



HAL
open science

Introduction to modeling disease spread in space

Vittoria Colizza

► **To cite this version:**

Vittoria Colizza. Introduction to modeling disease spread in space. Proceedings of the International School of Physics 'Enrico Fermi', 203, pp.141-170, 2019. hal-04632042

HAL Id: hal-04632042

<https://hal.science/hal-04632042>

Submitted on 2 Jul 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Introduction to modeling disease spread in space

Vittoria Colizza*

*INSERM, Sorbonne Université, Institut Pierre Louis d'Epidémiologie et de Santé Publique (IPLESP), 75012 Paris, France

Summary - The spatial structure of populations is a key element in the description and understanding of the spatiotemporal propagation of infectious diseases spread. Host population in space is often characterized by a highly fragmented environment where it is structured and localized in relatively isolated discrete patches or subpopulations connected by some degree of hosts movements. Metapopulation models provide the ideal theoretical framework to capture the separation of a host population into local communities, with strong homogeneous mixing within each community and weaker interactions between communities corresponding to the underlying substrate of commuting patterns, mobility networks and/or transportation infrastructures. This paradigm can be applied to model the spatiotemporal propagation of epidemics in structured populations at different scales, by considering for example families, city locations, hospital wards, farms, urban areas or regions as local communities connected by hosts' mobility processes.

Here we present the computational approach to the modeling of epidemic processes in spatially structured systems. We introduce metapopulation models as the standard modeling framework for the study of epidemic spread among localized communities of hosts. Taking into account the coupling provided by the interactions among localized populations, different modeling approaches are described, including mechanistic (i.e. microscopic) simulations and effective approaches, and the possible presence of memory effects. Topics like invasion dynamics and local vs. global containment of an emerging epidemic will be addressed, and the theoretical results will be put in relation with the design of possible intervention policies for epidemic control. These

notes represent a theoretical introduction for the development of data-driven realistic metapopulation models for application in public health.

1 Spatial spread of infectious disease epidemics

Space represents, in many circumstances, a relevant feature in the spread of an infectious disease. While this is intuitively clear if we think about recent outbreaks of directly transmitted diseases that spread throughout the world, such as the 2002-2003 epidemic of Severe Acute Respiratory Syndrome (SARS), the 2009 H1N1 influenza pandemic or the recent Zika epidemic, the importance of space must be noted in more localized processes as well. Consider for example outbreaks in close and localized populations (may them be a city, a village, a community, or a hospital, for example): the geographical structure of the population or of the space in which the epidemic takes place (e.g. different locations in the city, households in the village or community, wards in the hospital, and others) often represent key elements that shape the propagation pattern of the disease. The diffusion of the disease in a given environment is indeed dependent on the spatial distribution of susceptible hosts and on the geographic features of such environment that may facilitate the transmission along given routes or accesses, and prevent it across obstacles. Such features effectively embed the ability of hosts to move from one region to another of the space, connecting distinct localized communities and providing the means for the movement in space of the pathogen.

That human displacements and travel were related to carrying the disease across space has long been recognized in the history of human diseases. Historians recorded the initial spread of the plague of Justinian's reign (A.D. 542) affecting the lands bordering the Mediterranean Sea through travel and trade [1]. Historical records for the Black Death in 14th century Europe report the rapid infection of dock workers and sailors, as the immediate consequence of infected ships arriving in the ports of European cities, as in the cases of Messina and Marseille [2]. SARS rapidly disseminated from South-East Asia to Europe and North America brought by infected passengers traveling by plane [3, 4]. Historical examples not only provide the evidence supporting that hosts movements in a generally non-uniform environment represent an important mechanism for disease spatial spread, but also show how the intrinsic features of these movements are able to lead to completely different patterns of spatial spread, regarding both speed and geographical aspects.

In pre-industrial times, the spread of directly transmitted diseases was mainly a spatial diffusion phenomenon. Originating from Asia, the Black Death arrived in South-East Europe in the fall of 1347 and traveled across

most of continental Europe in less than three years reaching the far northern regions of the continent [2]. Despite the current debates on the origin of the plague, its epidemiological aspects emerging from historical records seem to indicate the presence of a direct contagion that could explain the high clustering of cases in households, and the vast spatial spread also to regions that would not favor the sustained epizootic transmission. In this perspective, the spatial propagation of the disease may be due mostly to human movements, and the diffusive pattern reported in contemporaries reports be the outcome of the relatively few traveling means available at that time to cover relatively short distances on the time scale of one day. Historical studies confirm that the disease diffused smoothly, generating an epidemic front that travelled through the continent as a continuous wave at a rate of about 200–400 miles/year (see Figure 1) [2].

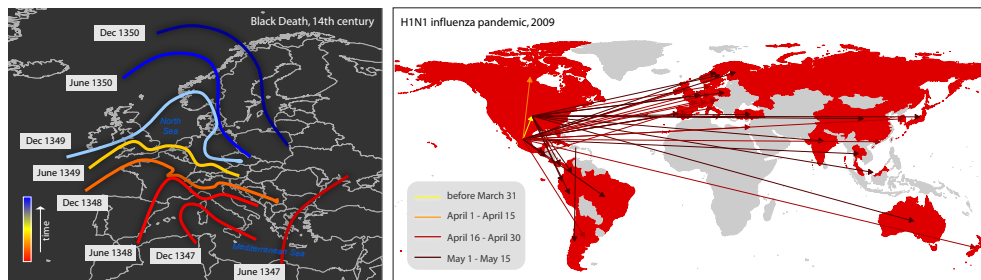


Figure 1: Epidemic spreading patterns changed dramatically after the development of modern transportation. In pre-industrial times, the spread of diseases was mainly a spatial-diffusion phenomenon (left). The speed and pattern of the 2009 H1N1 pandemic, in contrast, were radically different in terms of speed and non-local diffusion (right).

The advent of modern transportation has dramatically altered this picture, speeding up and favoring spatial disease transmission significantly. For example, already the dissemination of the 1889 influenza pandemic was extremely rapid. The so-called “Russian flu” moved along the travel routes by sea and along the network of railroads, and reached the United States through transatlantic travel [5]. The influenza pandemic of 1918 took less than one year to spread from its US or European source to isolated Pacific islands, while the 1957 flu virus swept the globe in about six months. For the first pandemic of the 21st century, the 2009 H1N1 influenza pandemic,

only two months passed between the first international alert and the WHO announcement that the epidemic had reached global proportions. Besides the speed of disease invasion that has rapidly changed with the advent of ever faster and efficient transportation systems, the same pattern of invasion has radically changed, from a continuous diffusive phenomenon to a discrete pattern of localized outbreaks in far away regions. This requires a change in the corresponding modeling framework as well. Although the spread of the Black Death can be adequately described mathematically using continuous differential equations with diffusive terms, the modeling of the spread of epidemics in modern times has to explicitly incorporate the spatial structure and deep interconnectedness of today’s modern society.

Going beyond the simple description of an epidemic in terms of “numbers appearing along the time horizon” [6], here we introduce spatial thinking in the modeling of disease transmission and provide the conceptual framework along with the corresponding qualitative and quantitative results of a class of spatial models – metapopulation models.

2 Spatially structured populations and metapopulation approach

The metapopulation modeling approach is an essential theoretical framework to describe a population dynamics whenever the spatial fragmentation of the host population is known to play a key role [7]. Conceptually, it relies on the basic assumption that the system under study is characterized by a highly fragmented landscape or environment in which the population is spatially organized in relatively isolated discrete *patches* connected by some coupling process. This means that, being structured into local populations, hosts have typically a higher degree of interaction when they are located in the same patch than when they belong to different patches. The interaction between different patches is however non-null, as it is provided by the coupling, given by hosts movements from one patch to another, which could take several different forms and generally represent migration or mobility processes. A typical example could be individuals located in different cities connected by transportation infrastructures along which they can travel and thus move from one city to another. Mobility represents the means by which people originally located in different cities can travel to the same location

and have the opportunity to interact. Departing from the homogeneous mixing assumption of basic compartmental models in single populations, the metapopulation approach is one of the simplest approaches that allows for the introduction of a spatial dimension in the modeling framework.

The approach was originally developed in ecology, to provide a spatial context to the studies of population biology [7]. Often, indeed, landscapes or environments can be seen as a network of connected habitat patches inhabited by local populations. This happens when the landscape itself is fragmented and prevents a continuous spatial distribution of the species, or when habitat loss occurs, i.e. when a natural habitat is destroyed and becomes functionally unable to support the survival of the species, or it is simply due to varying levels of fitness of the species to given locations of the environment. The application of the metapopulation approach to plants and other sessile organisms has been straightforward and has proved to be very effective, as the critical role of space and of neighborhoods (leading to coupling effects) is easily recognizable for organisms that do not move. Spatial positions became increasingly important also in ecological models focusing on mobile organisms, e.g. animals or individuals, due to the different capability of movements within subgroups of the same species (e.g. insect species with mobile adults but immobile larval stages), or to the emerging discontinuous structure of the species. With mobility not always being an adequate means to allow for efficient mixing in the population, the metapopulation approach offers a framework to analyze different ecological problems related to space, such as local extinction, recolonization and regional persistence [8, 9].

From ecology, this paradigm has been applied to genetics and evolution, and it is also very useful in the study of epidemics, aimed at the understanding of the influence of spatial structure in shaping the dynamics of infectious disease spreading in populations [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20]. The approach allows taking into account those situations in which space prevents the random mixing of individuals throughout the whole system, and different approximations need to be considered to define the interactions among hosts. From this point of view, space can be viewed as an additional heterogeneity of the host population that breaks down the assumption of homogeneous mixing. Similarly to age and social structuring of the population introduced to go beyond simple homogeneous mixing approximations, from a modeling perspective this translates into grouping the individuals according to different structuring criteria and in defining appropriate infection mechanisms among these groups of individuals, where the groups are now

identified by the individuals' spatial locations. However, there is a subtle difference between the age/social structure of the population and its spatial structure. Age and social features relevant to the transmission of infectious disease do not typically change in time or at most are characterized by some deterministic trend behavior. This is the case, e.g., of gender considered in the propagation of sexually transmitted diseases, or age for the transmission of childhood infectious diseases where the age of each person is updated year after year in the model following a deterministic increasing trend. On the other hand, the location of an individual, and thus her/his belonging to a specific group in the population substructure, can change repeatedly in time, and even explore several different locations depending on the specific mobility process considered. In migration processes, e.g., individuals may move permanently from one center to another. Temporary and recurrent movements are instead associated to commuting processes, from 'home' patch to 'workplace' patch, whereas temporary and non-recurrent movements behavior are typical of other hosts, such as e.g. livestock, and are dictated by trade patterns among animal holdings. The changing nature of the location of hosts adds a level of complexity that can be addressed by modeling through different approximations, as we will see in the following section. In general, differently from age and social structure where the interaction process requires a single level of description (given by the fixed or deterministically changing nature of the groups), the explicit description of movements make metapopulation models intrinsically hierarchical, representing a population of so-called *subpopulations*, each identified with a patch in the system and displaying a substructure that describes a set of interacting individuals who can move from patch to patch and enter different subpopulations (see Figure 2). From the individual, to the subpopulation, to the metapopulation level, the approach is characterized at least by two different scales of interactions: the *intra-patch* scale describing the hosts' interaction and contagion processes, and the *inter-patch* scale describing the interactions among subpopulations of hosts at the spatial level. Each level of description corresponds to specific time and length scales. Additional scales could be considered, e.g., to describe different processes of host movements, such as when dealing with multiple transportations and mobility dataset. For this reason, metapopulation models are also generally referred to as multiscale models.

