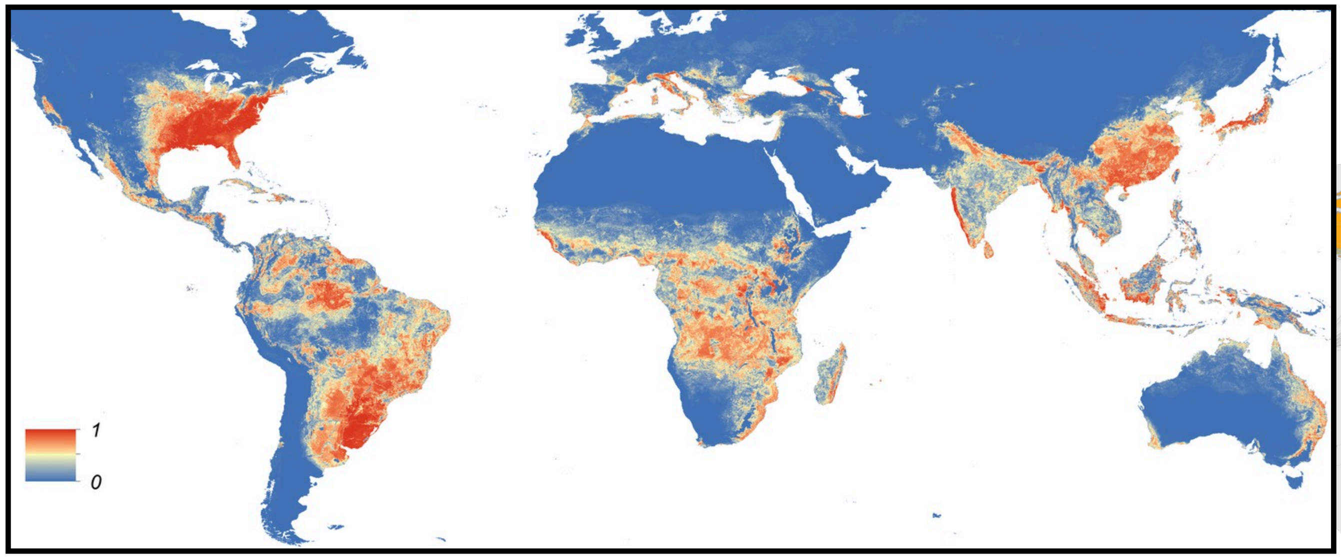
# LIFE DATA EPIDEMIOLOGY

*lect. 2: Introduction to metapopulation models*

Chiara Poletto
polettoc@gmail.com

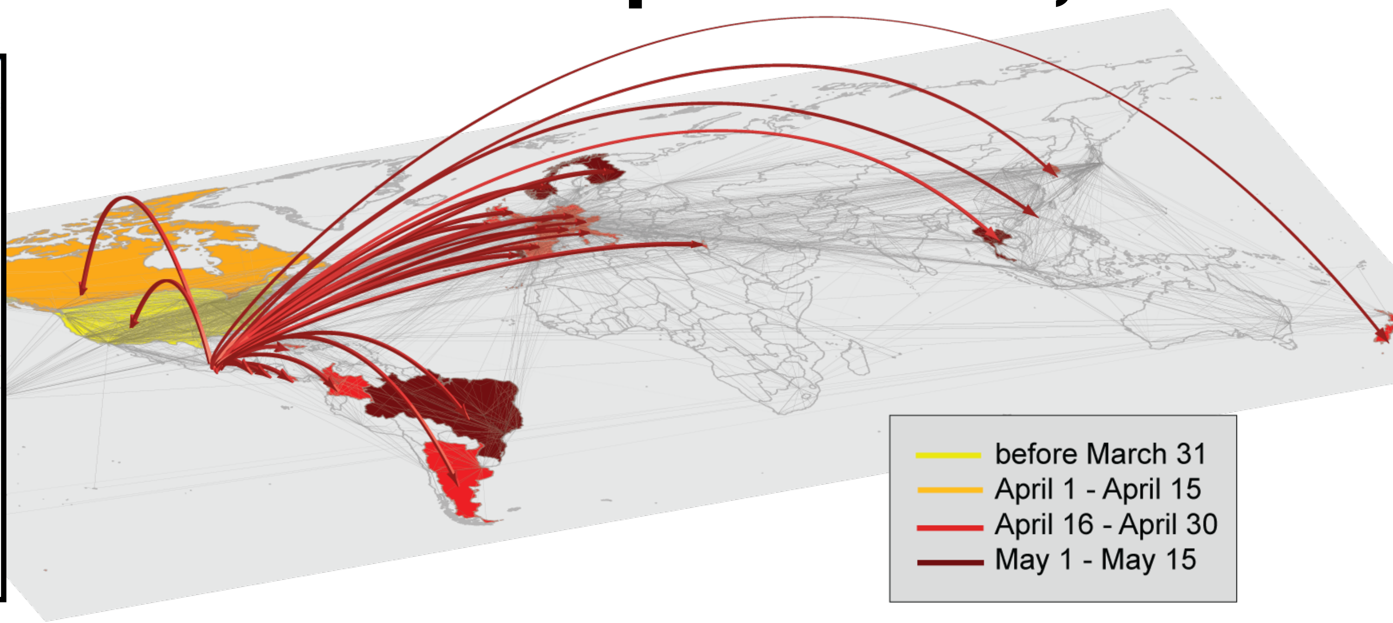
# Spatial spread of epidemics

**Aedes-transmitted infections**




[Kraemer et al eLife, 2015]

**flu pandemic, 2009**



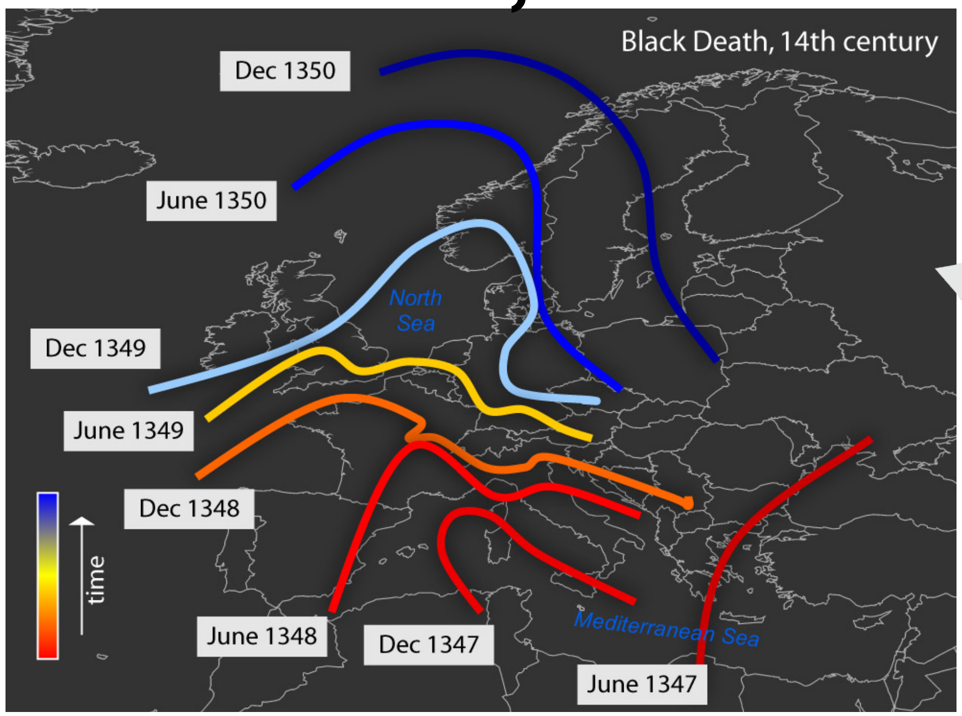

[Bajardi et al PLOS ONE, 2011]

**Drivers of spatial transmission:**

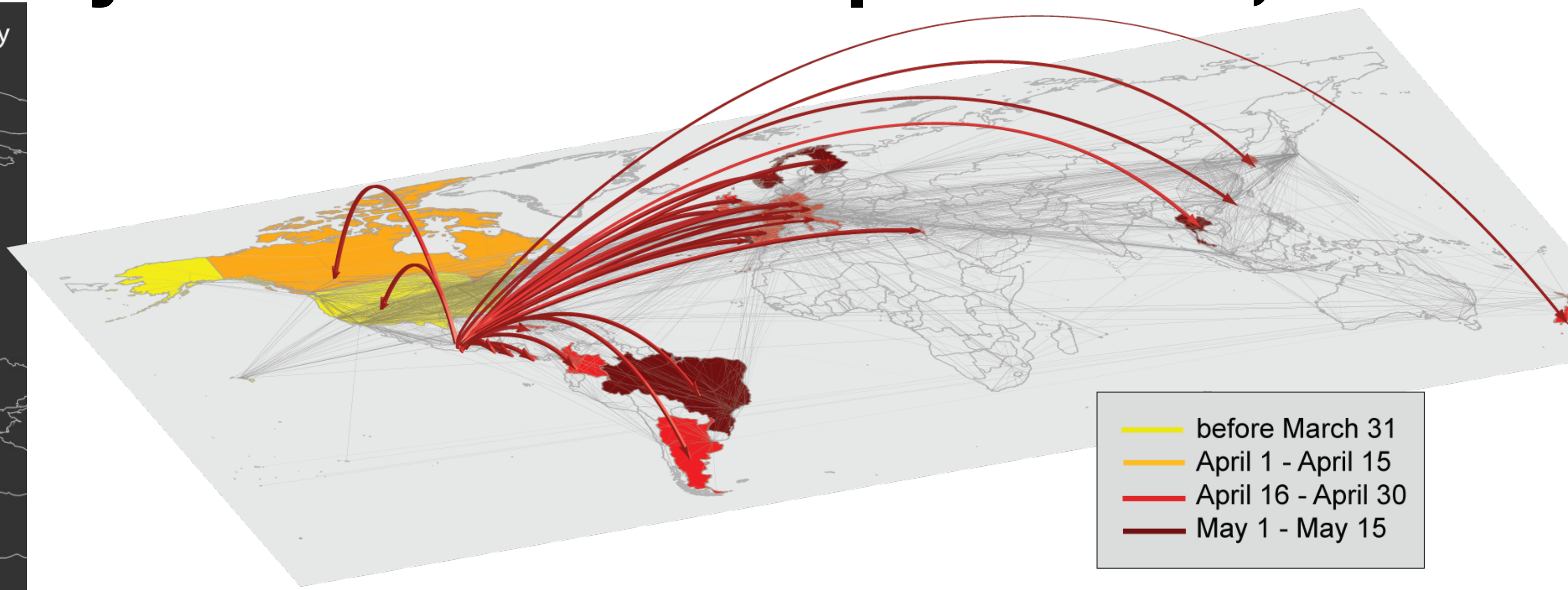
- directly transmitted human diseases: spatial spread of epidemics is determined by human mobility. The pathogen spreads carried by traveling individuals
- Vector borne diseases: the spatial propagation requires human mobility but also the local presence of the competent vector
- food borne, environmental diseases, zoonotic pathogens, etc.: different drivers

# human mobility and epidemics

**black death,14th century**




**flu pandemic, 2009**




[Bajardi et al PLOS ONE, 2011]

- human mobility behaviour determines the spatio-temporal pattern of spread.
- Different kinds of mobility become relevant according to the epidemic of interest and the epidemiological question

# human mobility

## data & models

[Human mobility: Models and applications, Barbosa et al. Physics Reports 734 (2018)]

# mobility network data: air travelling

data collected by the International Air Transport Association.

