Uncovering the spatial structure of urban mobility networks

Methods and applications to Spanish cities

T. LOUAIL

M. LENORMAND, JJ. RAMASCO, M. BARTHELEMY
M. PICORNELL, O. GARCIA-CANTU, R. HERRANZ, E. FRIAS-MARTINEZ
J. MURILLO ARIAS, M. SAN MIGUEL

- Mobility Data
- Types of mobility networks
  (regular daily flows & individual trajectories)
- Extracting an expressive signature of
  the structure of large mobility networks
  - Method
  - Application to the comparison of Spanish cities
- Discussion

Why?

- Mobility is the vector of other processes (epidemiology, city growth, etc.)
- Many phenomena we don't properly understand (mobility and social networks, individual strategies of space exploration, invisible networks of familiar networks, etc.)
- Mobility is interesting by itself !

Why?

- Mobility is the vector of other processes (epidemiology, city growth, etc.)
- Many phenomena we don't properly understand (mobility and social networks, individual strategies of space exploration, invisible networks of familiar networks, etc.)
- Mobility is interesting by itself !

What changed?

- Data data data data data data data data data data data...
- … suggested new questions/topics
- … bring new actors
- Complex systems/networks approach

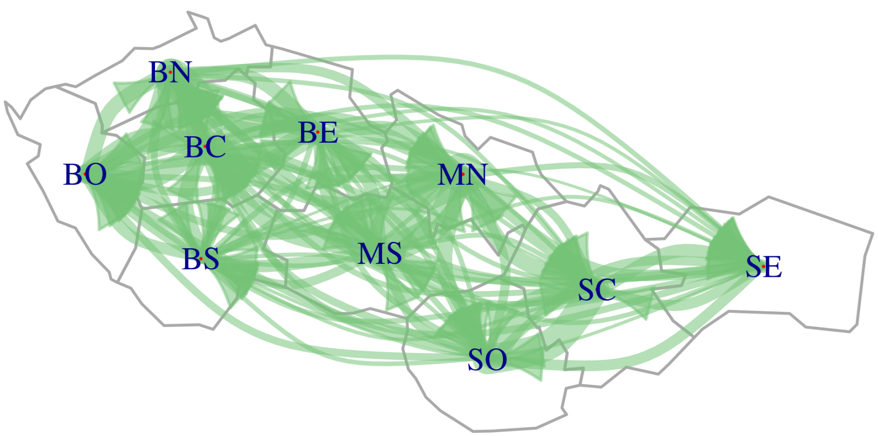
Collective, voluntary geographic information, public data, e.g. infrastructure networks (OSM), transport timetables, locations of public infrastructures, etc.

Collective, voluntary geographic information, public data, e.g. infrastructure networks (OSM), transport timetables, locations of public infrastructures, etc.

Geotagged (meta)data passively generated by individuals when using their mobile devices (mobile phones, social networking apps, transport smart cards, car GPS, credit cards, etc.)

1st type of mobility network
Origin-Destination (OD) matrices

	BC	BE	BN	BO	BS	MN	MS	SC	SE	SO
BC	0	13200	22600	15100	7900	7500	7100	2200	1500	3300
BE	19100	0	10800	5600	3000	9900	9600	1200	1200	2300
BN	30000	10100	0	13400	3700	6300	5100	2100	2400	3600
BO	18300	3700	11100	0	6000	6700	6000	2000	2200	3400
BS	11600	2900	3900	8200	0	3200	5800	700	600	1700
MN	8900	5600	4900	4600	2000	0	22000	5000	3400	6200
MS	11500	7800	6200	8900	6000	35600	0	2400	1500	6100
SC	3200	1700	2900	2900	1300	12000	3500	0	9500	23900
SE	3100	2200	5800	4200	1000	8100	2200	7500	0	7500
SO	5000	3400	4600	4600	2100	11500	6500	20400	6300	0



Classic object to study commuting patterns

F(i,j) is the number of individuals that live in i and work in j

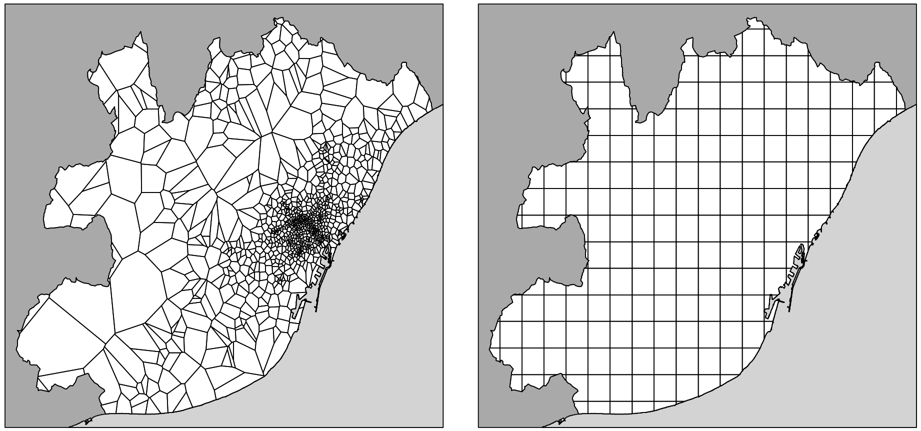
Provides the complete information on the commuting flows

Simple hypothesis used to determine
the home & work locations
of each individual

Home
is assimilated to
the most frequent location
between 8 p.m. and 8 a.m.

Work/School/Main activity
is assimilated to
the most frequent location
between 8 a.m. and 8 p.m.

(a) (b)



a=500m

a=1km

a=2km

from CDR data

HOME

Location	Frequency
1	**0.63**
2	0.2
3	0.13
4	0.04
...	...

=Hp

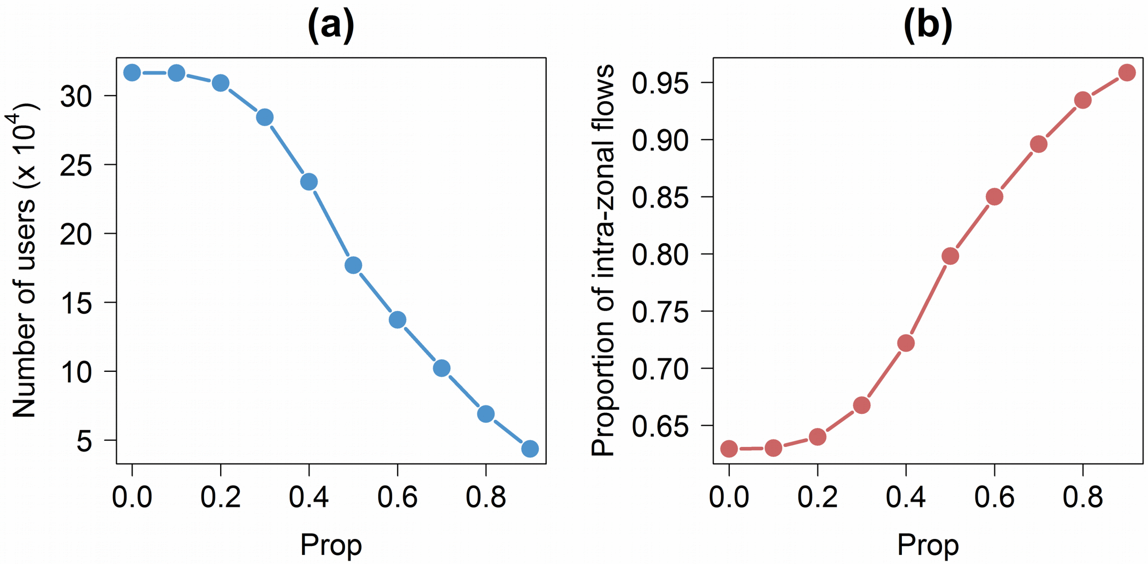
WORK

Location	Frequency
1	**0.73**
2	0.15
3	0.02
4	0.01
...	...

