Statistical analysis of urban mobility networks

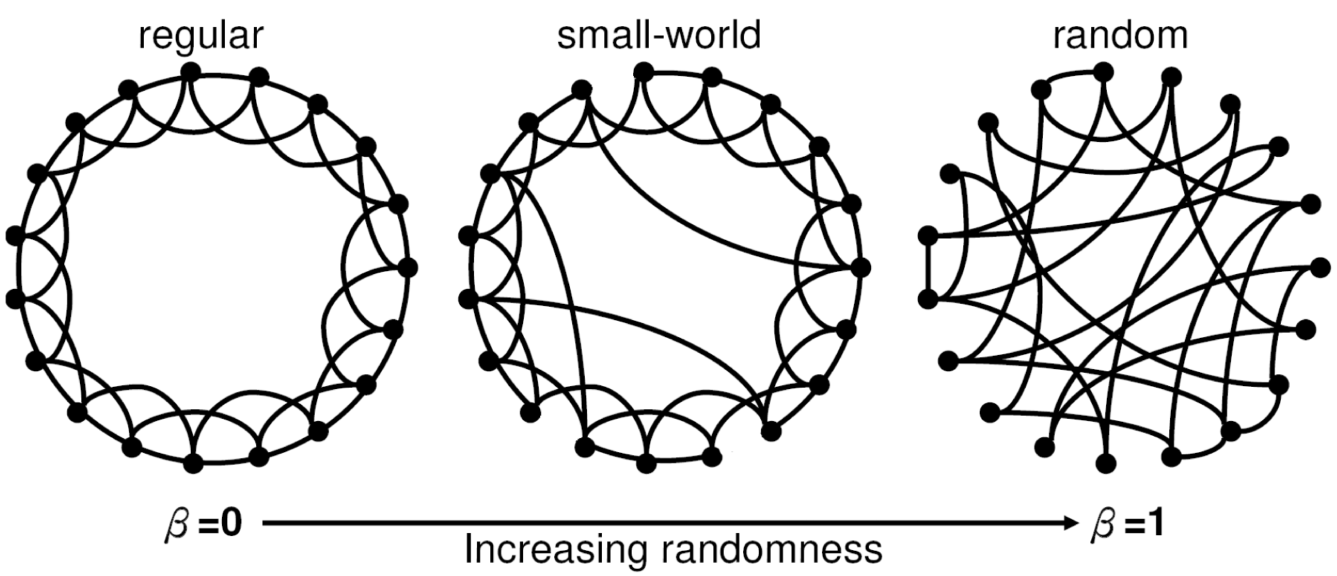
(A few examples of network science applied to human mobility)

Thomas Louail
CNRS, Géographie-cités

« Network science »?

Late 1990's turn: availability of data about large, real-world networks
> emergence of a « network science » made by statistical physicists

(Watts and Strogatz, 1998, *Nature*)

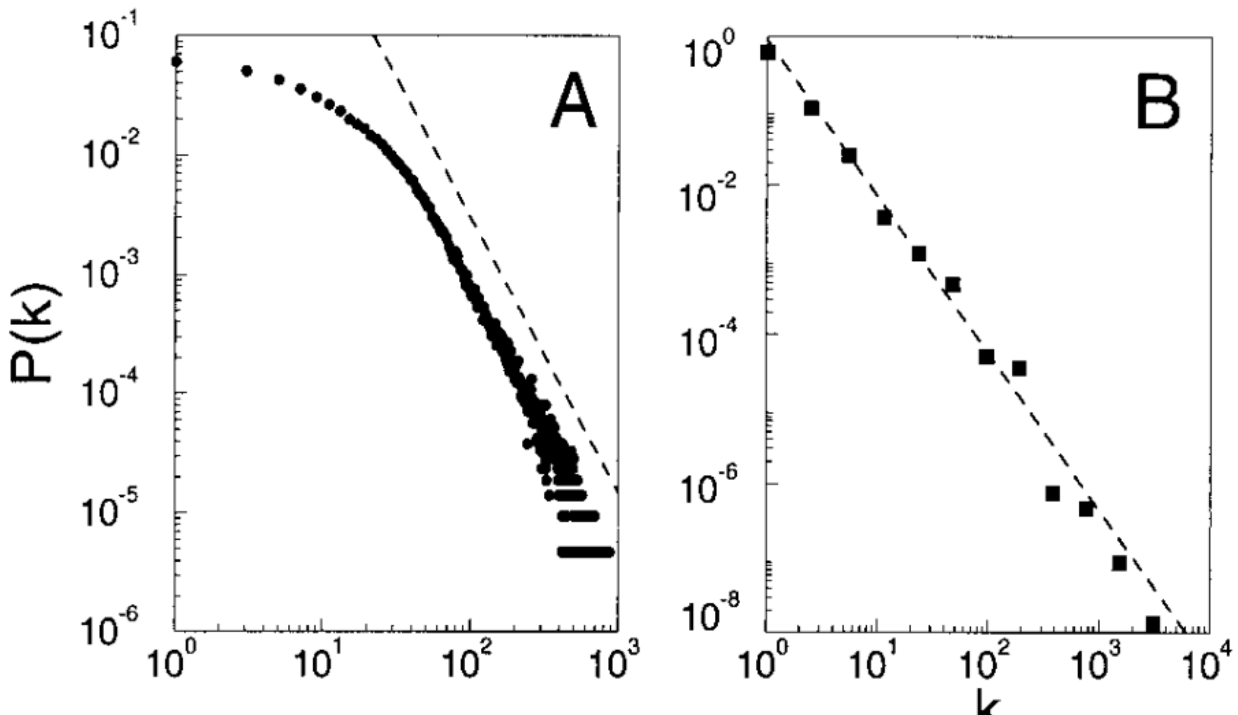


small-world networks
> high clustering coefficient
(like regular lattices)
+ short characteristic distance
(like random networks)

scale-free networks
(node connectivity exhibits
no typical scale
i.e. the degree distribution is unbounded
over several order of magnitudes)

Models, e.g.
continuous growth
+ preferential attachment
= scale-free net

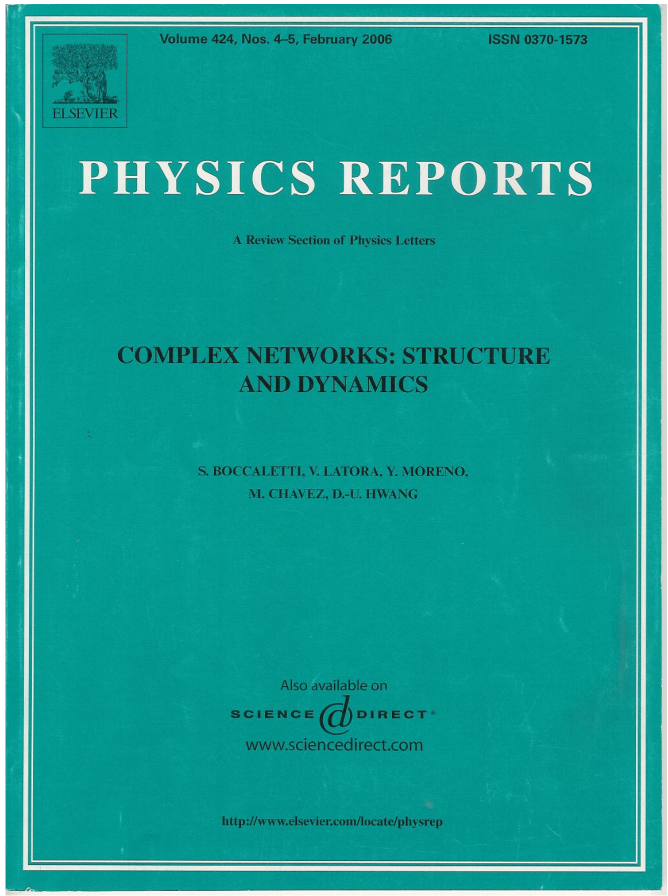
(Barabási and Albert, 1999, Science)

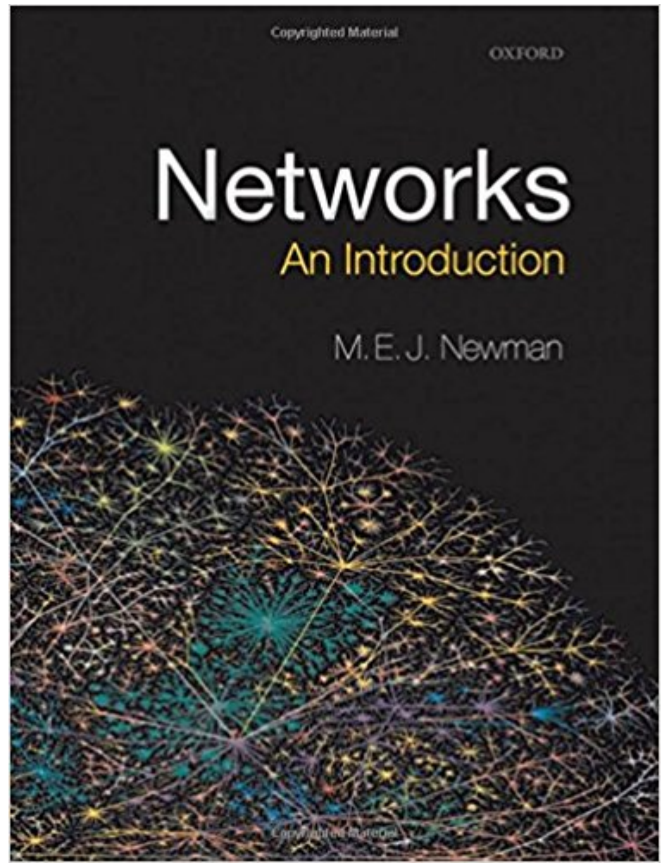


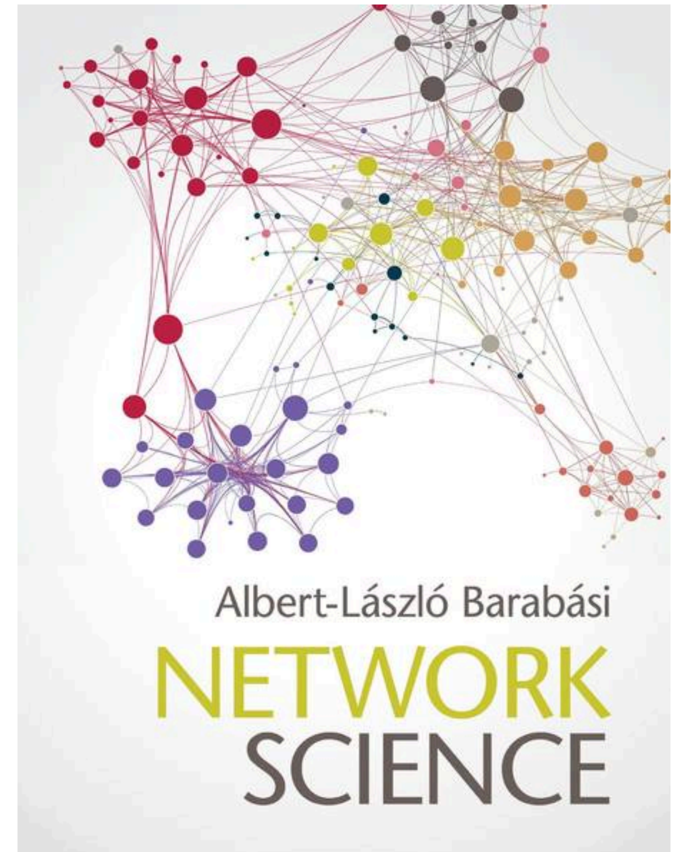
Actors collaborations **WWW pages**

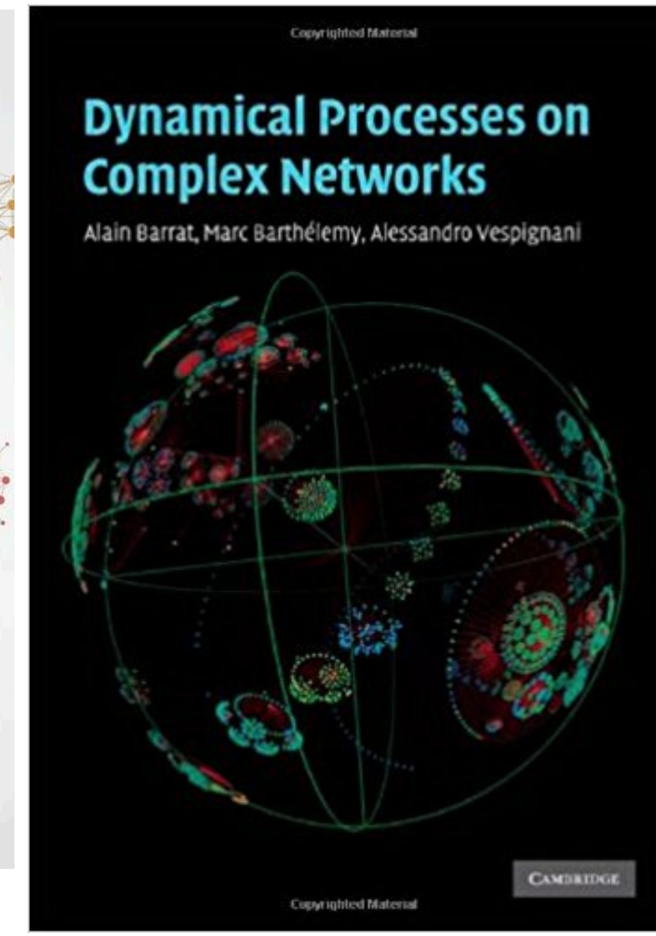
> Hidalgo (2017) Disconnected, fragmented, or united? a trans-disciplinary review of network science

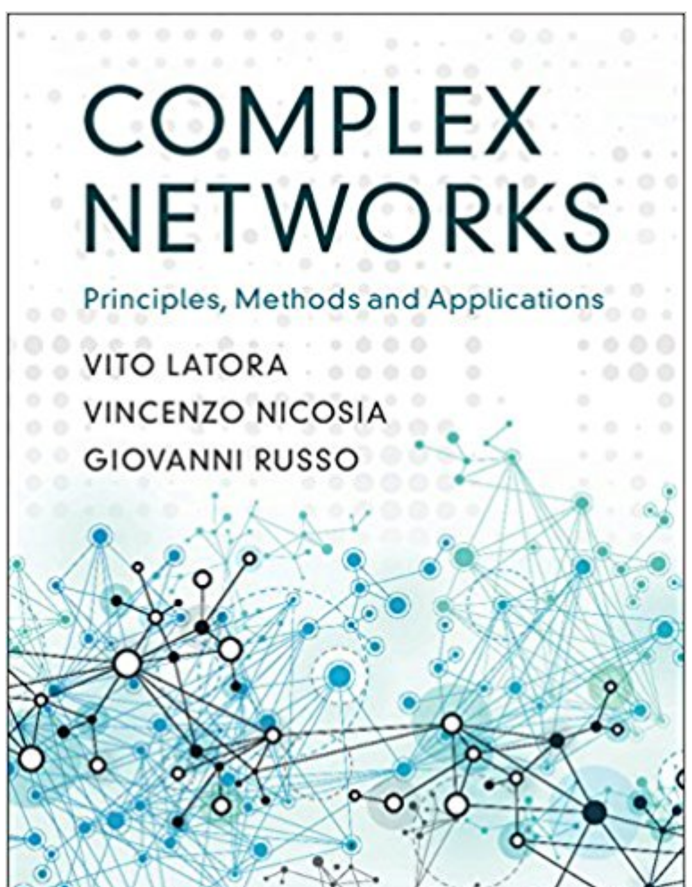
Network science?

