

Introduction to metapopulation models

1.1 Spatial spread of epidemics

Let us start now to dig deeper adding some complexity to our models. In particular now we want to understand why *spatial spread* of epidemics is important, and its possible effect over public policies. It allows us to estimate the **invasion risk** for a given territory, hence understanding whether a place is more likely to develop an epidemic. In this way we are able to model and realize the **conditions for containment**, since containing an epidemic spatially helps in its management. Moreover, we will discuss about the so called **spatial coupling**, that is to say how the epidemic in a given area influences the epidemic in another. This should have explained us why **spatial information** is an **essential ingredient** in epidemiology: we need to know where the epidemic is at a given moment in order to take the proper countermeasures.

There might be different **drivers** of spatial transmission:

- **direct** transmission among humans: the spatial spreading of epidemics is strongly affected by the *human mobility*. Hence the pathogen spreads carried by traveling individuals;
- **vector borne**: the spatial propagation requires both *human mobility* and the local presence of *competent vector* (mosquitos, rats...). Mobility for vectors is also possible, too;
- **different drivers**, such as food borne, environmental diseases, zoonotic pathogens, etc.

As said, **human mobility**¹ behavior determines the spatiotemporal pattern of spreads. This should take into account that there exist **different types** of mobility, and the actual kind becomes relevant according to the epidemic and the epidemiological questions we are facing.

1.1.1 Human mobility

There are actually different types of data for the human mobility network. For instance, **air travelling** data is collected by the International Air Transport Association (IATA). It can be actually purchased, since the information publicly available is limited. There are two **types** of this data:

- **segment**: number of seats for each company between two airports;

¹Human mobility: Models and applications, Barbosa et al. Physics Reports 734 (2018)

- **origin-destination**: number of passengers travelling between origin-destination, obtained from the tickets purchased.

Doing a similar analysis to the one we have done so far for the “air” network, one can note two facts related to the number of connections and the number of passengers: both **topology** and **traffic distribution** are **heterogeneous**.

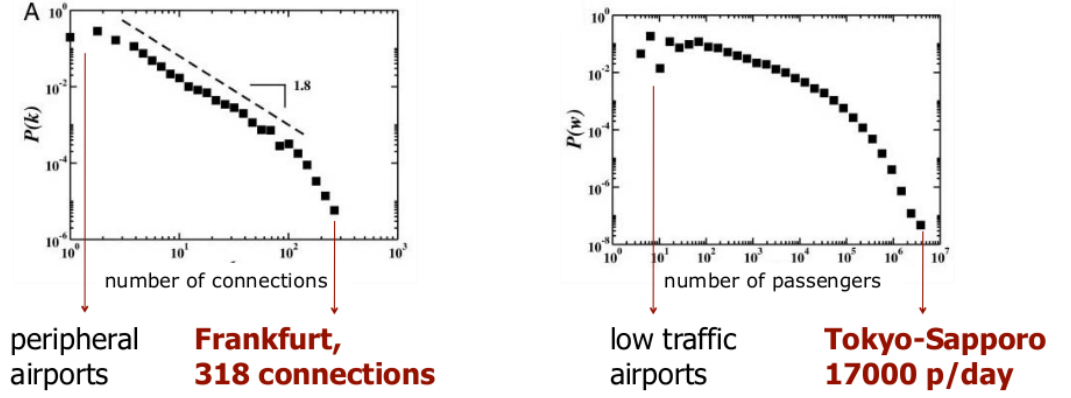


Figure 1.1: Whole segment network worldwide, 2002. Topology of the network and flux of passengers graphical analysis.

We can indeed find some **scaling relations** between fluxes, number of connections and population, as one can see from Fig. 1.2. The average number of route $i \rightarrow j$ is determined by a non linear function of the traffic in airports, and it has form $w_{ij} \sim (k_i k_j)^\theta$ with $\theta = 0.5$ and $N_i \sim k_i^\phi$ with $0.5 \leq \phi \leq 1.5$, where k_i is the traffic at the origin and k_j at the destination. Values for θ and ϕ are computed empirically.