=Wp

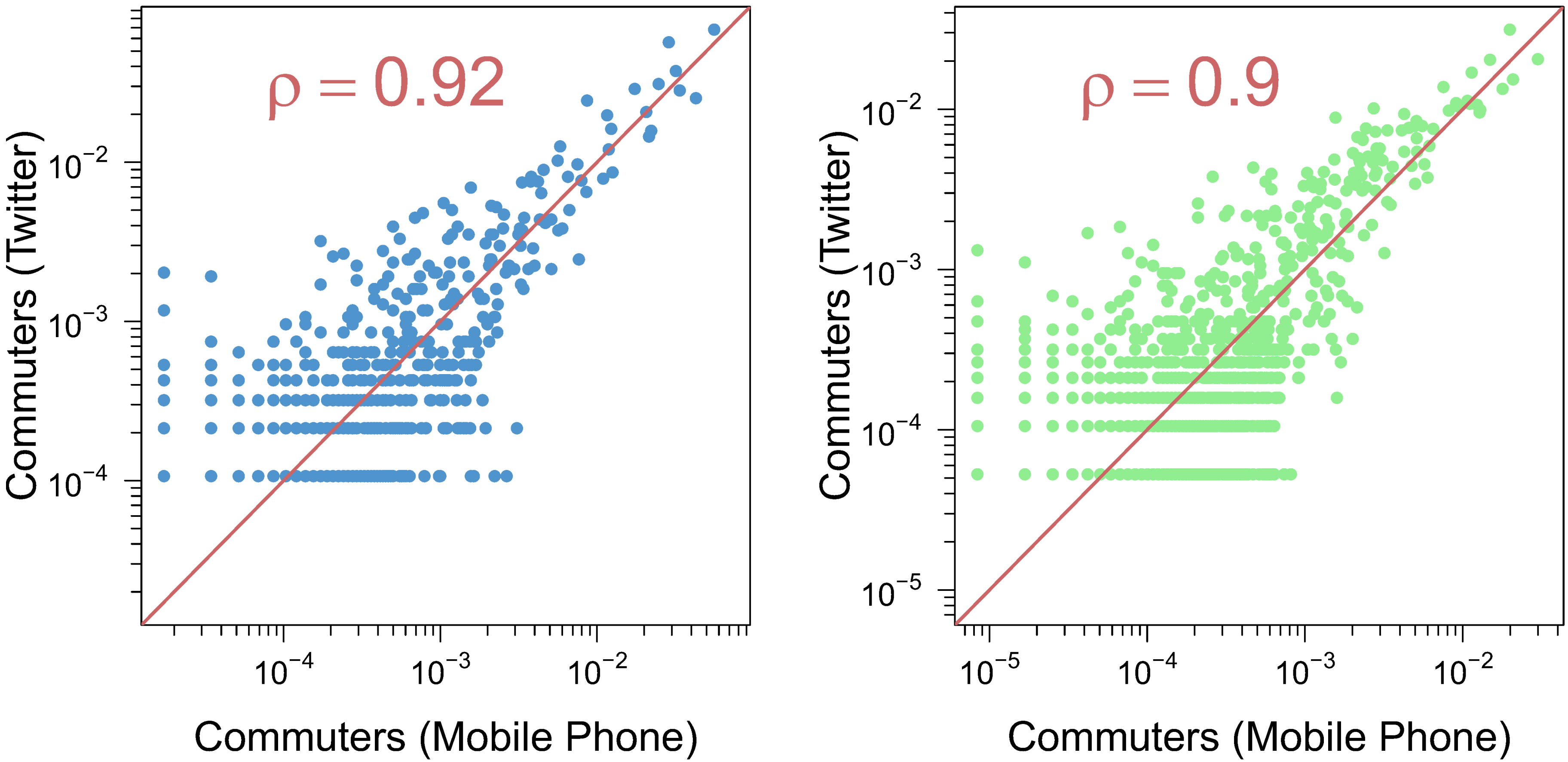
Sorted by frequency

Hp > Prop & Wp > Prop



Madrid
(a)

Barcelona
(b)

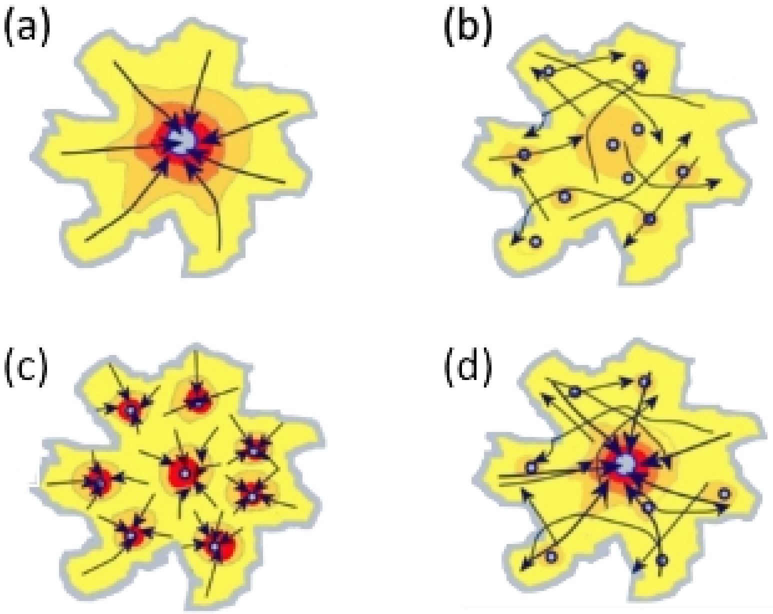


→ *(Lenormand* et *al. Plos One 2014)*

How to extract
a **simple and expressive footprint**
of a large mobility network?

Can we build a typology of cities based
on the spatial structure
of their commuting patterns?