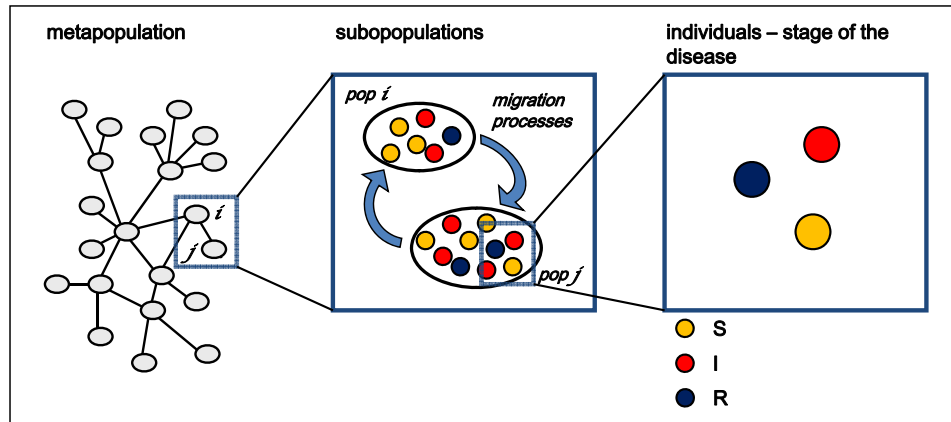


Figure 2: Schematic representation of a metapopulation model (from Ref. [20]). The system is composed of a set of subpopulations or patches, connected by coupling processes. Each patch contains a population of individuals who are characterized according to their stage of the disease (e.g. susceptible (S), infected (I), removed (R)), and identified with a different color in the picture.

2.1 Patches and coupling

Given the very large applicability of the metapopulation approach to a wide range of phenomena, the definition of a patch is intrinsically flexible. As we will see with illustrative examples, it is indeed hard to provide a clear-cut definition and classification of the various instances, and the choice of the correct scale and modeling approach – i.e. which patch definition and which interactions to consider – will mostly depend on the problem under study, on the data availability, and on the questions we aim to answer. By turning our attention to human hosts, several different modeling scales are possible. Along with the introduction of specific examples, in the following we will outline some general features and requirements that are valid at all scales.

Let us consider the study of the spatial spread of human infectious diseases, through two different scales and examples: seasonal influenza in a country, and an emergent disease such as pandemic influenza or SARS at the global scale. The overall scales of interest are automatically defined by the focus of our studies, thus models need to be centered on country geographical scales in the first example, and on the worldwide scale in the second exam-

ple. Plausible definitions of patches are urban aggregations, such as cities or large metropolitan areas where individuals are assumed to be homogeneously mixed. Coupling among patches is provided by the human movement patterns describing the mobility of individuals from one city to another. The coupling highlights the first difference between the two examples: which human movements are relevant for the description of an infectious disease like seasonal influenza in a country and of SARS invasion in the world? When the reference scale is the global one, it is rather straightforward to include the movements of individuals by air travel, as this is the relevant and crucial mean of transportation, given its ability to rapidly connect regions of the world that are very far apart. As discussed at the beginning of these notes, the features of high interconnectedness and rapidity that characterize modern transportation systems lead to the possibility of rapidly disseminating an emerging virus across the globe by means of infectious travelers, as experienced during the 2002-2003 outbreak of SARS and the 2009 H1N1 pandemic. Not considering air travel, and limiting the coupling to ground transportation only would clearly provide an inadequate modeling structure to describe the process. In the within-country description, on the other hand, it is reasonable to assume that air transportation by itself would not be enough to properly capture the movements of individuals relevant for the disease spreading, as this would neglect a large fraction of human movements that occur at a scale (short-range scale) comparable to the scale of the problem. Thus additional transportation means, such as ground transportation should be included as well. In this respect, many countries routinely collect data on the commuting behaviors of individuals [21]. Commuting generally include different means of transportation (from bicycles, to cars, buses, trains, and flights) and refer to the recurrent mobility behavior of individuals who live in a location ('home' location) and regularly travel to a different location for work or study ('workplace' location). These data provide the interactions among patches within a country, and the role of separate transportation modes can also be investigated.

A second difference can be noted in the identification and definition of patches. The commuting data, collected from national census or transportation surveys, are typically defined into georeferenced administrative areas – e.g. counties or municipalities, provinces, states – that map the whole territory of the country. If commuting is included in the metapopulation model as the coupling process, then administrative areas become the natural choice for the identification of patches for the description of an infectious disease within

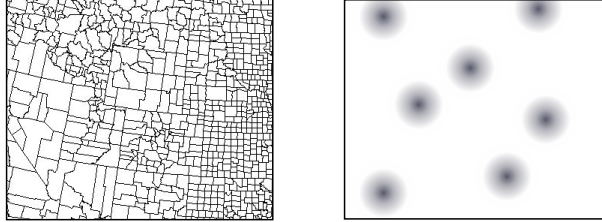


Figure 3: Schematic representation of patches and coupling where patches extension and boundaries are contiguous (left) or not (right).

the country. This corresponds to a demographically and geographically defined set of patches, with clearly defined extension and boundaries. In the metapopulation model characterizing the worldwide population distributed into cities connected by air travel, instead, patches may have a demographic definition only, obtained from census data to indicate the population of each patch, without a formal geographical extension. This is induced by the choice of the coupling, the air travel, and the lack of a spatial extension for the centers of mobility data. Commuting data is defined at the county or municipality level, representing a region, whereas air traffic data is defined at the airport level, which corresponds to a set of discrete points in space. The analogous of the administrative regions would be the definition of *basins of attraction* for each airport, representing the geographical boundaries of the area whose population would travel from/to the given airport. In general, if patches are well separated in space, a basic demographic definition of the patch with no spatial reference is adopted. This is the case for example of models that consider a sample of the worldwide air transportation network, with 52, 150 or 500 top airports spread across the globe. Given that the distance between two airports, e.g., Rome and Paris, is so large, the patches corresponding to associated airports do not specifically need a spatial definition (see Figure 3). This, however, also corresponds to neglecting the population not living in the patches included in the model, whereas in the case of administrative areas the whole population of the country is considered. If a higher level of detail is considered in the model aiming at covering the whole population of the system, the approximations needed in the approach may change. The large density of airports in given regions of the world coupled with high resolution demographic databases may require the

definition of the spatial extension of each patch, to identify geographically to which patch the population living in a given area belongs. This corresponds to a segmentation of the landscape which is not based on *a priori* definitions of administrative areas, but emerges directly from the definition of the metapopulation model. While georeferencing and definition of boundaries is a known problem in geography, its application to metapopulation models for infectious diseases was first introduced with a tessellation of the world into geographic census areas around main transportation hubs (i.e. airports) where short-range mobility between neighboring census areas is also considered in addition to air travel [22] (Figure 4).

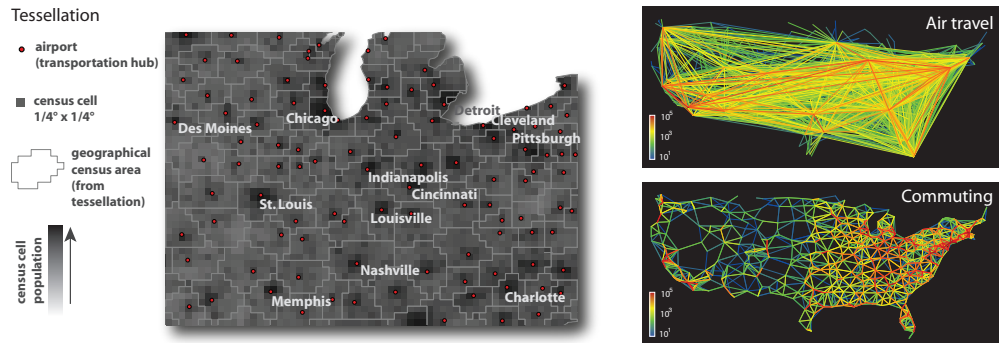


Figure 4: GLEAM – Global Epidemic and Mobility model –is an example of multiscale metapopulation model, with patches obtained through a tessellation procedure around major transportation hubs (left), and multiple modes of mobility including air travel and commuting (right, where a zoom on the US is shown, GLEAM extending on the entire globe). From Ref. [22].

In general, while it is often intuitive to identify a definition for a patch and the associated coupling adequate for the problem under study, the demographic and geographic definitions of the patches are not always straightforward. Very often, indeed, it is the identification of the coupling and the availability of the data quantifying it that allows the definition of the patch for the specific problem.

From the whole world to regions to countries, increasingly zooming in the spatial scales, metapopulation approaches can also be used to model epidemic outbreaks in small community settings, e.g. within a city. The increasing availability of data describing the movements of individuals within a city, from traditional traffic data to massive surveys to unsupervised data

collection, allows for the quantification of aggregated fluxes of individuals moving from location j to location l of the city. In this approach, patches can thus be identified by locations and coupling corresponds to the measured or estimated fluxes among locations [23]. While a metapopulation framework in essence, these modeling approaches are typically classified as agent-based approaches, since the high level of detail of the system and the creation of synthetic societies reproducing the real ones also require the inclusion of a vast number of features and parameters associated to each individual (e.g. age, employment/school activity, household structure, etc.), that go beyond the simple characterization of the population in terms of groups corresponding to locations. In addition, while individuals are indistinguishable in metapopulation approaches due to the homogeneous mixing approximation assumed inside subpopulations, their identity is tracked in time in agent-based model, as the specific features mentioned before need to be preserved and determine a set of behaviors at the individual level.

