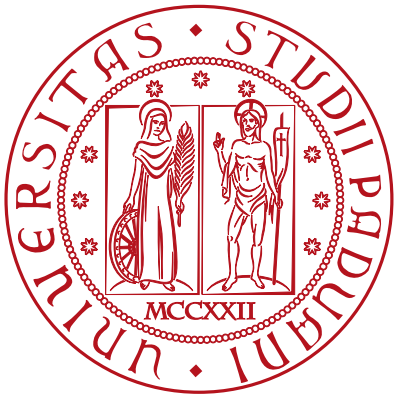


UNIVERSITÀ DEGLI STUDI DI PADOVA

Master Degree in Physics of Data

Final Dissertation

Physics of epidemics on contact networks with spatial and temporal features

---

Internal supervisor:
Prof. **Chiara Poletto**
External supervisor:
Dr. **Johannes Zierenberg**

Candidate:
Dario Barone

Academic Year 2022/2023

Contents

List of Figures

Chapter 1

Introduction

The 2020s decade opened with an invasive epidemic which quickly escalated in a global pandemic. The pandemic led to a perturbation that impacted on all aspects of human life, from economics and technology to society, culture and public health. Despite the technological level reached by humans, pathogens and diseases are a major threat to prosperity and stability. A deep understanding of causes and effects of epidemics is extremely important to mitigate tragic casualties, and it involves a broad set of disciplines, from social sciences to biochemistry, from medicine to physics.

Physics provides theoretical frameworks to tackle some key aspects of epidemics, at different scales [1]. At a small scale, it could provide models for the diffusion of pathogen molecules in fluids, useful to understand the spreading mechanics and the travelling capability of the pathogen [2]. These aspects are usually very dependent on the local properties of the environment, so that it becomes convenient to coarse-grain the results and obtain *effective* parameters that represent local details. More general models arise with the aim to measure the number of individuals that are infected or at risk in a community [3], and the propagation of the epidemic among societies [4]. The complexity rises as more and more characteristics of the reality are taken into account.

One of the most affirmed and solid landmarks in the development of more complex models is the introduction of networks as underlying structures on which the epidemics diffuse [5, 6]. Networks represent the ensemble of *contacts* that drive the infection, they play the role of the topological space of the diffusive process, so their properties can be very impactful on the epidemic dynamics. The analysis of the network properties falls into the domain of *network science*, a discipline that flourished in the last four decades, for purposes that extend well beyond the case of theoretical epidemiology.

The deep understanding of the characteristics of the real contact networks relies on data collection experiments, which can be challenging to perform for several

reasons. The advancement in technology has given more powerful tools to address the gathering process, like the recording of GPS positions and the bluetooth pairings [7]. Though, it is still debated how accurate it is to use these metrics as proxies for the epidemiological contact networks, often the data are limited in size and refer to particular settings [8, 9], and experiments must undergo restrictions due to privacy protection [10, 11].

This difficulty is a challenge for the research in epidemiology, so alternative routes are also being explored to understand real contact networks, with the aim to anticipate the epidemic spreading and adopt useful countermeasures. Generative models are a tool to create synthetic data that should replicate some characteristics of reality, so that analysis and simulations can provide information that can be generalized to real scenarios. Most of the refined generative models for networks used in theoretical epidemiology aim to recreate several statistical properties of real data [12–14]. However real world is far too complex to be entirely reproduced, especially when it comes to the mechanistic explanation of the properties observed. In this respect, starting from simple and solid models, small elements have to be introduced to incorporate more and more effects while keeping the causality under control.

One important challenge to tackle comes from the temporal dimension, which happens to be one main source of causal relations between events. Real human behaviour is prominently dependent on time, as everyone's life runs through an overlap of several cyclic routines on different timescales. The alternation between day and night is a daily cycle that affects the level of activity of an individual, the summer holiday is a yearly cycle that affects the type of contacts of the individuals, the attendance to big concerts increases the number of contacts of the individuals. These are just few examples in a huge variety that show the complexity of the human behaviour and its strict dependence on time [15]. It has been shown that temporal inhomogeneities have important consequences on epidemic dynamics [16–18], highlighting the importance to take them into consideration.

In this work, I try to address some of the challenges mentioned above, in particular with a reference to the mechanistic interplay between space and time that produces the contact networks, and its consequences on the physics of epidemics. I will propose a generative model to construct the network of contacts, not based on data driven considerations, but that relies on few first principles: an individual mobility model that follows a circadian pattern to establish contacts between individuals in an Euclidean spatial setting. The alternation between day and night is embodied in a change in the individuals' behaviour, that are free to explore their neighbourhood during the day while remaining confined at home during the night. I will then observe the consequences of such behaviour on the early epidemic growth of a spreading disease and the universality class of the system close to its dynamical phase transition. To contain the complexity of the model, I will consider a very simplistic model for the epidemic, both concerning the transmission and the

evolution of the disease, and I will neglect most of the social behavioural features that might be extremely variable and hard to model properly.