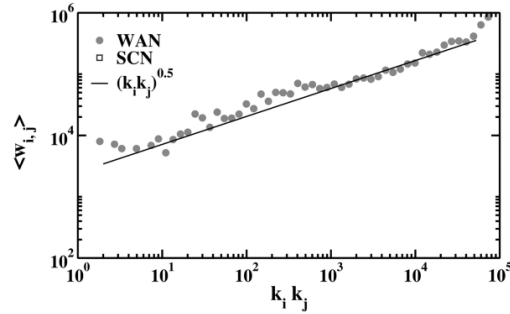


Figure 1.2: Whole segment network worldwide, 2002. Scaling relation.

Another type of mobility one may encounter is the so called **commuting**. This information is obtained from census of different countries and it mainly deals with locations of residence and work. For this reason, *spatial resolution* is highly *variable* and depends actually on the country: local/regional administrations can be actually organized differently within states. Moreover, we can make the aforementioned graphical analysis also for commuting network.

We want now to point out the **differences** between the *air travelling* and the *commuting* networks. In order to do so, first we obviously need to use the same spatial resolution: this is why researchers defined **macro urban areas** centered around airports. Now we are able to look at what differs one network from the other. The first feature one may want to analyze is the average **daily number of travellers** in a certain area: it is about 1000 for the air travel, while 20'000 for the commuting network. Order or magnitudes are indeed different. Another important point is the **daily traveling rate**; it is more probable to commute (10^{-2}days^{-1}) rather than

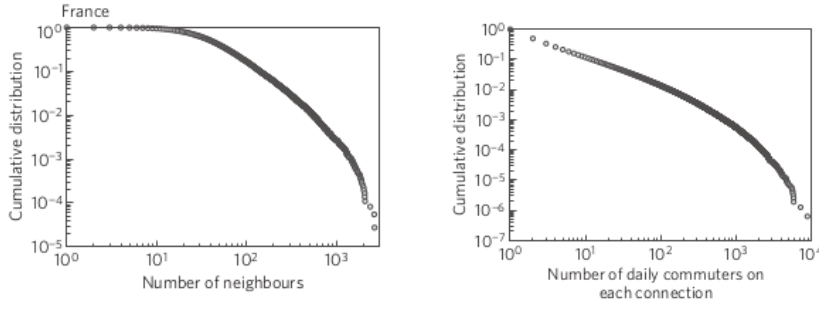


Figure 1.3: Topology of the network and flux of passengers graphical analysis for commuting network.

air travel (10^{-3}days^{-1}). Finally **time scales** are different: one usually travels by plane every days/weeks, whereas commuting is a matter of hour. In conclusion, **commuting** has a **faster dynamics** and leads to a **higher** level of **mixing**.

One another type of data that is available is the one shared privately by **telephone** providers. Information is recorded for each call and/or SMS: for instance time, caller ID, recipient ID, call duration and cellular tower. Using the position of the latters one may be able to reconstruct individual level trajectories. Obviously, for privacy, users are anonymized. Some **challenges** that may arise using this kind of data are the following: for **statistical reliability** the analysis is restricted to users that call *more frequently*, but this actually can turn out to be not precise since many locations can still be missed. Another issue is that the area covered by the cell tower is highly variable: towers are more dense in densely populated area, whereas the **spatial resolution** in *rural area* is very **poor**. This kind of data is really important and accurate: one can reconstruct individual trajectories and the mean of transport used and, eventually, why. For many low income countries this is actually the **main source of information** regarding mobility, despite it is available worldwide. Some drawbacks are that statistically this data must be treated carefully and poses statistical challenges, also from a numerical point of view. Moreover, data cannot be shared across groups for a matter of validation.

Other data can be collected from **GPS**, exploited by some apps or from research projects. This comes with the greatest level of accuracy on movement trajectories: **spatial resolution** is about few *meters* and **temporal resolution** is about *seconds*. However, devices with GPS are a really small subset (10^3) compared to $\sim 10^6$ mobile phones. Data can be collected by other **mobile application services** (e.g. Google, Twitter, Facebook...), it can give high spatial resolution, being it based on GPS, but the population may not be representative. Note that, because of some special events, data can be **donated**: Google, Apple have been sharing their data for good initiatives in helping against the fight or COVID19. An other **historical way** to collect migration was to trace the position of some US Federal banknotes ²: their trajectories are likely a convolution of the mobility of several individuals. Finally, annual information of residence from individual tax return files in the US can describe **migration**, which has a time scale that spans over years.