Another example in which spatial transmission is very relevant is the spread of infectious diseases among animals. Here the natural definition of the patch is the farm or premises where animals are being held, and the coupling represents the animals movements from one farm to the other [24, 25]. A common approximation that has generally been used in the literature is to consider the farm as the unit of the study and to assume an SIR dynamics at the level of the farms. This corresponds to a Levins-type approximation [8, 9] of the metapopulation system that ignores the possible impact of the within-farm epidemic dynamics. The implicit assumption considers that the disease spreads so fast in the animal population that, as soon as an animal becomes infected, then the whole farm becomes infected as in an instantaneous process. Besides biological reasoning, this assumption was also historically induced by the lack of data to characterize the animal populations within the farms and their movements from farm to farm. Also for this reason, the coupling in the contact network at the farm level approximation generally assumes an interaction between farms modulated on the distance [24]. This can be computed from the georeference data of the locations of the farms, usually available, and hypothesizes a higher transmission between farms on short distances and a weaker transmission on longer distances. Under this Levins-type approximation, models for livestock disease spreading at the farm level use the same description for an epidemic on contact networks, once the coupling is defined on data or on distance kernel models. However, recent availability of detailed data on animal census

and movements, allow for an explicit representation of the animal population in the model through a metapopulation approach, and few preliminary studies are available in this direction. Here we would like to note that this case represents yet another example of mobility. Differently from permanent migration, or recurrent patterns as typically observed in the travel and commuting of individuals, animal movements are characterized by highly non-trivial evolutionary dynamics induced by economic and trade driving forces, that prevent the definition of stable patterns across time [25]. In the absence of stationary or quasi-stationary processes that prevents the system to reach an equilibrium, modeling approaches should take into account the specific temporal aspects encoded in the data as these have a strong impact on the propagation of the epidemic [26, 25, 27, 28].

Finally it is important to discuss the role of timescales. Till now we have presented the role of different spatial scales to appropriately capture the hosts movement dynamics, their spatial structuring into patches and the relevance of these aspects with respect to specific infectious diseases. Implicitly, we have considered coupling processes whose timescales are relevant with respect to the characteristic timescale of the disease. In the cases described before, the timescales of the various processes range from fractions of a day (in the daily commuting process) to several days (in the air travel) and are therefore comparable to the timescales of the infectious diseases considered, influenza or SARS, whose evolution occurs on the timescales of few days. On the other hand, the fluxes of daily commuting or air travel appear to be less relevant to capture the spatial spreading of an infectious disease with much longer characteristic timescales, as e.g. HIV infection. The description at the global level on such long timescales would require different modeling ingredients – the movements of interest being not so much represented by the daily travel of people, but more importantly by the long-term massive permanent immigration of populations from one part of the globe to another. Empirical evidence have indeed shown the emergence of new HIV subtypes in Western Europe, due to migrations from Africa [29]. Metapopulation approaches can help explaining this phenomenon by using estimates of immigrating fluxes and rough approximations for the spreading dynamics at the level of large population.

2.2 Relevant spatial effects

The spatial structure of the population in relatively separated communities can have a strong impact on the dynamics of infectious diseases. Some of the effects known for the epidemics in single populations may be suppressed or enhanced, depending on the interplay between the spatial separation, the coupling and the population distribution into the patches. Metapopulation systems are indeed characterized by many subpopulations that interact through the coupling process. The strength of the coupling allows for different phenomena to occur, such as isolation, local or global extinction, persistence, synchronization [8, 9, 15, 30]. In the extreme case of absence of coupling, the epidemics in each single population are totally independent and isolated, and their behavior depend only on the local conditions of the population, the force of infection (which may be patch-dependent), and on the time of the start of the epidemic. If the coupling is very small, then the subpopulations behave almost independently as the interactions by means of movements of infectious individuals is very small compared to the local evolution of the disease within the patch, and the main effect of the interaction is the onset of the epidemic in each patch. In this situation, some patches may still be isolated though they are coupled to the rest of the system. This occurs if the patches have a very small exchange of individuals with neighboring patches and thus their populations have few contacts with the other communities where an epidemic may be unfolding. The spatial separation thus behaves as a mean of protection against the risk of transmission. If the coupling is very large, then the subpopulations are strongly interacting by exchanging a large fraction of individuals and thus increasing their possibility to mix almost homogeneously beyond the spatial constraints. This induces a strong synchronization in the time behavior of the epidemics within the patches that results as if the multiple subpopulations act like one single large homogeneously mixed population [30]. Another important aspect induced by spatial separation concerns the persistence of the disease in the global population. While in a single population the time of full duration of an epidemic is given by the epidemic and local conditions only, the same epidemic may persist for a longer time period if the population is spatially structured. Once the epidemic is fading out in a subpopulation, movements of individuals may allow the so-called *(re)-colonization events*, i.e. the new infection of a subpopulation not yet infected that enables the survival of the pathogen in a population of newly susceptibles [30]. Under given conditions,

host mobility may indeed favor the possibility of spreading of the epidemic. Very important are also the stochastic events that lead to the extinction of the disease in a subpopulation. When the population is spatially structured, such extinction events may occur in each subpopulation, and the conditions under which this can occur are extremely important from a public health point of view for the eradication of the disease. Being the system spatial, we have to distinguish between local extinctions and global extinction. If the coupling is very small, recolonization events are very rare and the patches behave almost independently also concerning their possibility of extinction – rescue effects to reestablish an epidemic in a given subpopulation have a small chance to occur and epidemics may go locally extinct. However, this does not necessarily correspond to a global eradication of the disease, as it may well circulate in other communities of the metapopulation system. If the coupling is very large, on the other hand, the local epidemics are well synchronized and extinction events at the local level may lead to the extinction at the metapopulation level as they all occur at the same time without the possibility of survival for the pathogen in relatively isolated communities. Intermediate values of coupling are expected to lead to more complicated conditions for the spread of the disease from one population to another, where the lack of synchronization is accompanied by the occurrence or recolonization events, so that disease persistence is enhanced [30]. The only addition of the spatial dimension in the population under study allows therefore for an interesting spectrum of possibilities regarding the persistence of an epidemic, with a crucial trade-off induced by the coupling strength. Additional effects such as seasonal forcing increase the complexity of the problem because of their interplay with synchronization effects. These aspects, along with stochastic resonance, have been vastly studied in the analysis of measles where they were recognized to play a major role in the maintenance of recurrent epidemics [14, 15, 31]. In the following, we will restrict our attention to the role that discrete and stochastic effects have in shaping the disease spread in the metapopulation system and the computational approaches that can be considered, but we will always assume a constant transmissibility that does not vary with time.

3 The stochastic discrete metapopulation scheme

In order to present the mathematical and computational descriptions of an epidemic spatial spreading through a metapopulation approach, we consider the simple susceptible-infectious-recovered (SIR) model [32, 33]. Differently from a single population approach, the global dynamics of a metapopulation system is described by a set of coupled SIR schemes, each associated to a single subpopulation. We need therefore to define a set of variables to denote the susceptible, infectious and recovered individuals in each patch, as a function of time. The variables S_t , I_t , R_t thus become:

$$\begin{aligned}
 S_t &\rightarrow S_{1,t}, S_{2,t}, S_{3,t}, S_{4,t}, \dots, S_{V,t} \\
 I_t &\rightarrow I_{1,t}, I_{2,t}, I_{3,t}, I_{4,t}, \dots, I_{V,t} \\
 R_t &\rightarrow R_{1,t}, R_{2,t}, R_{3,t}, R_{4,t}, \dots, R_{V,t} \\
 N &\rightarrow N_{1,t}, N_{2,t}, N_{3,t}, N_{4,t}, \dots, N_{V,t}; N_{tot} = \sum_j N_{j,t}
 \end{aligned}
 \tag{1}$$

where the suffix j of $S_{j,t}$ ($I_{j,t}$ or $R_{j,t}$) indicates the number of susceptible (infected or recovered, respectively) individuals in subpopulation j , with $j = 1, \dots, V$ and V representing the total number of patches in the system. Analogously to the single population case, the population size at each patch is given by $N_{j,t} = S_{j,t} + I_{j,t} + R_{j,t}$, and in principle it may change in time depending on the mobility process considered. The total population considered in the metapopulation system, N_{tot} , is given by the sum of the populations of each patch of the system.

Similarly to the SIR in a single population, we can write Markov chain relations describing the time behavior of the SIR in the metapopulation system:

$$\begin{aligned}
 S_{j,t+\Delta t} &= S_{j,t} + \Delta S_{j,t} \\
 I_{j,t+\Delta t} &= I_{j,t} + \Delta I_{j,t} \\
 R_{j,t+\Delta t} &= R_{j,t} + \Delta R_{j,t},
 \end{aligned}
 \tag{2}$$

where the quantities $\Delta S_{j,t}$ ($\Delta I_{j,t}$ and $\Delta R_{j,t}$) indicate the variations in the population of the susceptible (infected and recovered, respectively) individuals during the time step Δt . These variations are due to two processes: the *within-patch* dynamics and the *inter-patch* dynamics, i.e. the coupling. Let us start from the processes that occur within each patch. The change in the

populations of each compartment for each time step Δt are given by the generation I_+ of new infectious individuals through the successful transmission of the pathogen from infectious to susceptibles, and by the recovery of R_+ infectious individuals that at time $t + \Delta t$ enter the R class (see Eqs. (2.2)). These variations are defined by the disease parameters, i.e. the transmission rate β and the recovery rate μ , and by the homogeneous mixing approximation assumed in the model, and are modeled with binomial stochastic variables to ensure the discrete stochastic nature of the processes. The same expressions can be written also for the metapopulation approach and represent the variations of S_j , I_j and R_j for subpopulation j . In addition, we have to consider the coupling process that allow changes in the compartment sizes due to the mixing of the individuals in subpopulation j with neighboring subpopulation l . This contribution can be separated from the epidemic process occurring exclusively within the patch, so that for each compartment X in subpopulation j we can write $\Delta X_{j,t}$ as the sum of the appropriate variations expressed in terms of the various X_+ and the coupling factor. For the moment, let us indicate the coupling factor with the variable $\Omega_{j,l,t}^X$ for the compartment X in subpopulation j , so that we can write Markov chain relations that explicitly distinguish the within-patch from the inter-patch dynamics:

$$\begin{aligned}
S_{j,t+\Delta t} &= S_{j,t} - I_{j,+} + \Omega_{j,l,t}^S \\
I_{j,t+\Delta t} &= I_{j,t} + I_{j,+} - R_{j,+} + \Omega_{j,l,t}^I \\
R_{j,t+\Delta t} &= R_{j,t} + R_{j,+} + \Omega_{j,l,t}^R.
\end{aligned} \tag{3}$$

Here $I_{j,+} \sim \text{Bin}(S_{j,t}, \beta \Delta t I_{j,t} / N_j)$ is a binomially distributed random variable that represents the new infections generated in subpopulation j from the encounter of susceptible $S_{j,t}$ and infected individuals $I_{j,t}$ within the population N_j at time t , and $R_{j,+} \sim \text{Bin}(I_{j,t}, \mu \Delta t)$ is a binomially distributed random variable that represents the new recovered individuals at time $t + \Delta t$ from previously infectious individuals in j . Here we also assume that the transmission rate β is kept constant at the metapopulation level. A more general version of the metapopulation approach may consider a spatial dependence in the transmissibility that may change from subpopulation to subpopulation, i.e. β_j , to take into account for local aspects (such as environmental factors or population structure within the patch) that enhance or suppress the transmission. Additional complications may also include the dependence on time, such as e.g. in the case of the seasonal forcing we discussed before where the transmission rate may depend on both the location and the time, $\beta_{j,t}$.