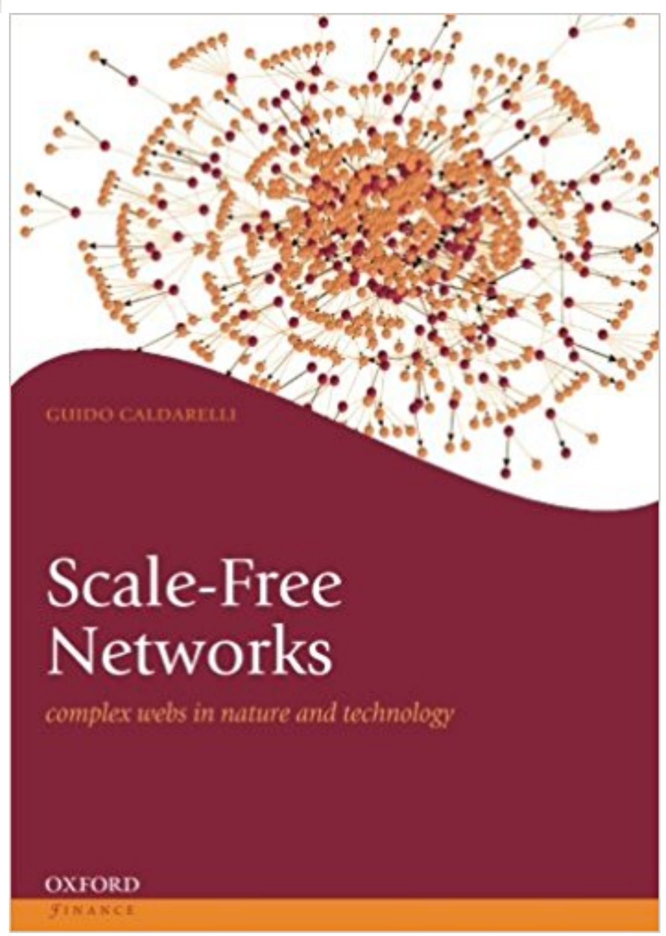


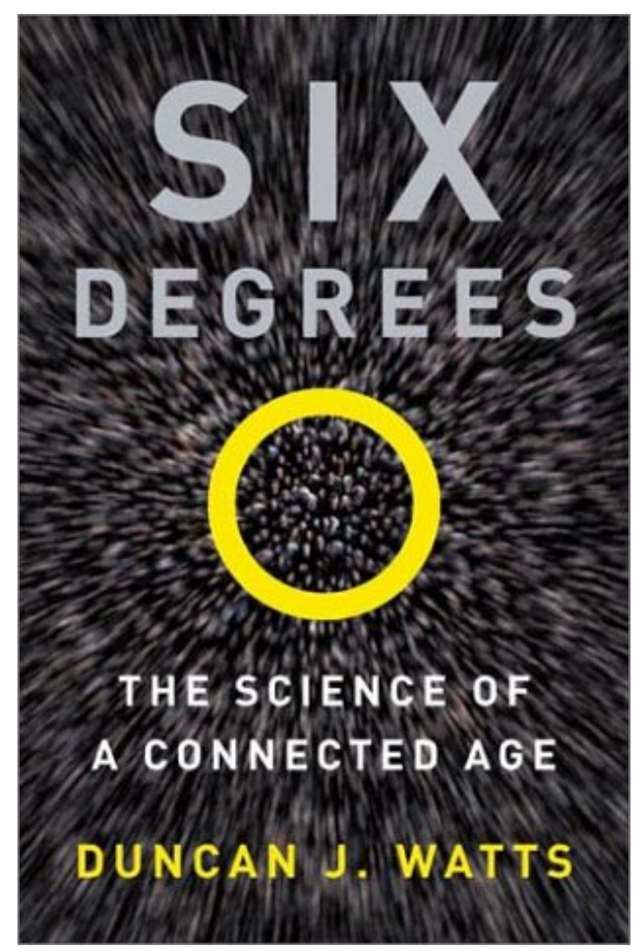


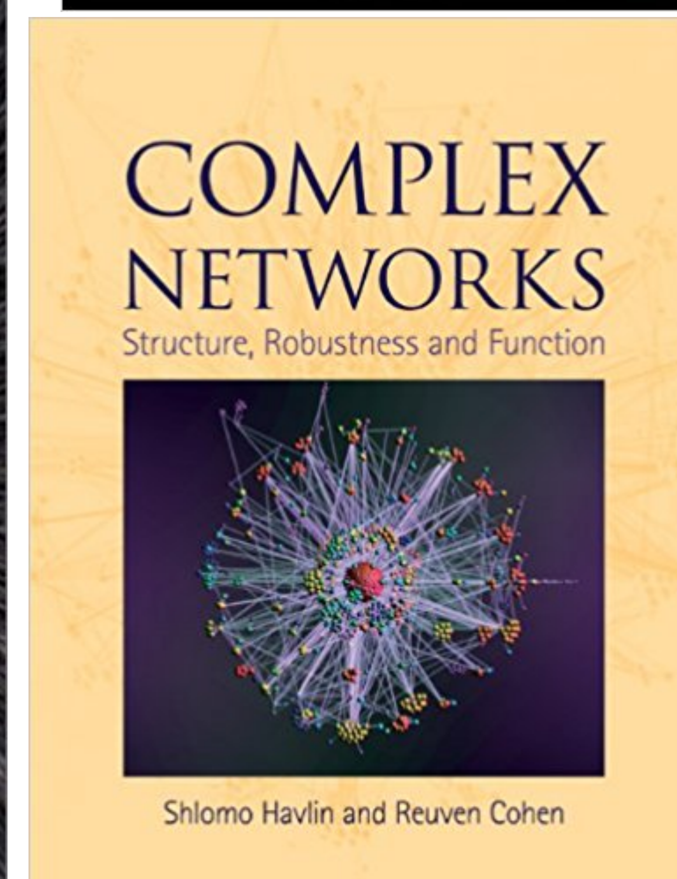




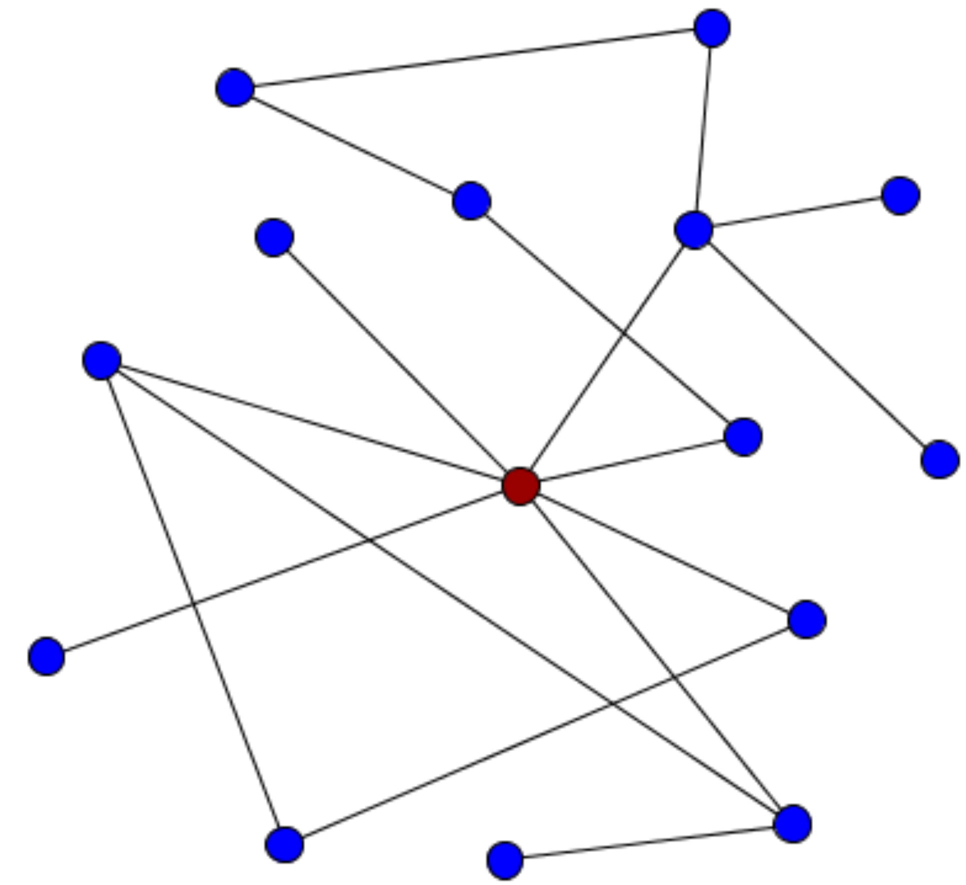








Some quantities of interest for (large) networks



Node degree: number of neighbors k_i

Node degree distribution $P(k_i)$

Clustering coefficient
(to what extent neighbors are connected)

$$C(i) = \frac{E_i}{k_i(k_i - 1)/2}$$

Distance between nodes

$$l(i,j) = \min_{paths(i \to j)} |path|$$

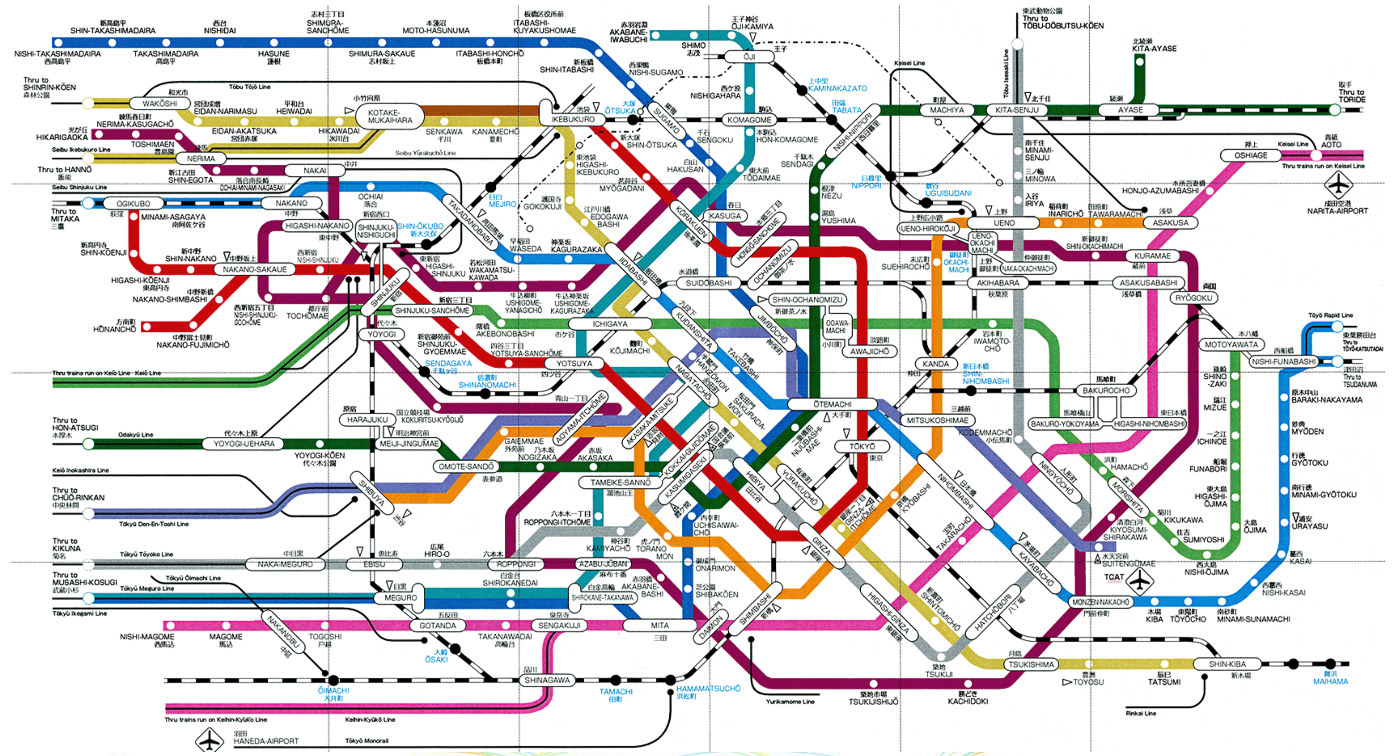
Betweenness centrality

$$g(i) = \sum_{s \neq t} \frac{\sigma_{st}(i)}{\sigma_{st}}$$

Mobility networks?

