatial polycentrism: many activity centers, basins spatially coherent



Monocentric-polycentric transition

- Start with one center $\eta_1 > \eta_j$
- All other subcenters have a zero traffic T(j)=0
- The number of individuals P increases, T(1) increases and at a certain point there is another j such that:

$$Z(i,j) > Z(i,1)$$

Or:

$$\eta_j - \frac{d_{ij}}{\ell} > \eta_1 - \frac{d_{i1}}{\ell}\left[1 + \left(\frac{P}{c}\right)^{\mu}\right]$$

Monocentric-polycentric transition

$$\eta_j - \frac{d_{ij}}{\ell} > \eta_1 - \frac{d_{i1}}{\ell}\left[1 + \left(\frac{P}{c}\right)^{\mu}\right]$$

- Mean-field type argument
  - $d_{i1} \sim d_{ij} \sim L \quad (L = \sqrt{A})$
  - The new subcenter has the second largest attractivity η_2
  - on average

$$\overline{\eta_1 - \eta_2} \simeq 1/N_c$$

- We obtain a 'critical' value for the population

$$P > P^{*} = c\left(\frac{\ell}{LN_c}\right)^{1/\mu}$$

Monocentric-polycentric transition

- Critical value for the population: effect of congestion !

$$P > P^* = c \left(\frac{\ell}{LN_c} \right)^{1/\mu}$$

- c sets the scale

- If ℓ is too small, P*<1 and the monocentric regime is never stable

Monocentric-polycentric transition

- If the population continues to increase, other subcenters will appear. We assume that for P, we have k-1 subcenters:

$$\eta_1 \geq \eta_2 \geq ... \geq \eta_{k-1}$$

with traffic: $$T(1) \sim T(2) \sim ...T(k-1) \sim \frac{P}{k-1}$$

- The next individual will choose a new subcenter k if:

$$Z_{ik} > \max_{1,..,k-1} Z_{ij}$$

$$\eta_k - \frac{d_{ik}}{\ell} > \max_{1,..,k-1}\{\eta_j - \frac{d_{ij}}{\ell}\left[1 + \left(\frac{T(j)}{c}\right)^{\mu}\right]\}$$

- We assume: $d_{ik} \sim d_{ij} \sim L$

Results: scaling for the number of centers

- We obtain the average population for which a k[th] subcenter appears is:

$$\overline{P}_k = P^*(k-1)^{\frac{\mu+1}{\mu}}$$

- Which implies:

$$k \sim \left(\frac{P}{P^*}\right)^{\frac{\mu}{\mu+1}}$$

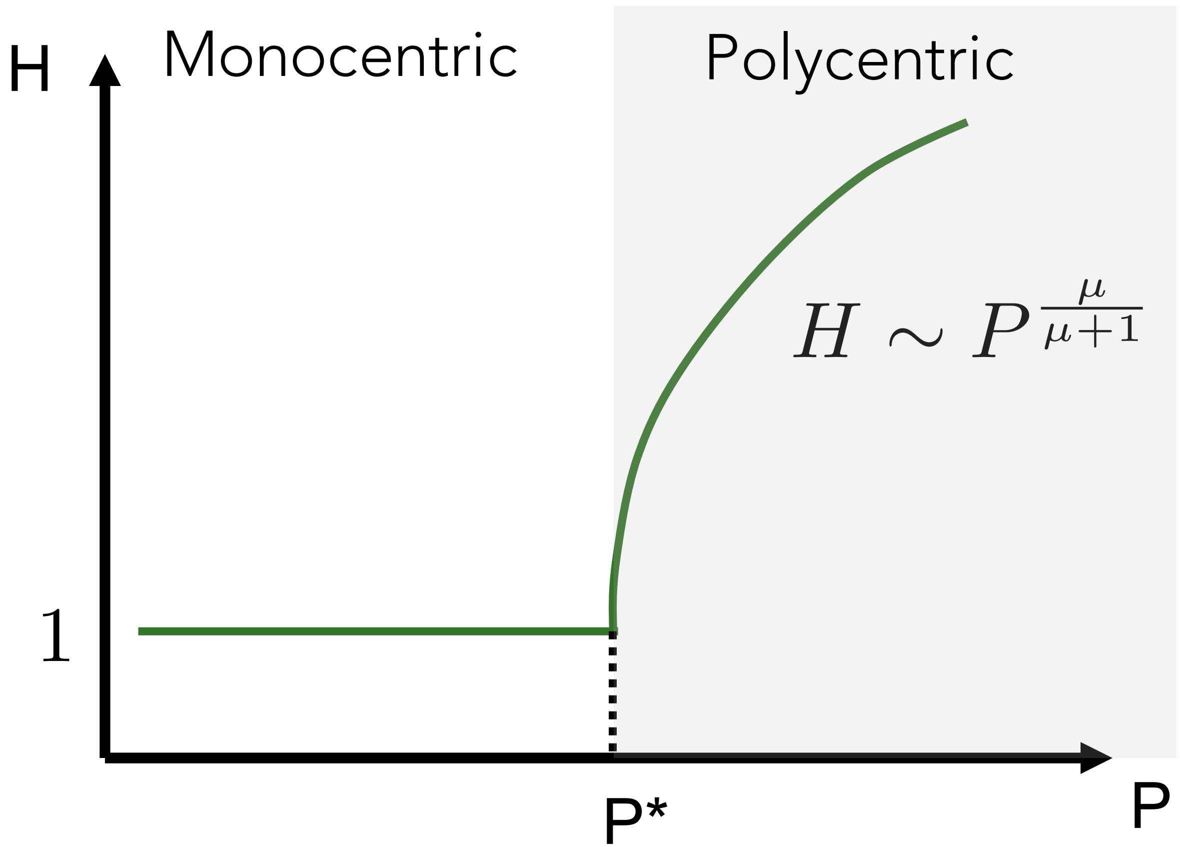
Sublinear relation !