The data can be purchased. The information publicly available is limited

Two type of data:

segment : number of seats for each company between two airports

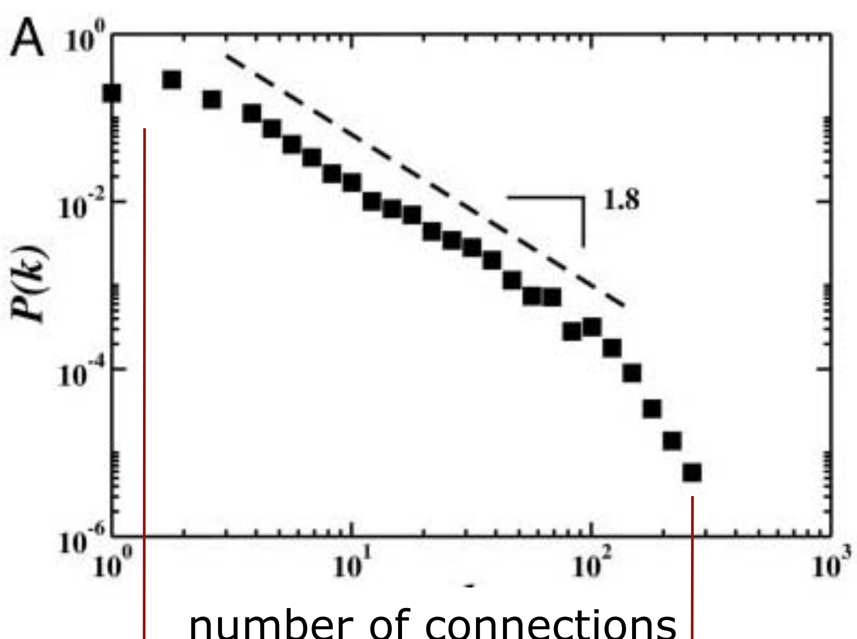
origin-destination: number of passengers between each (origin, destination), obtained from the ticket purchased

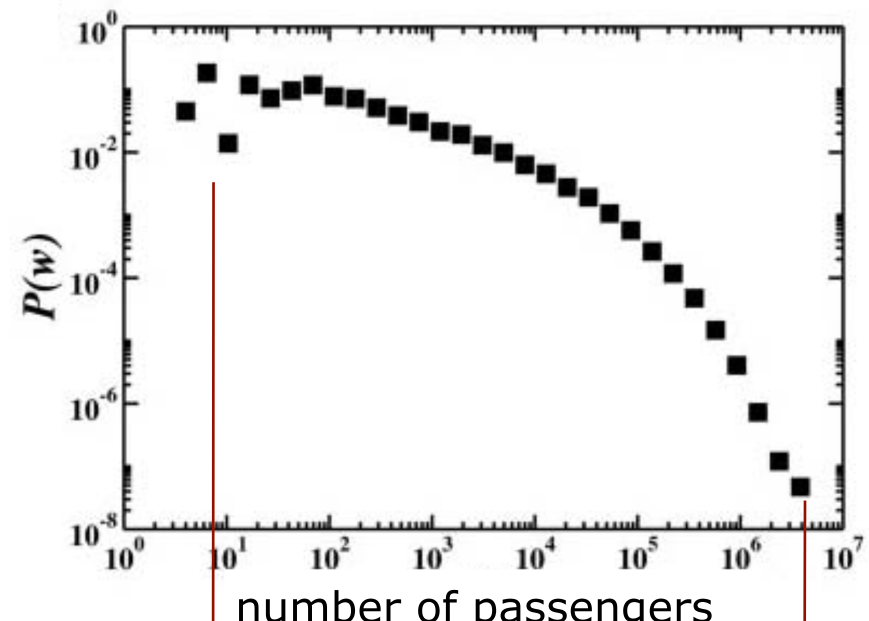
[Hufnagel et al. PNAS 2004; Colizza et al. PLoS Med 2007; Balcan et al. PNAS 2009]

# air travelling: network properties

- heterogeneous topology
- heterogeneous traffic distribution

**whole segment network worldwide, 2002**




peripheral airports

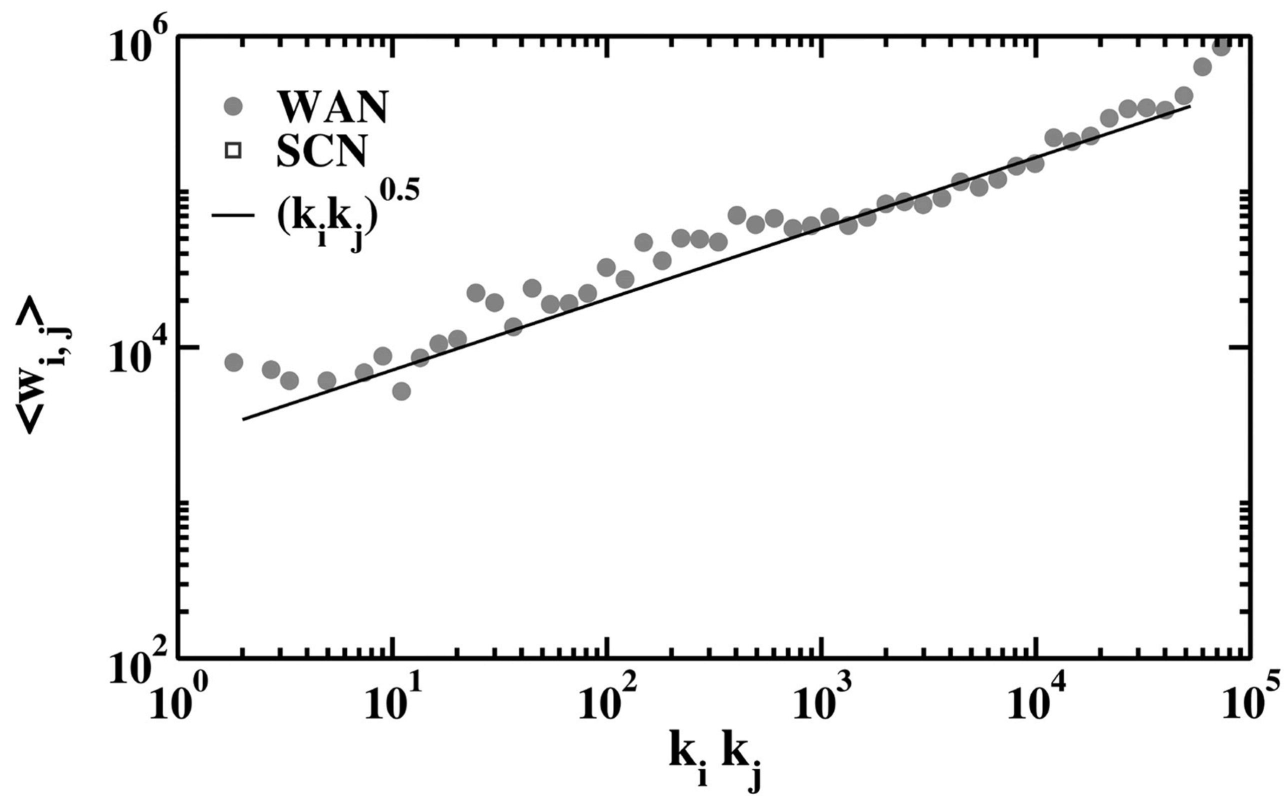
**Frankfurt, 318 connections**

low traffic airports

**Tokyo-Sapporo 17000 p/day**

# air travelling: network properties

**whole segment network worldwide, 2002**




scaling relations between fluxes, number of connections and population

$$w_{ij} \sim (k_i k_j)^{\theta}, \quad \theta = 0.5$$
$$N_i \sim k_i^{\phi}, \quad 0.5 \leq \phi \leq 1.5$$

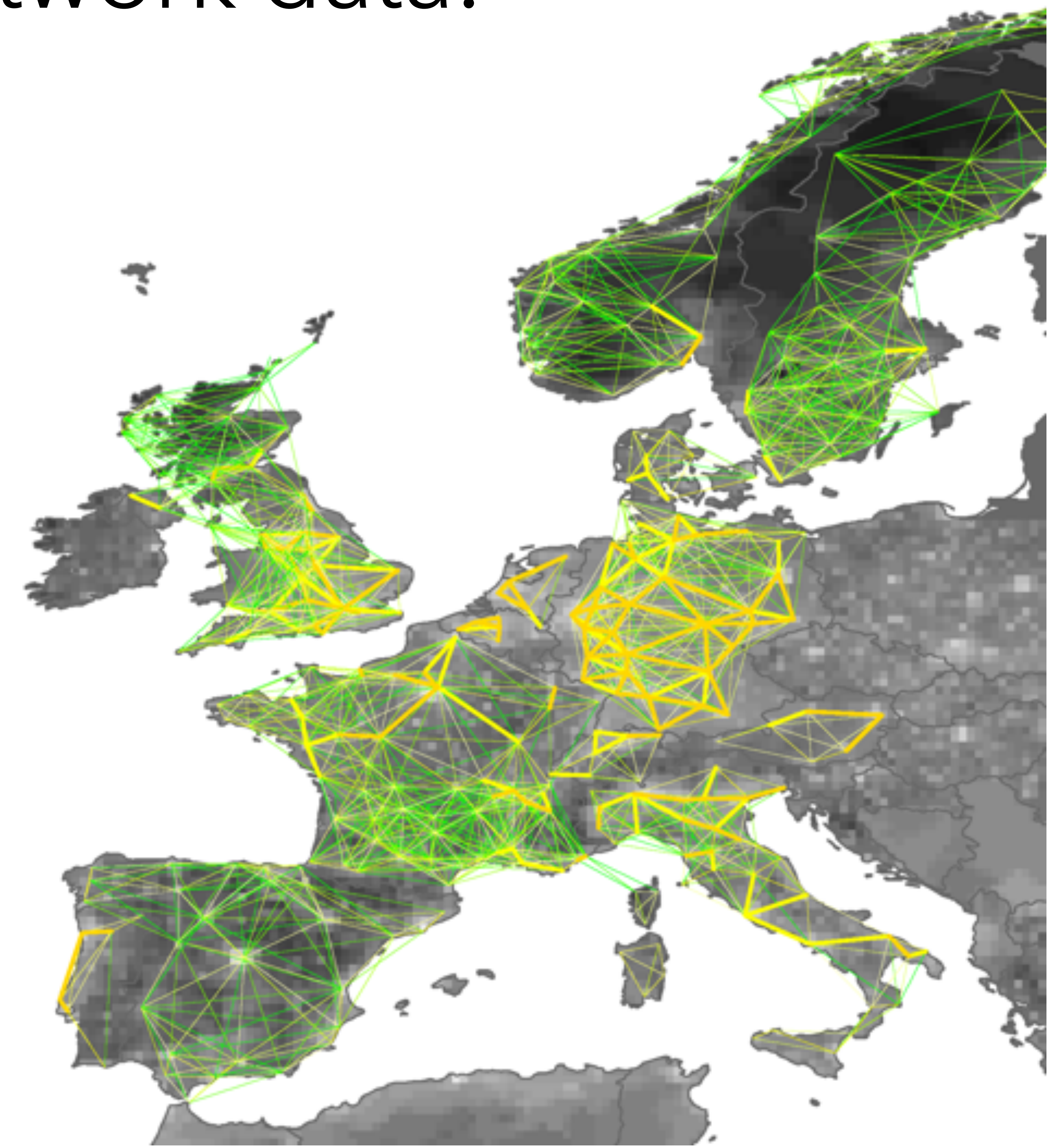
[Barrat, et al PNAS 2004]