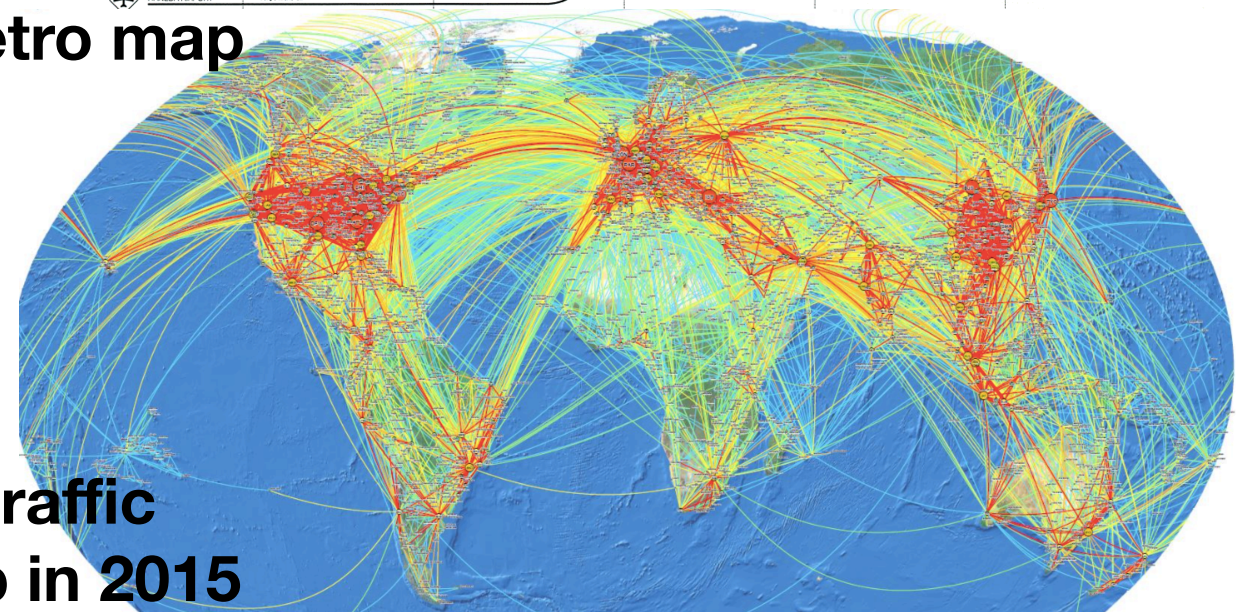
Are spatial networks

1) Infrastructure / Transport
(street/subway/bus/…)
-> multiple layers



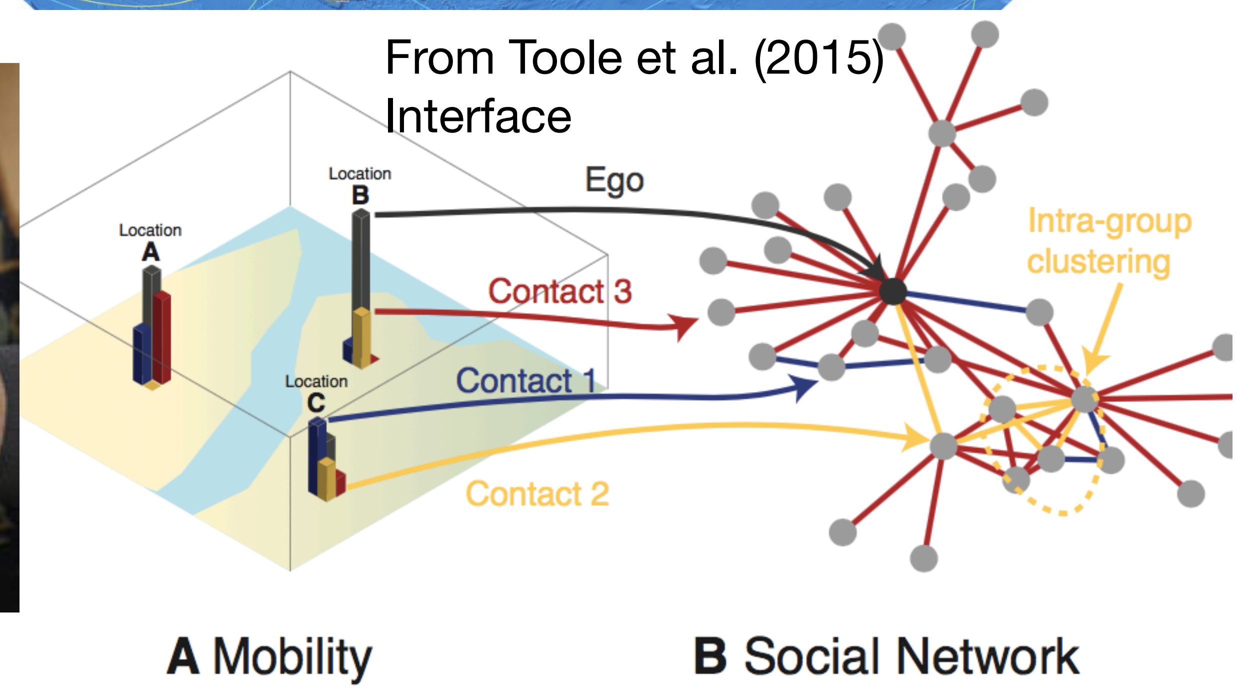
Tokyo metro map

2) Travel behavior
on top of them
(flows + individual trajectories)



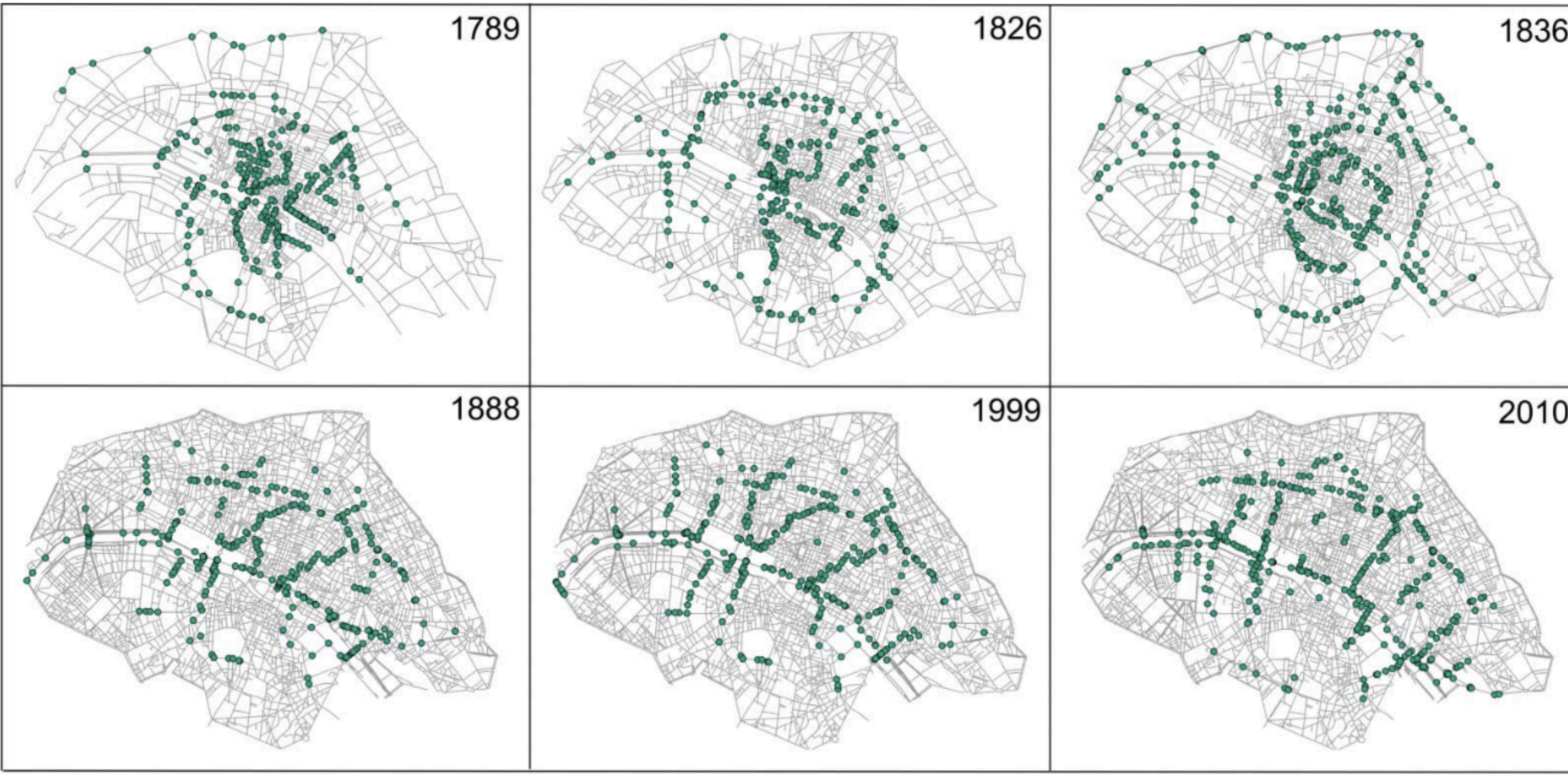
Air traffic map in 2015

3) Social networks
on the move



From Toole et al. (2015)

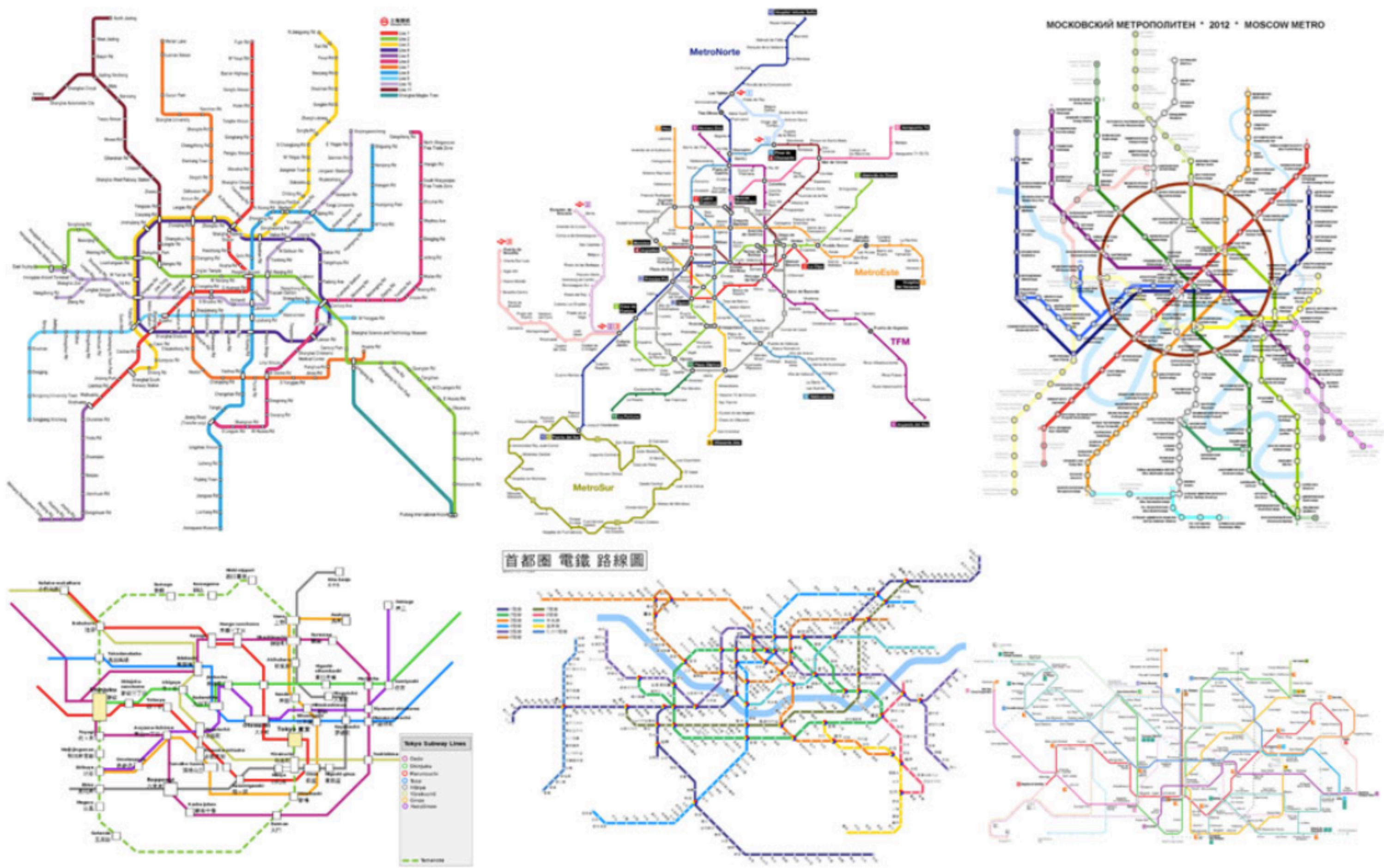
1) Transport networks - street networks

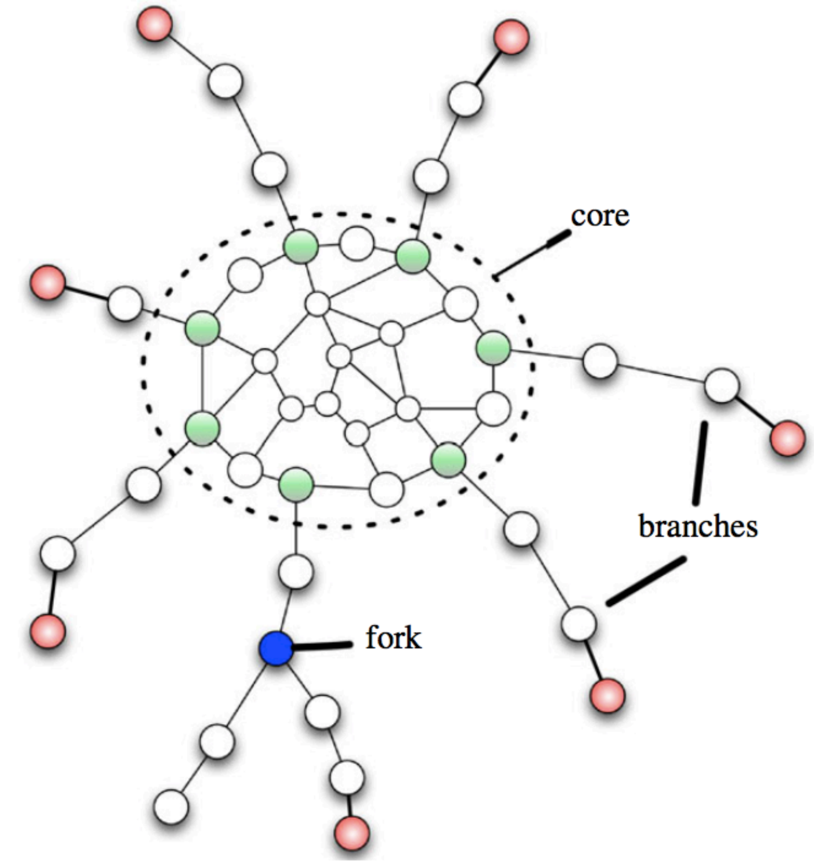


Spatial distribution of Paris most central nodes (with centrality g such that g > max(g)/10)