- From US employment data (9000 cities)

$$k \sim P^{0.64} \qquad (\Rightarrow \mu \simeq 2)$$

'Urban transition: Phase diagram'

Number of hotspots H versus population P (Mean-Field analysis)



- From US employment data (9000 cities)

$$H \sim P^{0.64} \; (\Rightarrow \mu \simeq 2)$$

III. Scaling of socio-economical quantities

Scaling

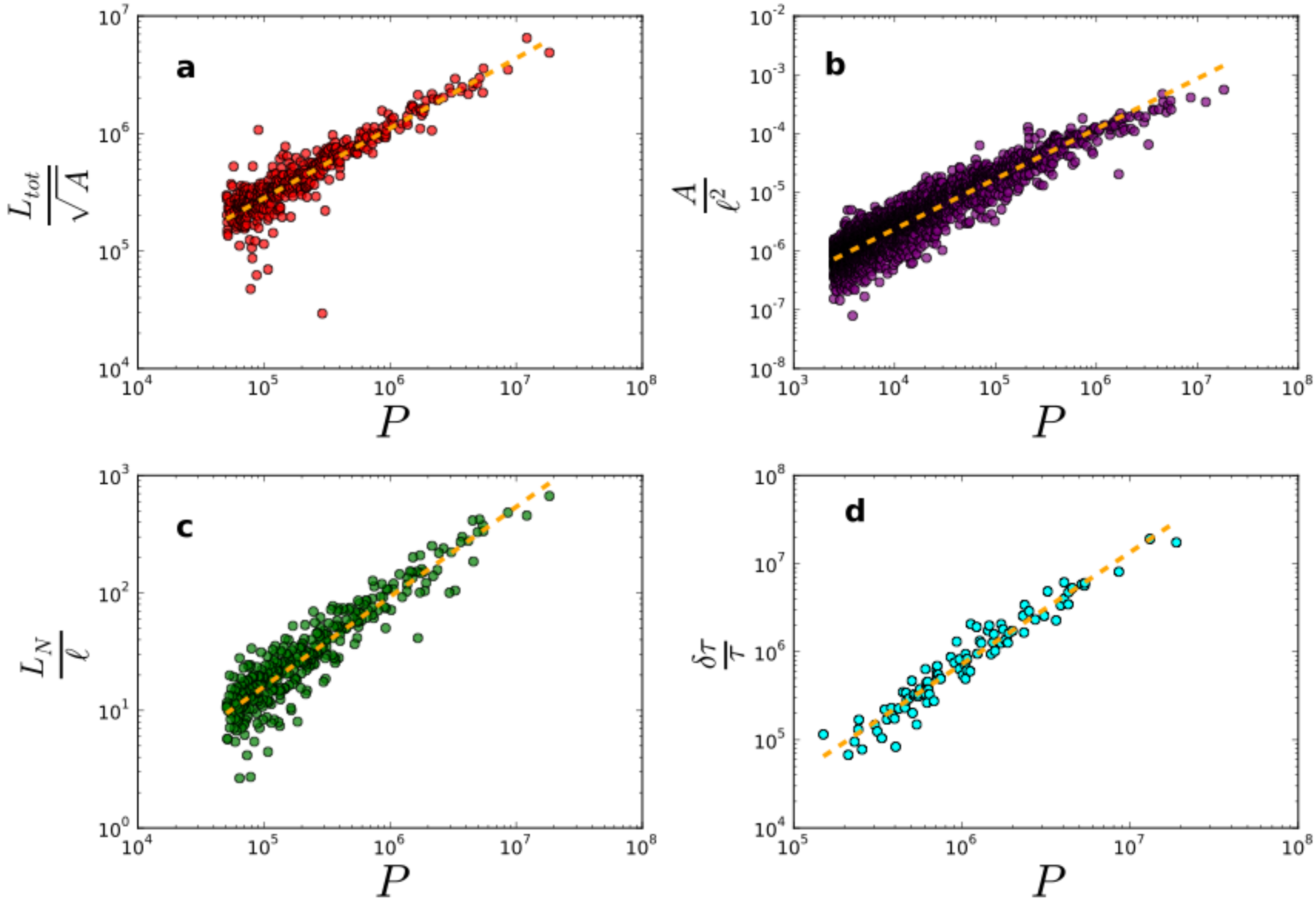
$$Y \sim P^{\beta}$$

Table 1. Scaling exponents for urban indicators vs. city size

Y	*β*	95% CI	Adj-R^2	Observations	Country–year
New patents	1.27	[1.25,1.29]	0.72	331	U.S. 2001
Inventors	1.25	[1.22,1.27]	0.76	331	U.S. 2001
Private R&D employment	1.34	[1.29,1.39]	0.92	266	U.S. 2002
"Supercreative" employment	1.15	[1.11,1.18]	0.89	287	U.S. 2003
R&D establishments	1.19	[1.14,1.22]	0.77	287	U.S. 1997
R&D employment	1.26	[1.18,1.43]	0.93	295	China 2002
Total wages	1.12	[1.09,1.13]	0.96	361	U.S. 2002
Total bank deposits	1.08	[1.03,1.11]	0.91	267	U.S. 1996