As one may imagine **data** is very **heterogeneous**, being heterogeneous the sources where we collect it from. Heterogeneity can be seen in spatial resolution, individuals-level/origin-destination fluxes/seats, broken down per transportation media or per purpose of the trip. Since every dataset provides **partial information**, one may think to try to **combine** some of them. For instance, this cannot be done for air-travel and commuting network, being the spatial ranges very different. In addition,

²www.wheresgeorge.com

combining cell-phones data and commuting we are able to extract commuting proxies from cell-phone data. Finally, we should state that we cannot spot clearly the differences that occur for people of different ages: among air travellers indeed there are few children and old people, and statistically they look like the same and cannot be distinguished.

1.1.2 Modelling Human Mobility

Let us now discuss what models all the data collected so far can lead to. There are many types of model we can think of, starting from *individuals-level* models or *population-level* ones. As one can imagine, in the **individuals-level** models we model trajectories of individual using mathematical tools that might include stochasticity: random walk, brownian motion, Levy flight or preferential return, but there are actually many others.

Regarding instead **population level models** we try to model fluxes of individuals, therefore adding some layers of abstraction and generalization: we want to find for instance the *Origin-Destination* matrices. There are two main families for these kind of models: **gravity models**, or **intervening opportunity models**.

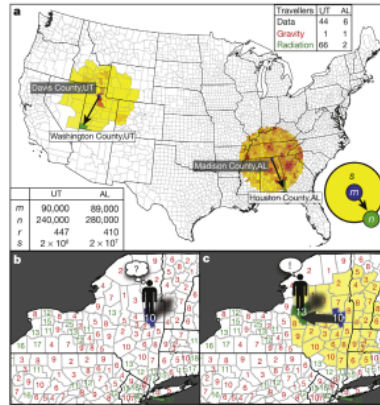
The **Gravity Model** was first introduced by G.K. Zipf. He took inspiration from Newton's law of gravitation in order to describe **mobility flows**:

$$T_{ij} \propto \frac{N_i N_j}{d_{ij}} \quad (1.1)$$

where N_i is the population in i -th site and d_{ij} is the distance between nodes i and j . The last formula can be written in a more general formula as it follows:

$$T_{ij} = CM_i M_j F(d_{ij}), \quad M_i = N_i^\alpha \quad M_j = N_j^\gamma \quad (1.2)$$

where $F(d_{ij})$ is a general function that is either a power law of the kind d_{ij}^β or exponential form $e^{-\beta d_{ij}}$ or a combination of both. This model actually permits to fit very well the data, despite there are no general values for the fitting parameters: they do vary according to the spatial granularity. A possible workaround for this problem was to introduce the formula of *Balcan et al PNAS 2009* ($T_{ij} = C \frac{N_i^\alpha N_j^\gamma}{e^{\beta d_{ij}}}$) which was fitted to 29 different countries spread across all continents. The main **result** was that, when data is aggregated at the *same level of spatial resolution*, the same parameters are able to model well the mobility fluxes in all countries.



[Simini et al Nature 2012]

Figure 1.4: Graphical description for the radiation model. Individual is more likely to move where opportunities are more, since we use their number as a metric.

Another historical model for the mobility was the so called **radiation model** (see Fig. 1.4). It was introduced by Stouffer (1940). He noticed that the *key* driver of migration was the number of **intervening opportunities** or the **cumulative number** of opportunities between the origin and the destination. However, definition of "*opportunities*" was intentionally left vague and they assume a different meaning with respect to the system we are dealing with.