Barthelemy, Bordin, Berestycki and Gribaudi (2013)
Self-organization versus top-down planning in the evolution of a city, *Sci. rep.*

1) Transport networks - subway networks

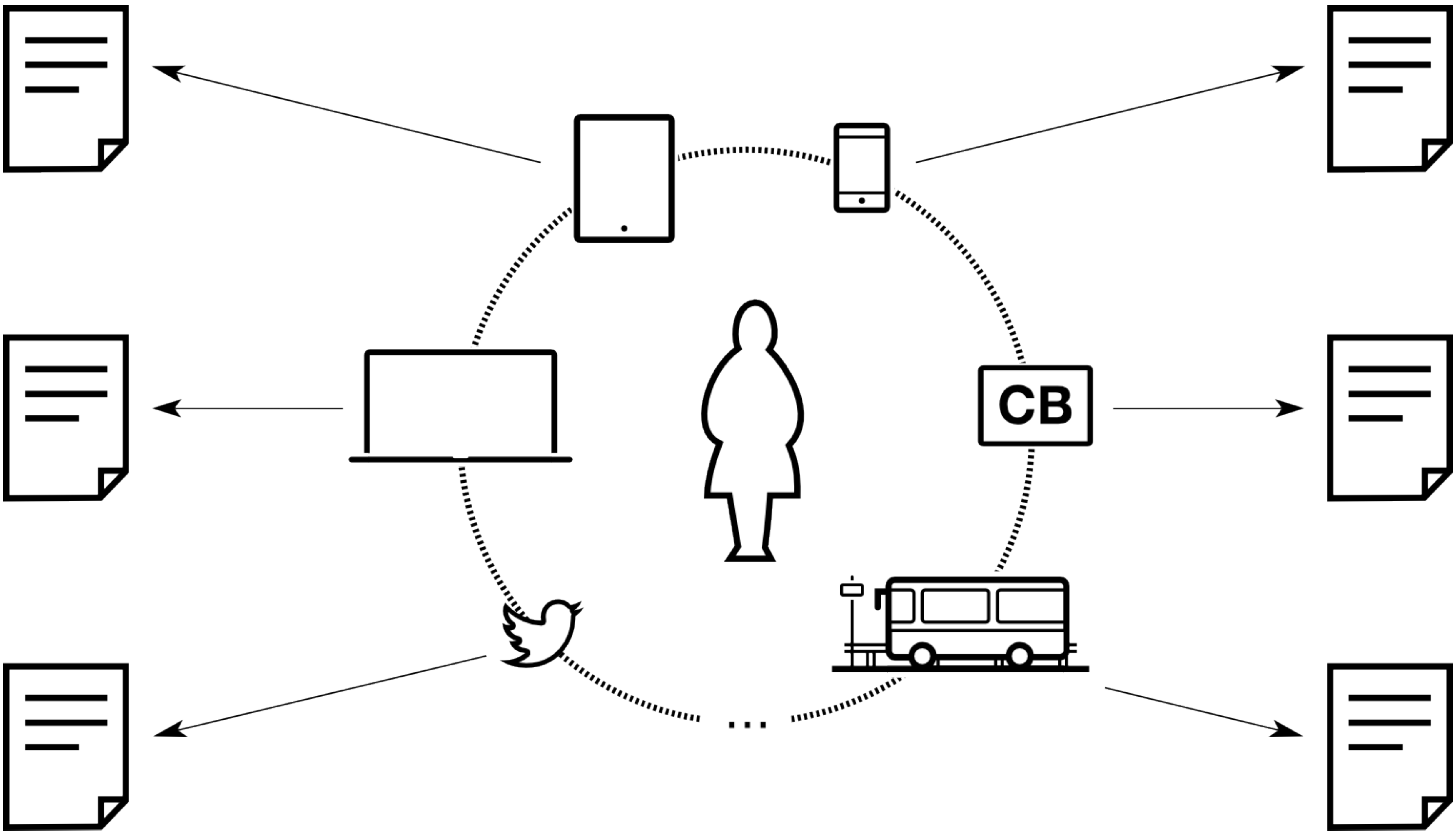




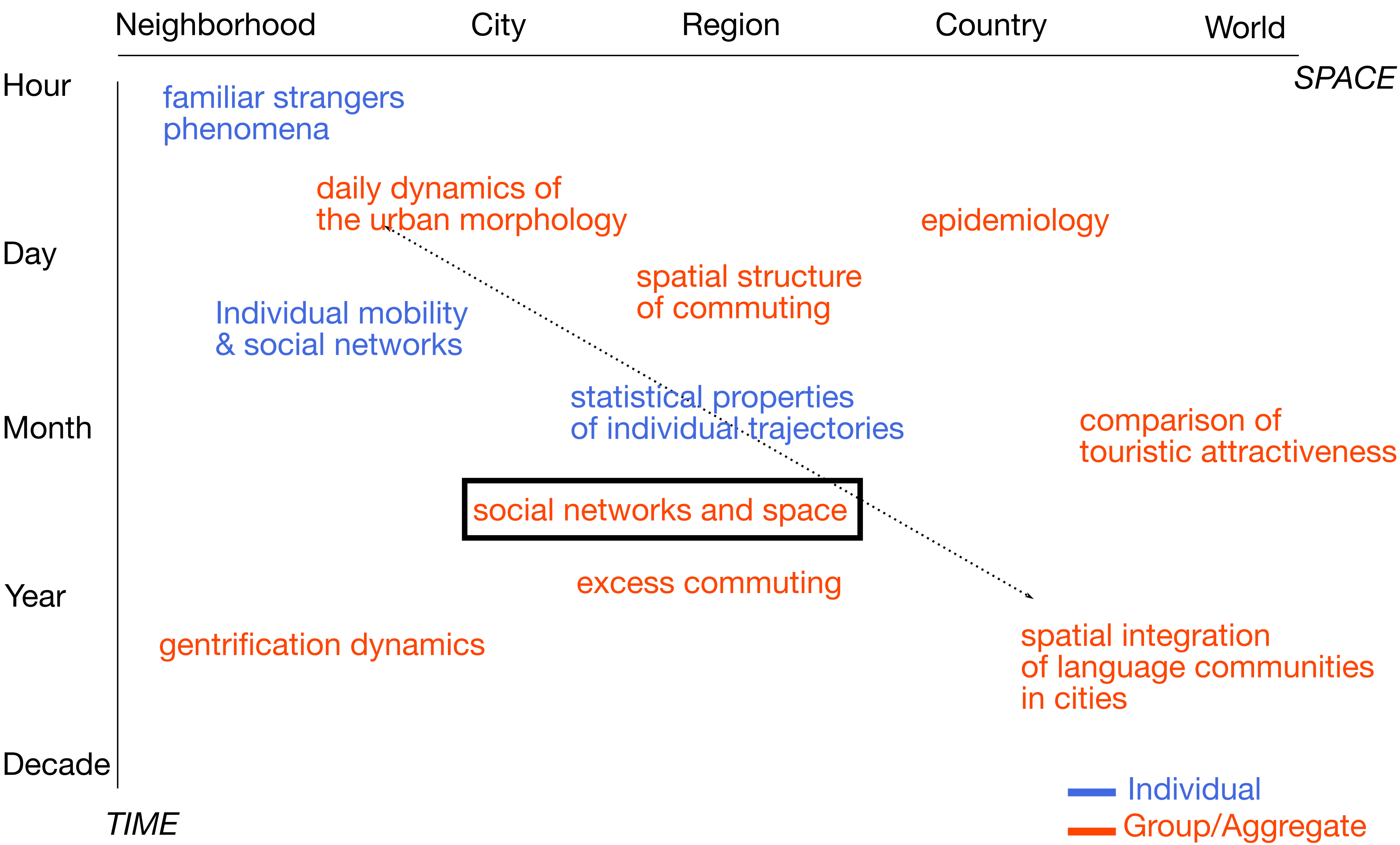
city	P (millions)	N_L	N	$\overline{\ell_1}$ (km)	ℓ_T (km)	ℓ_T/ℓ_T^{reg}	β (%)
Beijing	19.6	9	104	1.79	204	0.14	39
Tokyo	12.6	13	217	1.06	279	0.13	43
Seoul	10.5	9	392	1.39	609	0.39	38
Paris	9.6	16	299	0.57	205	0.18	38
Mexico City	8.8	11	147	1.04	170	0.15	39
NYC	8.4	24	433	0.78	373	0.12	36
Chicago	8.3	11	141	1.18	176	0.08	71
London	8.2	11	266	1.29	397	0.20	47
Shanghai	6.9	11	148	1.47	233	0.21	61
Moscow	5.5	12	134	1.67	260	0.16	71
Berlin	3.4	10	170	0.77	141	0.30	60
Madrid	3.2	13	209	0.90	215	0.42	46
Osaka	2.6	9	108	1.12	137	0.88	43
Barcelona	1.6	11	128	0.72	103	0.32	38

> Roth et al. (2012) A long time limit for world subway networks; *Interface*

ICT data



MULTI-SCALE DYNAMICS MEASURED WITH ICT DATA



Social aspects of mobility

OPEN ACCESS Freely available online PLoS one

[Bagrow & Lin, PlosOne, 2012]

Mesoscopic Structure and Social Aspects of Human Mobility

James P. Bagrow[1,2*], **Yu-Ru Lin**[3,4]

1 Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, Illinois, United States of America, **2** Center for Complex Network Research, Northeastern University, Boston, Massachusetts, United States of America, **3** College of Computer and Information Science, Northeastern University, Boston, Massachusetts, United States of America, **4** Institute for Quantitative Social Science, Harvard University, Cambridge, Massachusetts, United States of America

[Crandall et al., PNAS, 2010]

Inferring social ties from geographic coincidences

David J. Crandall[a], **Lars Backstrom**[b,1], **Dan Cosley**[c], **Siddharth Suri**[b,2], **Daniel Huttenlocher**[b], and **Jon Kleinberg**[b,3]

[a]School of Informatics and Computing, Indiana University, Bloomington, IN 47403; [b]Department of Computer Science, Cornell University, Ithaca, NY 14853; and [c]Department of Information Science, Cornell University, Ithaca, NY 14853

Edited by Ronald L. Graham, University of California, San Diego, La Jolla, CA, and approved October 25, 2010 (received for review May 16, 2010)

[Toole et al., Interface, 2015]

Coupling human mobility and social ties

Jameson L. Toole , Carlos Herrera-Yaqüe , Christian M. Schneider , Marta C. González

DOI: 10.1098/rsif.2014.1128 . Published 25 February 2015

Social aspects of mobility

[Herrera-Yagüe et al., Sci.rep., 2010]

OPEN

The anatomy of urban social networks and its implications in the searchability problem

Received: 23 October 2014
Accepted: 02 April 2015
Published: 02 June 2015

C. Herrera-Yagüe[1,2,3], C. M. Schneider[1], T. Couronné[4], Z. Smoreda[4], R. M. Benito[1,5], P. J. Zufiria[2,3] & M. C. González[1]

[Sun et al., PNAS, 2013]

PNAS

Understanding metropolitan patterns of daily encounters

Lijun Sun[a,b], Kay W. Axhausen[a,c,1], Der-Horng Lee[b], and Xianfeng Huang[a,d]

[a]Future Cities Laboratory, Singapore–ETH Centre for Global Environmental Sustainability, Singapore 138602; [b]Department of Civil and Environmental Engineering, National University of Singapore, Singapore 117576; [c]Institute for Transport Planning and Systems, Swiss Federal Institute of Technology, CH-8093 Zürich, Switzerland; and [d]State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan 430079, China

Edited by Susan Hanson, Clark University, Worcester, MA, and approved July 3, 2013 (received for review April 5, 2013)

[Sun et al., Interface, 2015]

INTERFACE

rsif.royalsocietypublishing.org

Quantifying long-term evolution of intra-urban spatial interactions

Lijun Sun[1,2], Jian Gang Jin[3], Kay W. Axhausen[1,4], Der-Horng Lee[2] and Manuel Cebrian[5]

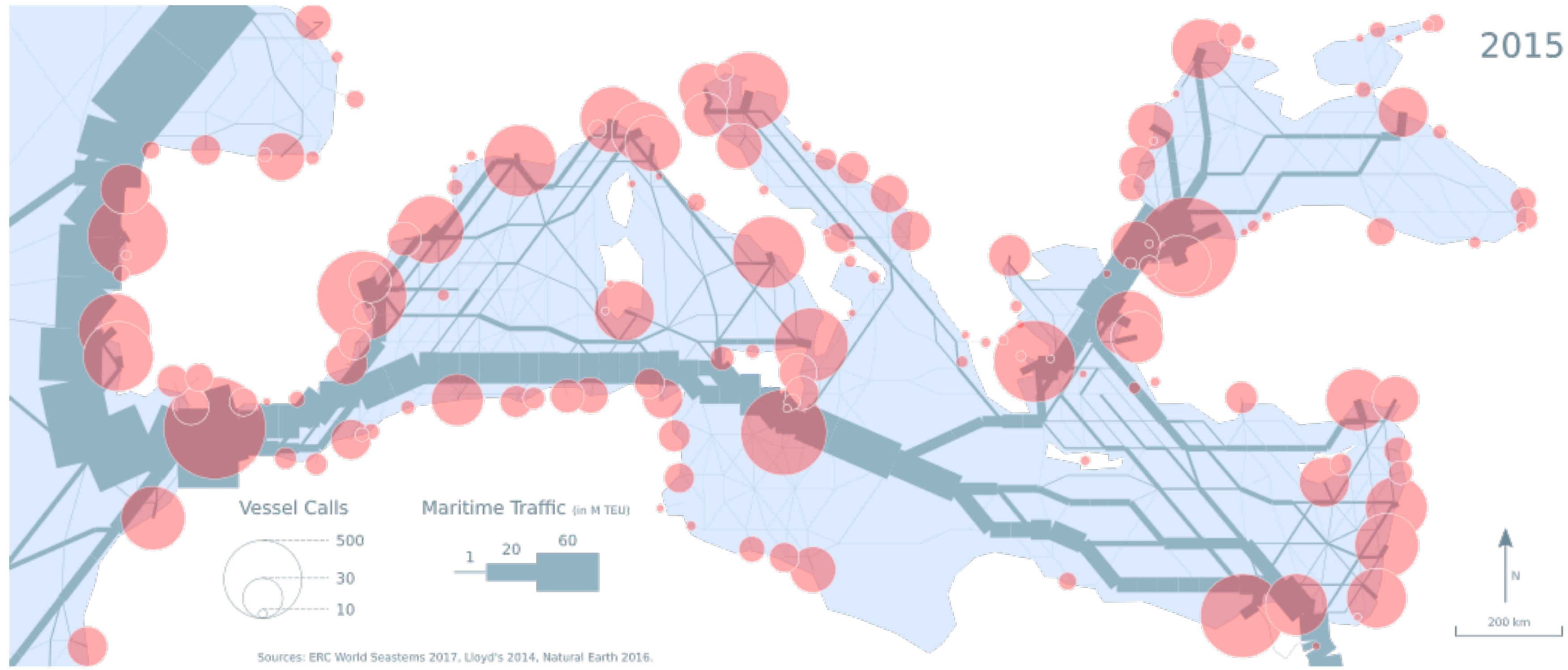
Spatial networks?