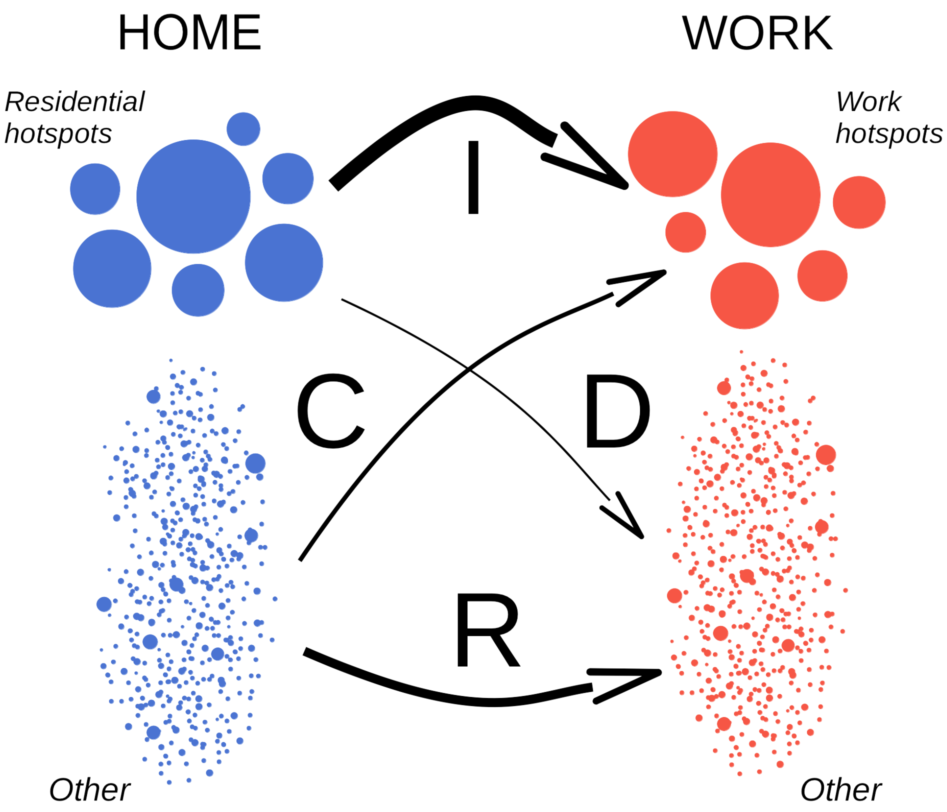
Inspiration:
schematic/simple figures
of commuting flows
in idealized city forms



(from Bertaud & Malpezzi 2003)

1. Determine
   residential hotspots
   and **employment hotspots**

2. Separate **4 categories of flows**
   - **I**ntegrated :
     from residential hotspots
     to work hotspots
   - **C**onvergent :
     from elsewhere
     to work hotspots
   - **D**ivergent :
     from residential hotspots
     to elsewhere
   - **R**andom :
     from elsewhere
     to elsewhere



(*Louail* et *al. Sci. Rep. 2014*)

1.

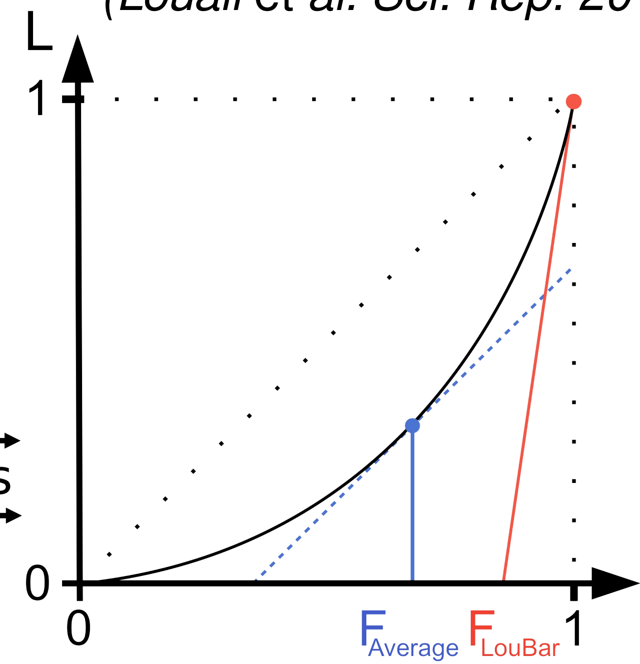
$$OD = \begin{bmatrix} C_{1,1} & \cdots & C_{1,n} \\ \vdots & C_{i,j} & \vdots \\ C_{n,1} & \cdots & C_{n,n} \end{bmatrix} \qquad \sum_{j\in 1..n} C_{i,j} = C_i^{out}$$

$$\sum_{i\in 1..n} C_{i,j} = C_j^{in}$$

2.

$$C^{out} = (C_1^{out}, \ldots, C_n^{out})$$
$$C^{in} = (C_1^{in}, \ldots, C_n^{in})$$

Determine hotspots



3.

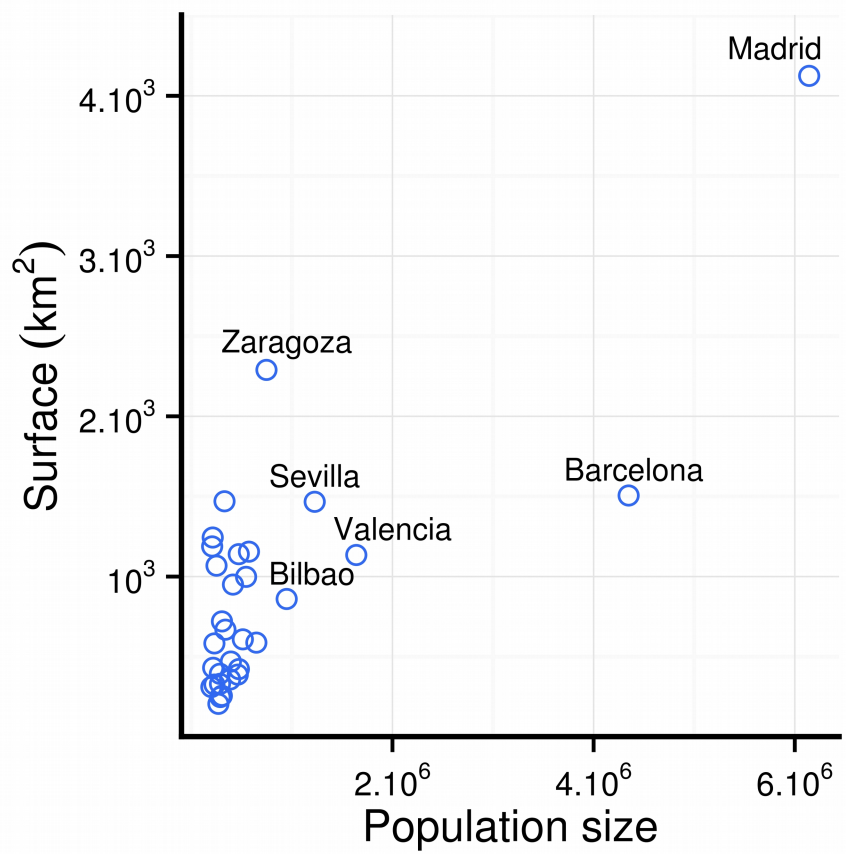
$$OD = \begin{array}{c} \mathbf{I} \qquad\qquad\qquad\qquad \mathbf{D} \\ \left[\begin{array}{cc} \sum_{\substack{i\in 1..m \\ j\in 1..p}} C_{i,j}/\sum C_{i,j} & \sum_{\substack{i\in 1..m \\ j\in p+1..n}} C_{i,j}/\sum C_{i,j} \\ \sum_{\substack{i\in m+1..n \\ j\in 1..p}} C_{i,j}/\sum C_{i,j} & \sum_{\substack{i\in m+1..n \\ j\in p+1..n}} C_{i,j}/\sum C_{i,j} \end{array}\right] \\ \mathbf{C} \qquad\qquad\qquad\qquad \mathbf{R} \end{array}$$