This is a mechanism commonly used to model the seasonality of influenza in different regions of the world, i.e. the empirical evidence that transmission is enhanced during winter and suppressed during summer times, taking into account the season (i.e. time) dependence and the different behavior observed in the Northern and Southern Hemispheres of the world due to the counter-phase of seasons [22].

The coupling factors $\Omega_{j,l,t}^X$ depend not only on the compartment X in the subpopulation j at time t , but clearly they depend also on the compartments in subpopulation l , neighbors of j , where neighbors means connected through coupling processes. The explicit expression of $\Omega_{j,l,t}^X$ depends on the modeling approximation used for the description of the coupling process. In the following, we will see the phenomenological approach that treats the coupling as an effective force, and the mechanistic approach that explicitly models the movement of hosts from one subpopulation through another.

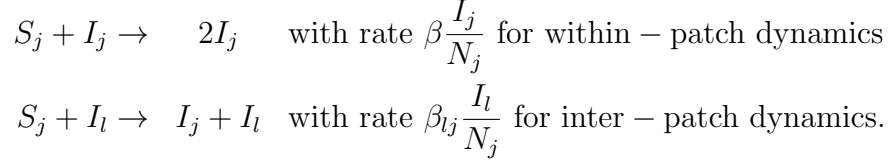
3.1 Effective approach

The most basic approach to take into account for coupling effects is achieved by treating the coupling in a phenomenological way, i.e. allowing for the definition of a likelihood of contracting the infection for susceptibles of a given subpopulations j that effectively accounts for the infectious individuals present in the metapopulation system, besides those present in j [31, 34, 35, 24, 36]. In each subpopulation j , the Markov chain equations can be written as in Eqs. (3), however the quantity $\Omega_{j,l,t}^X$ defined in the previous section is now expressed as a force of infection arising from the other subpopulations which are in contact with the subpopulation under study. In other words, a susceptible individual in subpopulation j is exposed to a total force of infection λ , which now includes the force of infection induced by the presence of infectious individuals in the same subpopulation j (i.e. the factor leading to the new generation of infectious individuals, $I_{j,+}$), plus the one induced by the infectious individuals who are in subpopulations l connected to j . The latter is defined in a way analogous to the standard definition, $\beta I_t/N$, of the force of infection in a single population:

$$\sum_l \beta_{lj} \frac{I_{l,t}}{N_j}, \quad (4)$$

where l is a subpopulation coupled with j and β_{lj} is the effective transmission rate from l to j . Clearly, this formulation of the coupling process affects only

the change in the compartment populations of susceptible and infectious individuals, since the recovered compartment is not influenced by the force of infection. In terms of reaction processes we thus have for population j :



If we include the subpopulation j in the sum over the index l , the overall force of infection for a susceptible individual in j is given by a weighted sum of forces of infection, $\sum_l \beta_{lj} I_l / N_j$, where it is commonly assumed that the transmission rate within each community has one value, β , whereas a smaller value is assumed for contacts between patches:

$$\begin{aligned} \beta_{lj} &= \beta \text{ for } l = j \\ \beta_{lj} &= \epsilon \beta \text{ for } l \neq j, \end{aligned} \tag{5}$$

with $\epsilon < 1$ defines the coupling strength. Another formulation of the effective coupling instead assumes the following values for the transmission rates within and inter-patches:

$$\begin{aligned} \beta_{lj} &= \beta(1 - \sigma) \text{ for } l = j \\ \beta_{lj} &= \beta\sigma \text{ for } l \neq j, \end{aligned} \tag{6}$$

in order to keep the reproductive number $R_0 = \beta/\mu$ independent of variations of the coupling strength σ to allow for the comparison between models with different coupling strengths. In both formulations, for the sake of simplicity, the dependence of the coupling strength from origin and destination (i.e. subpopulations j and l) is disregarded and the coupling is considered homogeneous across the metapopulation system. More in general, the coupling strength depends on the specific subpopulations under study as the movements of hosts from l to j is found to dramatically vary in real datasets [20]. In the general formulation where we retain the dependence of the coupling strength on the specific connections between subpopulation, i.e. ϵ_{ij} (or, analogously σ_{ij}), we can write the following equations regulating the behavior of an SIR epidemic in a metapopulation system with effective coupling, in terms of the values $s_{j,t}$, $i_{j,t}$ and $r_{j,t}$ of the stochastic variables $S_{j,t}$, $I_{j,t}$ and $R_{j,t}$:

$$s_{j,t+\Delta t} = s_{j,t} - \text{Bin} \left(s_{j,t}, \beta \Delta t \frac{i_{j,t} + \sum_l \epsilon_{lj} i_{l,t}}{N_j} \right)$$

$$\begin{aligned}
i_{j,t+\Delta t} &= i_{j,t} + \text{Bin}\left(s_{j,t}, \beta\Delta t \frac{i_{j,t} + \sum_l \epsilon_{lj} i_{l,t}}{N_j}\right) - \text{Bin}(i_{j,t}, \mu\Delta t) \quad (7) \\
r_{j,t+\Delta t} &= r_{j,t} + \text{Bin}(i_{j,t}, \mu\Delta t).
\end{aligned}$$

Finally, it is important to note that the effective approximation described here attempts to estimate the actual mixing of individuals between subpopulations, without explicitly allowing for the movements of individuals from one patch to another. Individuals always stay in one patch and are exposed to the epidemics undergoing in neighboring patches. From this point of view, this approach clearly oversimplifies the two distinct processes of contagion and movements, and describes the movement process as an effective additional force of infection. The result is completely analogous to the structuring of the population in social or age classes of basic compartmental models where individuals belong to one single class only and their interactions with the other individuals of the populations is described through an infection matrix β_{lj} between classes l and j [32, 33]. As in the social structuring, also in the metapopulation system with effective approach, V^2 parameters of the matrix β_{lj} need in principle to be estimated, with V being the number of subpopulations. However, several assumptions can be employed to simplify this task, by assuming that all the values on the diagonal (i.e. within-patch) are equal to a constant value β and that all off-diagonal values (i.e. inter-patch) are equal to a fraction of this value, e.g. $\epsilon\beta$ as in Eqs. (5). Additional simplifications may arise from a more detailed knowledge of the coupling among the patches, where at least the topology of the interactions between the subpopulations are known, or are based on modeling assumptions (as e.g. distance kernel assumptions), or are inferred from incidence data. For this reasons, the effective approach to describe an epidemic in a metapopulation system, besides simplifying the dynamics, was also the first to be historically introduced given the lack of detailed data on hosts movements that could allow the explicit modeling of the coupling dynamics.

3.2 Mechanistic approach

Mechanistic approaches consider the explicit movements of individuals who leave a subpopulation j and enter a subpopulation l , so that, at time t , they are therefore exposed only to the local risk of transmission, i.e. the one experienced in the subpopulation they are located in at time t . If we indicate with $X_{jl,+}$ the stochastic variable quantifying the number of individuals in

compartment X traveling from subpopulation j to subpopulation l at time t , the coupling variables $\Omega_{l,j,t}^X$ introduced in Eqs. (3) can be simply expressed as:

$$\begin{aligned}\Omega_{j,l,t}^S &= \sum_j (S_{lj,+} - S_{jl,+}) \\ \Omega_{j,l,t}^I &= \sum_j (I_{lj,+} - I_{jl,+}) \\ \Omega_{j,l,t}^R &= \sum_j (R_{lj,+} - R_{jl,+}).\end{aligned}\tag{8}$$

In other words, for population j , they represent the balance of the incoming flows of individuals in compartment X from neighboring subpopulations l ($X_{lj,+}$ with positive sign in the above equations) and the outgoing flows of individuals in compartment X from subpopulation j to neighboring subpopulations l ($X_{jl,+}$ with negative sign).

The full formulation of the model in the mechanistic approach now requires the definition of the mobility process of individuals and in particular of the mobility rates at which individuals leave each patch and the probability to reach a particular destination [20]. In principle, this approach requires V^2 parameters of mobility rates p_{jl} times the number of compartments to define the movement of each individual in the population from every origin j to each possible destination l . As in the previous approach, however, several possible approximations can be adopted to simplify the problem. Typically compartments are characterized by the same mobility rate, or, in the case of more complicated compartmental structures including e.g. hospitalizations, quarantine, and others, the model may assume that these individuals are restricted from traveling to ensure realism in the simulations. Simplified mobility assumptions may also assume that the rate of leaving a patch is uniform and that the movements of individuals are equally distributed on all possible destinations, where the set of destinations is identified by available data (mobility networks) or modeling assumptions (such as nearest neighbors on a spatially embedded topology). Or, having fixed a uniform leaving rate from each patch, the distribution of the moving individuals across destinations can be assumed to be heterogeneous, based on statistical properties extracted from the data or based on modeling assumptions that consider a varying coupling induced e.g. by the distance, as in the distance kernel approximation. For the presentation of different mobility processes and their inclusion in metapopulation epidemic models, we refer the reader to Ref. [20].

Besides modeling assumptions as the ones just discussed that allow the definition of the metapopulation model in absence of mobility data or its study from a theoretical point of view (also enabling an analytical treatment under certain conditions, as shown in Section 4) the power of the mechanistic approach is fully exploited when data on the movements of hosts are available. The recent availability of large spatiotemporal data on human behavior, interactions and movements have lifted the constraints that were limiting the design of realistic data-driven metapopulation models, and allow the definition of mobility rates based on the data. In this framework, the data typically provide the number of hosts traveling from subpopulation j to subpopulation l in a unitary timescale, w_{jl} . This number is independent of the population size of the patch of origin, N_j , so that by considering a mean-field approximation we can define the diffusion rate $\sum_l w_{jl}/N_j$ as the probability per unit time that an individual located in j leaves the subpopulation, where $\sum_l w_{jl}$ is the total number of people traveling out of j provided by the data. The approximation is valid for the case of indistinguishable individuals, where no individual heterogeneity that may change the mobility behavior of hosts (such as e.g. age, social status, profession, and others) is considered. By defining the mobility rate of individuals from subpopulation j to subpopulation l as $p_{jl} = w_{jl}/N_j$, the outgoing flow of individuals $X_{jl_i,+}$ in compartment X from j to any patch l_i connected to j (with $i = 1, \dots, k$, and k being the degree of the patch j) is thus given by a stochastic variable that follows a multinomial distribution $\sim Mult(X_{j,t}, w_{jl_1}\Delta t/N_j, w_{jl_2}\Delta t/N_j, \dots, w_{jl_k}\Delta t/N_j, 1 - \sum_i w_{jl_i}\Delta t/N_j)$. The multinomial distribution is needed because the process now involves multiple possible outcomes: an individual in compartment X and location j at time t may travel to connected patch l_1 with probability $w_{jl_1}\Delta t/N_j$, or to l_2 with probability $w_{jl_2}\Delta t/N_j$, and similarly for all connected patches l_i , or may remain in her/his original location j with probability $1 - \sum_i w_{jl_i}\Delta t/N_j$.