# mobility network data: commuting

recovered from census of different countries

location of residence and of work

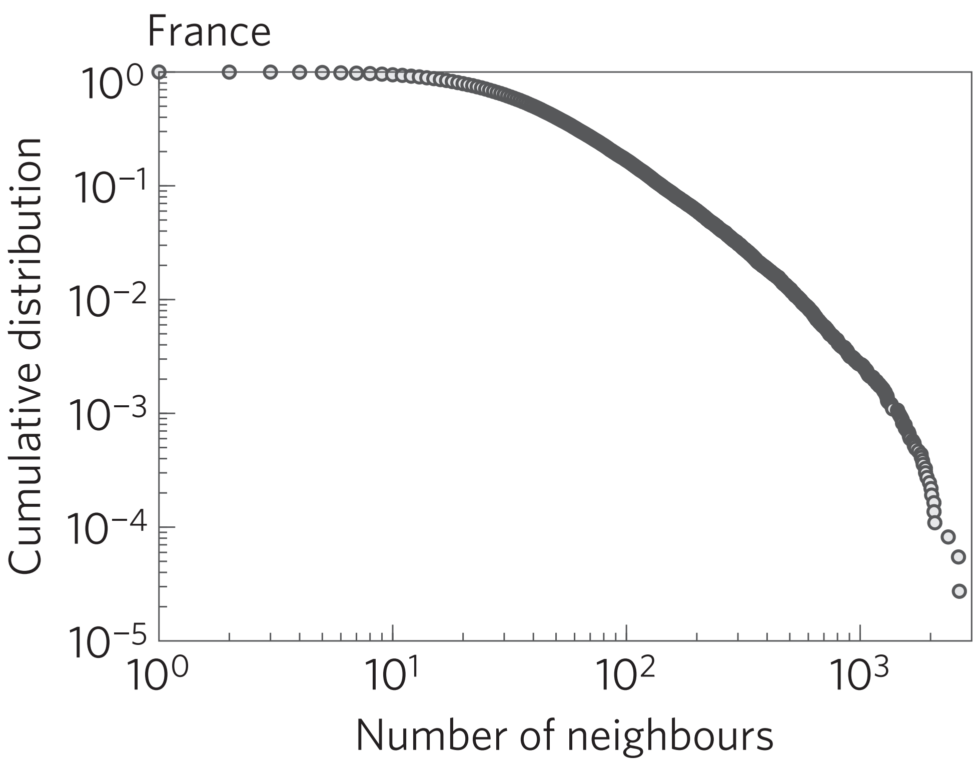
spatial resolution (administrative level of the data) highly variable by country

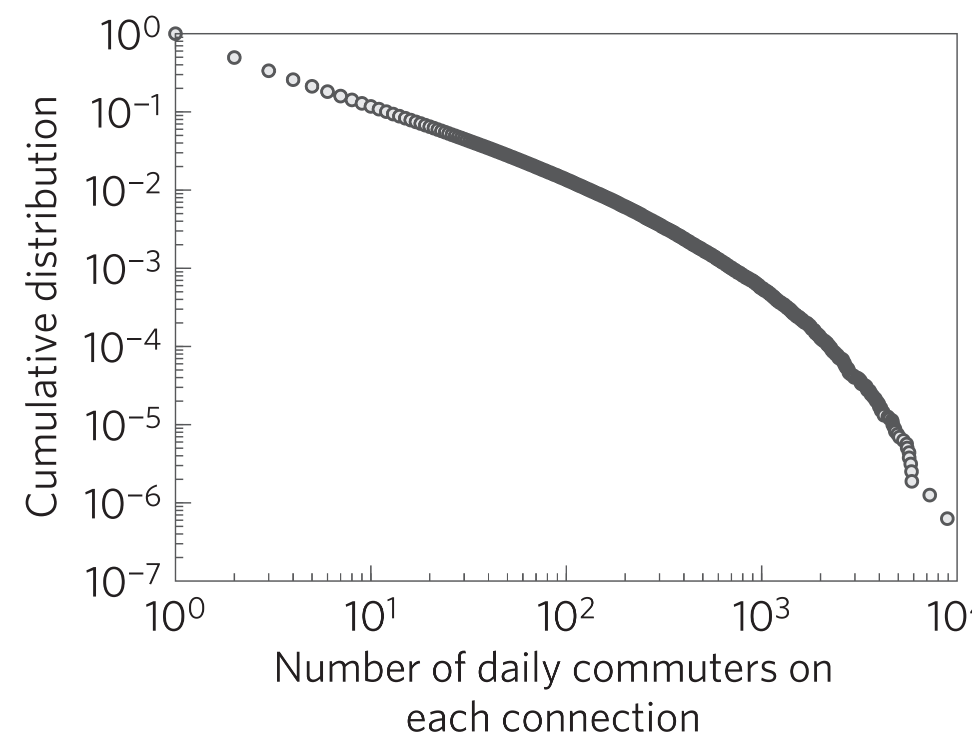


[Balcan, Colizza et al. PNAS (2009)]

# mobility network data: commuting

- heterogeneous topology
- heterogeneous traffic distribution



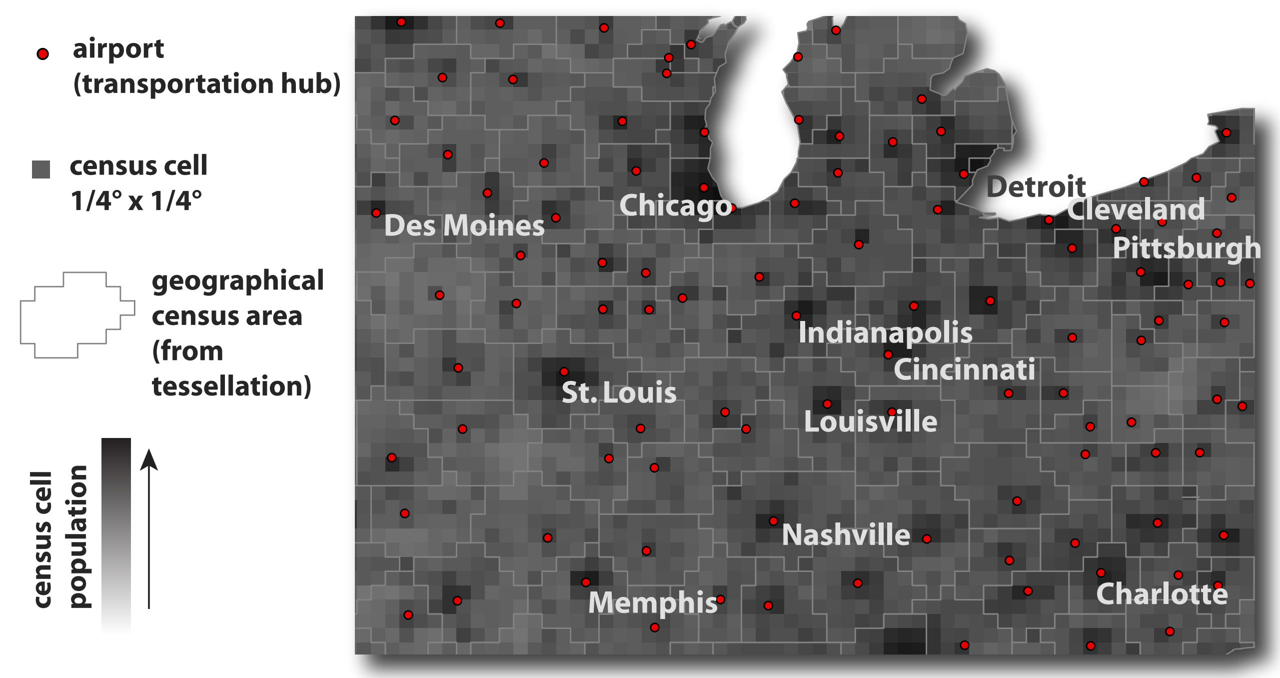




[Balcan, Vespignani Nature Physics (2011)]

# air travelling vs. commuting

To be compared the two networks must be defined at the same spatial resolution: Balcan et al. defined macro urban areas centred around airports

- daily number of travellers
  - ~1,000 air travel
  - ~20,000 commuting
- traveling rate
  - ~$10^{-3}$ days$^{-1}$ air travel
  - ~$10^{-2}$ days$^{-1}$ commuting
- time scales
  - days /weeks
  - hours




Commuting faster dynamics and higher level of mixing

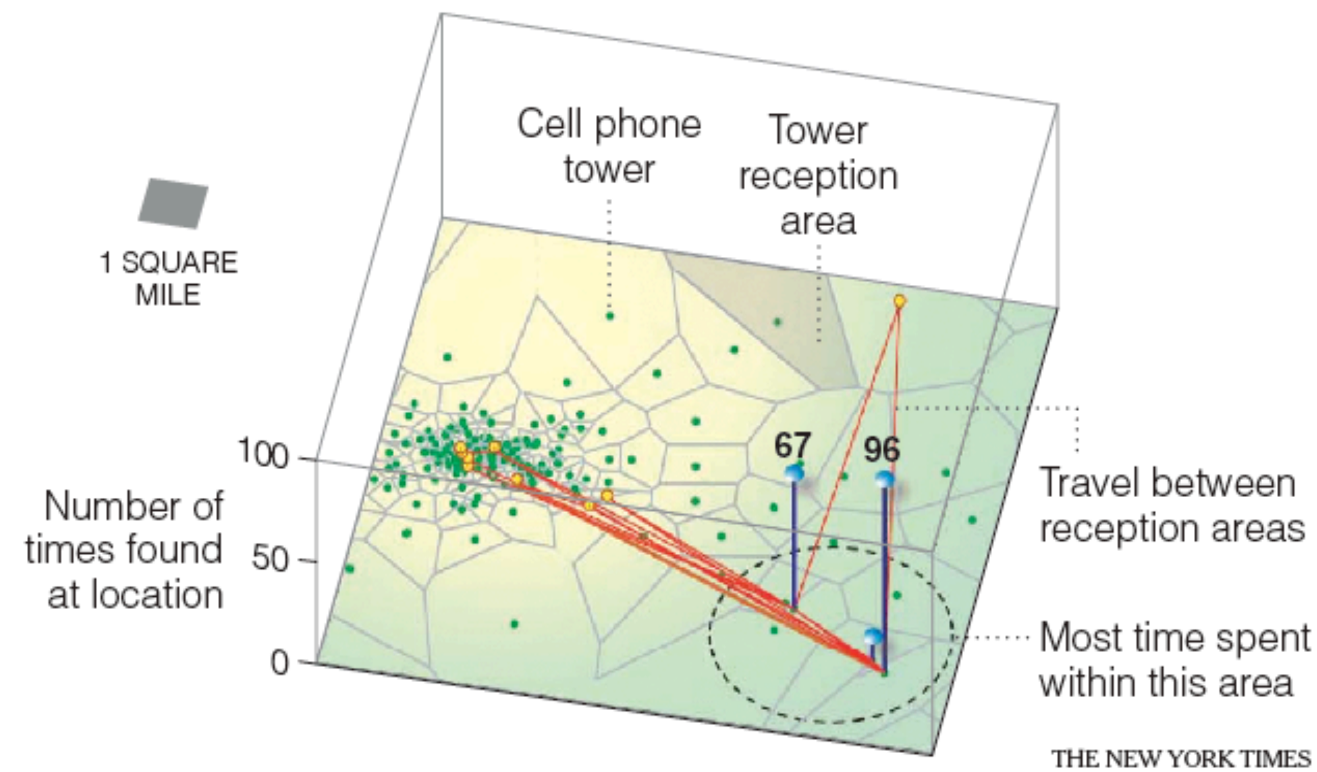
[Balcan, et al PNAS 2009]

# mobility network data: mobile phone

data shared privately by the telephone providers

information recorded for each call & SMS:
time, caller ID, recipient ID, call duration, cellular tower

individual level trajectories (users are anonymised



[Gonzalez et al, Nature (2008)]

Challenges in the data analysis:

- for statistical reliability the analysis is restricted to users that call with high frequency (still many locations may be missed)
- Area covered by the cell tower highly variable: Towers more dense in densely populated area –> spatial resolution in rural areas very poor