To comprehensibly explain the results of this work, in chapter 2 I will give an overview of the theoretical setting, starting with concepts of network science and ending up with epidemiological models, pointing out all the milestones that are a necessary conceptual prelude to this work. Equipped with the theoretical foundations, in chapter 3 I will describe the conceptual details of the model adopted, and discuss the motivations underlying the technical solutions implemented in the numerical simulations. Chapter 4 will show all the analytical calculations used to set meaningful values to the free parameters, and the effects that the parameters project onto the system's behaviour. This will drive us to chapter 5, where I will discuss the importance of the relation between the spatio-temporal scales of mobility and epidemic. The interplay between these scales will be shown to lead to a super-quadratic epidemic growth and to a change of the universality class of the absorbing phase transition. Finally, in chapter 6 I will summarize the findings achieved and the possible improvements to increase the complexity of the model.

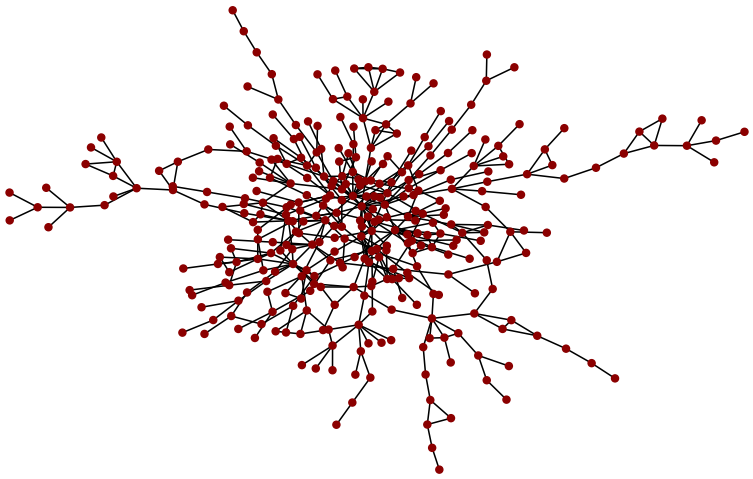
Chapter 2

Theoretical Background

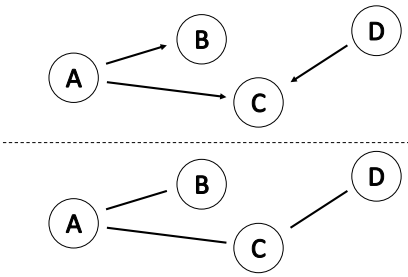
The understanding of the driving motivation and key results of this work require an introductory chapter about the extensive surrounding scientific literature, with a focus on the elements that will turn particularly useful for the following discussion. The chapter is divided into three main sections. The first section will be devoted to present some network models and their main properties. The fundamental role of contact heterogeneity in the spreading of a disease is the reason why temporal networks are a relevant framework to describe contacts between individuals. In the second section I will discuss the causal relation between the mobility and the contact network, as face-to-face contacts are inherently defined in a spatial setting, in which individuals move and meet. In the third and last section, I will present some phenomena that are observed when a disease spreads onto a network, with a reference to the properties of the network models that have causally affected the epidemic.

2.1 Network science

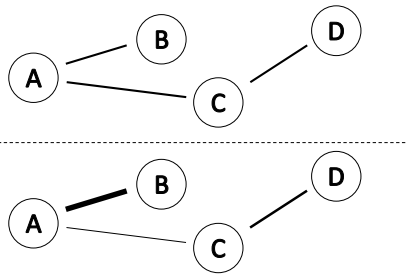
Every system composed by individual elements that have pairwise interactions can be framed as a network. Network science is a discipline that was born at the beginning of this century after some seminal works published by Albert-László Barabási. As interestingly narrated in the author's introduction of his book [19], network science developed after realizing that real networks, independently from the context, present a similar structure and show a common emergent behaviour [20]. From there on, many scientists contributed to the topic, as the applications are extremely versatile, covering a variety of very different subjects. Infrastructures, intended as rail/road/naval or power grids, are naturally modelled as networks. Also the financial structure of agencies providing loans and insurances can be modelled as a network. So can be the connection structure of telecommunications, or the machines in an IT infrastructure, or websites on the internet. Biologically speaking, it is straightforward to mention the neural network, vital for animals;



(a) *Undirected network of phone calls among the subjects of the Copenhagen Network Study [8].*



(b) *A comparison between a directed (above) and undirected (below) network edges.*



(c) *A comparison between a unweighted (above) and weighted (below) network edges.*

Figure 2.1: **Introductory examples of networks.**

also in society, the network of social contacts is what defines the structure of the society and of everyone's life. All these different fields of application contribute to make network science a blooming topic of research, and in the following sections I will present some milestone concepts that will be useful for the thesis.

2.1.1 The general framework

Networks are composed of nodes, the individual elements of the system, and links, the connections between different elements (Figure 2.1a). Depending on the application context, nodes and links can embody different objects. The interest for this kind of structure in mathematics was born long before the emergence of network science, and is realized in a discipline called graph theory. *Graph* is the mathematical term that describes the network, and nodes and links are called *vertices* and *edges*.