(Barthelemy, 2011, Spatial Networks, *Physics Reports*)

> Nodes and edges are embedded in space
This strongly constrains the creation of links

- Planar
(e.g. road networks, subways, mobility networks)

- Non-planar (intersecting links)
(e.g. airline net, cargo ship nets, the Internet)



M.Bunel, ERC Worldseastems (2017)

Effect of geographical space on networks properties include:

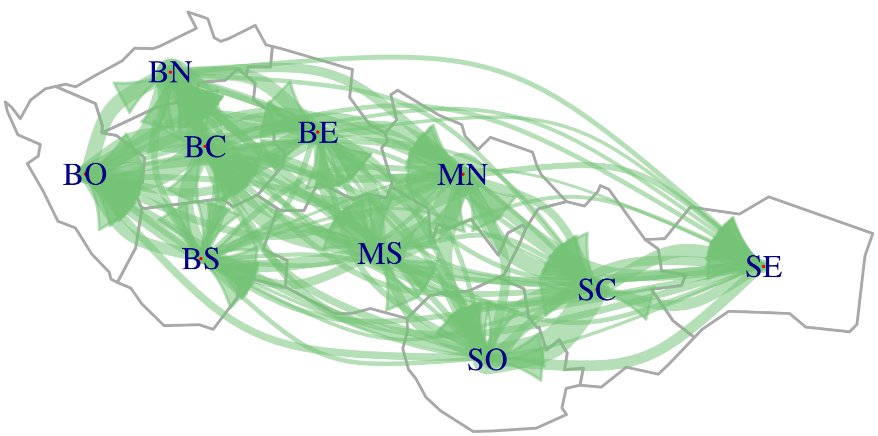
> Peaked P(k) in most cases, but can be broad for non-planar networks
> clustering coefficient ++
> $\langle l \rangle \sim N^{1/2}$
> Large betweenness centrality (BC) fluctuations;

> Ducruet and Beauguitte (2014) Spatial science and network science: Review and outcomes of a complex relationship

Uncovering the spatial structure of mobility networks

Louail et al. (2015) Nature Comm.

	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

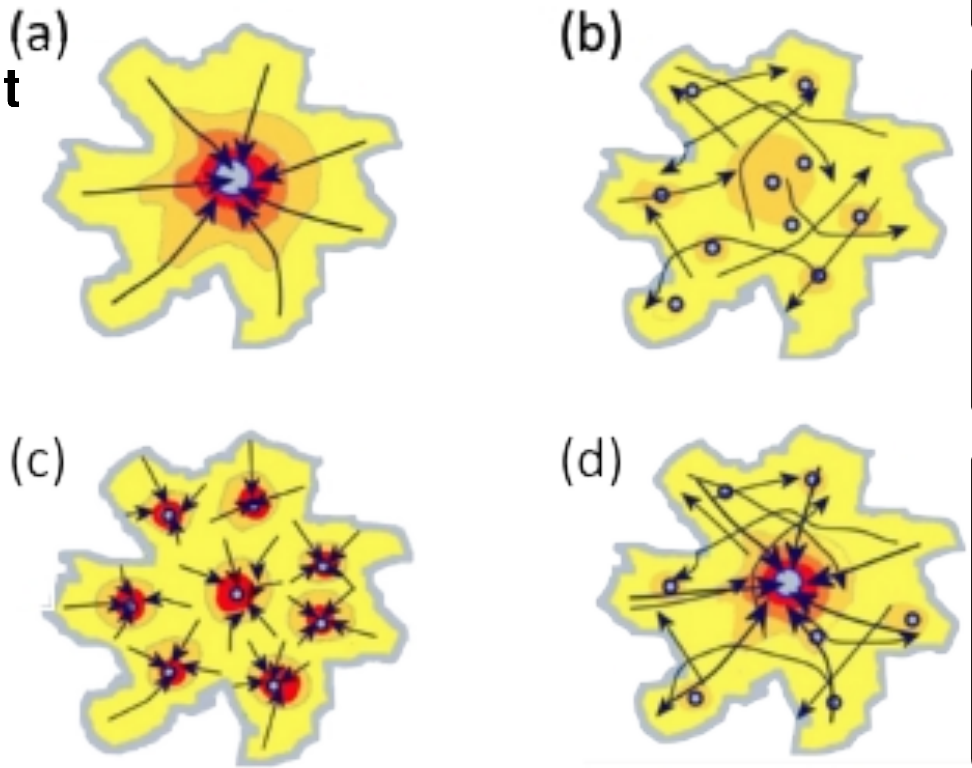
F(i,j) is the number of individuals that live in i and work in j

Provides the complete information on the commuting flows

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?

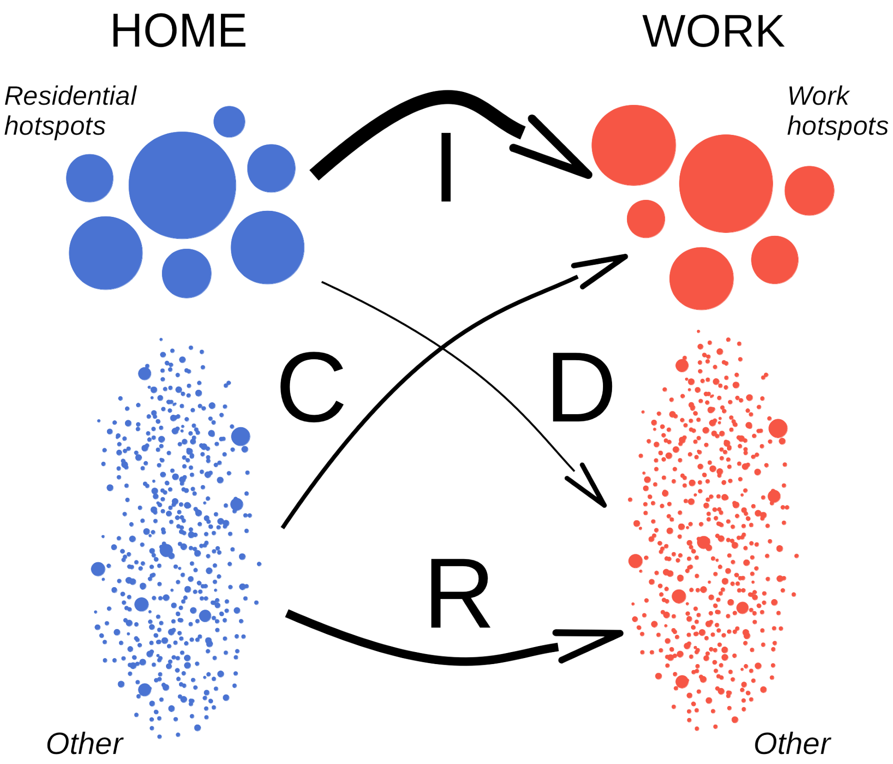
Inspiration:
schematic/simple figures
of commuting flows
in idealized city forms



(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



1.

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

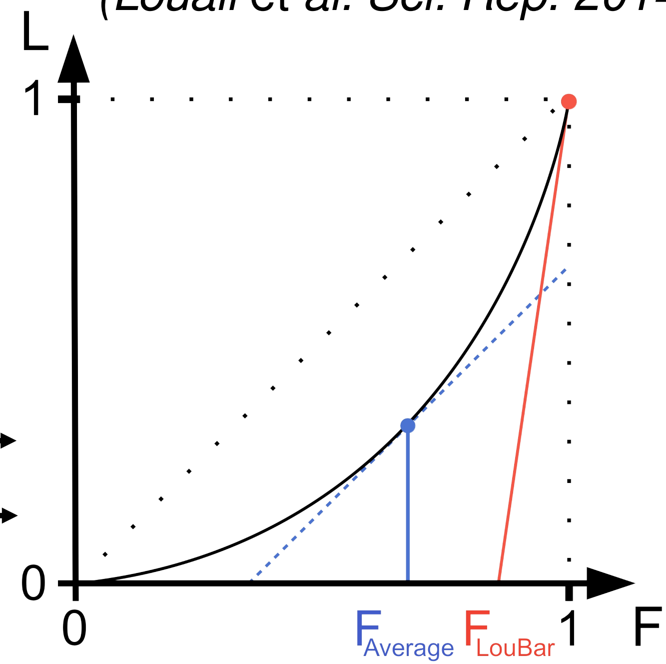
$$\sum_{i \in 1..n} C_{i,j} = C_j^{in}$$

2.

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

Determine hotspots

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



3.

$$OD = \begin{bmatrix} \sum\limits_{\substack{i \in 1..m \\ j \in 1..p}} C_{i,j} / \sum C_{i,j} & \sum\limits_{\substack{i \in 1..m \\ j \in p+1..n}} C_{i,j} / \sum C_{i,j} \\ \sum\limits_{\substack{i \in m+1..n \\ j \in 1..p}} C_{i,j} / \sum C_{i,j} & \sum\limits_{\substack{i \in m+1..n \\ j \in p+1..n}} C_{i,j} / \sum C_{i,j} \end{bmatrix}$$

I D C R

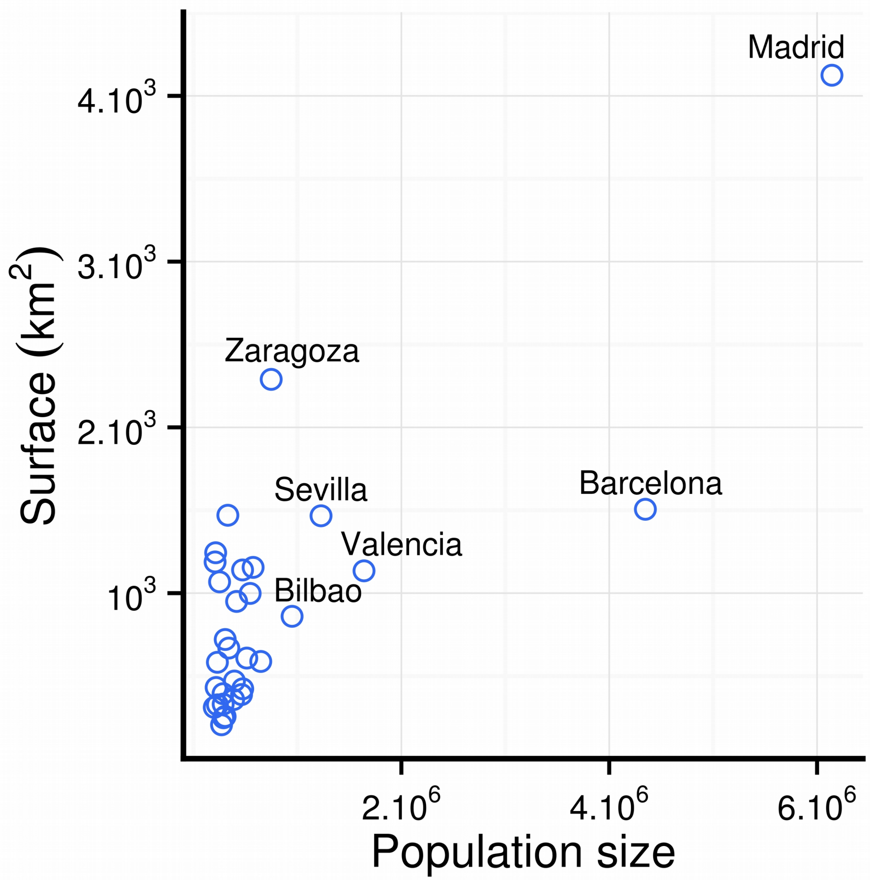
p residential (out) hotspots

m work (in) hotspots

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

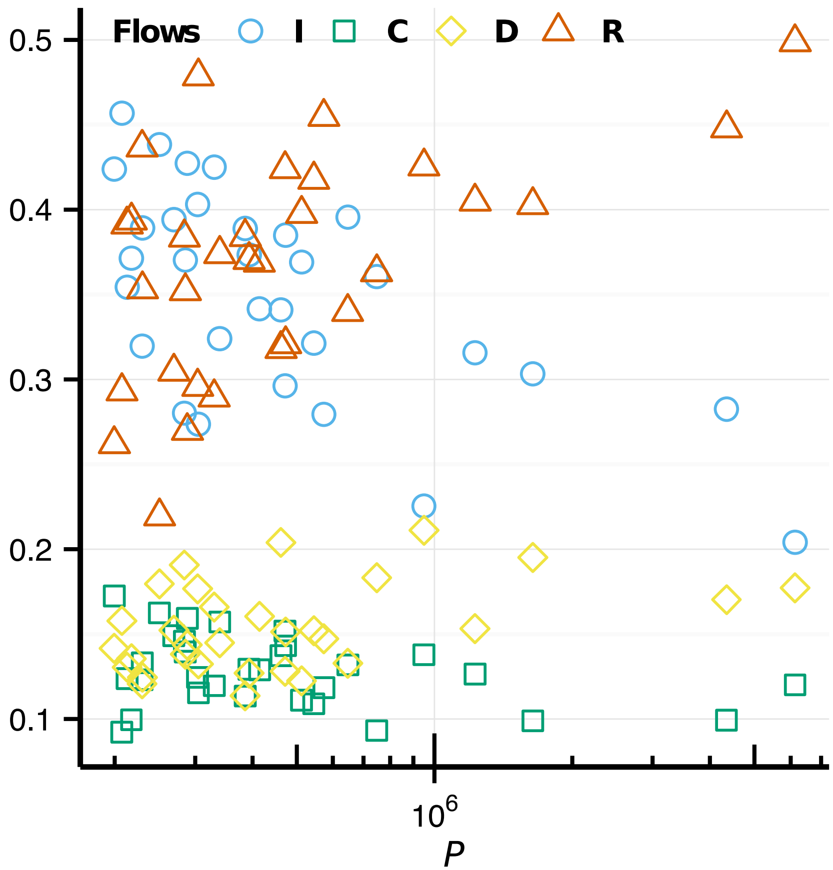
with I + D + C +R = 1

Comparing the commuting structure of 31 Spanish cities



Harmonized functionnal definition
(same for all cities)