p residential (out) hotspots

m work (in) hotspots

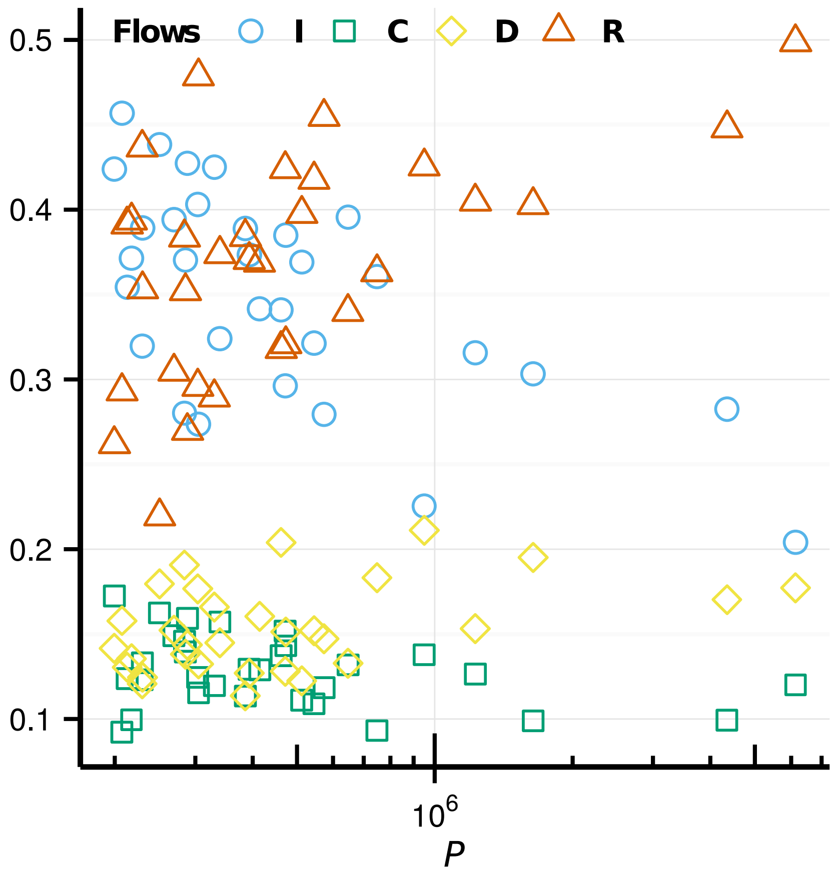
$$OD = \begin{bmatrix} I & D \\ C & R \end{bmatrix}$$

with I + D + C +R = 1



Harmonized functionnal definition
(same for all cities)

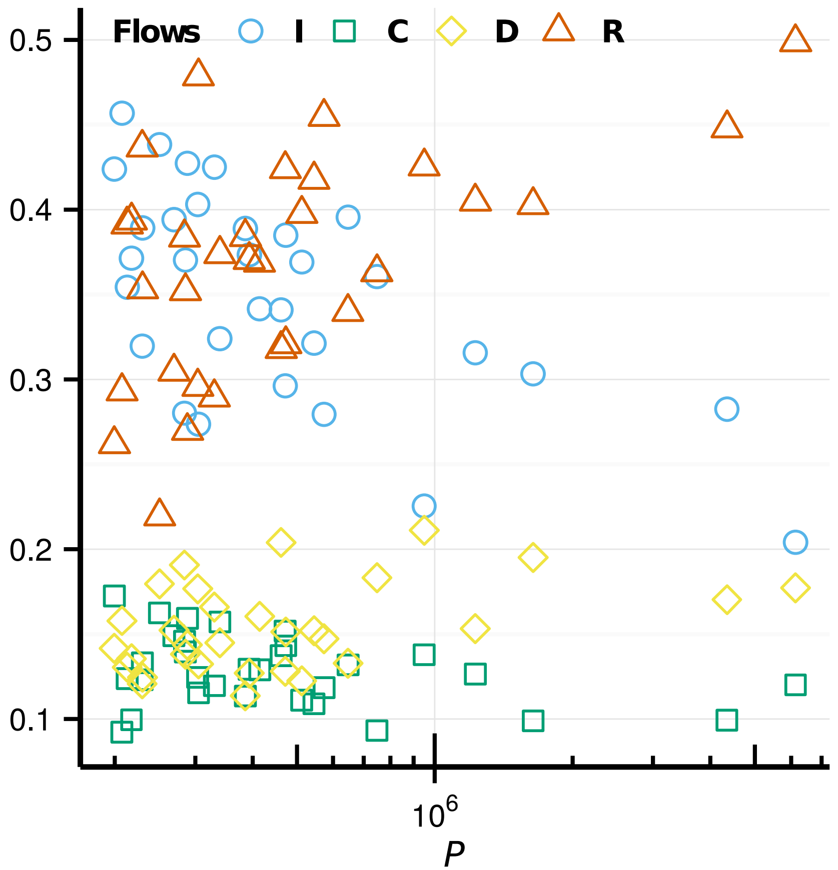
ICDR values vs. P



Weight of **I**ntegrated flows decreases
when population size P increases,
in favor of of **R**andom flows

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

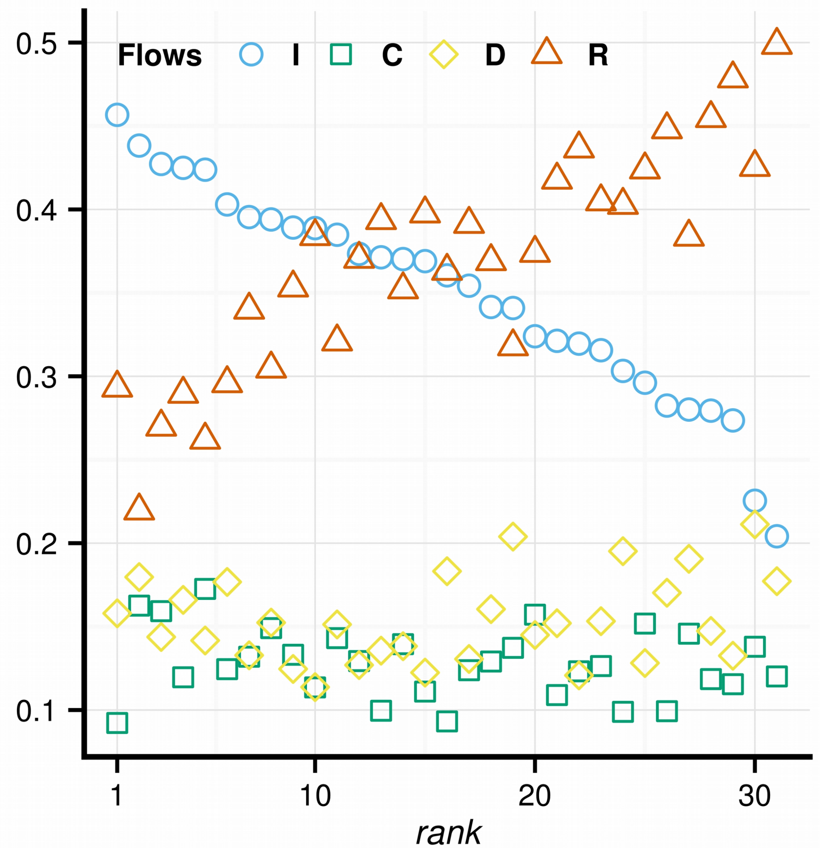
ICDR values vs. P



Weight of **I**ntegrated flows decreases
when population size P increases,
in favor of of **R**andom flows