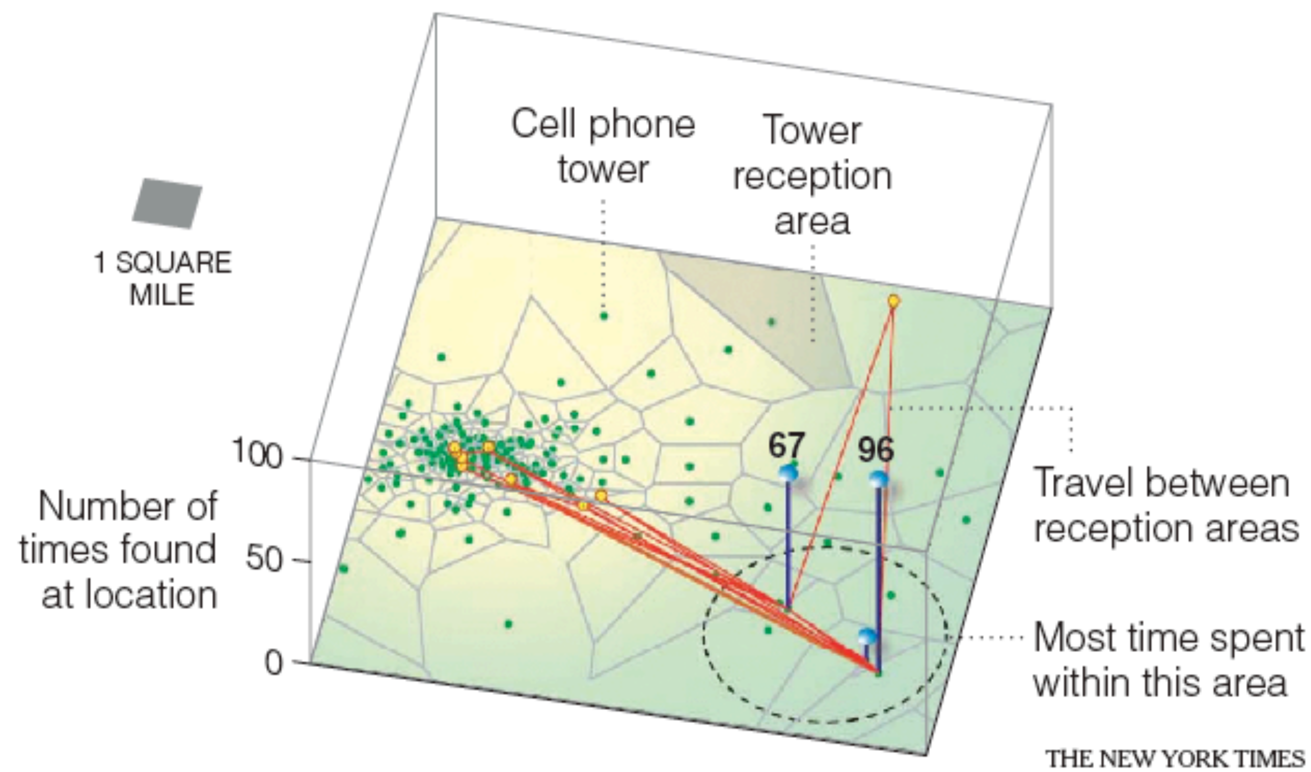
# mobility network data: mobile phone

PROS:

Individual trajectories combining all transportation media and purposes

high temporal and spatial resolution

available at large geographical extent (main source of information regarding mobility for many low income countries)




[Gonzalez et al, Nature (2008)]

CONS:

data cannot be shared across groups (problems with validation)

data-analysis poses statistical challenges (also numerical)

# mobility network data: others

**GPS (available from apps or from research projects)**

PROS: greatest level of accuracy on movement trajectories: spatial resolution few meters, temporal resolution seconds

CONS: a smaller number of individual users: ~$10^3$ GPS vs. ~$10^6$ mobile phones

**Online social network services (e.g. Twitter, Facebook, …)**

PROS:

high spatial resolution (based on GPS)

CONS:

the population may not be representative

# mobility network data: others


**bills:**

www.wheresgeorge.com

- analysis of bank note dispersal in the United States (excluding Hawaii and Alaska)
- trajectories of 464,670 dollar bills
- around 11% of the bills are reported multiple times (often 3–5 entries per bill)
- the trajectories of bank notes are likely a convolution of the mobility of several individuals

[Brockmann, et al, Nature 2006]

**migration:**

E.g. annual information of residence from individual tax return files. Available from the Statistics of Income Division (SOI) of the Internal Revenue Service (IRS) in the United States

# mobility network data

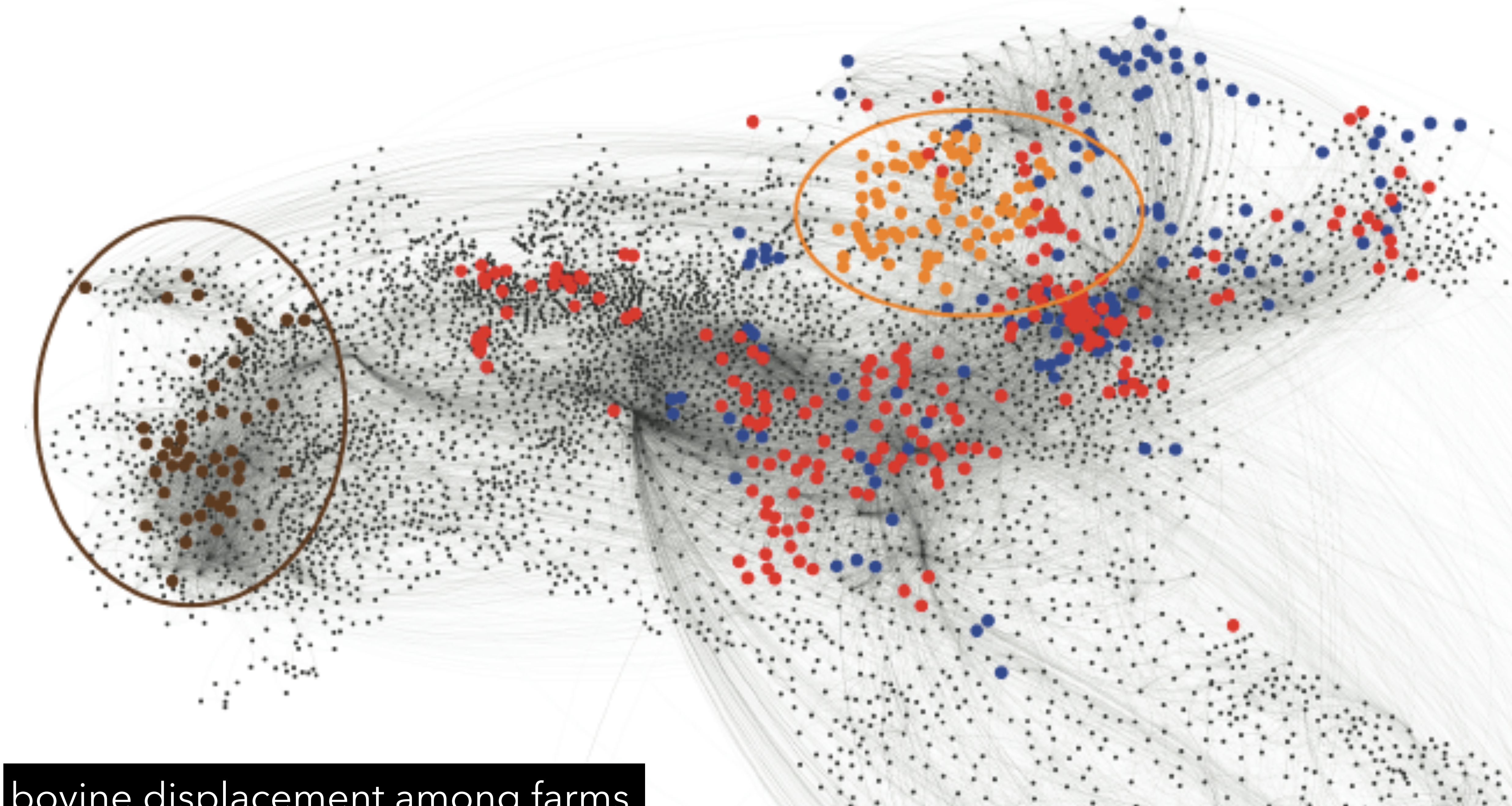
Data of heterogenous nature: heterogenous spatial resolution, individuals-level/ origin-destination fluxes/seats, broken down per transportation media or per purpose of the trip

All datasets provide partial information, can we combine them?

- air-travel & commuting: spatial range are very different
[Balcan et al PNAS 2009]
- cell-phone data & commuting: we can extract commuting proxies from cell-phone data
[Tizzoni et al PLoS Comp Biol 2014]

Differences by age are not well characterised (proportion of children or elderly among air-travellers?)

# if we were caws …



bovine displacement among farms

[Bajardi, PLoS ONE (2011)]

# modelling human mobility

**Individuals-level models**

modelling trajectories of individuals:  random walk, brownian motion, Levy flight, preferential return, …

**Population level model**

modelling fluxes, i.e. the Origin-Destination matrices.

- Two main families: gravity models, intervening opportunities models

# gravity model

Introduces by G. K. Zipf (1946) . Equation to calculate mobility flows inspired by Newton's law of gravitation

$T_{ij} \propto \frac{N_i N_j}{d_{ij}}$, $N_i$ population of $i$, $d_{ij}$ distance between $i$ and $j$

More general form:

$T_{ij} = C\ M_i M_j\ F(d_{ij})$

$M_i = N_i^{\alpha},\ M_j = N_j^{\gamma}$

$F(d_{ij}) =$ either power low $d_{ij}^{\beta}$ or exponential form $e^{-\beta\, d_{ij}}$ or combination of both

PROS: Is able to fit well the data

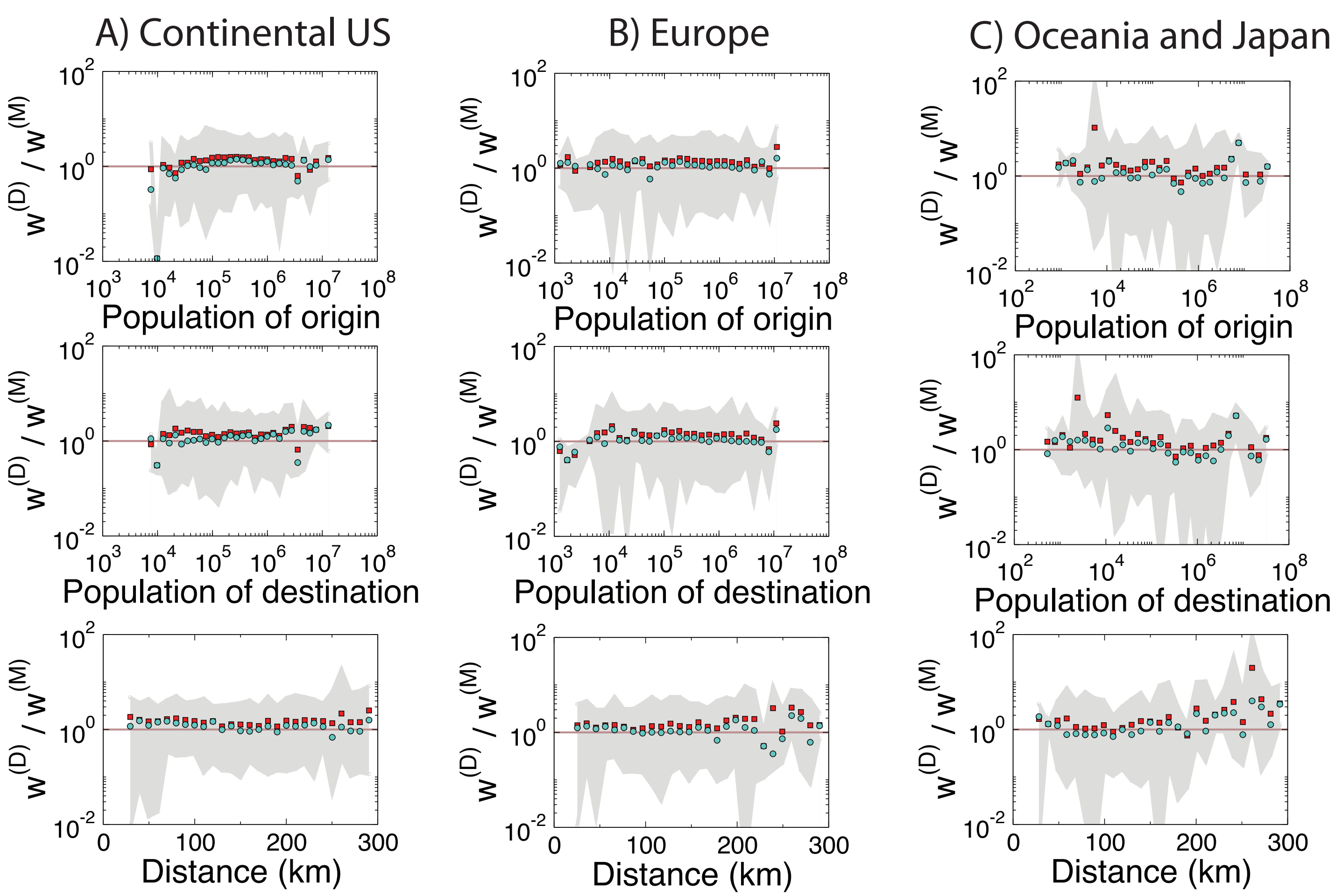
CONS: fitted parameters vary according to the spatial granularity