The stochastic set of equations regulating the behavior of the epidemic in the mechanistic metapopulation system under this approximation can be written as:

$$\begin{aligned}
s_{j,t+\Delta t} &= s_{j,t} - Bin(s_{j,t}, \beta\Delta t i_{j,t}/N_j) + \sum_j (s_{lj,+} - s_{jl,+}) \\
i_{j,t+\Delta t} &= i_{j,t} + Bin(s_{j,t}, \beta\Delta t i_{j,t}/N_j) - Bin(i_{j,t}, \mu\Delta t) + \\
&\quad + \sum_j (i_{lj,+} - i_{jl,+})
\end{aligned} \tag{9}$$

$$r_{j,t+\Delta t} = r_{j,t} + Bin(i_{j,t}, \mu\Delta t) + \sum_j (r_{lj,+} - r_{jl,+}),$$

where $s_{lj,+}$ and $s_{jl,+}$ (and analogously for r and i) are the random values obtained from the multinomial distribution described above for the compartment $S_{j,t}$ ($R_{j,t}$, $I_{j,t}$).

It is clear from Eqs. (9) that the mechanistic approach does not lead to a formulation that is analogous to the one adopted for the structuring of the population in social or age classes. Differently from the effective approach, the force of infection is the same as in the single population and is accompanied by the description of the movement of individuals. Only through the explicit movement, individuals may become exposed to an outbreak undergoing in a given location. As discussed at the beginning of the chapter, the possibility to move and visit different communities in time (thus changing the group to which the individual belongs to) represents a major modeling difference with respect to the structuring of the population by age or social aspects, and thus requires a true multi-level description of the process.

Mobility processes that can be included in the metapopulation model, can be of very different nature. As discussed when introducing patches and coupling, the movement of hosts may be a permanent migration from one patch to another (as it occurs in the case of livestock movements from one farm to another) or a transient mobility process (as typically in the case of travels of individuals who return to their home location after the trip) with recurrent patterns to the same destination and almost fixed visiting times (such as for the commuting form of mobility where individuals typically spend a duration equal to a working day in a specific work location), or varying destinations and visit lengths (as in the case of business or pleasure trips to different destinations). The mobility type has also important implications in the time behavior of specific quantities as e.g. the population sizes of the patches. While these quantities are conserved in time for transient mobility processes, as the mobility is by definition symmetric (traveling individuals will come back to their home locations), population sizes are found to strongly change in time for permanent mobility processes (e.g. cattle are moved from one type of animal holding to another for trade, fattening, etc. without necessarily go back to their initial location). If recurrent, though not symmetric, patterns are considered, as in the case of homogeneous rate of diffusion out of a patch but heterogeneous rate of movement to the available possible destinations, a stationary state for the population sizes can be reached after an

initial transient of the dynamics [20]. This stationarity greatly simplifies the process and is equivalent to drop the time dependence of $N_{j,t}$ as we have done in the previous equations.

The above equations are also based on another assumption. Till now we have illustrated a process that defines the individuals through compartments that depend on the location j and the time t only. The location j indicates the patch where the individual is found at that given time t . Clearly, if we want to model recurrent mobility processes (as e.g. home to workplace and return), we would need to add two ingredients: the identification of the origin patch of the individuals (the home patch), and the definition of the return rates, depending on the visiting times. This is accomplished by structurally modifying Eqs. (3) to incorporate a recurrent dynamics and by defining new variables for the compartments that now explicitly have two location indices besides the dependence on time – $X_{jl,t}$, indicating the number of individuals in compartment X of population j (home patch) who currently are located in l (workplace patch). This approach has been originally introduced in Refs. [37, 38] and also incorporated into a data-driven metapopulation model with multi-scale mobility [22]. Within the recurrent mobility framework, moreover, it is possible to reconcile the effective approach and the mechanistic approach of a metapopulation model by demonstrating an equivalence based on the appropriate estimation of the parameters on one formulation from the parameters of the other [37]. In the following, we will adopt the markovian assumption for the mobility process presented in Eqs. (3), in that at each time step the rate of movement of individuals is given by p_{jl} and it applies to all individuals in the subpopulation without having memory of their origin. We will focus on this approximation for the computational implementation of the model and for the definition of properties of the metapopulation system. At the end of these notes we will also address non-markovian mobility processes and discuss the effects on those properties.

4 Local vs. global invasion

So far we have focused on the definition, characterization, and implementation of a metapopulation system and its various modeling assumptions concerning mobility processes and effective vs. mechanistic approaches. An infectious disease with a given reproductive number R_0 can however invade a closed and fully susceptible population only if $R_0 > 1$ [32, 33]. In the

simple case of an SIR model we have $R_0 = \beta/\mu$ and the threshold condition translates into stating that the transmission rate needs to be larger than the recovery rate, i.e. the epidemic needs to generate a number of new infections larger than the ones who recover. It is natural to wonder whether this condition still applies to a set of fully susceptible populations which are coupled to each other in a metapopulation system (i.e. in an overall closed population, though spatially structured). If we consider the effective approach, the same expression of the reproductive number can be easily recovered if we adopt the formulation of Eqs. (6) where also the within-patch transmission rate β_{jj} is rescaled by a function of the coupling strength σ in order to preserve the R_0 value for varying values of σ , as previously noted. In the case of mechanistic approach, however, the calculation of R_0 is complicated by the explicit representation of the mobility process that is absent in the effective approach. The analytical treatment of a metapopulation model, however, is not an easy task since we need to solve a system of V sets of non-linear differential equations (one for each compartment) coupled by the mobility process (even in the simplifying deterministic approximation). Prompted by the study of real-world mobility networks and leveraging on some of the statistical properties recurrently found in these systems, an approach was recently introduced based on the reaction-diffusion (RD) framework (already used to model e.g. chemical and physical phenomena) that provides an analytic solution of a metapopulation epidemic model with explicit movements [39, 40, 20]. Mobility networks are indeed found to exhibit important variability in the number of connections of each patch, a property that can be mathematically encoded in a heavy-tailed probability distribution $P(k)$ for the degree k of a given patch, i.e. the number of its connected neighbor patches [41, 21]. The approach is based on the statistical equivalence for subpopulations of similar degree. This is a mean-field approximation that considers all subpopulations with a given degree k as statistically equivalent, thus allowing the introduction of degree-block variables that depend only on the subpopulation degree. That is, instead of formulating the model by writing equations in terms of individual subpopulations j and l as we have done in the previous section, the subpopulations of the system are grouped according to one relevant topological feature, their degree k , and all other features specific of a given subpopulation j (such as e.g. its spatial position, the cultural aspects of its population, etc.) are disregarded. This is analogous to the degree-block approximation introduced for the study of the epidemic on contact networks [42]. The difference is that here the nodes have a substructure (i.e. the scale of the

individuals inside each subpopulation) where the epidemic unfolds and may then spread throughout the system through mobility fluxes. Although this is an obvious approximation of the system description, it has been successfully applied to many dynamical processes on complex networks and it is rooted in the empirical evidence gathered in previous works. Besides the large fluctuations observed in k , several scaling properties as functions of the degree k have been uncovered in the study of real-world demographic and mobility datasets, including population size, travel fluxes, and others [41, 21].

4.1 Local epidemic threshold

By changing the description of the epidemic metapopulation process from quantities that depend on the specific subpopulation j (S_j, I_j, R_j, p_{jl}) to quantities in the degree-block approximation that depend only on the degree k ($S_k, I_k, R_k, p_{kk'}$, where $p_{kk'}$ indicates the mobility rate of an individual from subpopulation of origin with degree k to subpopulation of destination with degree k') it is possible to express the time behavior of the epidemic at the metapopulation level in terms of RD rate equations for the quantities $S_k(t)$, $I_k(t)$ and $R_k(t)$ for each degree class k .

The dynamical rate equations for the number of infectious individuals in the deterministic continuous limit can be written as

$$\begin{aligned} \partial I_k = & - [p_k I_k + (1 - p_k) \mu I_k] + \\ & + (1 - p_k) \beta \Gamma_k + k \sum_{k'} P(k'|k) p_{k'k} [(1 - \mu) I_{k'} + \beta \Gamma_{k'}], \end{aligned} \quad (10)$$

where $\Gamma_k = S_k I_k / N_k$ indicates the transmission kernel, $p_k = \sum_{k'} p_{kk'}$ represents the diffusion rate out of a given subpopulation with degree k , and $P(k'|k)$ is the conditional probability that a patch with degree k is connected to a patch with degree k' . The first term, in [], on the r.h.s of the equation represents the depletion term of infectious individuals as the sum of those who diffuse away of the subpopulation ($p_k I_k$) and those who stay in the subpopulation and recover ($(1 - p_k) \mu I_k$). The remaining terms of the r.h.s take into account both the new infected individuals generated by the disease dynamics within the subpopulation ($(1 - p_k) \beta \Gamma_k$) and the infected individuals who diffuse from the neighboring subpopulations with diffusion rate $p_{kk'}$. Similar expressions can be written also for the evolution of S_k and R_k .