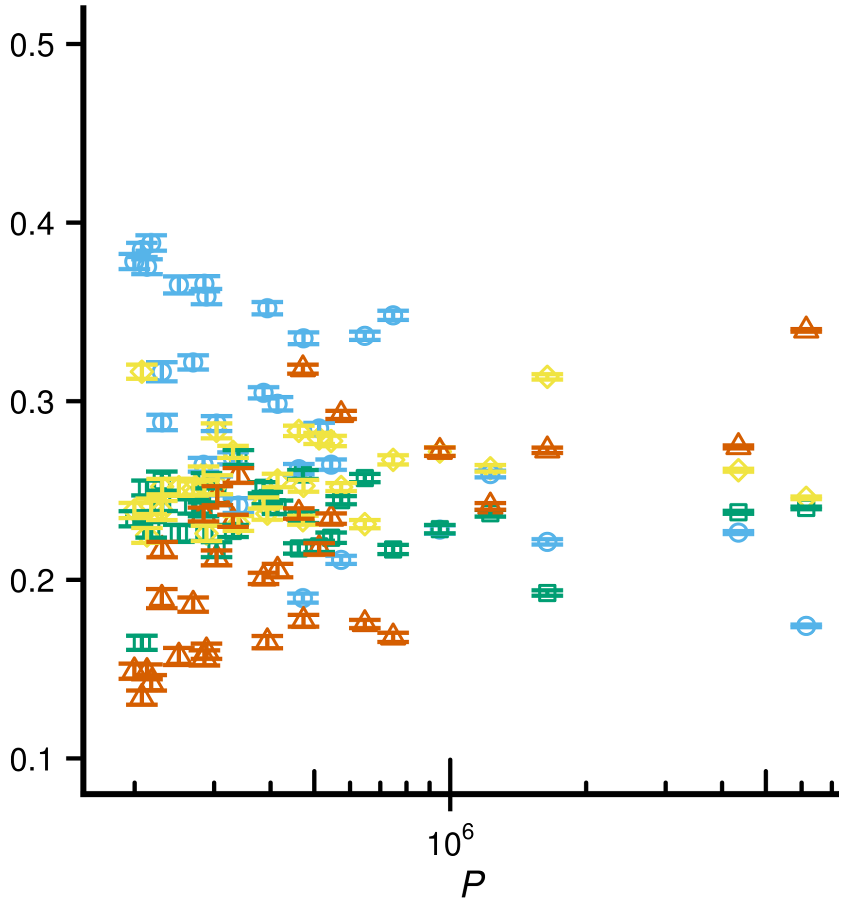
Weights of **D**ivergent and **C**onvergent flows
almost constant whatever the city size