# gravity model

$T_{ij} = C \frac{N_i^\alpha M_j^\gamma}{e^{\beta d_{ij}}}$ fitted to 29 countries spread across all continents

<u>result</u>: same parameters model well the mobility fluxes in all countries

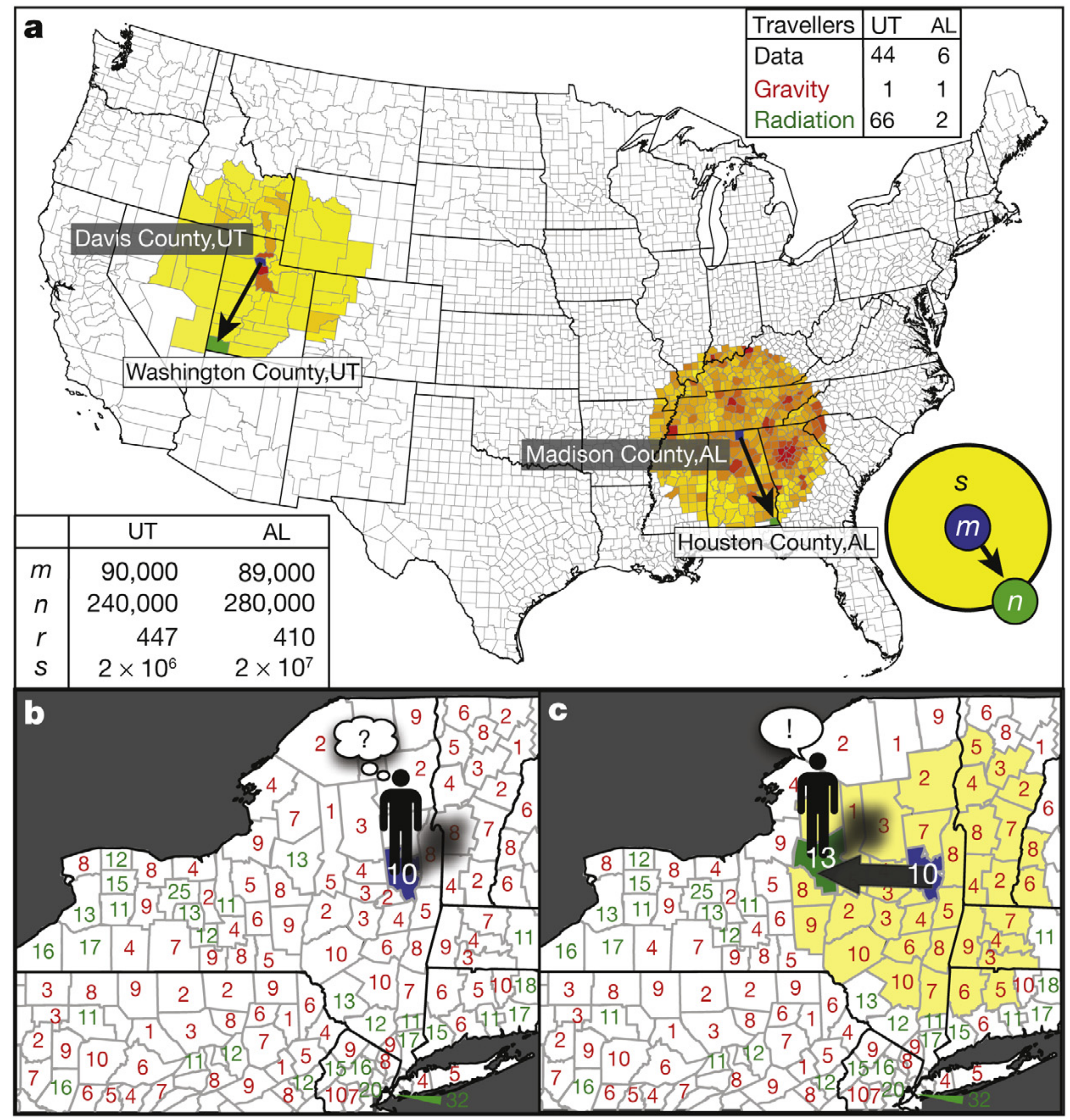
<u>key ingredient</u>: data were aggregated at the same level of spatial resolution




[Balcan et al PNAS 2009]

# radiation model

Introduces by Stouffer (1940). A key driver of migration is the number of intervening opportunities or the cumulative number of opportunities between the origin and the destination. Definition of “Opportunities” intentionally *vague*.




[Simini et al Nature 2012]

# radiation model

Resulting fluxes are independent of $p(z)$ and parameter free

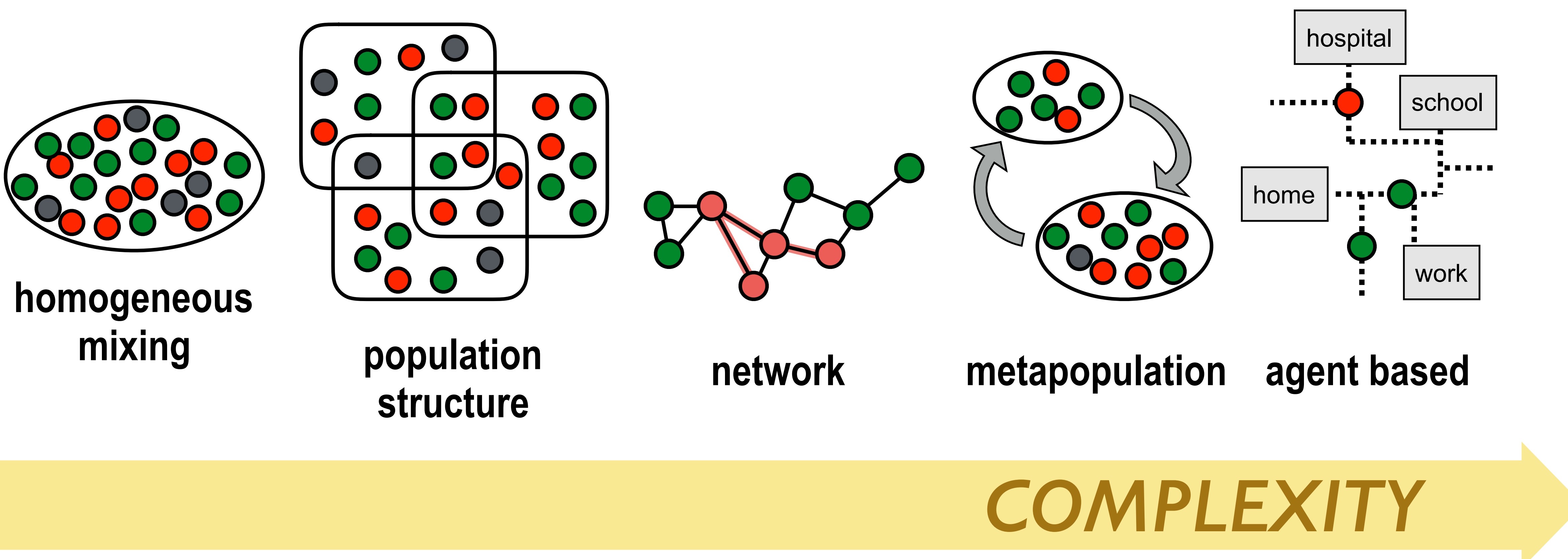
$$T_{ij} = O_i \frac{1}{1 - \frac{N_i}{M}} \frac{N_i N_j}{(N_i + S_{ij})(N_i + N_j + S_{ij})}$$

$S_{ij}$ = population in the radius $d_{ij}$ ; $M = \sum_i N_i$

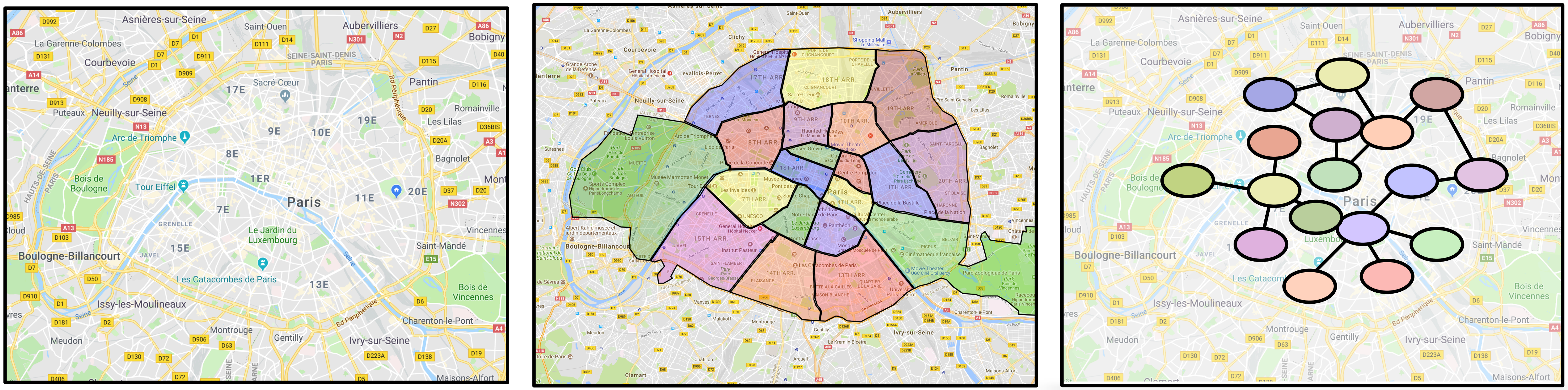
PROS: parameters free. Useful in epidemiology were we have only information of population distribution (low developed countries)