Resulting fluxes in this way are independent of $p(z)$ and are parameters free:

$$T_{ij} = O_i \frac{1}{1 - \frac{N_i}{M}} \frac{N_i N_j}{(N_i + S_{ij})(N_j + S_{ij})} \quad (1.3)$$

where S_{ij} is the population in the radius d_{ij} and $M = \sum_i N_i$. The advantage that we do not need any parameters in order to use this model: it is useful in epidemiology when the only information we have is the population distribution (it usually happens for low developed countries). However, the goodness of fit depends on the spatial resolution.

1.2 Integrating Human Mobility in Epidemic Models

Now we want to use the knowledge we have acquired so far to describe better how Human Mobility affects epidemic spreading. In order to do it, we will borrow from ecology the concept of **metapopulation models**. The last were introduced to study the interplay between stochasticity and spatial heterogeneities³: the entire population was divided in **patches** which are discrete entities (see 1.5). It follows that there may be two different levels of mixing: *local*, that occurs within a patch, or *global* that occurs among patches. It is a **coarse grained** description, and patches can be seen as the new elementary units for our network.

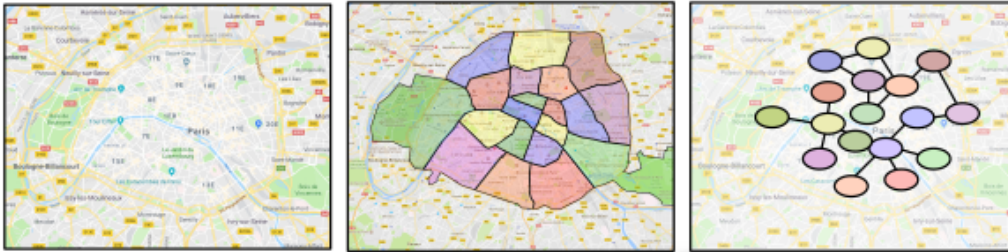


Figure 1.5: Population is divided in discrete entities, the so called *patches* that will become our new elementary units when dealing with network.

In ecology the dynamics is therefore driven by stochastic effects, which may lead for instance to extinction or recolonisation. This, obviously, will suggest us to find an analogy when dealing with epidemic models. However, one should keep into account that the **discrete nature of individual** is one of the most essential ingredients to describe the dynamics: it is meaningless to state that half an individual travel between two patches. The first models assumed that the mixing between patches occurred homogeneously, while more recently more complexity has been added: **mixing** among patches has started to be **mediated** by the **human mobility network** hence coupling the metapopulation perspective with network theory.

1.2.1 SIR metapopulation model

Let us discuss now how we can introduce the metapopulation concept into a model we have studied so many times: the *SIR* model (see fig. 1.6). The only difference

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

here is that it is **not** an **individual-based** model any more. now we do not keep track of every individual, but we just monitor the occupation number of patches and compartments.

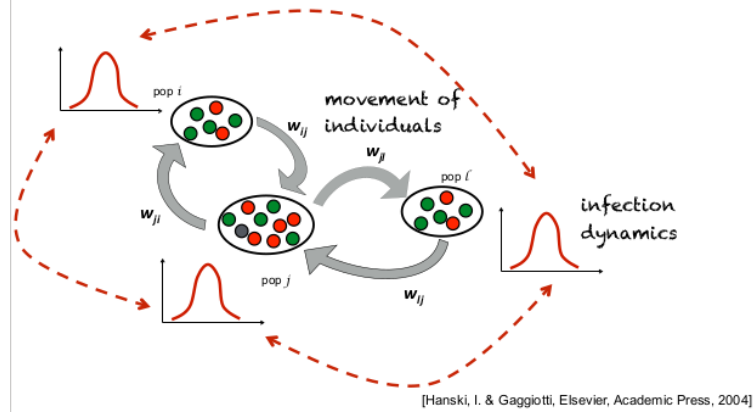


Figure 1.6: How we can model mobility and transmission dynamics using SIR metapopulation model