GDP	1.15	[1.06,1.23]	0.96	295	China 2002
GDP	1.26	[1.09,1.46]	0.64	196	EU 1999–2003
GDP	1.13	[1.03,1.23]	0.94	37	Germany 2003
Total electrical consumption	1.07	[1.03,1.11]	0.88	392	Germany 2002
New AIDS cases	1.23	[1.18,1.29]	0.76	93	U.S. 2002–2003
Serious crimes	1.16	[1.11, 1.18]	0.89	287	U.S. 2003
Total housing	1.00	[0.99,1.01]	0.99	316	U.S. 1990
Total employment	1.01	[0.99,1.02]	0.98	331	U.S. 2001
Household electrical consumption	1.00	[0.94,1.06]	0.88	377	Germany 2002
Household electrical consumption	1.05	[0.89,1.22]	0.91	295	China 2002
Household water consumption	1.01	[0.89,1.11]	0.96	295	China 2002
Gasoline stations	0.77	[0.74,0.81]	0.93	318	U.S. 2001
Gasoline sales	0.79	[0.73,0.80]	0.94	318	U.S. 2001
Length of electrical cables	0.87	[0.82,0.92]	0.75	380	Germany 2002
Road surface	0.83	[0.74,0.92]	0.87	29	Germany 2002

Bettencourt et al, PNAS 2007

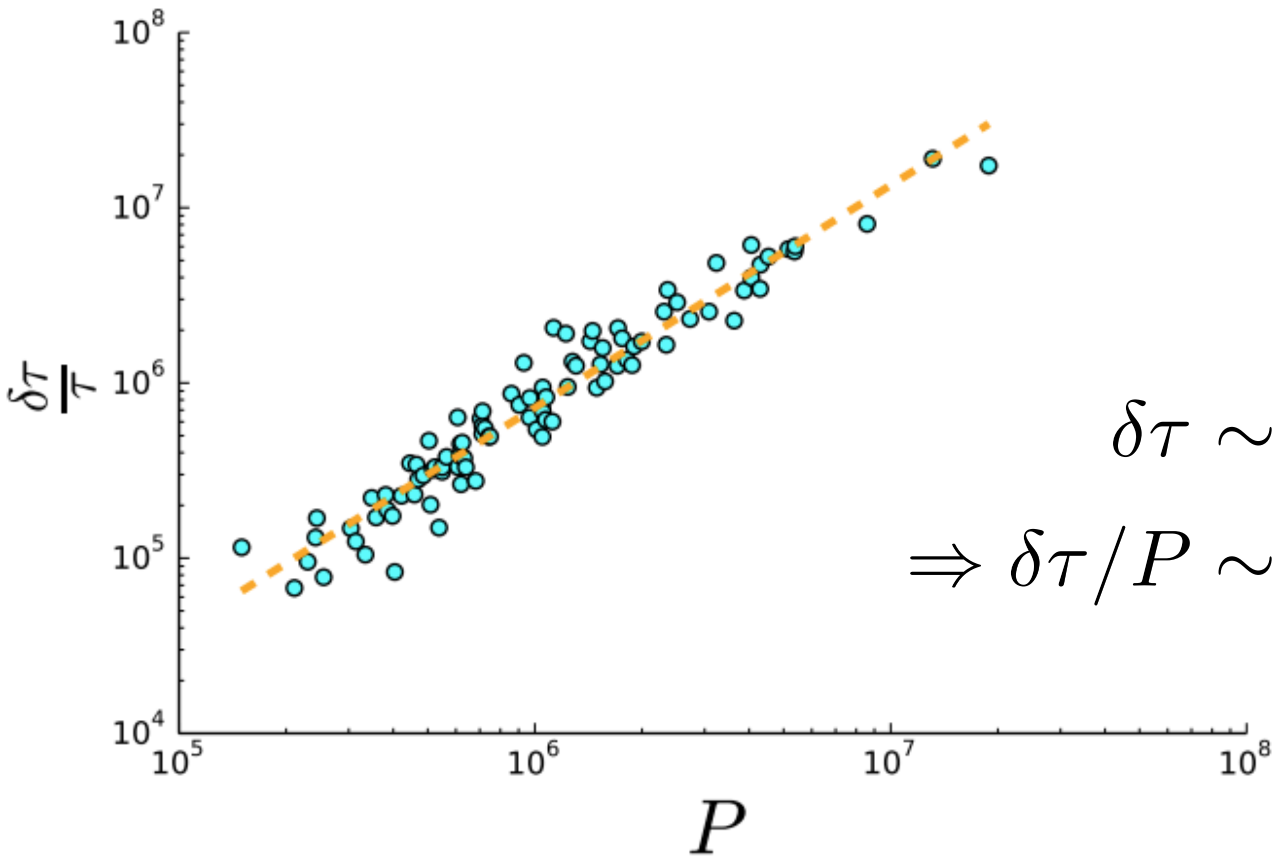
Scaling in cities: measures



US cities+ some OECD data (Louf, MB, 2013)