CONS: goodness of fit depends on the spatial resolution

# integrating human mobility in epidemic models



# metapopulation models: a compromise

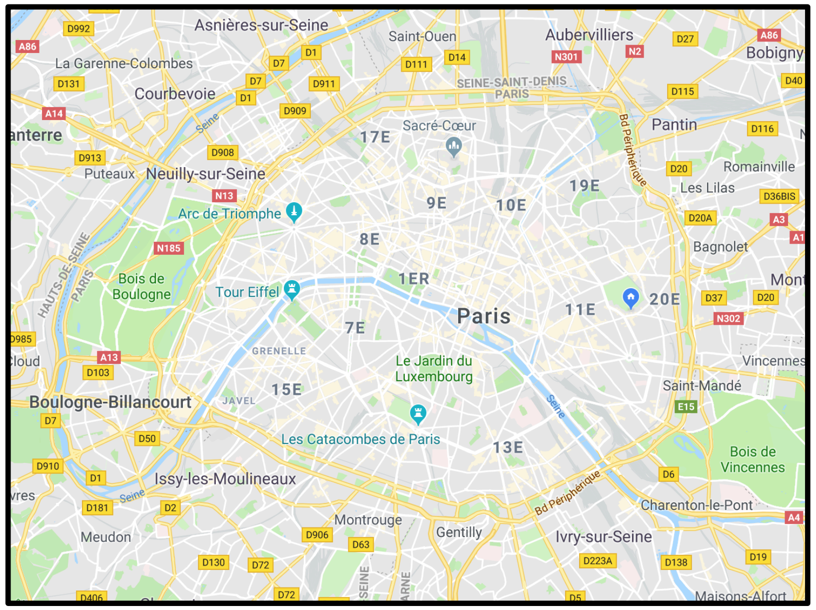


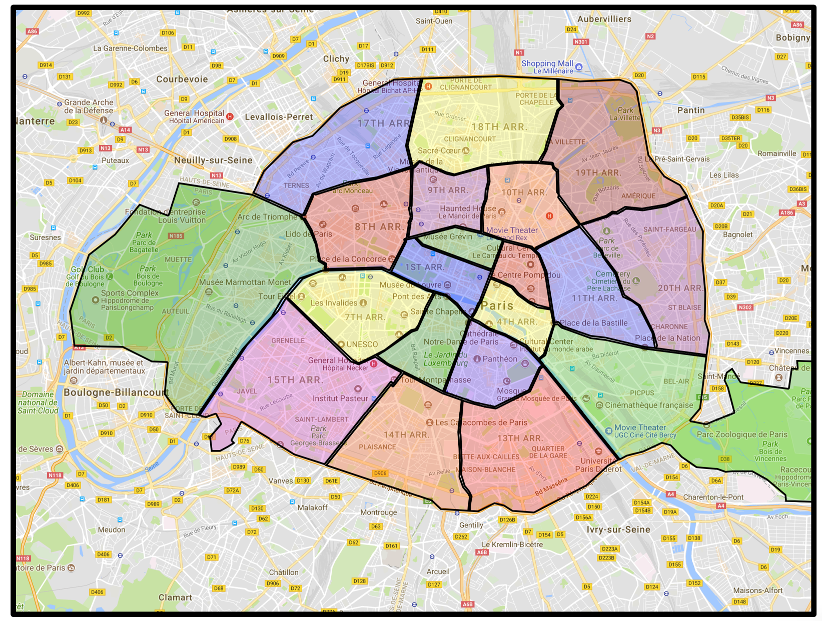

Introduced in ecology to study the interplay between stochasticity and spatial heterogeneities

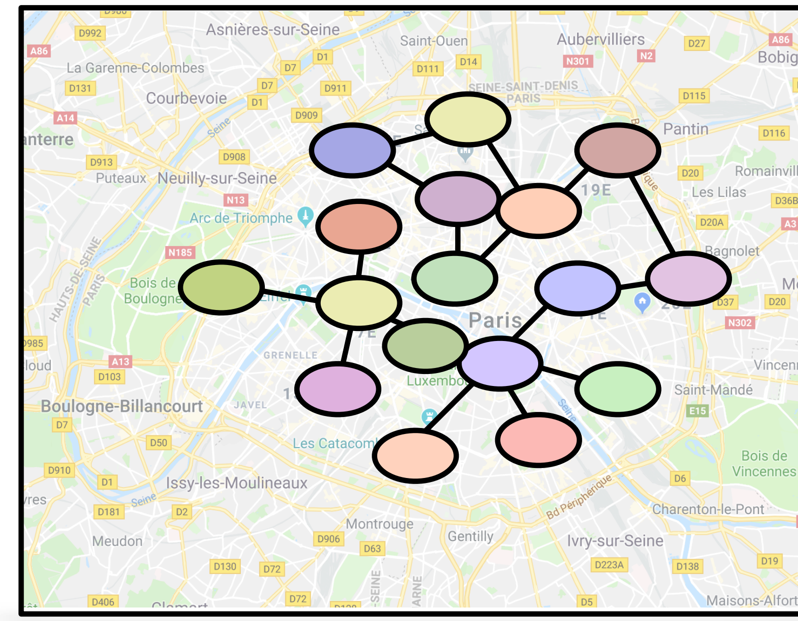
[Levins Bull. Entomol. Soc. Am., 15 (3) (1969)]

- population divided in discrete entities, *patches*
- two level of mixing: *local* (within a patch), *global* (among patches)
- coarse grained description: patches can be seen as elementary units

# metapopulation models: a compromise



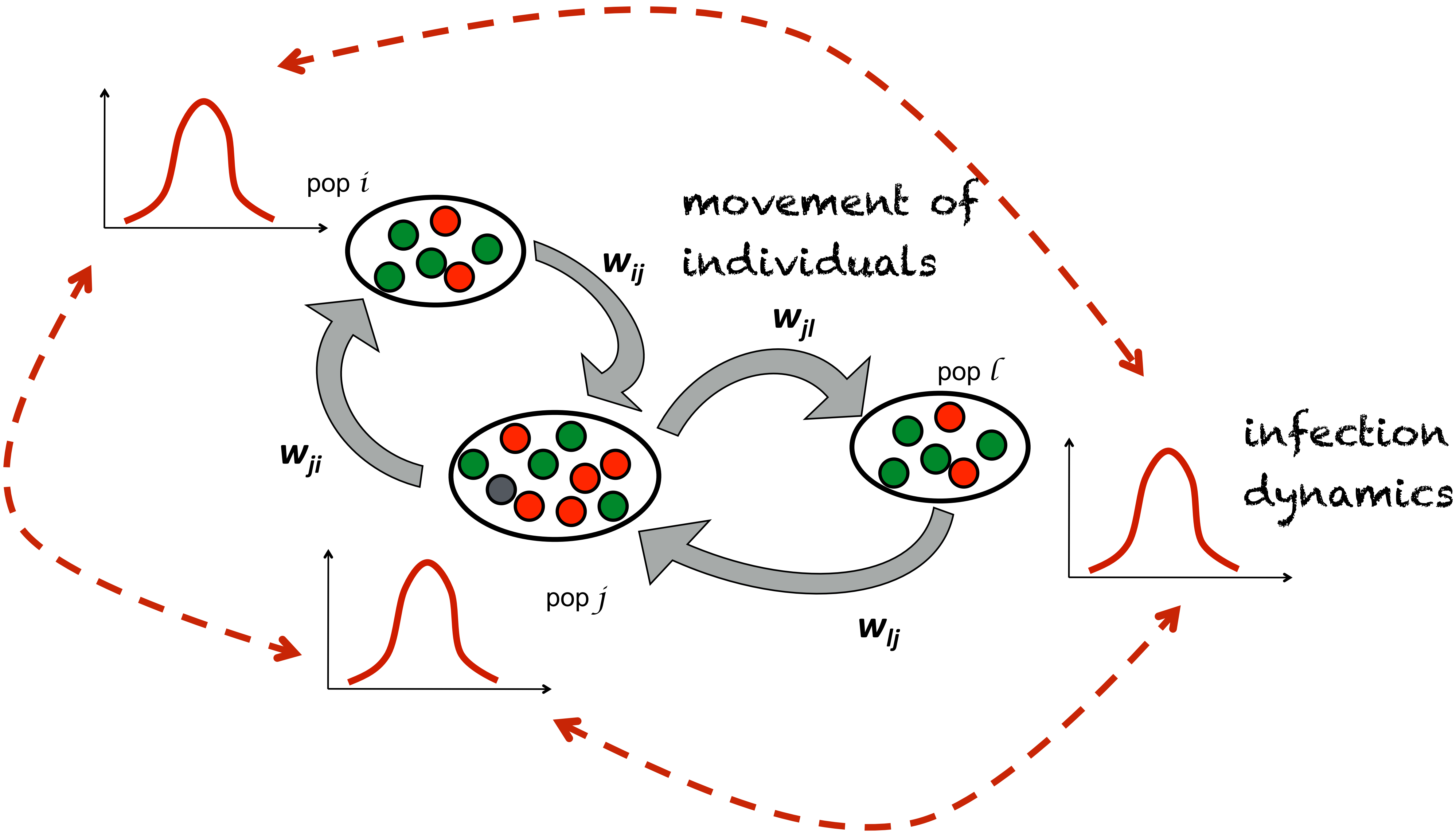






- dynamics driven by stochastic effects: extinction, recolonisation
- discrete nature of individuals essential ingredients to describe the dynamics - If I let half an individual travel I obtain an unrealistic mixing
- early works: mixing among patches homogeneous
- more recently: mixing among patches mediated by the human mobility network: coupling the metapopulation perspective with network theory

# SIR metapopulation model

## modelling of mobility AND transmission dynamics

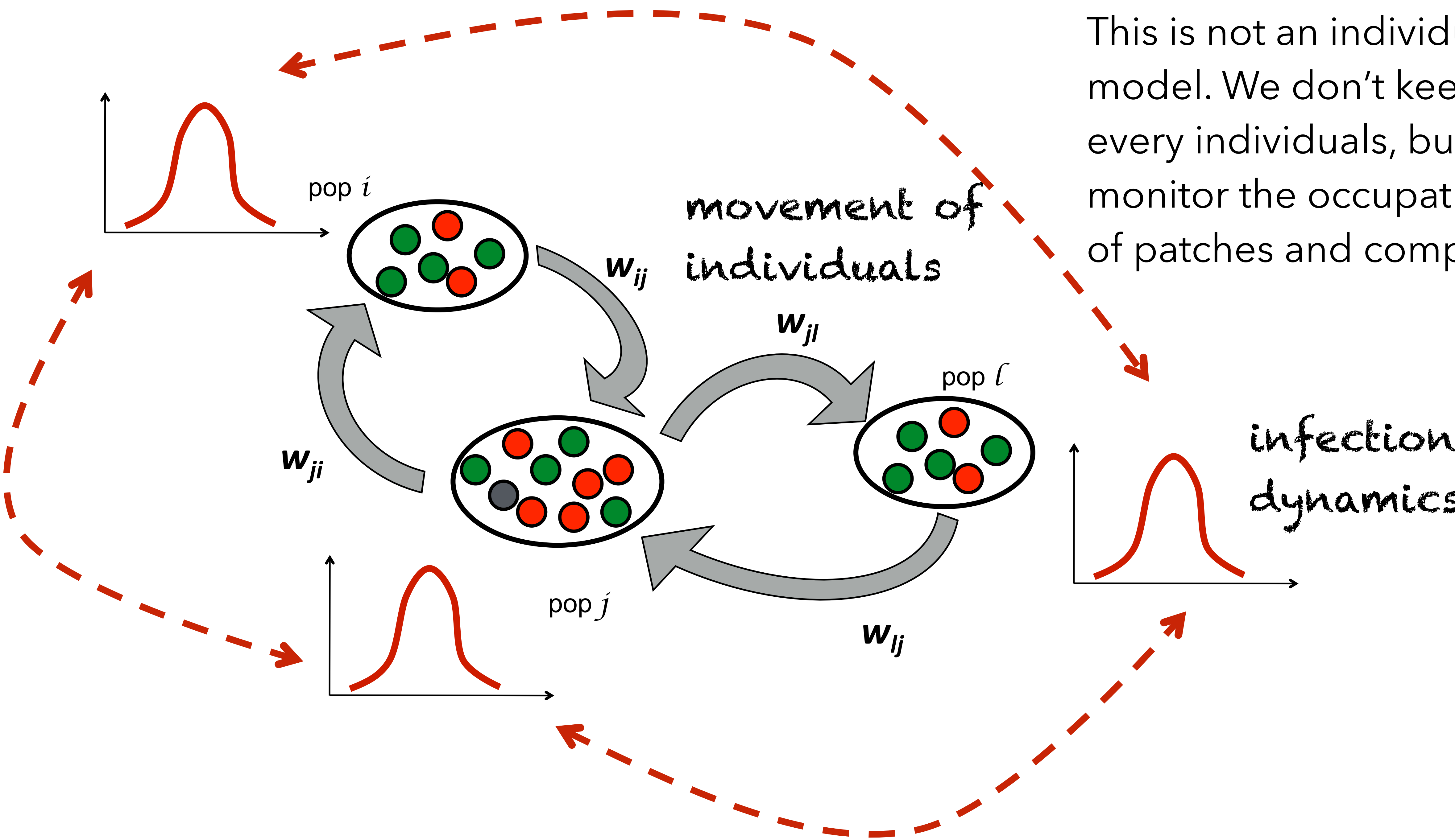



[Hanski, I. & Gaggiotti, Elsevier, Academic Press, 2004]