An explicit solution to the previous equations can be obtained for the early stages of the epidemic [20], when we can assume very small densities

of infectious individuals in the metapopulation system so that contributions of order I_k^2 can be neglected. In this setting the reaction kernel can be approximated as $\Gamma_k = \frac{(N_k - I_k - R_k)I_k}{N_k} \simeq I_k$, where we have neglected all terms of order I_k^2 and considered that R_k is of the same order of I_k in the early stage of the dynamics. By plugging this expression into the above equations and assuming an uncorrelated system (i.e. $P(k'|k) = k'P(k')/\langle k \rangle$), we obtain:

$$\begin{aligned} \partial I_k = & -p_k I_k + (1 - p_k)(\beta - \mu)I_k + \\ & + \frac{k}{\langle k \rangle} \sum_{k'} k' P(k') p_{k'k} [(1 - \mu + \beta)I_{k'}]. \end{aligned} \quad (11)$$

At this point we need to characterize the mobility process and define $p_{kk'}$ to obtain explicit solutions for the early dynamics. Consider the mobility process described in Subsection 3.2 where $p_{kk'} = w_{kk'}/N_k$, and let us assume the scaling relation found in real networks for the traffic flows $w_{kk'}$ as a function of the degrees of the origin and destination subpopulations, $w_{kk'} = w_0(kk')^\theta$. The diffusion rate out of a patch with degree k is then $p_k = \sum_{k'} w_0(kk')^\theta/N_k = T_k/N_k$ where T_k is the total traffic flow of the patch and from normalization conditions it is possible to obtain the relation $T_k = k^{1+\theta}w_0\langle k^{1+\theta} \rangle/\langle k \rangle$. The equations for the early stage thus become:

$$\partial I_k = -p_k I_k + (1 - p_k)(\beta - \mu)I_k + \frac{k^{1+\theta}}{\langle k^{1+\theta} \rangle} (1 + \beta - \mu)\Omega, \quad (12)$$

where $\Omega = \sum_k P(k)p_k I_k$. By averaging both sides of the equation over $P(k)$, it is possible to write an equation for the early stage behavior of the average number of infectious individuals $\bar{I} = \sum P(k)I_k$, $\partial \bar{I} = (\beta - \mu)\bar{I}$, yielding

$$\bar{I} = \bar{I}(0)e^{(\beta - \mu)t} \quad (13)$$

and thus recovering the epidemic threshold condition $R_0 = \beta/\mu > 1$. The condition depends only on the reaction rates and is not affected by the spatial structure of the population [20]. Intuitively this is stating that if the epidemic is not able to proliferate in each local subpopulation, then it cannot produce a major outbreak at the metapopulation level. Analogous analytic treatments can be done for different mobility processes that yield the same solution for the expression of the reproductive number [20].

4.2 Global invasion threshold

Although the $R_0 > 1$ result is valid at the level of each subpopulation, the epidemic behavior at the metapopulation level is determined also by the diffusion process of individuals. In particular, the effects due to the finite size of subpopulations, the discrete nature of the individuals and the stochastic nature of the diffusion may have a crucial role in the problem of resurgent epidemics, extinction and eradication [43, 44, 45, 46, 47]. Each subpopulation may or may not transmit the infection to another subpopulation it is in contact with, depending on the occurrence or not of the travel event of at least one infected individual to the non-infected subpopulation during the entire epidemic duration (see Subsection 2.2). Intuitively, this must depend on the level of mixing among the subpopulations. If there is no mixing at all, then the epidemic is restricted to the seeded subpopulation. Varying coupling strength has important effects in the synchronization of the epidemics at the subpopulation level. Here we would like to assess under which conditions a single initially infected subpopulation is able to lead to an epidemic that invades the system at the global level. This is captured by the definition of a new predictor of disease invasion, R_* , regulating the number of subpopulations that become infected from a single initially infected subpopulation [43, 44, 45], i.e. the analogous of the reproductive number at the subpopulation level [32, 33]. The emergence of the global invasion is due to the interplay of the timing and size of the epidemic in a given subpopulation, and the diffusion rate that must be large enough to ensure the timely diffusion of infected individuals to other subpopulations of the metapopulation system, before the local epidemic outbreak dies out. If we want to capture this phenomenon, we clearly need to consider the discrete stochastic nature of hosts in their movements, since deterministic continuous description would always lead to an invasion.

Given the analogy between the R_0 and R_* parameters in determining the conditions for the spreading at the individual and patch level, respectively, it is possible to study the invasion dynamics considering the subpopulation model in a coarse-grained view and provide a characterization of the invasion dynamics at the level of the subpopulations [39, 20], translating epidemiological and demographic parameters into Levins-type metapopulation parameters of extinction and invasion rate. More in detail, abandoning the description in terms of $S_k(t)$, $I_k(t)$, $R_k(t)$, it is possible to define analogous quantities at the level of subpopulations, and in particular define the number

of infected subpopulations of degree k . The notion of time in the microscopic evolution of the disease at the level of individuals within each subpopulation translates into the definition of steps or generations n , where $n = 0$ indicates the infected subpopulations at the beginning of the process, $n = 1$ indicates the set of subpopulations directly infected by those of generation 0, and so on for the following generations (see Figure 5 for a schematic example of the invasion process). This corresponds to a basic branching process at the subpopulation level, similar to the chain of transmission occurring at the individual level. Under the assumption that during the early stage of the epidemic the number of subpopulations affected by an outbreak (with $R_0 > 1$) is small, the branching process can adopt a tree-like approximation that allows relating the number of infected subpopulations at generation n to the ones at generation $n - 1$.

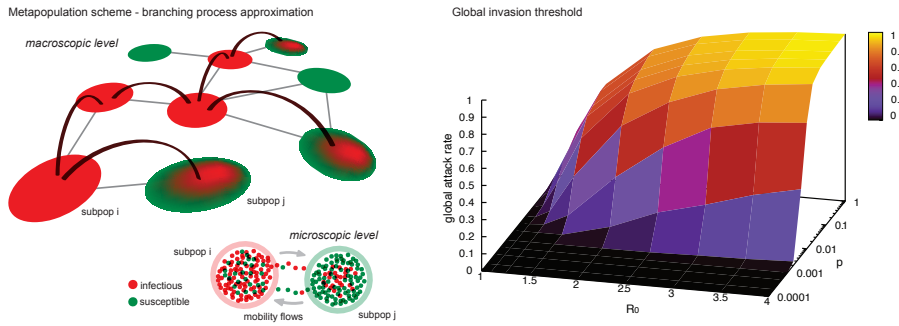


Figure 5: Schematic illustration of the branching process of epidemic diffusion at the subpopulation level, where subpopulation are identified as infected or not (left, from Ref. [48]). Global invasion threshold in a heterogeneous metapopulation system: the global attack rate is shown as a function of R_0 and the diffusion rate p (right, from Ref. [39]).

If we denote with D_k^n the number of infected subpopulations with degree k at generation n of the branching process, it is possible to describe the evolution of this quantity in terms of the number of infected subpopulation at the previous generation $n - 1$ and the seeding events [39, 20]:

$$D_k^n = \sum_{k'} D_{k'}^{n-1} (k' - 1) P(k|k') \left(1 - R^{-\varphi_{k'k}}\right) \left(1 - \sum_{m=0}^{n-1} \frac{D_k^m}{V_k}\right). \quad (14)$$

The r.h.s. of the equation describes the contribution of the subpopulations of degree k' at generation $n - 1$ to D_k^n . Each of the D_k^{n-1} has $(k' - 1)$ possible connections along which the infection can proceed (-1 takes into account the link through which each of those subpopulations received the infection). In order to infect a subpopulation of degree k , three conditions need to occur: (i) the connections departing from nodes with degree k' point to subpopulations of degree k , as indicated by the conditional probability; (ii) the reached subpopulations are not yet infected, as indicated by the probability $(1 - \sum_{m=0}^{n-1} \frac{D_k^m}{V_k})$, where V_k is the total number of subpopulations with degree k ; (iii) the outbreak seeded by $\varphi_{k'k}$ infectious individuals traveling from k' to k takes place, and the probability for this event to happen is given by $(1 - R^{-\varphi_{k'k}})$. The latter term is the one that relates the microscopic dynamics of the local infection occurring within a subpopulation to the coarse-grained view that describes the disease invasion at the metapopulation level. It depends on the details of the diffusion process of individuals as well as the individual travel behavior and its interplay with the disease stages. In the case of an SIR model we can write $\varphi_{kk'} = p_{kk'} \alpha N_k / \mu$, where α is the epidemic size in a single subpopulation, and the expression quantifies the fraction of the total number of cases (αN_k) who diffuse ($p_{kk'}$) while infectious (average infectious period being μ^{-1}). This relation, defined at the macroscopic level, contains a quantity that links the two scales of the metapopulation model from the individual to the patch level, and measures the number of infected individuals that move from an infected patch at generation $n - 1$ to a not yet infected patch at the following generation, thus seeding the outbreak with a probability related to the value of R_0 . By considering the same mobility process as before, described by the rate $p_{kk'} = w_{kk'} / N_k = w_0 (kk')^\theta / N_k$, and an uncorrelated network, the previous equation reads as

$$D_k^n = (R_0 - 1) \frac{k^{1+\theta} P(k)}{\langle k \rangle} \frac{w_0 \alpha}{\mu} \sum_{k'} D_{k'}^{n-1} k'^\theta (k' - 1), \quad (15)$$

where we have assumed R_0 close to the epidemic threshold so that we can write $(1 - R_0^{-\varphi_{kk'}}) \simeq \varphi_{kk'} (R_0 - 1)$, and we consider that at the early stage of the epidemic $(1 - \sum_{m=0}^{n-1} \frac{D_k^m}{V_k}) \simeq 1$. By using the auxiliary function $\Theta^n = \sum_{k'} D_{k'}^n k'^\theta (k' - 1)$, we obtain the recursive relation $\Theta^n = (R_0 - 1) \frac{\langle k^{2+2\theta} \rangle - \langle k^{1+2\theta} \rangle}{\langle k \rangle} \frac{w_0 \alpha}{\mu} \Theta^{n-1}$, yielding for the global invasion the condition

$$R_* = \frac{2(R_0 - 1)^2}{R_0^2} \mu^{-1} w_0 \frac{\langle k^{2+2\theta} \rangle - \langle k^{1+2\theta} \rangle}{\langle k \rangle} > 1. \quad (16)$$

We rewrite the expression of R_* in a simplified form:

$$R_* = G(R_0) \cdot H(\mu) \cdot F(w_0; P(k)) > 1, \quad (17)$$

where the functions G , H , and F take the form

$$\begin{aligned} G(R_0) &= \frac{2(R_0 - 1)^2}{R_0^2} \\ H(\mu) &= \mu^{-1} \\ F(w_0; P(k)) &= w_0 \frac{\langle k^{2+2\theta} \rangle - \langle k^{1+2\theta} \rangle}{\langle k \rangle}, \end{aligned}$$

respectively.

R_* is therefore equivalent to a basic reproductive number at the subpopulation level, defining the average number of subpopulations to which each infected subpopulation will spread the contagion process. R_* thus defines the invasion threshold, as any contagion process will spread globally in the network system only if $R_* > 1$. Given the dependence of R_* on the mobility scale w_0 , once the topology of the metapopulation substrate and the epidemic parameters are fixed, Eq. (17) explicitly defines the critical value for the mobility above which an epidemic at the global level takes place, ensuring that on average each subpopulation can seed more than one neighboring subpopulation [39, 20].