Other quantities

- We know the location of home and office => we can compute other mobility-related quantities
- Scaling of delay due to traffic jams (US cities)

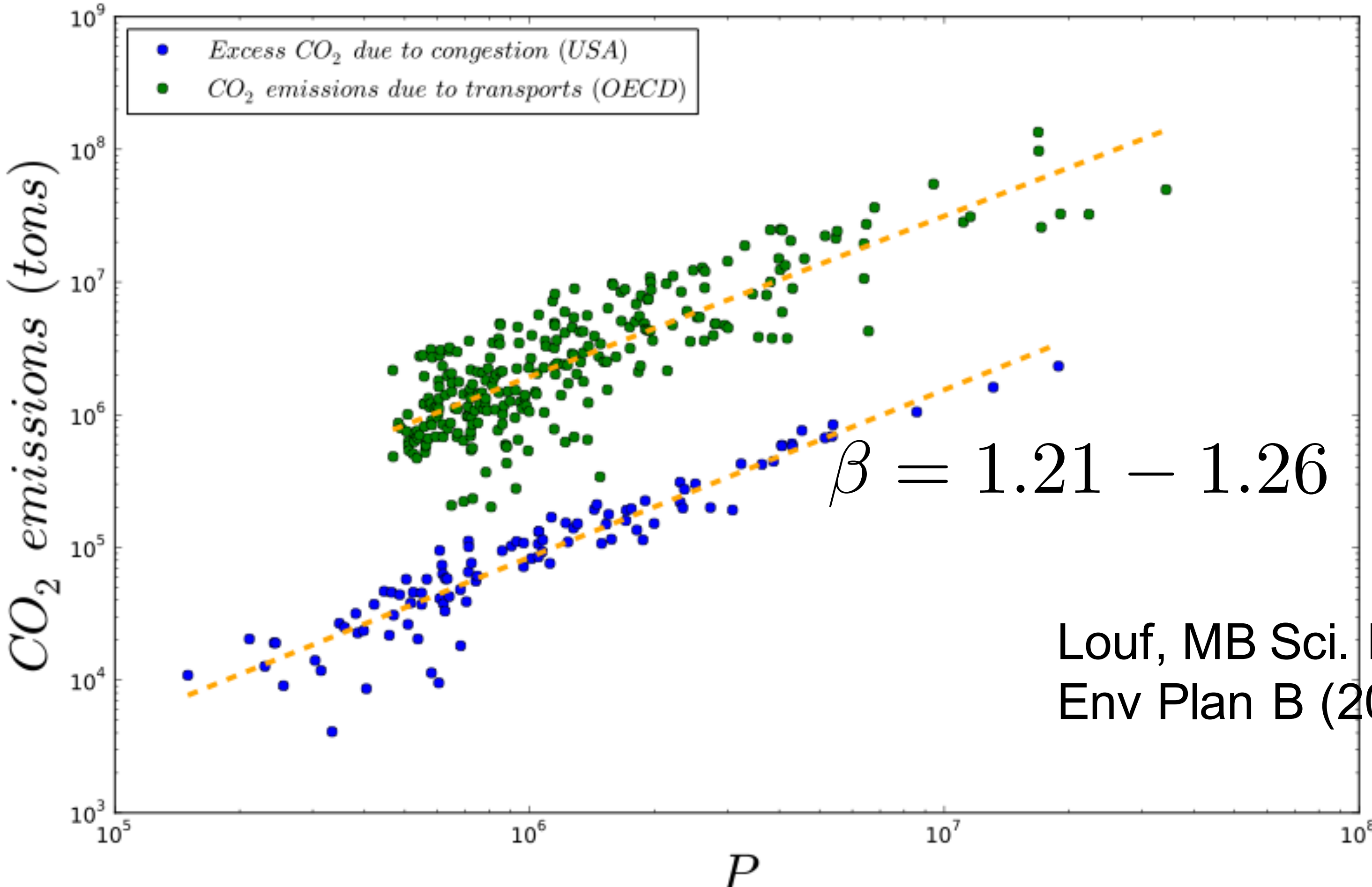


$$\delta\tau \sim P^{1.3}$$

$$\Rightarrow \delta\tau / P \sim P^{0.3}$$

Other quantities

- Emitted CO2 (transport-related)



Louf, MB Sci. Rep (2013); Env Plan B (2014)

- Superlinear !