# SIR metapopulation model

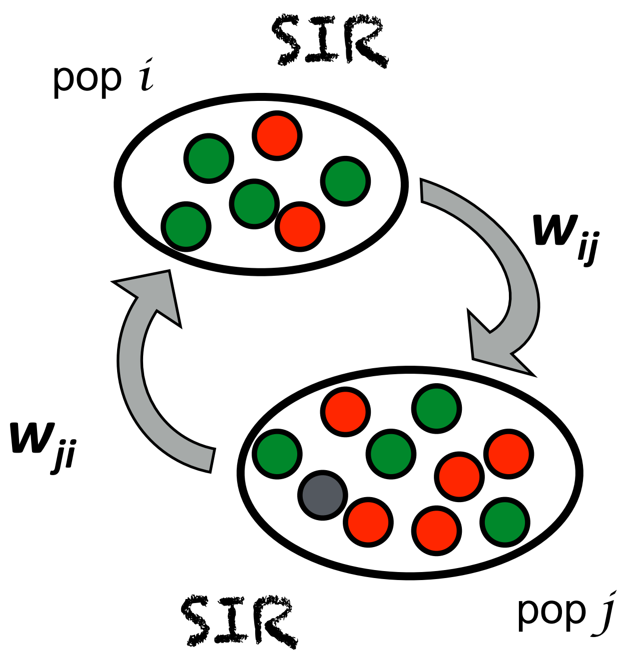
## modelling of mobility AND transmission dynamics

This is not an individual-based model. We don't keep track of every individuals, but we just monitor the occupation number of patches and compartments




[Hanski, I. & Gaggiotti, Elsevier, Academic Press, 2004]

# SIR metapopulation model




$S(t)$
$I(t)$
$R(t)$
$N(t)$

$\Rightarrow$

*V: # populations*
$S_i(t)$
$I_i(t)$
$R_i(t)$
$N_i(t) = S_i(t) + I_i(t) + R_i(t)$

global variables

$$S(t) = S_1(t) + S_2(t) + S_3(t) + \ldots + S_V(t) = \sum_i S_i(t)$$

$$I(t) = I_1(t) + I_2(t) + I_3(t) + \ldots + I_V(t) = \sum_i I_i(t)$$

$$R(t) = R_1(t) + R_2(t) + R_3(t) + \ldots + R_V(t) = \sum_i R_i(t)$$

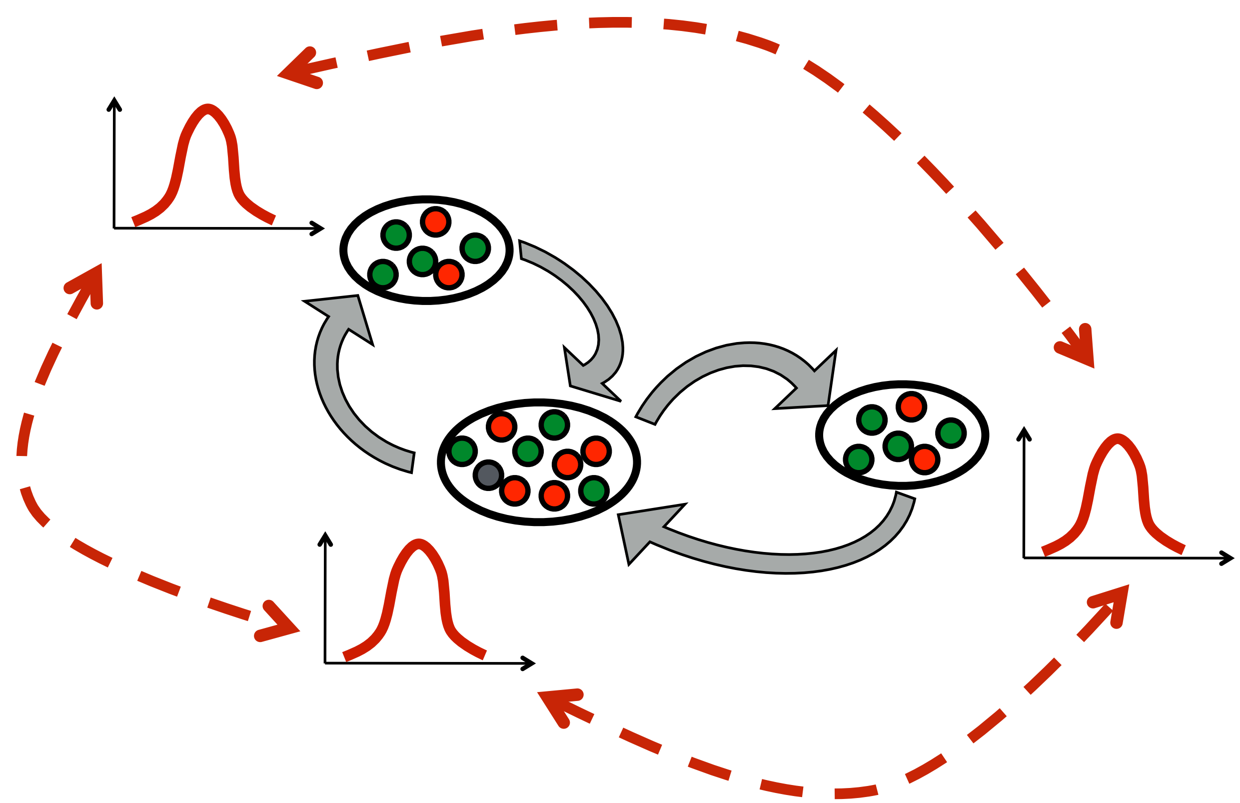
$$N(t) = N_1(t) + N_2(t) + N_3(t) + \ldots + N_V(t) = \sum_i N_i(t)$$

# SIR metapopulation model

$$\frac{dS_i}{dt} = -\beta \frac{I_i(t)S_i(t)}{N_i} + \Omega_i^S$$

$$\frac{dI_i}{dt} = \beta \frac{I_i(t)S_i(t)}{N_i} - \mu I_i(t) + \Omega_i^I$$

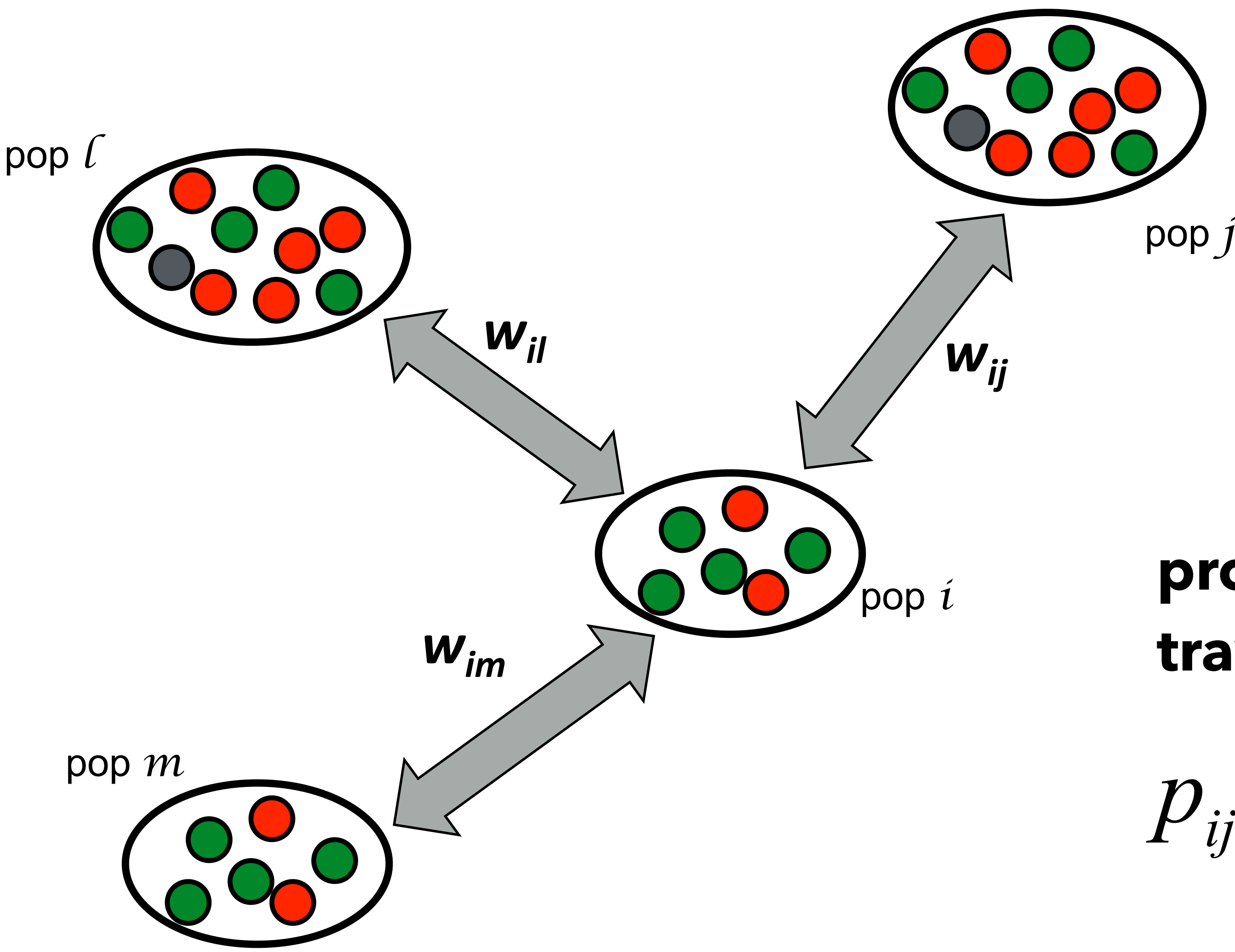
$$\frac{dR_i}{dt} = \mu I_i(t) + \Omega_i^R$$