ICDR values vs. P

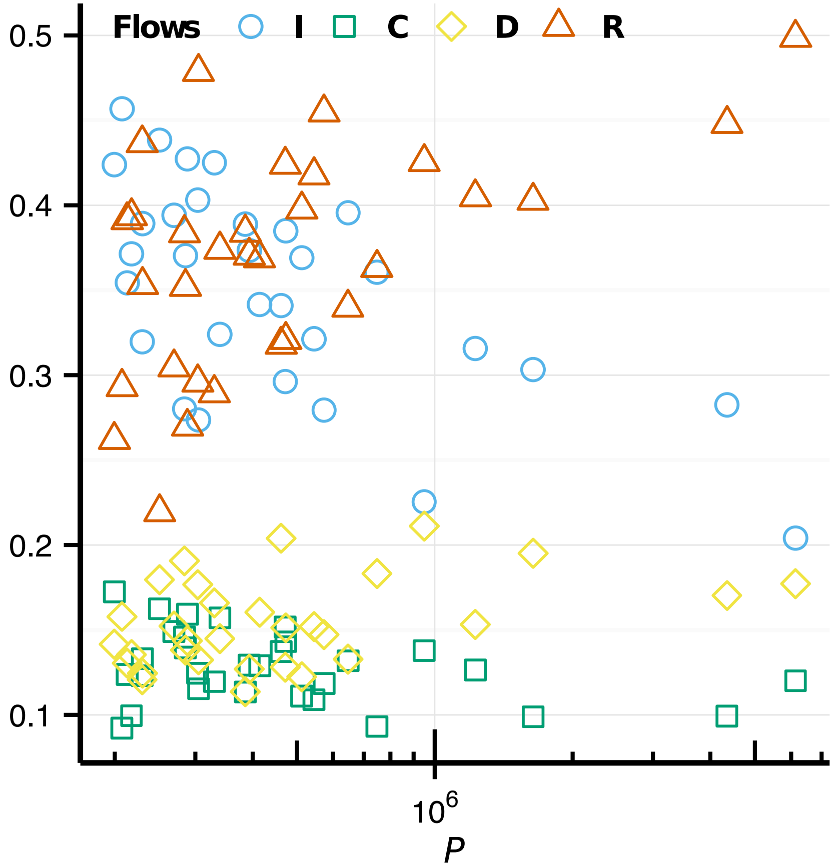


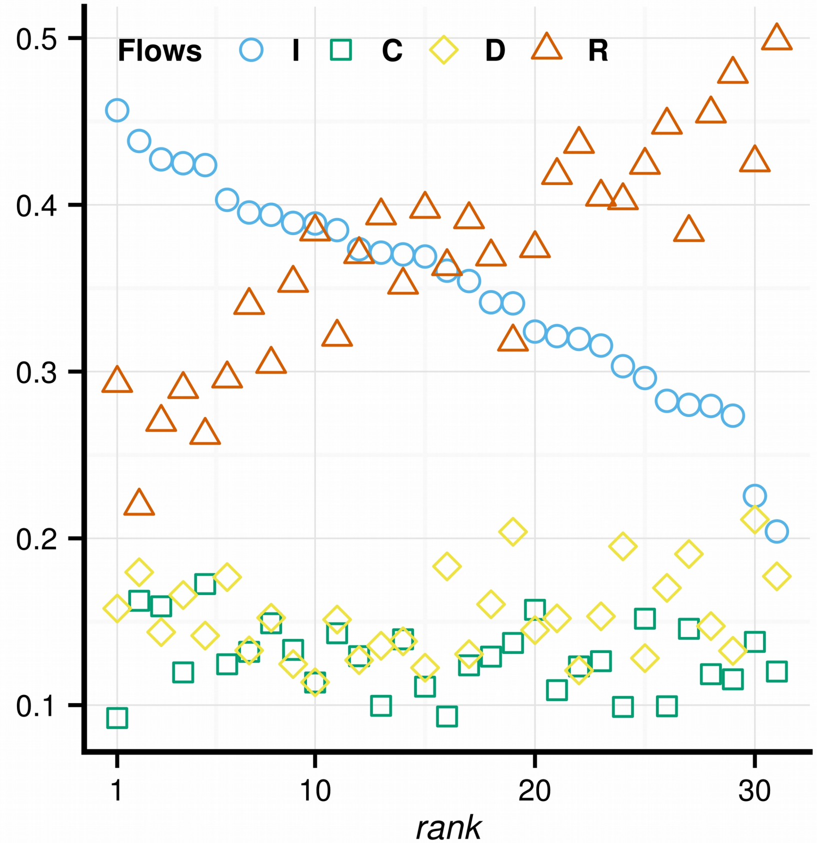
Weight of **I**ntegrated flows decreases
when population size P increases,
in favor of of **R**andom flows

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

ICDR values vs. P

Same values ranked by decreasing I



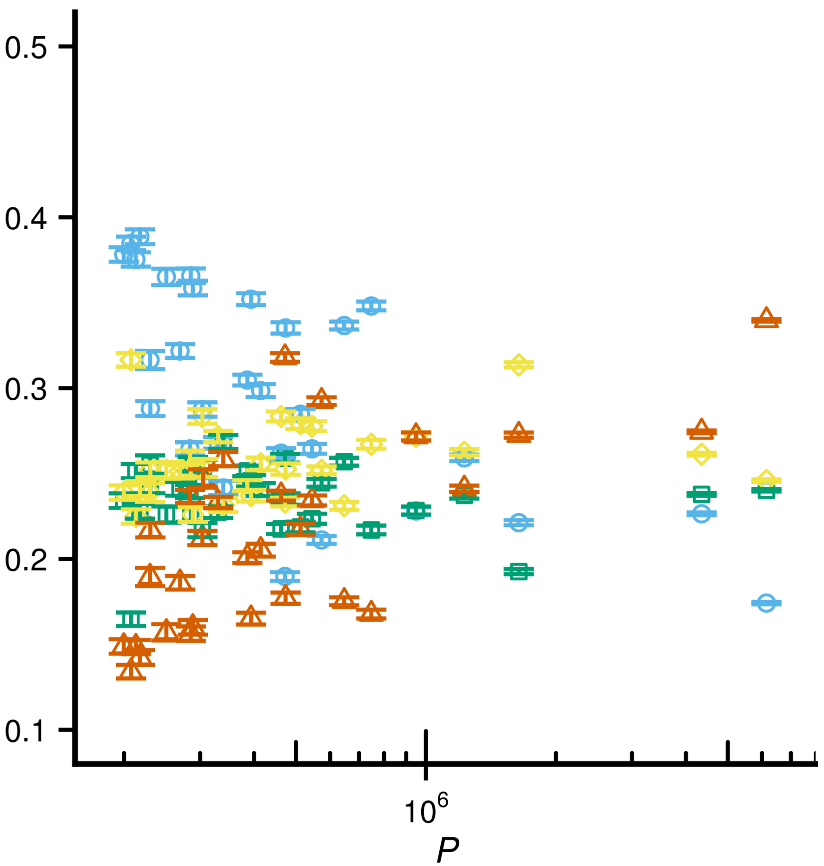


Weight of **I**ntegrated flows decreases when population size P increases, in favor of of **R**andom flows

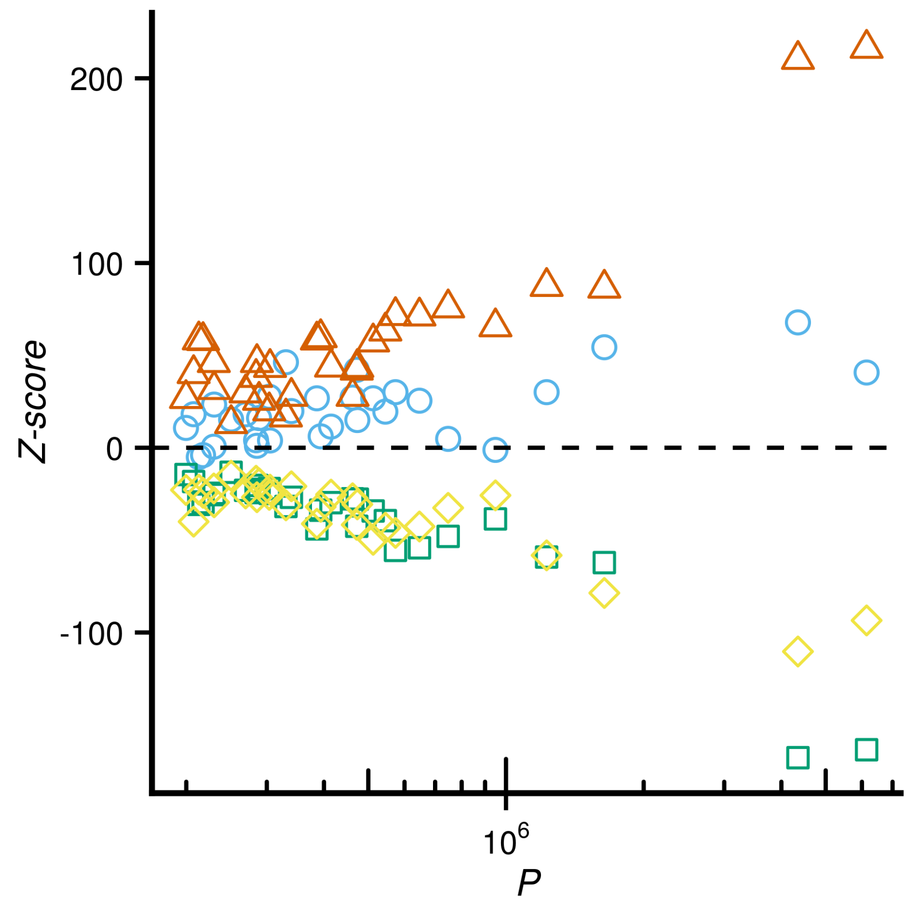
I and **R** alone seem sufficient to classify cities