For each i -th patch we can define the following variables, as we did for the entire population before: $S_i(t)$, $I_i(t)$, $R_i(t)$, and obviously it holds that $N_i(t) = S_i(t) + I_i(t) + R_i(t)$ with clear meanings for each of them. Note that, given a total number of V patches, it follows the definition of global variables:

$$S(t) = S_1(t) + S_2(t) + S_3(t) + \dots + S_V(t) = \sum_i S_i(t) \quad (1.4a)$$

$$I(t) = I_1(t) + I_2(t) + I_3(t) + \dots + I_V(t) = \sum_i I_i(t) \quad (1.4b)$$

$$R(t) = R_1(t) + R_2(t) + R_3(t) + \dots + R_V(t) = \sum_i R_i(t) \quad (1.4c)$$

$$N(t) = N_1(t) + N_2(t) + N_3(t) + \dots + N_V(t) = \sum_i N_i(t) \quad (1.4d)$$

The set of equations related to i -th node and the compartments is the following:

$$\frac{dS_i}{dt} = -\beta \frac{I_i(t)S_i(t)}{N_i} + \Omega_i^S \quad (1.5a)$$

$$\frac{dI_i}{dt} = \beta \frac{I_i(t)S_i(t)}{N_i} - \mu I_i(t) + \Omega_i^I \quad (1.5b)$$

$$\frac{dR_i}{dt} = \mu I_i(t) + \Omega_i^R \quad (1.5c)$$

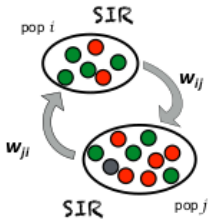


Figure 1.7: Mobility between two different patches occurs with probability $p_{ij} = w_{ij}/N_i$.

where we have introduced Ω_i^X that is a measure of the *in-flow* or *out-flow* of people in compartment X . Note as the set of equations really resembles the SIR model we previously studied.

We will now discuss how to compute Ω_i^X for which we need to model human mobility. The first assumption we make is that the mobility is a **Markovian process**. Indeed, this is the easiest model one can think of. We need now to model human mobility, and this can be done in the following way: we now that N_i people live in i -th patch, and w_{ij} is the number of people that travel $i \rightarrow j$. Therefore, the **probability** for an individual in i to travel from i to j is:

$$p_{ij} = \frac{w_{ij}}{N_i} \quad (1.6)$$

The simplest possible model is when p_{ij} is the same for all individuals **regardless** their *infectious status* (S,I,R) and their *travel history*. This is the Markovian assumption we stated above. As soon as an individual enters into a new population, she mixes completely within it and cannot be distinguished from other individuals any more. Moreover, she will be considered as part of that population from now on.

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

$$\langle T_{ij}^X \rangle = p_{ij} X_i(t) = \frac{w_{ij}}{N_i} X_i(t) \quad (1.7)$$

Therefore, a formula for Ω_i^X can be the following:

$$\Omega_i^X = \sum_j \left(\frac{w_{ji}}{N_j} X_j - \frac{w_{ij}}{N_i} X_i \right) \quad (1.8)$$

where, for the flux, we consider all the people that are entering patch i from all possible other patches j , as well as all individuals exiting the node i being their destination any other patch j .

We now recall the **assumptions** we have done so far. We have modelled mobility as a **Markovian process**. In other words, we assume that travellers mix with the population at destination and forget about travel origin. The implications hence are:

- **travel trajectory** is *random*: patch $i \rightarrow$ patch $j \rightarrow$ patch $l \rightarrow \dots$;
- we do *not* take into account the **residence location**;
- we do *not* take into account the *length of stay* while travelling.

At the end of the day we are modelling a **migration process**. The Markovian assumption for mobility actually works well as long as **travels** are **not frequent**, that is to say that travelling rate is *negligible* wrt epidemic time scales ($p_{ij} \ll \mu$). Moreover one should choose the Markovian assumption when we want to model the **short term dynamics** of an epidemic. In the real world, some situations where this hold at a first approximation are for instance:

- air-travel and acute infections: for example flu or COVID 19. We have seen that travelling rate is about 10^{-3}days^{-1} that way smaller than their recovery rate 10^{-1}days^{-1} ;
- the early spread of COVID-19 or a flu pandemic. However, it does not work when we want to model the spreading in the long run (i.e. for large times).