$\Omega_i^X$ Measure of *in-flow* and *out-flow* of people in compartment *X*

# SIR metapopulation model: markovian mobility

To $\Omega_I^X$ compute we need to model human mobility



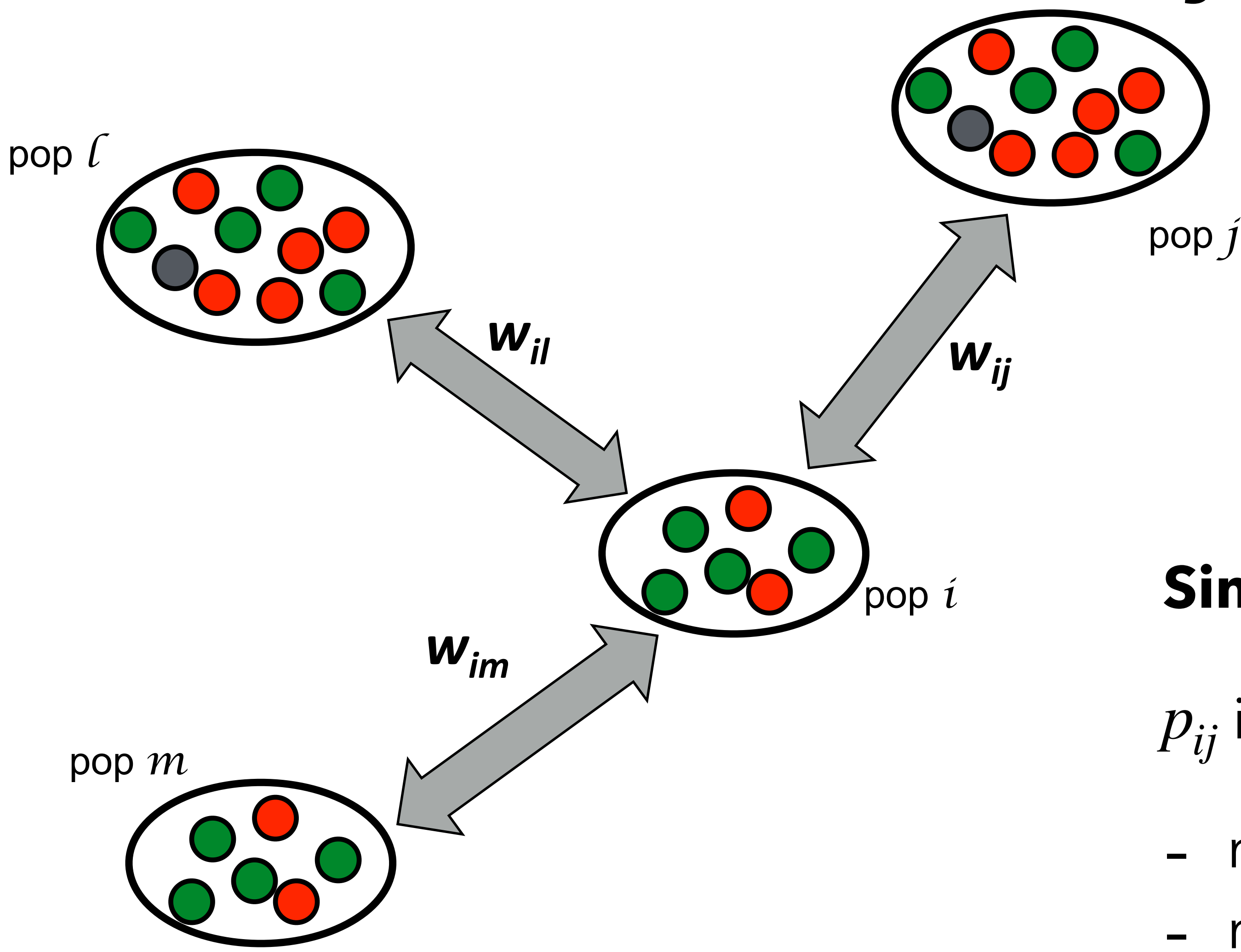

we know that:

- $N_i$ live in $i$
- $w_{ij}$ travel from $i$ to $j$

**probability for an individual in *i* to travel from *i* to *j* ?**

$$p_{ij} = \frac{w_{ij}}{N_i}$$

# SIR metapopulation model: markovian mobility




**probability for an individual in *i* to travel from *i* to *j* at each time?**

$$p_{ij} = \frac{w_{ij}}{N_i}$$

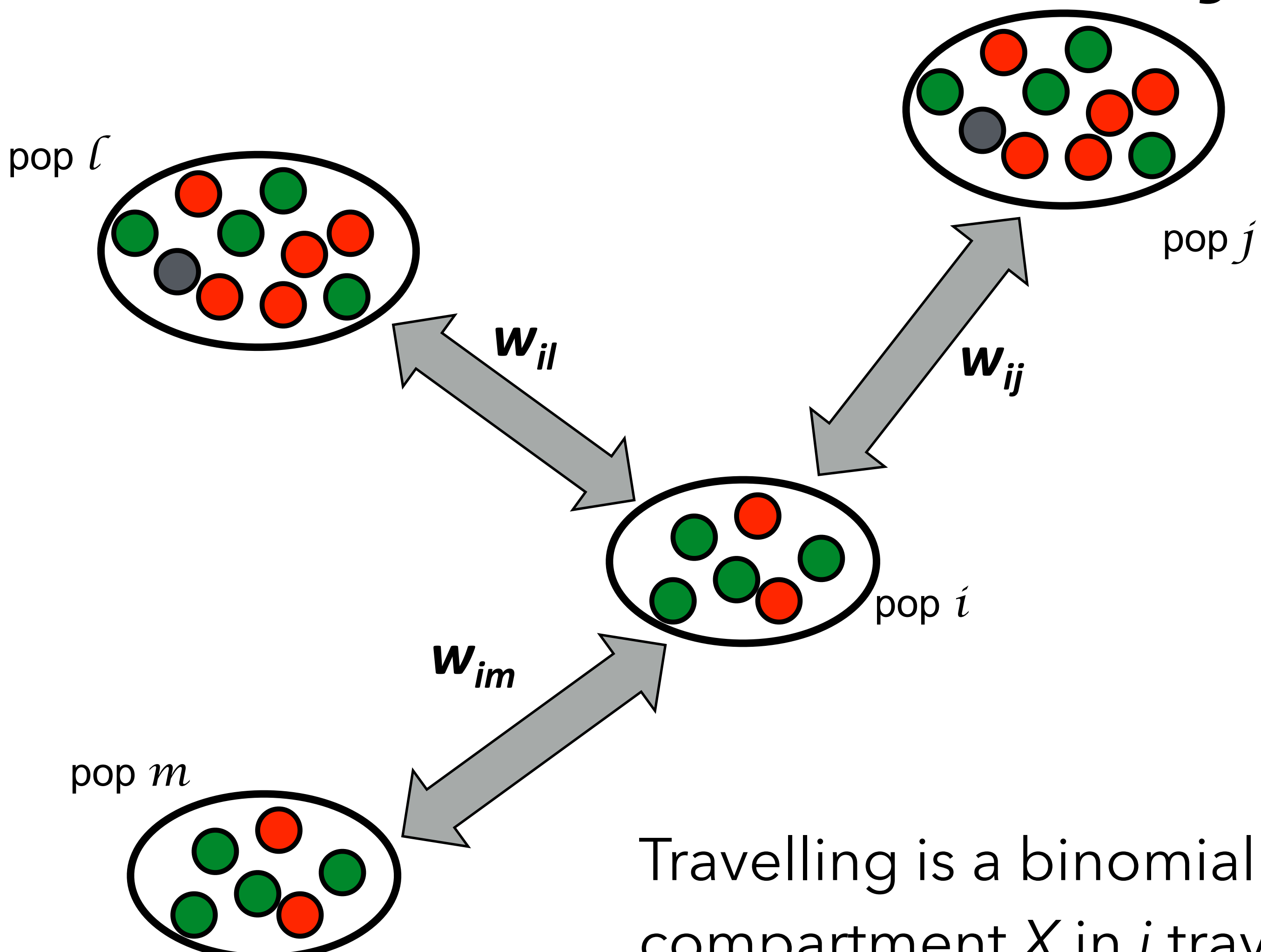
**Simplest possible model:**

$p_{ij}$ is the same for all individuals:

- regardless their infectious status (S,I,R)
- regardless their travel history (time since last travel, previous patch of origin)

as soon as an individuals enter in a new population, it mixes completely with the other individuals of that population

# SIR metapopulation model: markovian mobility




**probability for an individual in *i* to travel from *i* to *j* ?**

$$p_{ij} = \frac{w_{ij}}{N_i}$$

Travelling is a binomial process. Average number of individuals in compartment *X* in *i* traveling from *i* to *j* at each *t*:

$$\langle T_{ij}^{X} \rangle = p_{ij} X_i(t) = \frac{w_{ij}}{N_i} X_i(t)$$

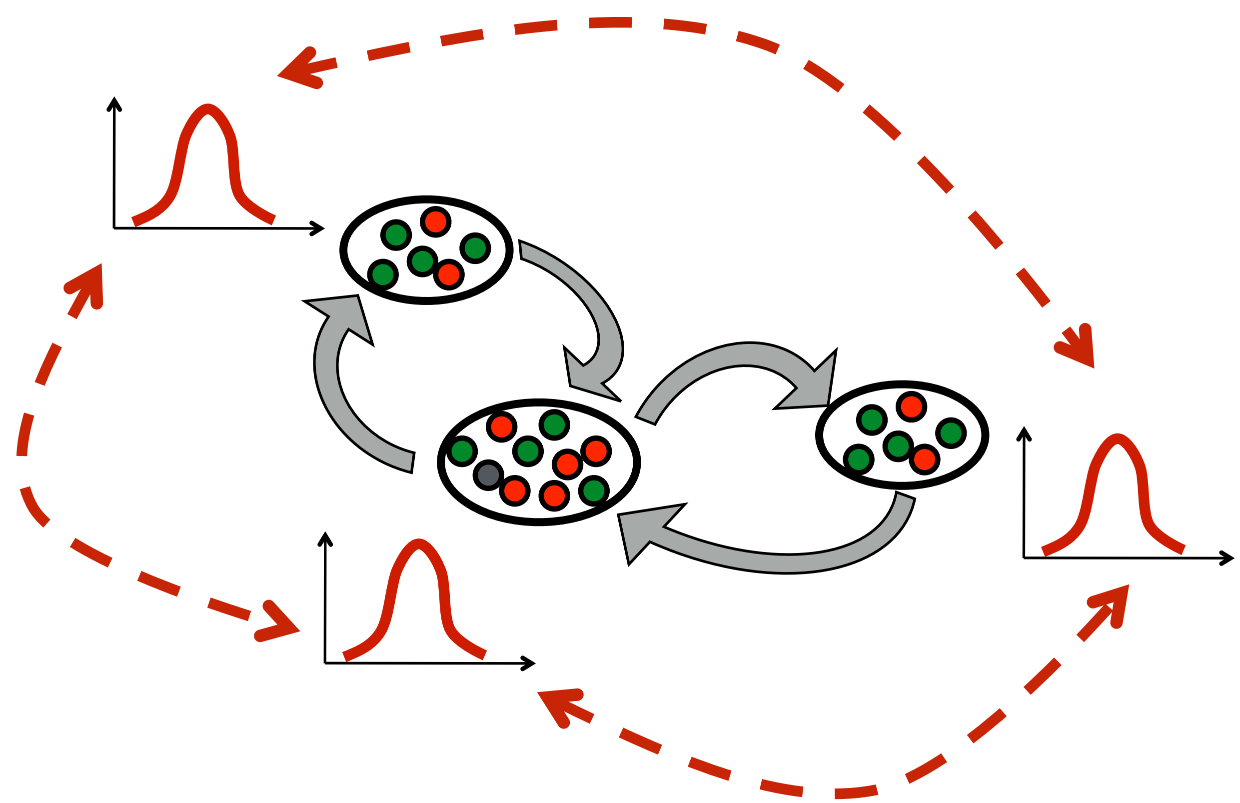
# SIR metapopulation model: markovian mobility

$$\frac{dS_i}{dt} = -\beta \frac{I_i(t) S_i(t)}{N_i} + \Omega_i^S$$

$$\frac{dI_i}{dt} = \beta \frac{I_i(t) S_i(t)}{N_i} - \mu I_i(t) + \Omega_i^I$$

$$\frac{dR_i}{dt} = \mu I_i(t) + \Omega_i^R$$

$$\Omega_i^X = \sum_j \frac{w_{ji}}{N_j} X_j - \frac{w_{ij}}{N_i} X_i$$



# SIR metapopulation model: markovian mobility

## Assumption so far:

we have modelled mobility as a Markovian process : travellers mix with the population at destination and forget about travel origin.

- The travel trajectory is random: patch $i$ –> patch $j$ –> patch $l$ –> …
- We do not account for the location of residence
- we do not account for the traveling length of stay.
- We are in fact modelling a migration process

# SIR metapopulation model: markovian mobility

The assumption works well as long as

- travels are not frequent, i.e. traveling rate negligible with respect to the epidemic time scales $p_{ij} \ll \mu$
- we want to model the short term dynamics of an epidemic

Situations for which this holds in first approximation:

- air-travel and acute infections. E.g. for flu: traveling rate= $10^{-3}$ days$^{-1}$ vs. recovery rate> 0.1 days$^{-1}$)
- early spread of a flu pandemic. It does not work well if I want to model the long term continuous circulation