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

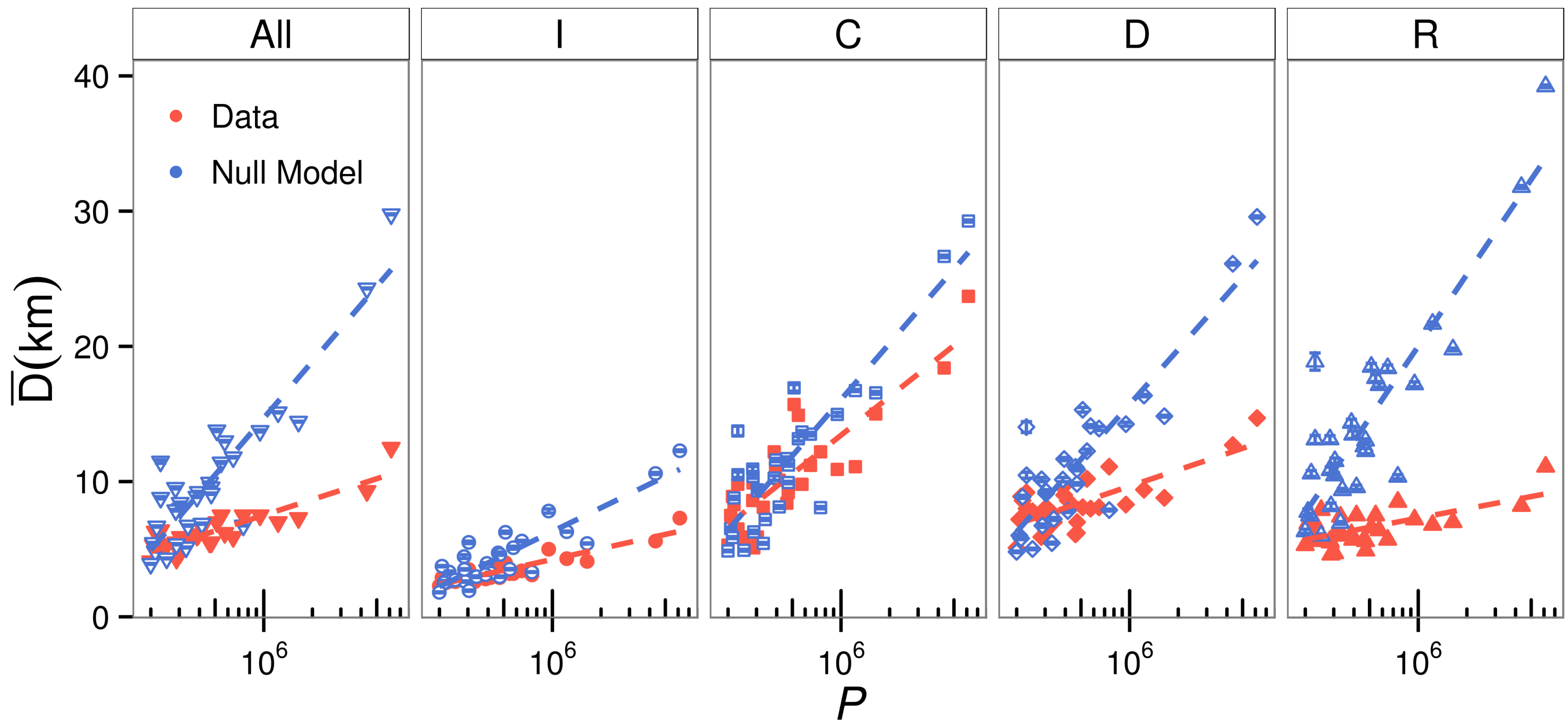
Null model : ICDR values vs. P



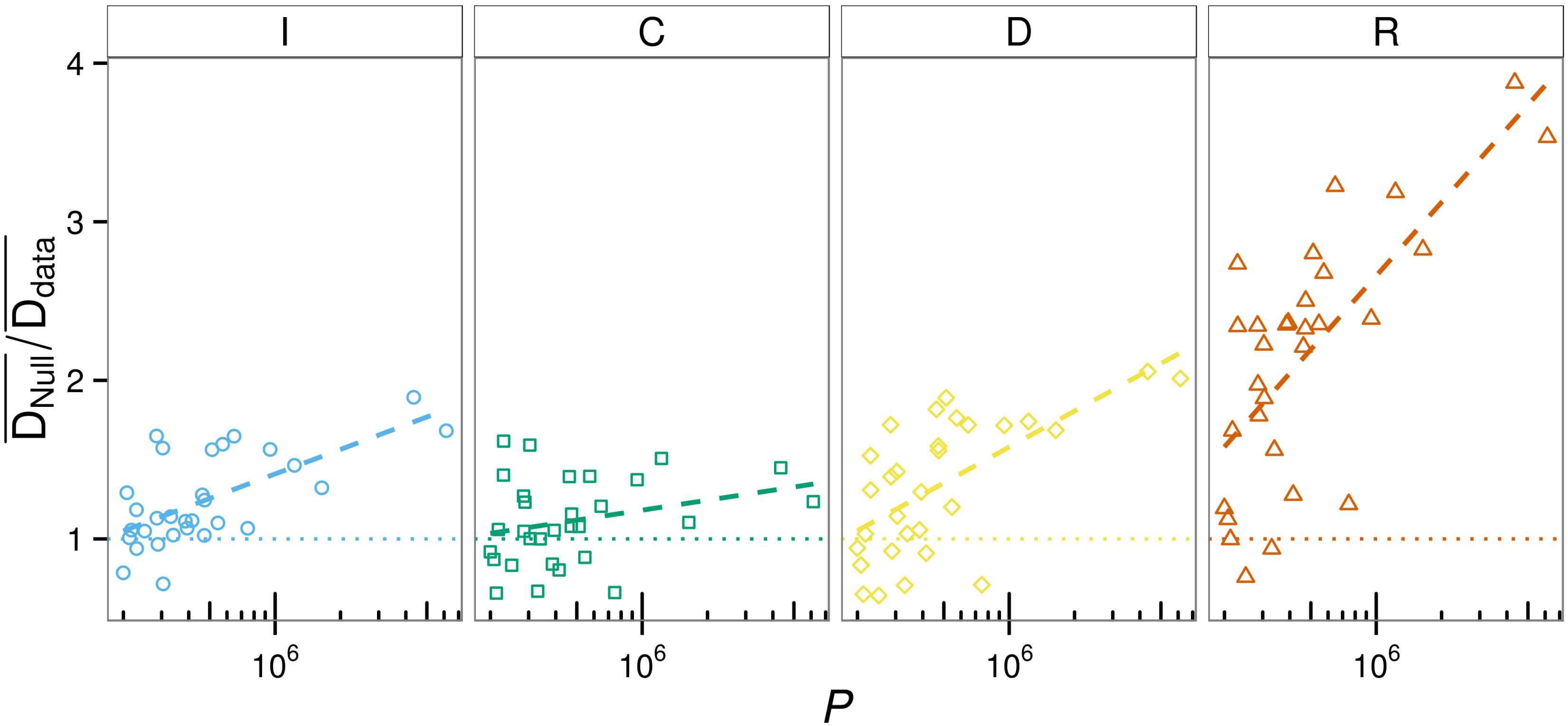
Z-scores



Average distance per type of flow



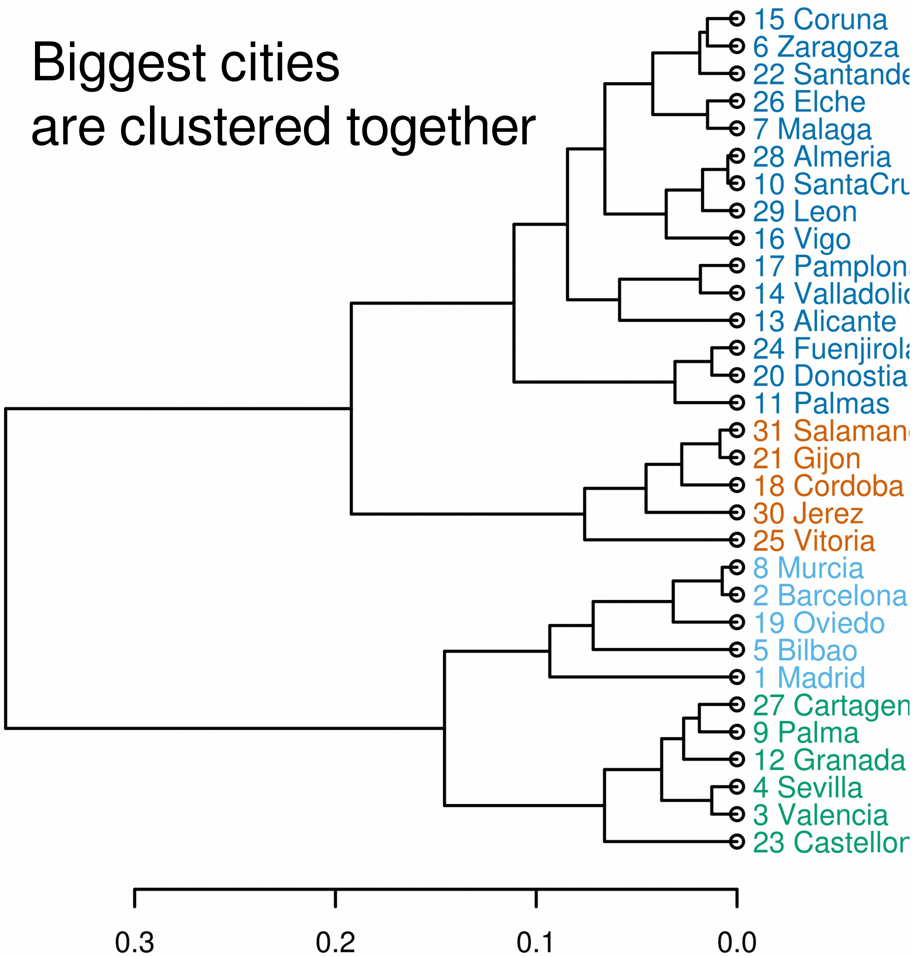
- For all types of flows distance increases with population size
- Convergent Flows (C) are the longest and the most penalized when P increases



- 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

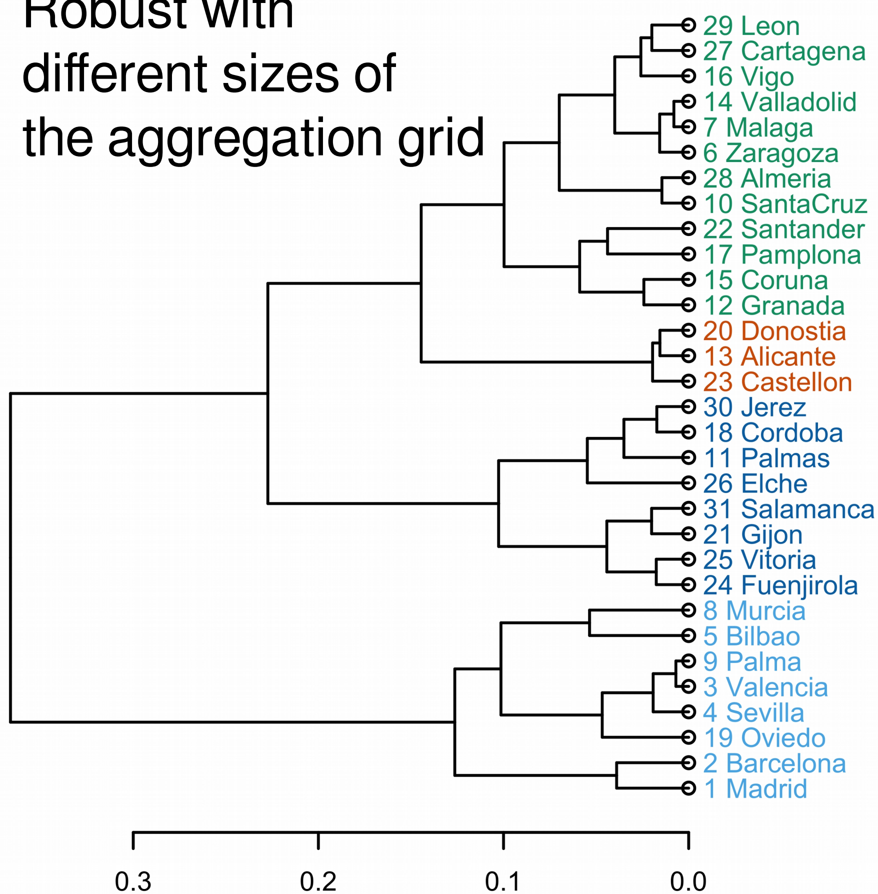
1km * 1km grid

Biggest cities
are clustered together



2km * 2km grid

Robust with
different sizes of
the aggregation grid



4 clusters of cities

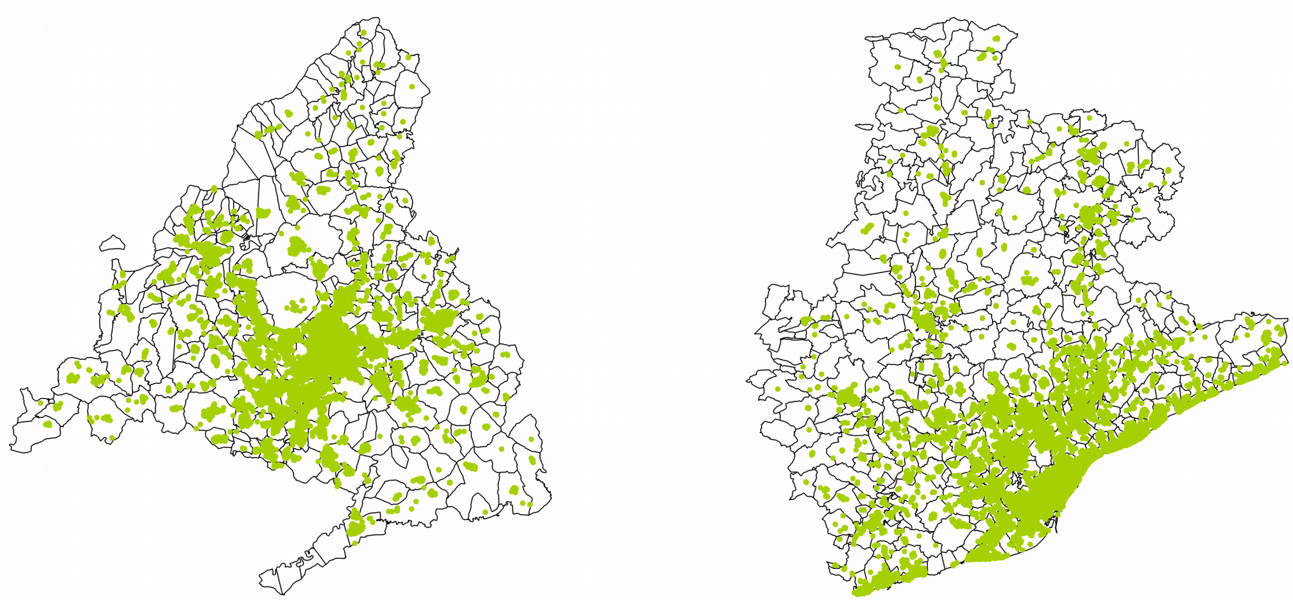
Cluster	Average population of cities in cluster	I	R	D	C
Cordoba, Gijon, Vitoria, etc	255,330	**0.43**	**0.27**	0.16	0.14
Zaragoza, Malaga, etc.	392,970	**0.37**	**0.36**	0.15	0.13
Valencia, Sevilla, etc.	732,992	**0.31**	**0.41**	0.16	0.13
Madrid Barcelona, etc.	2,463,551	**0.25**	**0.46**	0.17	0.12