Naive scaling: total area

Population density $\rho = \frac{P}{A}$

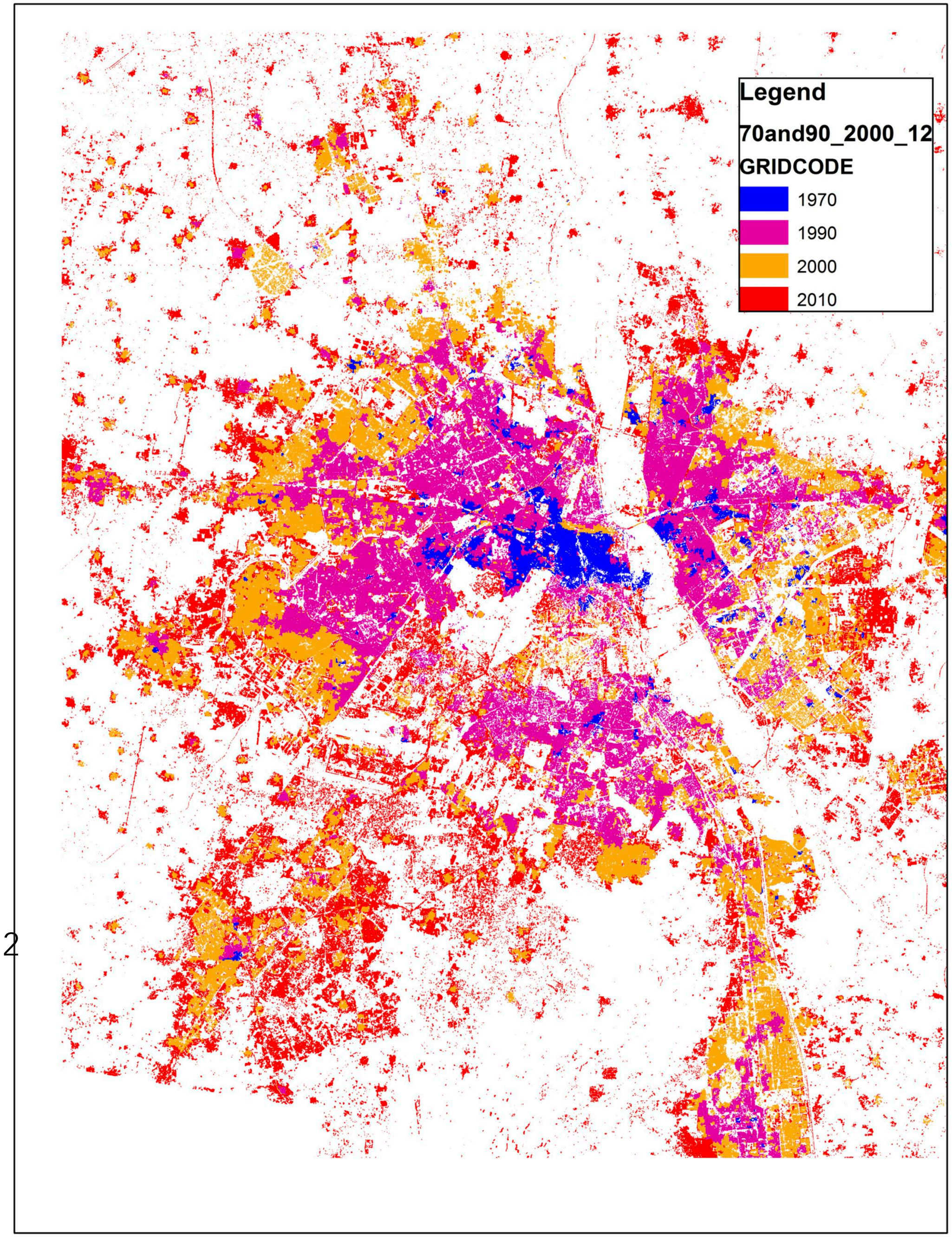
(Crude) Assumption:

$$\rho = \text{const.}$$

$$\Rightarrow A \sim \lambda^2 P^\alpha$$

Exponent $\alpha = 1$

Order of magnitude: 10^3-10^4hab/km^2
- North America: 2,000hab/km2
- Europe: 4,000-10,000
- Asia: 10,000-40,000
- Paris: 18000 (Region: 3000)



Naive scaling: total length of roads

Density of nodes

$$\rho_n \propto \frac{P}{A}$$

Length of road segments:

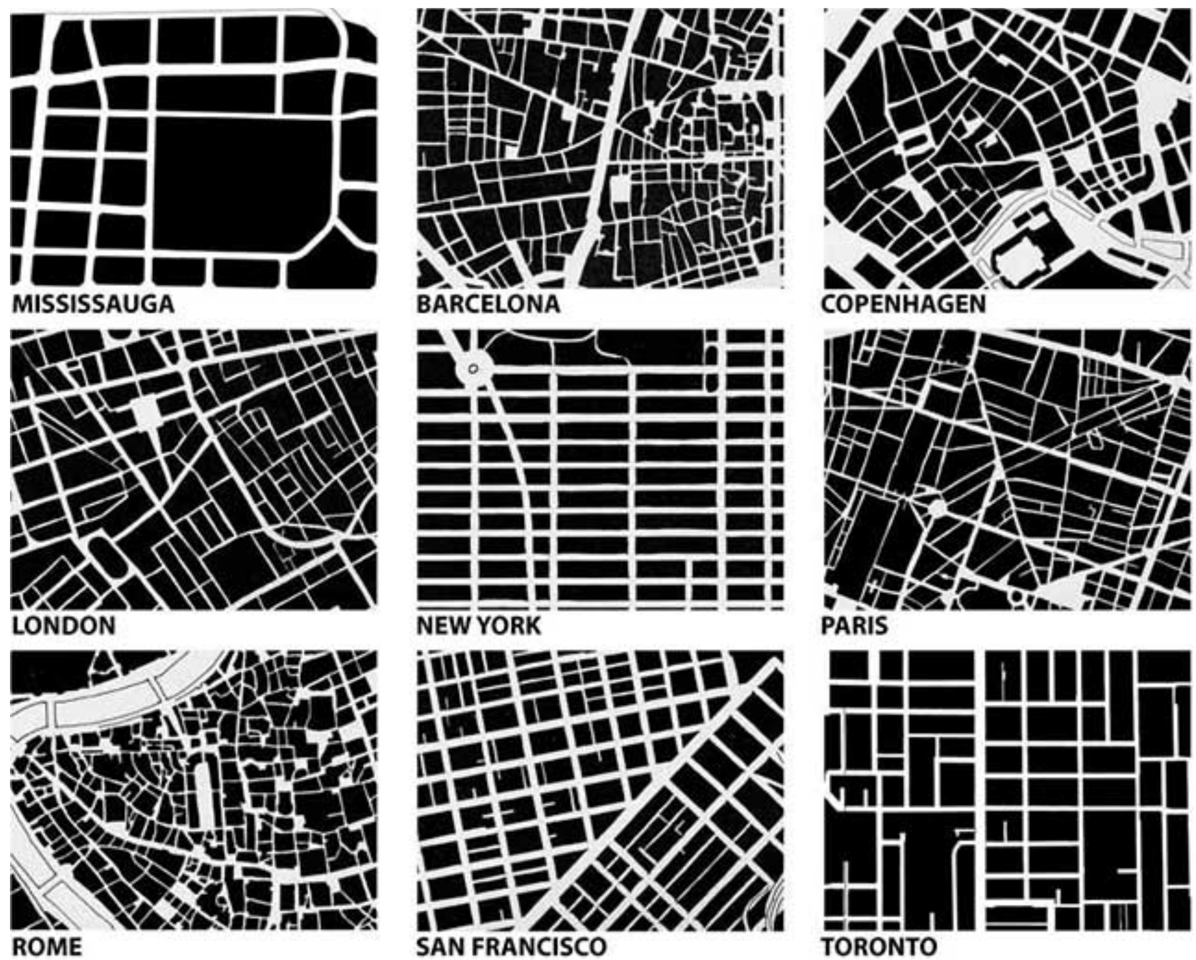
$$\ell_R \sim 1/\sqrt{\rho_n}$$

Total length:

$$L_N = P\ell_R$$

$$\frac{L_N}{\sqrt{A}} \sim \sqrt{P}$$

Exponent=1/2



Incidentally: Block area distribution

- Simple argument: density fluctuations

$$\ell_R \sim 1/\sqrt{\rho}$$
$$\Rightarrow a \sim \ell_R^2 \sim 1/\rho$$

- Assumption: density random

$$\rho \text{ follows } F(\rho)$$
$$\Rightarrow P(a) \sim \frac{1}{a^2} F\left(\frac{1}{a}\right)$$

- Fragmentation process-Master equation argument P(A,t)

Naive scaling: Total commuting distance (1)

- Simple argument:

Existence of a typical average journey-to-work distance, independent from the city