Same values ranked by decreasing I

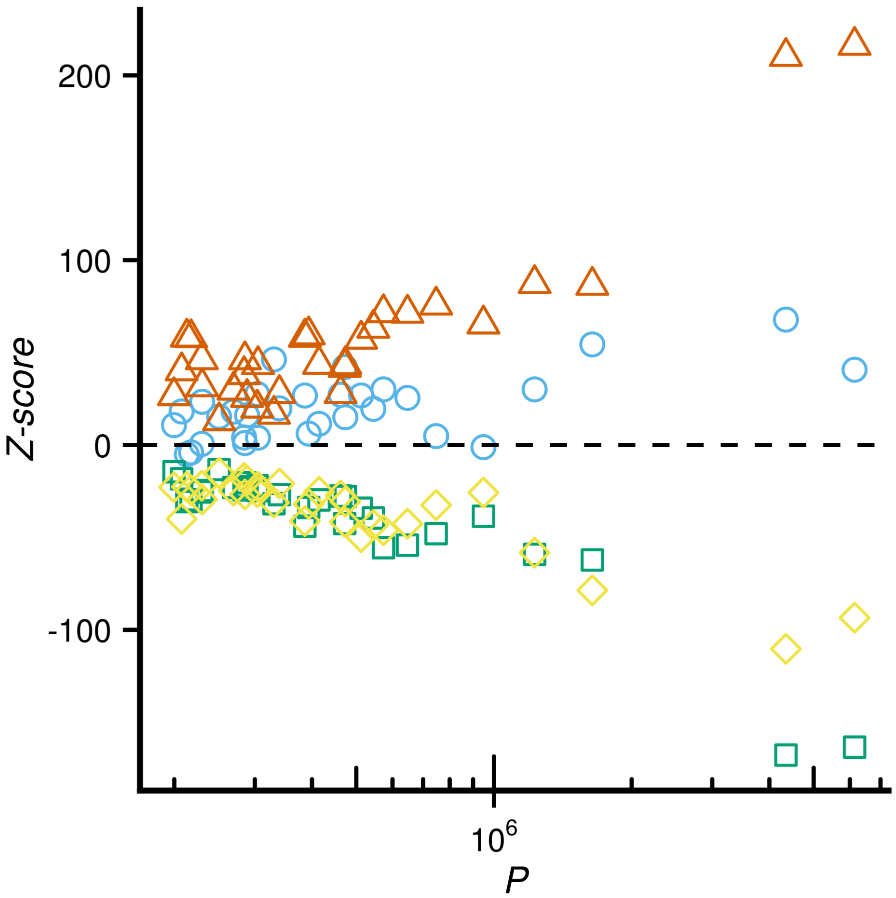


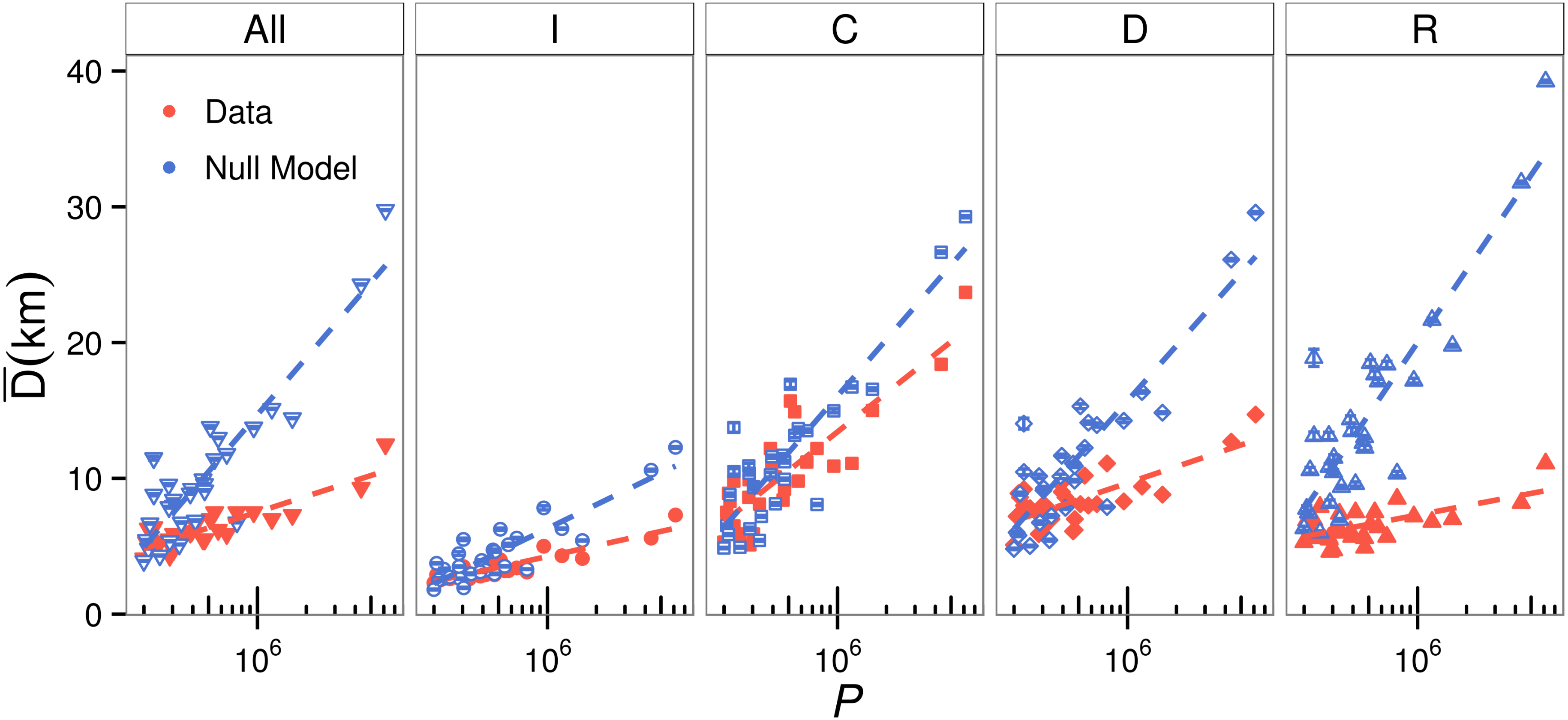
I and **R** alone seem sufficient
to classify cities

Null model : ICDR values vs. P

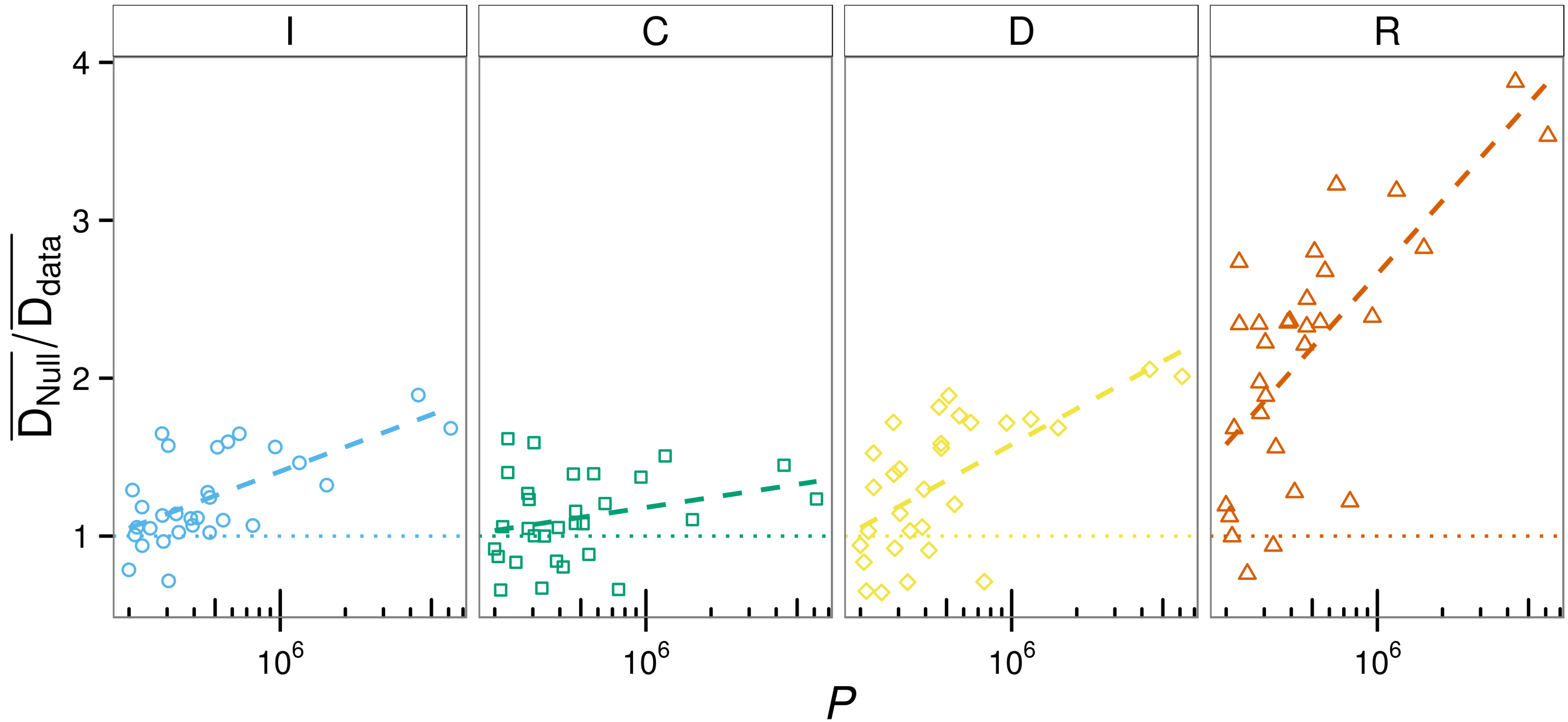


Z-scores





- For all types of flows distance increases with population size
- Convergent Flows (C) are the longest and the most penalized when P increases