The condition for the critical mobility yields [39, 20]:

$$w_0 \geq \frac{\langle k \rangle}{\langle k^{2+2\theta} \rangle - \langle k^{1+2\theta} \rangle} \frac{\mu R_0^2}{2(R_0 - 1)^2}. \quad (18)$$

These results have been generalized to different types of mobility processes and to more complicated compartmental structures to increase the realism of the disease description beyond a simple SIR model, and confirmed by extensive numerical simulations [20]. In all cases, R_* depends on a combination of functions of the reproductive number ($G(R_0)$ in Eq. (17)), of the disease parameter values ($H(\mu)$), and of mobility and topological aspects of the metapopulation network ($F(w_0; P(k))$). Particularly interesting is the explicit effect of the network topology, encoded in the moments of the degree distribution $P(k)$. Indeed, the heterogeneity of the network leads to very large values of the function $F(w_0; P(k))$ (much larger than in a homogeneous network with same size and average degree) that ensure very large

values of R_* , thus above the threshold value 1 (see Figure 5 where the R_* surface is obtained considering an SEIR model with asymptomatic infections for the description of influenza transmission). In the perspective of the threshold condition on mobility, this implies that the heterogeneity of the metapopulation network favors the global spread of epidemics by lowering the critical mobility value. In other words, the topological fluctuations favor the subpopulation invasion [39], similarly to what happens for heterogeneous contact networks at the individual level [42]. These findings have recently attracted attention because, besides enabling the understanding of a qualitative threshold behavior for the global invasion of a pathogen in a spatially structured population, they are able to provide an explicit expression that allows the calculation of the threshold condition on mobility in a realistic situation. In other words, they provide a theoretical framework and rationale for the evidence concerning the inefficacy of travel restrictions in the containment of global epidemics. Consider for example a metapopulation model for influenza spread built on the worldwide air transportation network. Taking into account the heterogeneous topological pattern of the network, the simple plug-in of the actual numbers for the transportation network, the population sizes, and realistic disease parameters in the threshold condition indicates that a reduction of almost 2 orders of magnitude of the mobility would be needed to bring the system below the invasion threshold [39]. This approach shows how simplified modeling frameworks may allow the explanation of effects found in simulation results, and, in the specific case, provide the understanding of mobility effects in the spread and containment of infectious diseases.

5 Going beyond basic assumptions

The study of the global invasion threshold presented in the previous Section is obtained under specific conditions of the mobility process. Various mobility types are considered that affect the functional form of the mobility rate p_{jl} and its dependence on the origin patch, destination patch or both. However the basic assumption underlying those framework is the Markovian nature of the mobility, i.e. the fact that individuals may leave subpopulation j to enter subpopulation l irrespectively of their previous movements. This is typically adopted to describe a range of random diffusive phenomena where no memory comes into play.

At the beginning of these notes, however, we have discussed the presence of different types of mobility modes in which memory is a crucial aspect of the dynamics that cannot be disregarded. This is especially true for the case of human hosts whose movements are typically characterized by recurrent patterns, from specific home locations to given destinations. In the specific case of commuting, moreover, the recurrent nature is also associated to a specific destination, i.e. the workplace location, and to a specific visiting time approximately equal to the amount of hours corresponding to an average working day. Recurrent mobility patterns such as commuting are clearly poorly modeled by the random diffusive dynamics presented so far, and specific methodologies and approximations capable of coping with non-Markovian diffusive processes in complex systems need to be introduced that go beyond the simple adaptation of previous theoretical frameworks.

The first framework to be introduced is based on the inclusion of memory effects through the description of the population with two indices – residence patch and patch of current location at time t [37, 38]. The individuals original from the subpopulation j , N_j , are divided in the contribution $N_{jj}(t)$, who are from j and are located in j at time t , and $N_{jl}(t)$, who are from j and are located in a neighboring subpopulation l at time t (see the schematic diagram of Figure 6). In addition to the leaving rate $p_{kk'}$ (that one can assume to be dependent from the degrees k and k' of the origin and destination patches, respectively, as obtained from empirical evidence), the commuting mobility needs the definition of an additional parameter, the return rate τ^{-1} with τ being the timescale of the commuting. The interplay between the different timescales of the processes – the timescale of the epidemic, the one of the commuting and the inverse of the leaving rate – opens the possibility to a simplification of the modeling description, under certain conditions. In particular, if $\sigma_k \ll \tau^{-1}$ and the epidemic characteristic timescale is much larger than the one of the commuting process, then it is possible to consider the system in a quasi-stationary state and solve the rate equations obtaining the stationary values of the compartment populations [22, 37, 38]. These conditions are met in the case of human commuting, as the leaving rate is typically very small and smaller compared to the inverse of the commuting time (generally few hours, representing a working day), whereas an outbreak caused by influenza epidemic, e.g., has a typical infectious period of about 3 days. Rate equations characterizing the commuting dynamics among subpopulations can then be defined in terms of the variables $N_{jj}(t)$ and $N_{jl}(t)$

as

$$\begin{aligned}\partial_t N_{jj} &= -\sum_l \sigma_{jl} N_{jj}(t) + \tau^{-1} \sum_l N_{jl}(t), \\ \partial_t N_{jl} &= \sigma_{jl} N_{jj}(t) - \tau^{-1} N_{jl}(t).\end{aligned}\tag{19}$$

If the conditions for the quasi-stationary approximation are satisfied, stationary solutions are found by imposing $\partial N_{jj}(t) = \partial N_{jl}(t) = 0$. By using the consistency relation $N_j = N_{jj}(t) + \sum_l N_{jl}(t)$, the stationary conditions yields the equilibrium values

$$N_{jj} = \frac{N_j}{1 + \sigma_j/\tau} \quad \text{and} \quad N_{jl} = \frac{N_j \sigma_{jl}/\tau}{1 + \sigma_j/\tau}.\tag{20}$$

Under this approximation, it is possible to write Levins-type equations as in the Markovian case, where the seeding events $\varphi_{kk'}$ are calculated on the obtained stationary values. Results provide a richer phase diagram than in the previous case, as now the global invasion threshold depends on both the mobility rates, i.e. the leaving rate and the return rate (Figure 6). The expression for the global invasion threshold provide the critical values of the mobility rates that distinguish between a regime in which the pathogen may invade a macroscopic fraction of the metapopulation system and a regime in which it is limited to few subpopulations (see Figure 6). Intuitively, this is explained by considering that if the diffusion rate approaches zero, the probability of the contagion entering neighboring subpopulations goes to zero, as there are no occasions for the carriers of the process to visit them. On the other hand if the return rate is very high, then the visit time of individuals in neighboring populations is so short that they do not have time to spread the infection in the visited subpopulations.

Between permanent movements to daily commuting, human mobility is also characterized by other types of non-Markovian movements where return rates are not constant. Available statistics indeed show that the time spent by travelers at destination is characterized by wide fluctuations, ranging from a single day up to several months (Figure 7). Such varying length of stay crucially affects the chance and duration of mixing events among hosts and may therefore have a strong impact on the spread of an emerging disease. To address this aspect, a novel analytical framework for a metapopulation epidemic model with heterogeneous lengths of stay at destination was introduced [49, 50]. While the economic literature considers the length of stay as one of the key elements in the visitor's decision-making process, its determinants remain however largely undefined. A possible assumption is to consider

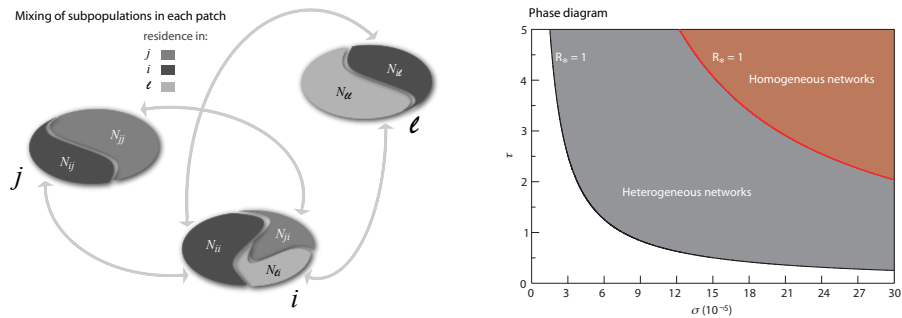


Figure 6: Schematic representation of the subdivision of the population in each patch into subpopulations N_{xy} , where x represents the subpopulation of residence and y represents the subpopulation of the actual location at time t (left). Phase diagram in terms of mobility rates, with continuous lines identifying the $R^* = 1$ relation and shaded areas corresponding to the invasion regime (from Ref. [37]).

a scaling function for the length of stay at destination that is proportional to the destination patch degree [49, 50], i.e.

$$\tau_k \sim k^\chi, \quad (21)$$

where the parameter χ defines different regimes of the mobility dynamics. For $\chi > 0$ the length of stay is positively correlated with the degree of the patch of destination, meaning that individuals traveling to a well-connected location will spend a longer time at destination, thus being longer exposed to the local population, with respect to individuals traveling to peripheral locations. This can be motivated by the attractiveness of popular locations, for which the pattern of connection is optimized through large degrees to manage large fluxes volumes of individuals, both at the within-city scale and at the larger geographical scale where airport hubs handle large traffic due to tourism or seasonal/temporary job opportunities. The opposite regime, obtained for $\chi < 0$, implies that the time spent at a location is larger for decreasing degree of the subpopulation of destination, and may correspond to an individual choice of optimization between the time spent to reach the destination and the time spent at destination. Low degree locations are indeed generally peripheral in the system, so that a trip from a given origin may take multiple steps to reach the final destination, and thus a longer length of stay at the hard-to-reach destination may then adequately balance

and justify the travel time.

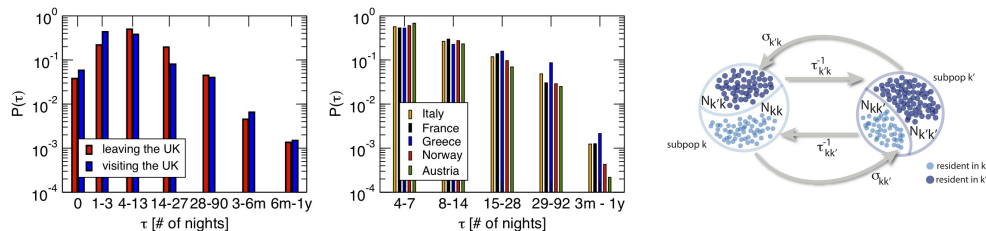


Figure 7: Heterogeneous length of stay at destination, for the UK (left) and for 5 European countries (center). Schematic representation of the subdivision of the population in each patch into subpopulations, considering return rates dependent on degree k . From Refs. [49, 50].

Under these conditions and the time-scale separation approximation, it is possible to solve analytically the dynamics of epidemic invasion in the metapopulation network and write an expression for the global invasion threshold, depending on epidemic and mobility parameters [49, 50]. The solution shows that large fluctuations of the length of stay, as observed in reality, can have a significant impact on the threshold conditions for the global epidemic invasion (see Figure 8), thus altering model predictions based on simpler assumptions, and displaying important public health implications. Two regimes are indeed found that may dramatically favor or hinder the invasion, induced by the positive or negative degree-correlation, respectively. The regimes alter not only the conditions for invasion, but also the epidemic invasion trees once the disease spreads in the metapopulation network (Figure 8). Despite its simplicity, the framework uncovered an important aspect that must be considered in interpreting and simulating epidemic spreading patterns, and in providing detailed model predictions.