L_{tot} : total distance travelled by all commuters

$$\frac{L_{tot}}{P} \sim \text{const.}$$

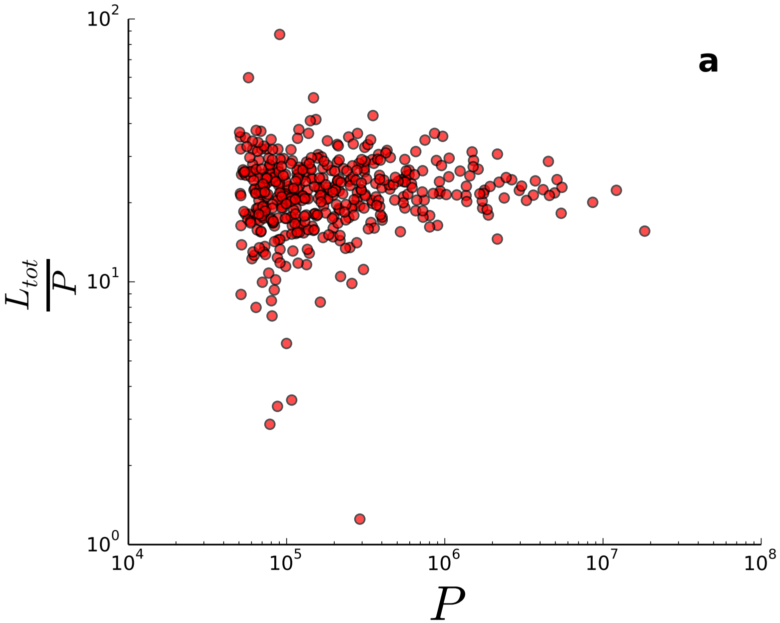
Naive scaling: Total commuting distance

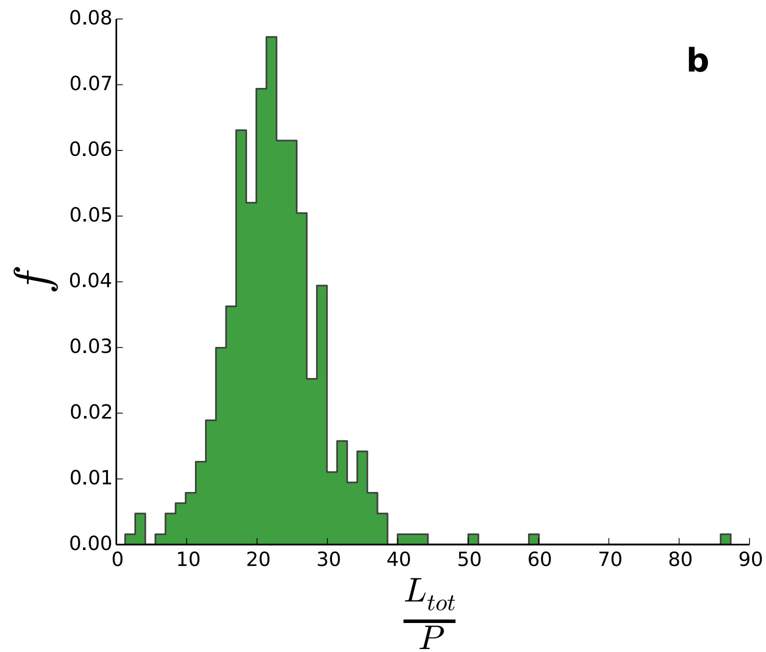
- Simple consistency relation

$$\begin{cases} L_{tot} \sim P \\ L_{tot}/\sqrt{A} \sim P^{\beta} \\ A \sim P^{\alpha} \end{cases}$$

$$\Rightarrow 1 - \frac{\alpha}{2} = \beta$$

Measures: Total commuting distance





Consistent with

$$\frac{L_{tot}}{P} \sim \text{const.}$$

US cities+ some OECD data (Louf, MB, 2013)

Scaling in cities: measures

Quantity	"Naive" scaling	Measured value for the exponent of P
$L_{tot}/\sqrt{A}$ L_{tot}/P	$1/2, 1$ 0	0.595 ± 0.026 $(r^2 = 0.91)$ [USA] 0.03 ± 0.02 $(r^2 = 0.1)$ [USA]
$L_N/\sqrt{A}$	$1/2$	$0.42 \pm 0.03(r^2 = 0.83)$ [USA]
A/ℓ^2	1	0.853 ± 0.011 $(r^2 = 0.93)$ [USA]
$\delta\tau/\tau$	?	1.270 ± 0.067 $(r^2 = 0.97)$ [USA]

- We have consistency: $1 - \frac{0.853}{2} = 0.574 \simeq 0.595$

- L_{tot} seems to scale as P

- Area A? Monocentric picture seems wrong

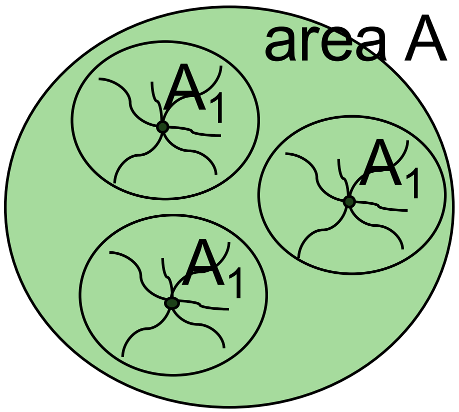
What is wrong with the naive scaling ?

- Assume k 'hotspots' or activity centers:

$$A = kA_1$$

$$L_{tot} = k\frac{P}{k}\sqrt{A_1}$$

$$\Rightarrow \frac{L_{tot}}{\sqrt{A}} = \frac{P}{\sqrt{k}}$$



- Can change scaling exponents if k varies with P !

Total delay due to congestion

- We have

$$\tau = \sum_{i,j} \frac{d_{ij}}{v_0} \left[1 + \left(\frac{T(j)}{c} \right)^{\mu} \right]$$

- Delay

$$\frac{\delta\tau}{\tau_0} \sim P^{1+\delta}, \quad \delta = \frac{\mu}{2\mu + 1}$$

- From the data

$$\frac{\delta\tau}{\tau_0} \sim P^{1.39} \quad (1.27)$$

Superlinear !