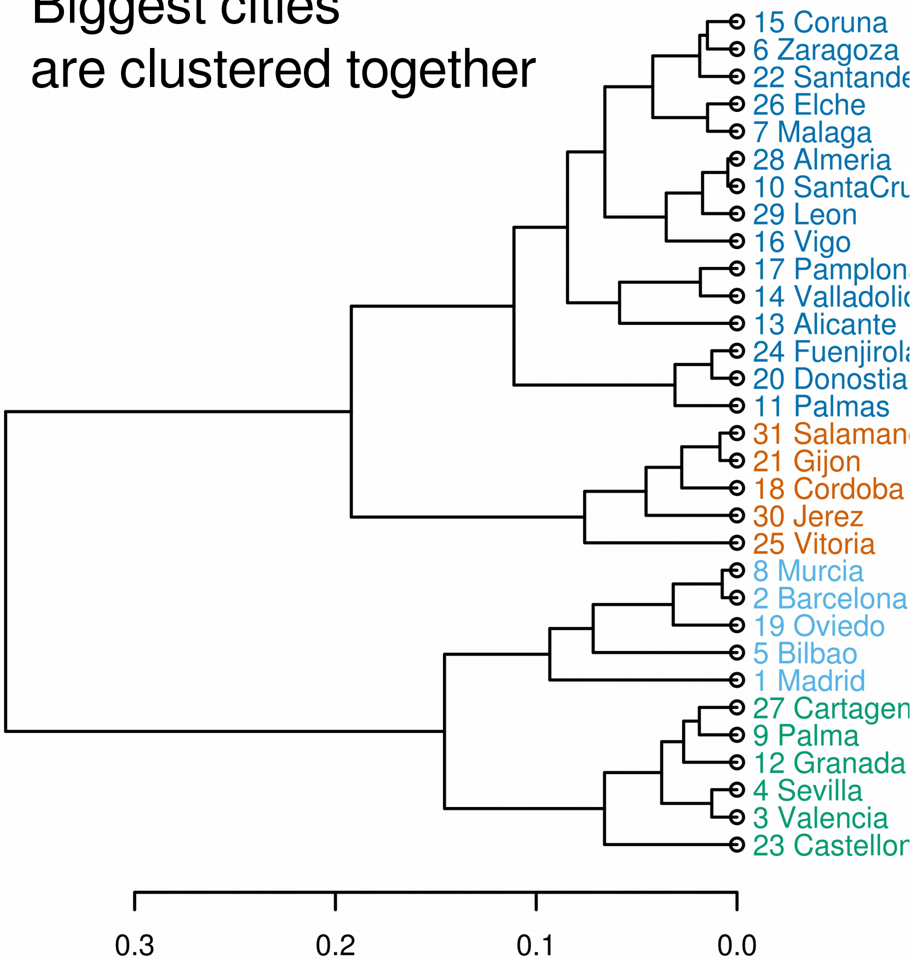


- As cities grow, the spatial organisation of commuting flows
is more and more rationale (i.e. advantageous when compared to random flows)
- Some small cities display a value less than 1,
indicating the lesser importance of space at shorter scales

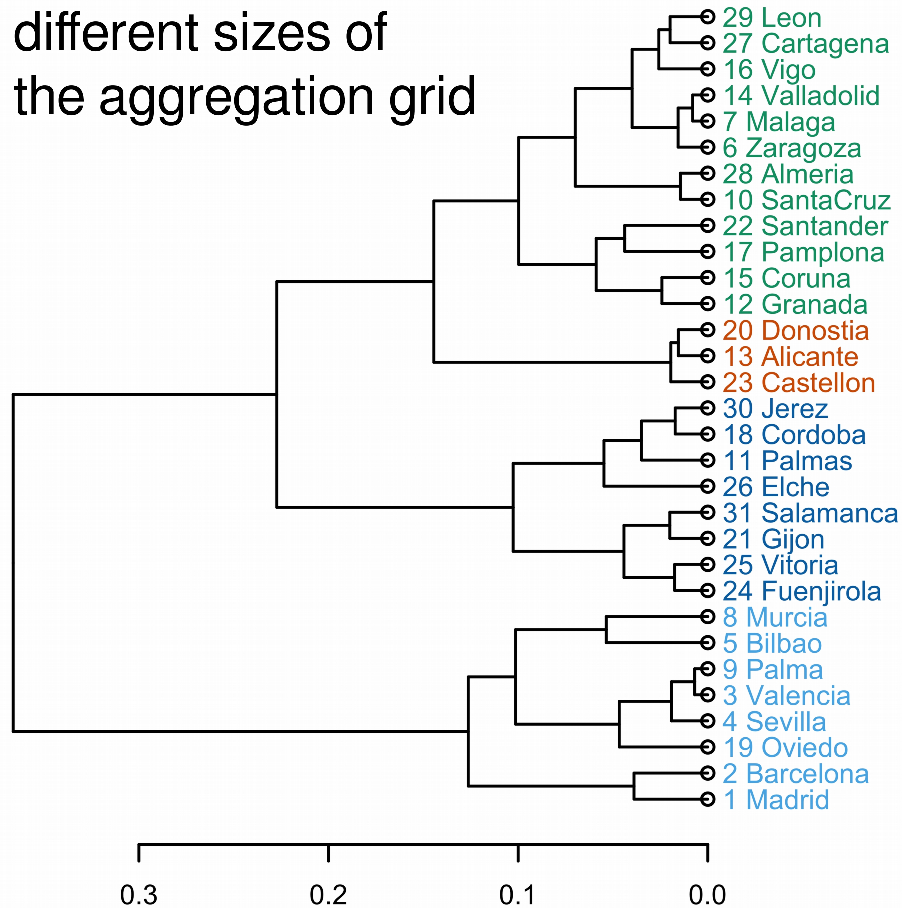
1km * 1km grid

2km * 2km grid

Biggest cities are clustered together



Robust with different sizes of the aggregation grid



Cluster	Average population of cities in cluster	I	R	D	C
Cordoba, Gijon, Vitoria, etc	255,330	**0.43**	**0.27**	0.16	0.14
Zaragoza, Malaga, etc.	392,970	**0.37**	**0.36**	0.15	0.13
Valencia, Sevilla, etc.	732,992	**0.31**	**0.41**	0.16	0.13
Madrid Barcelona, etc.	2,463,551	**0.25**	**0.46**	0.17	0.12

Tap at **start** and **end** of train journeys

Accepted at 695 Underground and rail stations, and on thousands of buses

A week of data (week starting 02/07/2012)