6 Conclusions

Our understanding of communicable disease prevention and control is rooted in the theory of host population transmission dynamics, where contacts between hosts and contacts between populations of hosts drive the epidemiology of infectious diseases, determining if and how quickly they spread, and who gets infected. As the spatial spread plays a crucial role in the management

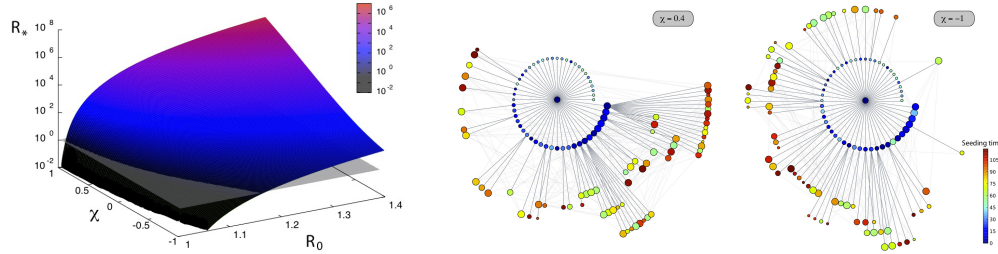


Figure 8: Analytical surface of the global invasion threshold R^* as a function of R_0 and χ (left, from Refs. [49, 50]). Epidemic invasion trees (see [49] for more details) when the length of stay is positively correlated (center) or negatively correlated (right) with the destination patch degree k . Nodes are colour coded according to the time of their seeding, and their size scales with their degree; nodes in the first layer are ordered according to their degree to highlight the role of different degree nodes in the hierarchical invasion pattern in the two cases. From Refs. [49, 50].

and control of a disease, the theory and results presented in these notes form the basis to build increasingly realistic metapopulation epidemic models for public health applications. Additional factors such as, e.g., the heterogeneity of the travel behavior of the population, the dependence of mobility rates on socio-economic or geographical aspects, non-homogeneous mixing, adaptive behavior, integration of high-resolution contact and mobility data, and others enable an increasingly accurate description of disease propagation.

References

- [1] Lester K Little. *Plague and the end of Antiquity: the pandemic of 541–750*. Cambridge University Press, 2007.
- [2] JD Murray. *Mathematical Biology II. Spatial models and biomedical applications*. Springer, 2003.
- [3] *SARS: A case study in emerging infections*. Oxford University Press, March 2005.
- [4] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of

- infectious diseases: the SARS case study. *BMC Medicine*, 5:34, November 2007.
- [5] Alain-Jacques Valleron, Anne Cori, Sophie Valtat, Sofia Meurisse, Fabrice Carrat, and Pierre-Yves Boille. Transmissibility and geographic spread of the 1889 influenza pandemic. *PNAS*, 107(19):8778–8781, May 2010.
- [6] Peter Gould, Joseph Kabel, Wilpen Gorr, and Andrew Golub. AIDS: Predicting the Next Map. *Interfaces*, 21(3):80–92, 1991.
- [7] Ilkka A. Hanski and Michael E. Gilpin, editors. *Metapopulation Biology: Ecology, Genetics, and Evolution*. Academic Press, San Diego, Calif., 1 edition edition, March 1997.
- [8] Richard Levins. Some Demographic and Genetic Consequences of Environmental Heterogeneity for Biological Control. *Bull Entomol Soc Am*, 15(3):237–240, September 1969.
- [9] Levins, Richard. Extinction. *Lecture Notes in Mathematics*, 2:75–107, 1970.
- [10] Herbert W. Hethcote. An immunization model for a heterogeneous population. *Theoretical Population Biology*, 14(3):338–349, December 1978.
- [11] Robert M. May and Roy M. Anderson. Population biology of infectious diseases: Part II. *Nature*, 280(5722):455–461, August 1979.
- [12] Anderson, RM and May, RM. Spatial, temporal and genetic heterogeneity in host populations and the design of immunization programs. *IMA J. Math. Appl. Med. Biol.*, 1:233–266, 1984.
- [13] Robert M. May and Roy M. Anderson. Spatial heterogeneity and the design of immunization programs. *Mathematical Biosciences*, 72(1):83–111, November 1984.
- [14] B. M. Bolker and Bryan Thomas Grenfell. Chaos and biological complexity in measles dynamics. *Proc. R. Soc. Lond. B*, 251(1330):75–81, January 1993.

- [15] B. Bolker and B. Grenfell. Space, persistence and dynamics of measles epidemics. *Philos. Trans. R. Soc. Lond., B, Biol. Sci.*, 348(1325):309–320, May 1995.
- [16] Matt J. Keeling and Pejman Rohani. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology Letters*, 5(1):20–29, 2002.
- [17] Bryan Grenfell and John Harwood. (Meta)population dynamics of infectious diseases. *Trends in Ecology & Evolution*, 12(10):395–399, October 1997.
- [18] Neil M. Ferguson, Matt J. Keeling, W. John Edmunds, Raymond Gani, Bryan T. Grenfell, Roy M. Anderson, and Steve Leach. Planning for smallpox outbreaks. *Nature*, 425(6959):681–685, October 2003.
- [19] Steven Riley. Large-scale spatial-transmission models of infectious disease. *Science*, 316(5829):1298–1301, June 2007.
- [20] Vittoria Colizza and Alessandro Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of Theoretical Biology*, 251(3):450–467, April 2008.
- [21] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *PNAS*, 103(7):2015–2020, February 2006.
- [22] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, Josée J. Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *PNAS*, 106(51):21484–21489, December 2009.
- [23] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, Madhav V. Marathe, Aravind Srinivasan, Zoltán Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, May 2004.
- [24] Matt J Keeling. Models of foot-and-mouth disease. *Proc Biol Sci*, 272(1569):1195–1202, June 2005.

- [25] Paolo Bajardi, Alain Barrat, Fabrizio Natale, Lara Savini, and Vittoria Colizza. Dynamical Patterns of Cattle Trade Movements. *PLOS ONE*, 6(5):e19869, May 2011.
- [26] Rowland R. Kao, Darren M. Green, Jethro Johnson, and Istvan Z. Kiss. Disease dynamics over very different time-scales: foot-and-mouth disease and scrapie on the network of livestock movements in the UK. *Journal of The Royal Society Interface*, 4(16):907–916, October 2007.
- [27] Paolo Bajardi, Alain Barrat, Lara Savini, and Vittoria Colizza. Optimizing surveillance for livestock disease spreading through animal movements. *Journal of The Royal Society Interface*, page rsif20120289, June 2012.
- [28] Eugenio Valdano, Chiara Poletto, Armando Giovannini, Diana Palma, Lara Savini, and Vittoria Colizza. Predicting Epidemic Risk from Past Temporal Contact Data. *PLOS Computational Biology*, 11(3):e1004152, March 2015.
- [29] Michael M. Thomson and Rafael N°jera. Travel and the Introduction of Human Immunodeficiency Virus Type 1 Non-B Subtype Genetic Forms into Western Countries. *Clinical Infectious Diseases*, 32(12):1732–1737, 2001.
- [30] Alberto Aleta, Andreia N. S. Hisi, Sandro Meloni, Chiara Poletto, Vittoria Colizza, and Yamir Moreno. Human mobility networks and persistence of rapidly mutating pathogens. *Royal Society Open Science*, 4(3):160914, March 2017.
- [31] B. T. Grenfell, O. N. Bjørnstad, and J. Kappey. Travelling waves and spatial hierarchies in measles epidemics. *Nature*, 414(6865):716–723, December 2001.
- [32] RM Anderson, RM May. *Infectious Diseases of Humans: Dynamics and Control*. Oxford University Press, Oxford; New York, reprint edition edition, September 1992.
- [33] Matt J Keeling, Pejman Rohani. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press, Princeton, 1 edition edition, October 2007.

- [34] I. M. Hall, J. R. Egan, I. Barrass, R. Gani, and S. Leach. Comparison of smallpox outbreak control strategies using a spatial metapopulation model. *Epidemiol Infect*, 135(7):1133–1144, October 2007.
- [35] Cécile Viboud, Ottar N. Bjørnstad, David L. Smith, Lone Simonsen, Mark A. Miller, and Bryan T. Grenfell. Synchrony, Waves, and Spatial Hierarchies in the Spread of Influenza. *Science*, 312(5772):447–451, April 2006.
- [36] Matt J. Keeling, Leon Danon, Matthew C. Vernon, and Thomas A. House. Individual identity and movement networks for disease metapopulations. *PNAS*, 107(19):8866–8870, May 2010.
- [37] Duygu Balcan and Alessandro Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature Physics*, 7(7):581–586, July 2011.
- [38] Duygu Balcan and Alessandro Vespignani. Invasion threshold in structured populations with recurrent mobility patterns. *J Theor Biol*, 293:87–100, January 2012.
- [39] Vittoria Colizza and Alessandro Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Phys. Rev. Lett.*, 99(14):148701, October 2007.
- [40] Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, April 2007.
- [41] A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani. The architecture of complex weighted networks. *PNAS*, 101(11):3747–3752, March 2004.
- [42] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Rev. Mod. Phys.*, 87(3):925–979, August 2015.
- [43] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with Two Levels of Mixing. *The Annals of Applied Probability*, 7(1):46–89, 1997.

- [44] Paul C. Cross, James O. Lloyd-Smith, Philip L. F. Johnson, and Wayne M. Getz. Duelling timescales of host movement and disease recovery determine invasion of disease in structured populations. *Ecology Letters*, 8(6):587–595, 2005.
- [45] Paul C Cross, Philip L.F Johnson, James O Lloyd-Smith, and Wayne M Getz. Utility of R_0 as a predictor of disease invasion in structured populations. *J R Soc Interface*, 4(13):315–324, April 2007.
- [46] Duncan J. Watts, Roby Muhamad, Daniel C. Medina, and Peter S. Dodds. Multiscale, resurgent epidemics in a hierarchical metapopulation model. *PNAS*, 102(32):11157–11162, August 2005.
- [47] Alexei Vazquez. Epidemic outbreaks on structured populations. *Journal of Theoretical Biology*, 245(1):125–129, March 2007.
- [48] Paolo Bajardi, Chiara Poletto, Jose J. Ramasco, Michele Tizzoni, Vittoria Colizza, and Alessandro Vespignani. Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1n1 Pandemic. *PLOS ONE*, 6(1):e16591, January 2011.
- [49] Chiara Poletto, Michele Tizzoni, and Vittoria Colizza. Heterogeneous length of stay of hosts’ movements and spatial epidemic spread. *Scientific Reports*, 2:476, June 2012.
- [50] Chiara Poletto, Michele Tizzoni, and Vittoria Colizza. Human mobility and time spent at destination: Impact on spatial epidemic spreading. *Journal of Theoretical Biology*, 338:41–58, December 2013.