Crowdsourcing the Robin Hood effect in cities

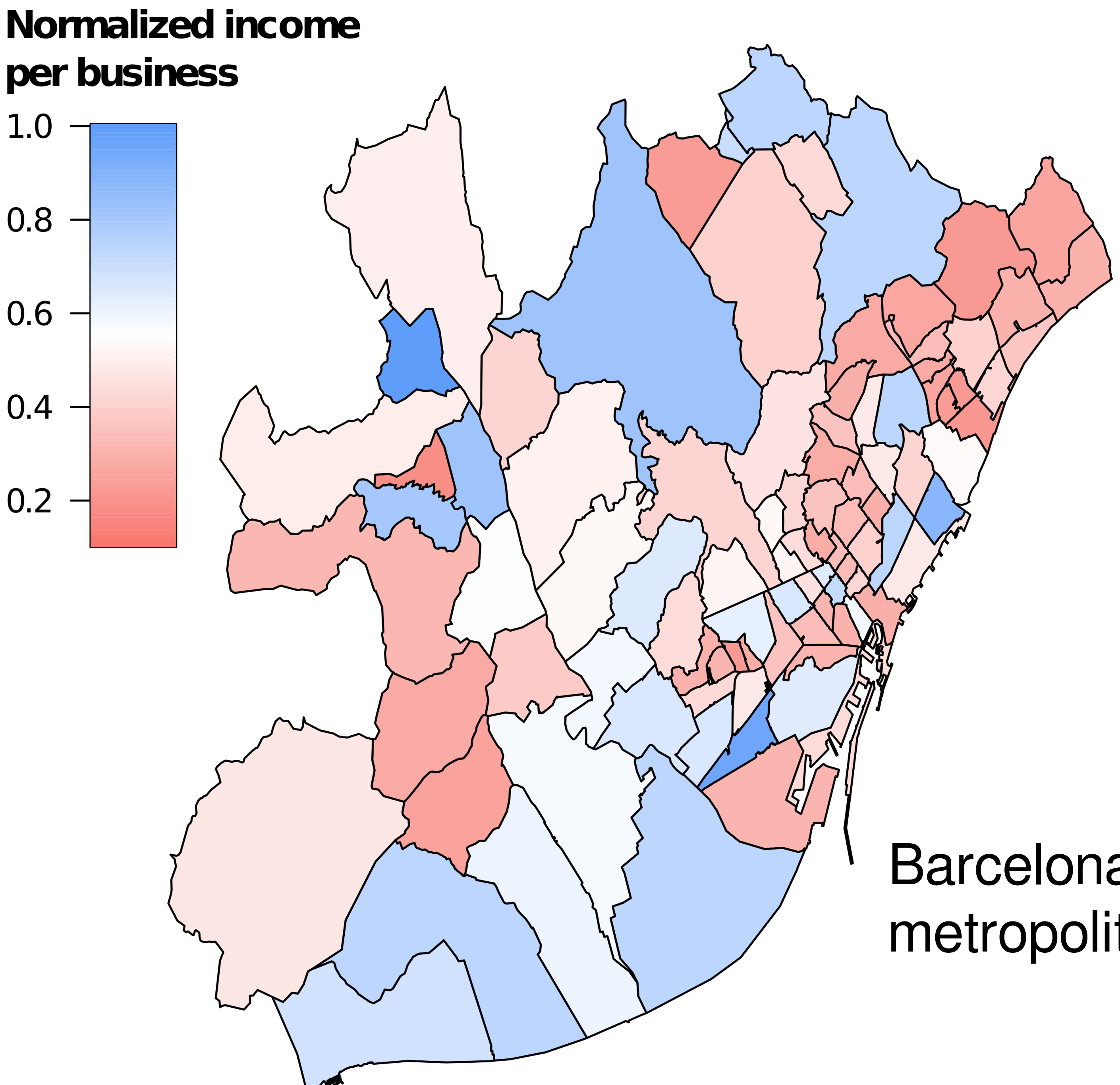
Louail et al. (2016) arxiv:1604.08394

Provinces of Madrid and Barcelona

- 130 M of transactions
- 3.5 M of anonymized BBVA customers
- 320 000 businesses classified in 16 categories
- Each transaction contains the entire information (customer, business, amount, date)

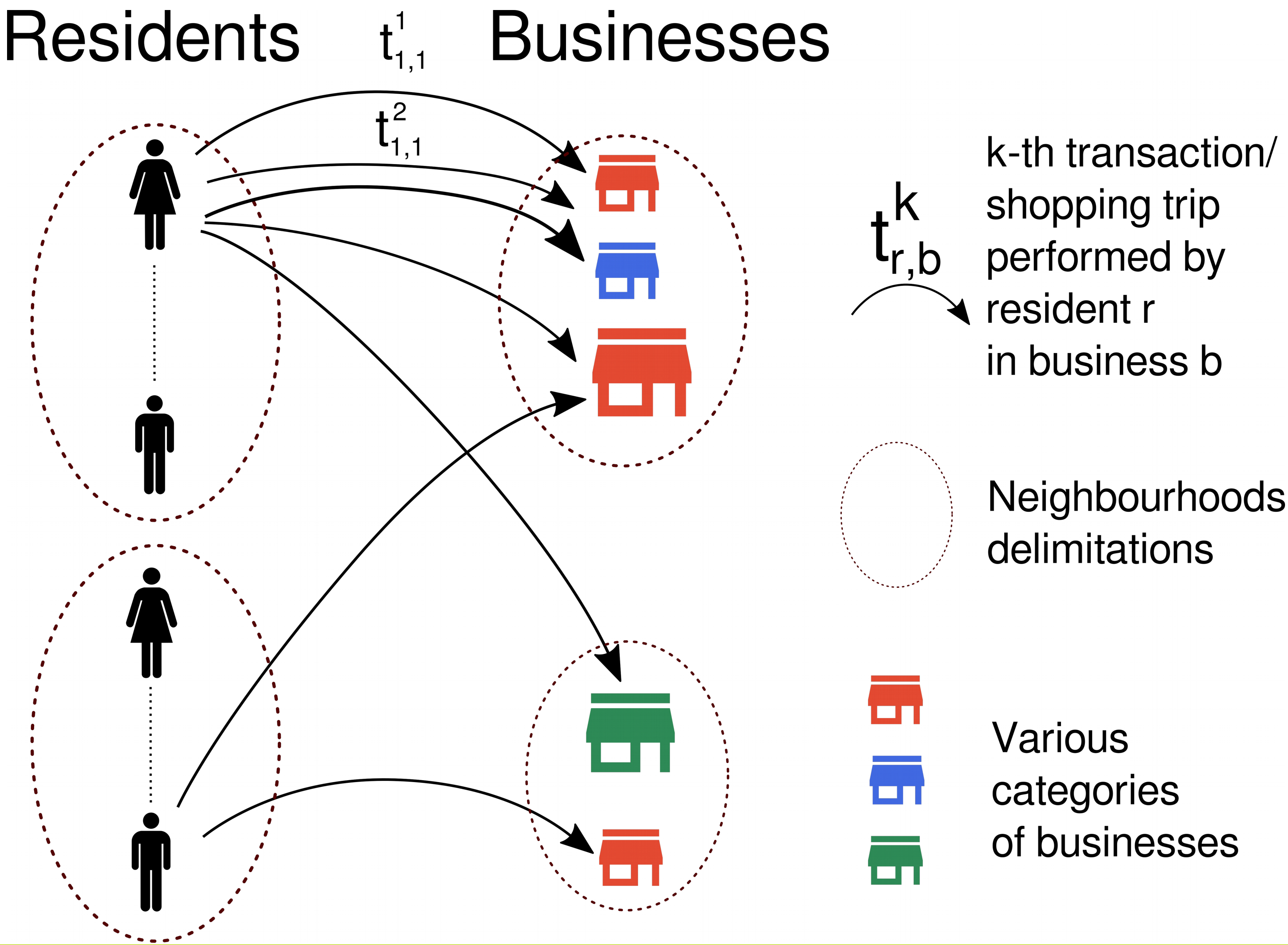


Spatial distribution of business income

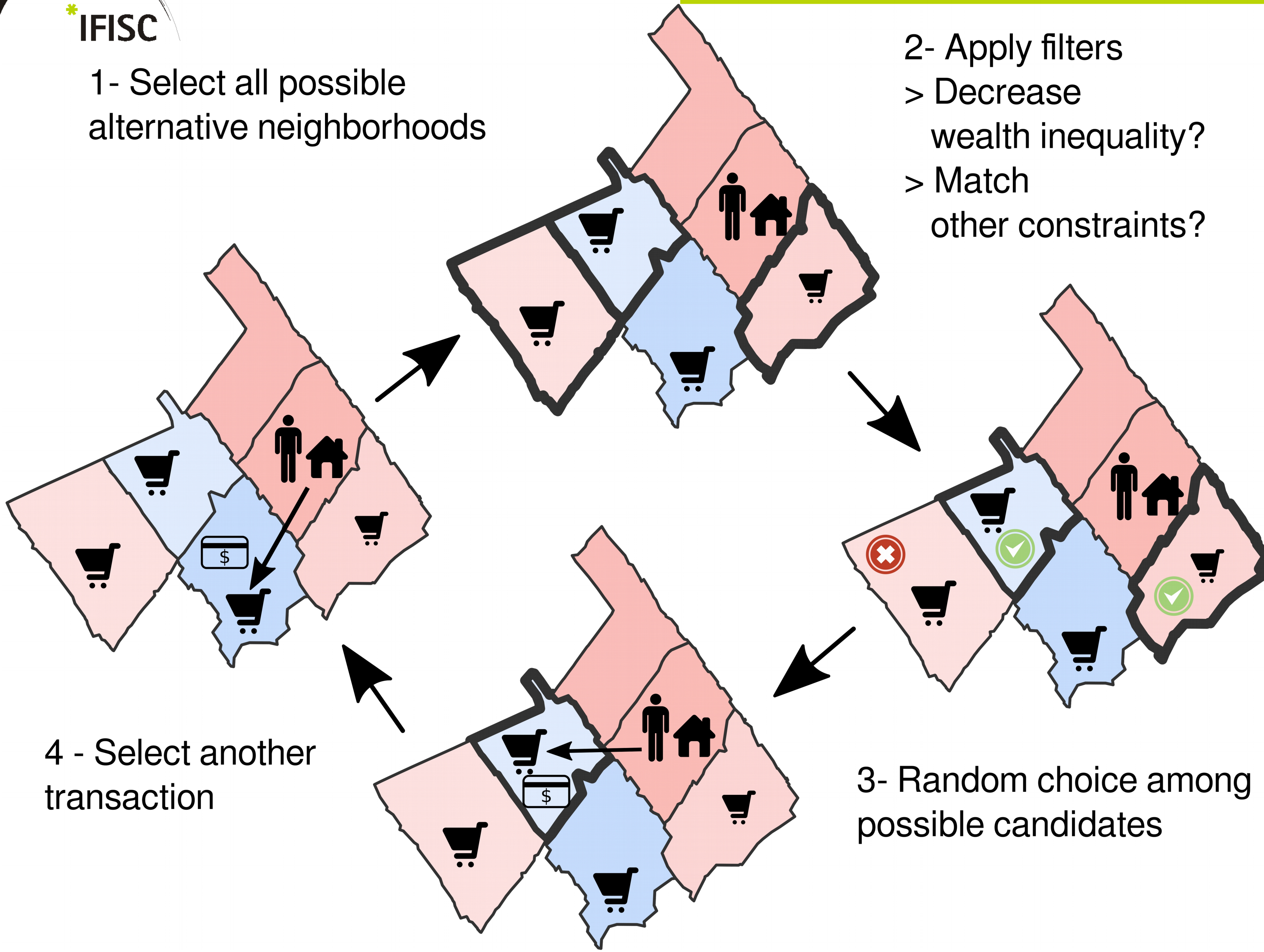


Barcelona metropolitan area

Graph representation of credit card transactions



Principles of rewiring methods



In addition to the spatial distribution of business income
and its distance to the egalitarian repartition (W),
we also take into consideration:

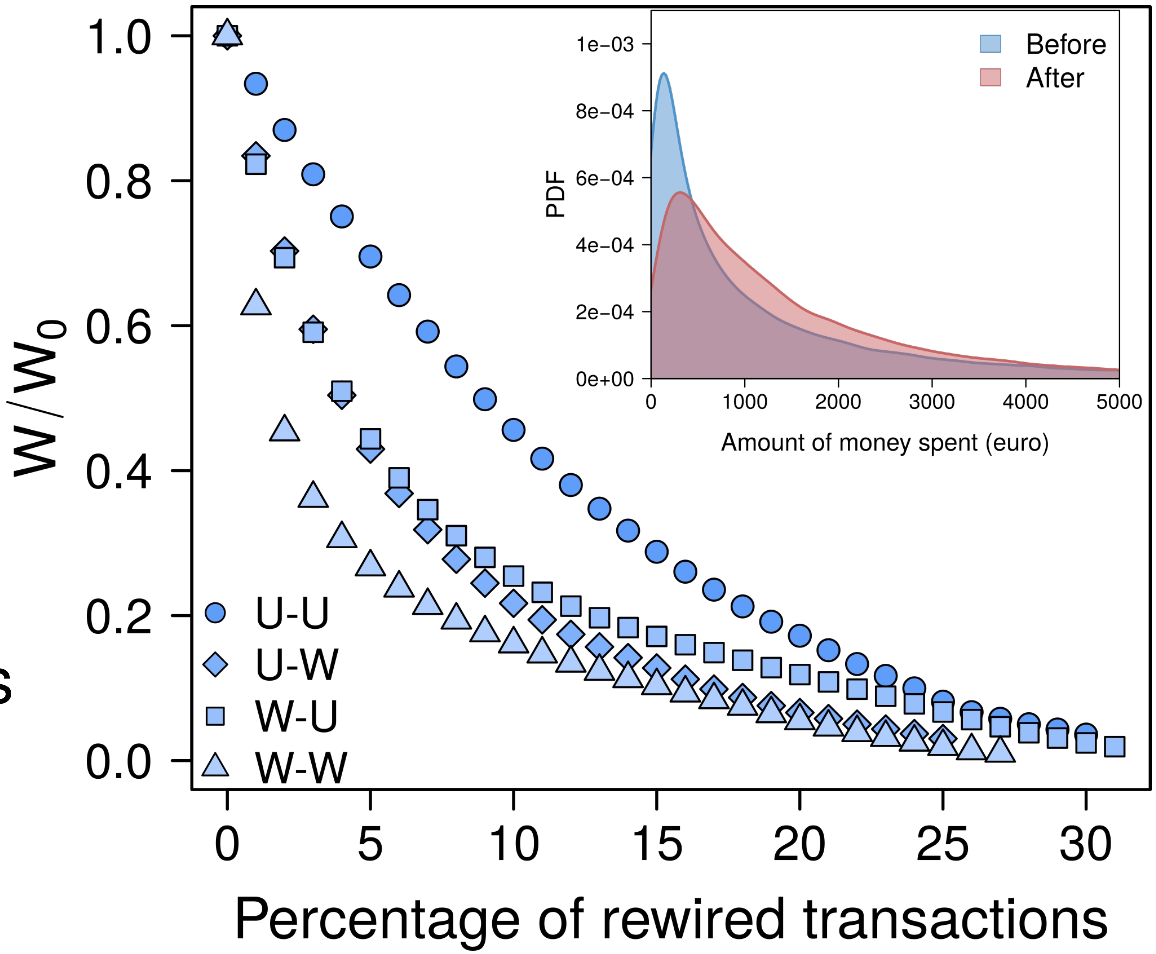
- **The distance traveled** (D)
- **The spatial routines of individuals** (ρ)
- **The spatial mixing of individuals** residing in different parts of the city evaluated as the distance to a « fully mixed » city (S)

Goal was
- To homogenize the spatial distribution of business income
- To preserve other key aspects

> **Wealth inequalities between neighborhoods are reduced by 95 %**

Many possible rewiring methods

« Clever » methods perform better
→ Room for optimization



To go further

Ghoshal et al. (2017) Human mobility: models and applications
Available on arxiv.org by the end of september

Thank you.