For each 15 minutes time-slice, from Monday to Sunday,

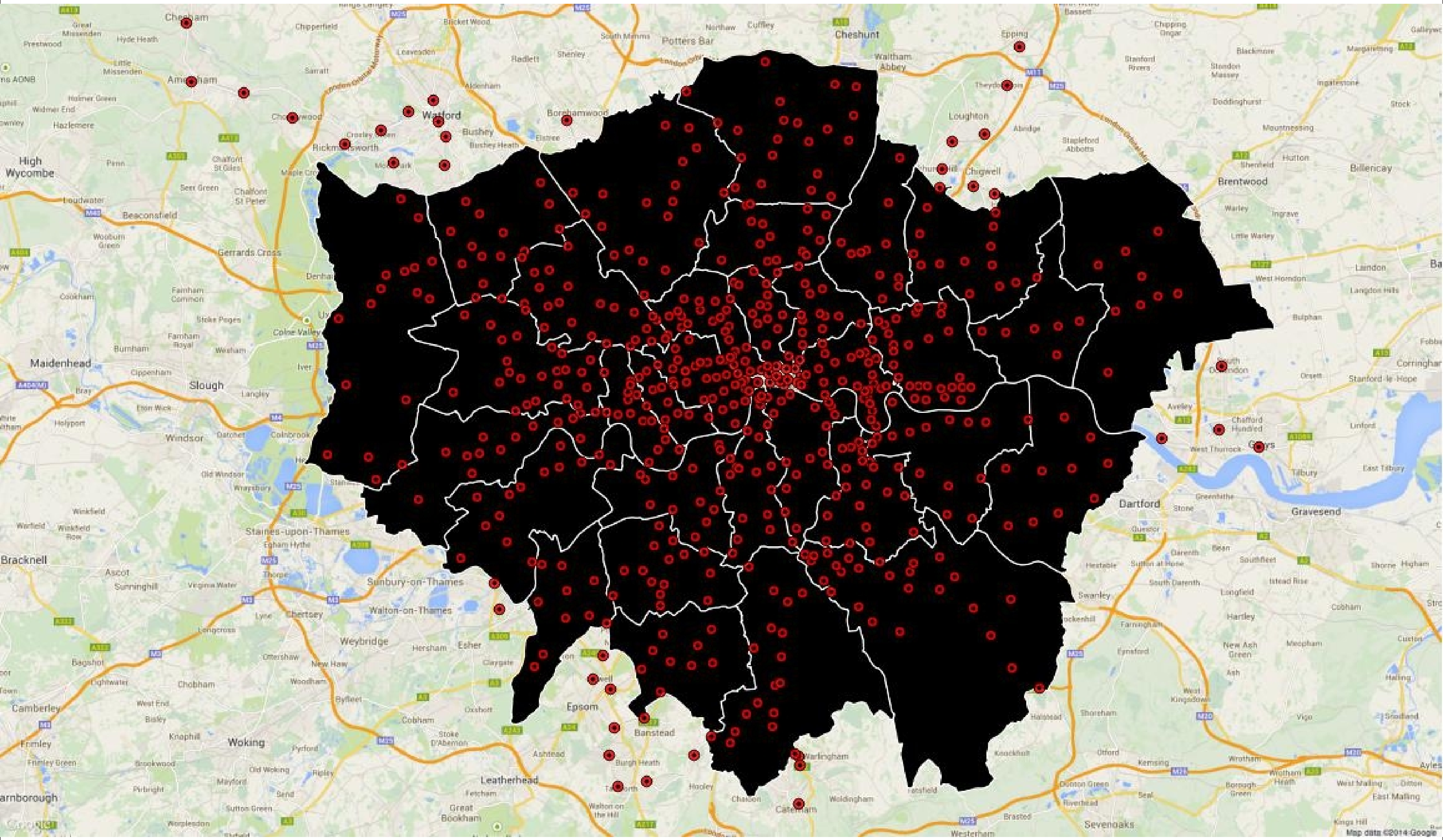
An O-D matrix containing the numbers of commuters:

•597 stations

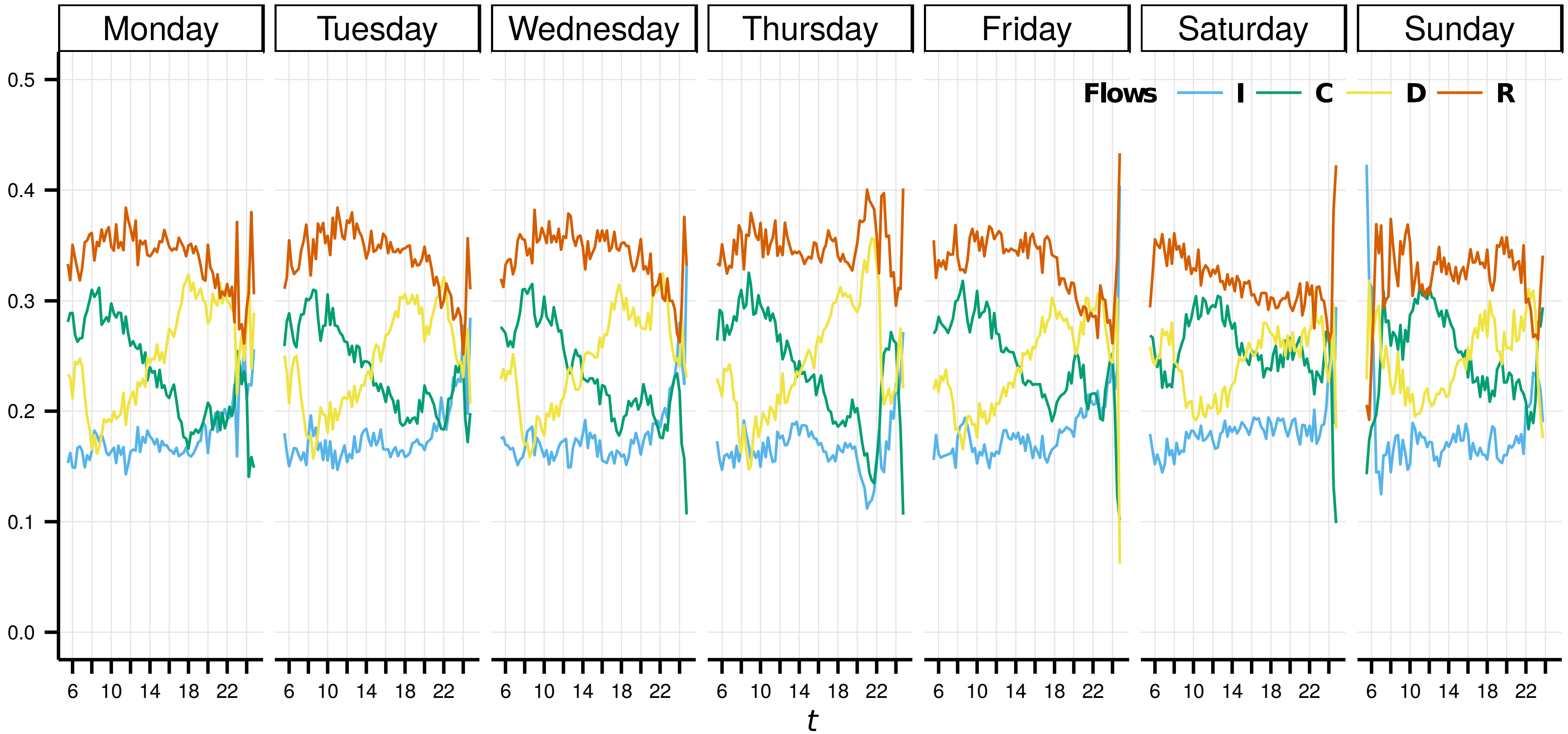
•22,938,721 taps

•18,858,038 Mon-Fri

•4,080,683 Sat-Sun



Aggregated stations used in our analysis (597 stations)



- **ICDR values are almost constant during the week**
- They constitute a signature of the mobility structure in the city

ICDR is a versatile method to extract a simple footprint of the structure of large, weighted and directed networks

The method is independent from any particular definition of hotspots and from the spatial resolution at which data are available

Application on 31 Spanish cities:

- As the population size increases, I decreases while R increases
- The comparison with a reasonable null model revealed that commuting flows are more organized in the largest cities
- Convergent flows are the more penalized when the population size increases

Application to London proved that ICDR values are a signature of the mobility structure in the city