- CO2 emitted (car related)

$$Q_{CO_2} \propto \tau \sim P^{1+\delta}$$

Predicting the exponent values

Quantity	Theoretical dependence on P ($\delta = \alpha/\alpha + 1$)	Predicted value	Measured value
A/ℓ^2	$\left(\frac{P}{c}\right)^{2\,\delta}$	$2\delta = 0.78 \pm 0.20$	0.853 ± 0.011 ($r^2 = 0.93$) [USA]
L_N/ℓ	$\sqrt{P}\ \left(\frac{P}{c}\right)^{\delta}$	$\frac{1}{2} + \delta = 0.89 \pm 0.10$	0.765 ± 0.033 ($r^2 = 0.92$) [USA]
$\delta\tau/\tau$	$P\ \left(\frac{P}{c}\right)^{\delta}$	$1 + \delta = 1.39 \pm 0.10$	1.270 ± 0.067 ($r^2 = 0.97$) [USA]
$Q_{gas,CO_2}/\ell$	$P\left(\frac{P}{c}\right)^{\delta}$	$1 + \delta = 1.39 \pm 0.10$	1.262 ± 0.089 ($r^2 = 0.94$) [USA] 1.212 ± 0.098 ($r^2 = 0.83$) [OECD]

Louf, MB (2013, 2014)

Predicting the exponent values

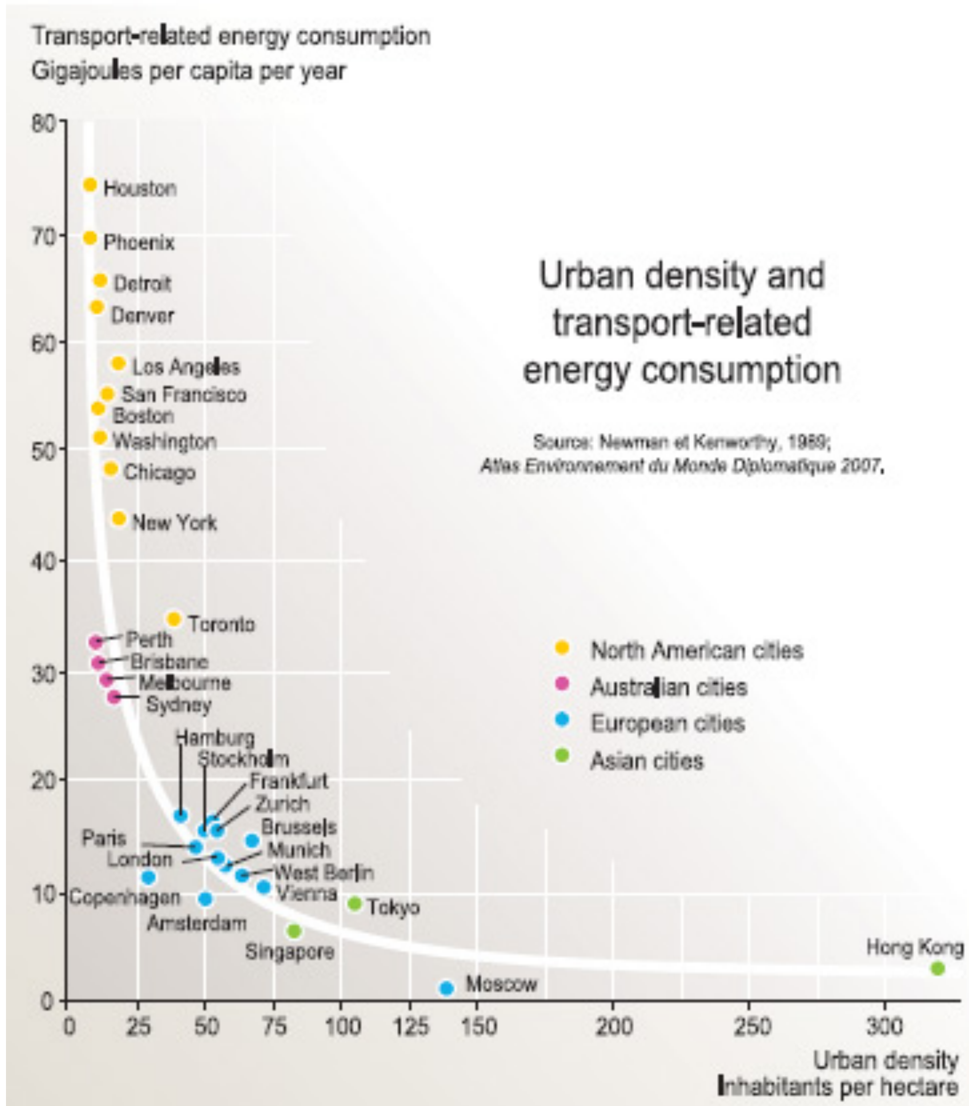
- Polycentrism is the natural response of cities to congestion, but not enough !

- For large P: Effect of congestion becomes very large
  => large cities based on individual cars are not economically sustainable !

Louf, MB (2013, 2014)

Predicting the exponent values

- Q_{gas}/P is not a simple function of density (cf. Newman & Kenworthy)



Discussion and outlook

- Pushing the models and compute predictions; testing predictions against data. Goal: understand the hierarchy of mechanisms (and a model with a minimal number of parameters).

- End of story ? Integrating socio-economical factors: rent, other transportation modes, income,…

- Discussion: importance of scaling exponents $Y \sim P^{\beta}$
  - 'close' to 1: the value depends on assumptions
  - one cannot rule out the linear behavior

Is this scaling nonlinear?
Leitao, Miotto, Gerlach, Altmann, arXiv preprint arXiv:1604.02872 (2016)