The first taxonomic levels to classify networks are related to the properties of them being *directed/undirected* and *weighted/unweighted*. Referring to the example in Figure 2.1b, there is a link connecting two nodes A and B; this link can be undirected (or bi-directed) if it only matters whether A and B are in contact or not; or it can be directed if it matters which node affects the other $A \rightarrow B$ or $A \leftarrow B$. The whole network is defined directed if it has *at least one* directed link and undirected if *all* links are undirected. An example of a directed network is the network of phone calls, where there is a caller and a callee; an undirected network can be an electrical scheme, where each device is connected to others without any directionality. In principle, different links may represent bonds with different intensities, as in the already-mentioned case of the phone calls. If node A spends long time chatting with B and much less with C, this information might be worth to keep in the network representation. So, it is convenient to associate a *weight* to each link, proportional to the time spent calling, as exemplified in Figure 2.1c. Thus, a network is defined weighted if all its links have a weight, while it is unweighted if they all have the same weight (which gets usually normalized to 1). An example of unweighted network might be the network of scientific citations, as there is no citation stronger or weaker than others.

Networks can be represented graphically as in Figure 2.1a, but computations make use of two other common types of representations: edge lists and adjacency matrices. An edge list is, as the name suggests, a list of edges, which are usually defined as the structure (a, b, w), where a represents the identifier for the output node, b the identifier for the input node and w the weight of the link connecting the two. In case of an undirected unweighted network, the order of the identifiers does not matter and the weight is usually omitted. An adjacency matrix can be seen as the pivoted version of an edge list, where the pivoting occurs on the input node. Thus, it is a $N \times N$ matrix, where each entry w_{ij} is the weight of the link coming out from node i and entering in node j. In the case of an undirected network, the adjacency matrix is symmetric, and in case of an unweighted network, the matrix is boolean. For sparse networks, an edge list representation is more efficient than the adjacency matrix (that would be filled with 0s), while the adjacency matrix is more intuitive and efficient when dealing with linear algebra operations.

2.1.2 Network metrics

As network science grew in late 20th century, much effort has been put in the invention of metrics that could characterize networks [21]. In this subsection, I will present a selection of metrics that will be useful to follow the later chapters. The definitions could be extended for directed or weighted graphs, but for the sake of simplicity I will stick to undirected unweighted networks relevant to this work.

Degree and degree distribution A first characterizing property of a network is the collection of node degrees. The degree of a node i is the number of links

that have i as partner $k_i = \sum_{j \in N_i} 1$ where N_i is the set of neighbours of node i. The higher the degree of a node, the easier its communication with the rest of the network. The structure of the entire network is strongly influenced by the degree distribution $P(k)$, so are the processes occurring onto the network.

Distance and diameter Consider two nodes A and B in a network. A *path* from A to B is a collection of contiguous edges that starts in A and finishes in B without any discontinuity. The *path length* is the number of edges that compose the path. Usually there are several ways to reach B from A, but the most interesting metric is the *shortest* path from A to B. This gives an information about the *distance* between these two nodes. In case there is no path connecting the two nodes, the distance is considered infinite. It is possible to compute the distance between every pair of nodes (it is usually an expensive computation as there are $N(N-1)$ pairs in the network). So, the *diameter* of the network is defined to be the largest distance between any two nodes in the network. It provides information about the number of steps that some process have to cross before possibly reaching all the nodes.

Connected components A network can be made of one or several connected components. A connected component is a sub-graph (a portion of the entire graph) containing nodes for which a path always exists. Using the notion of diameter, a connected component is the sub-graph that has a finite diameter. The *connected components* of a network is the collection of sub-graphs that have finite diameter, have null intersection and which union is the full network. Usually it happens that connected components are very unbalanced in their size, as the bigger one is, the more chance it has to connect to another connected component. The largest connected component is usually called *giant component*.

Connectivity and percolation threshold Consider the connectivity as the coefficient $0 \leq c \leq 1$ that represents the fraction of actual link in the network over the possible ones $c \coloneqq \frac{\sum_i k_i}{2N(N-1)}$. For very low values of the connectivity, the network is very likely to be divided in many small connected components; as the connectivity increases, the connected components join into fewer and bigger ones. The connected and disconnected phases are usually separated by a sharp transition, which occurs at a connectivity value defined as the *percolation threshold*. More specifically, the exact definition of the percolation threshold may rely on different metrics: the transition can be referred to the number of connected components, to their average diameter, to the size of the giant component. In the specific case of this work, the definition of percolation threshold is the connectivity over which only one giant component includes more than 50% of the nodes.

Clustering coefficient In real networks, especially in case of social networks, a tendency to create closed groups is often observed. If nodes i and j are both

linked to the same node m, it is more likely that i and j are also linked than if they did not have any common neighbour. This property is called *clustering*. The *local clustering coefficient* is a node-specific metric that represents amount of linkage among its neighbours. If the heterogeneity of local clustering is not dominant, it is possible to define a *global* clustering coefficient as the arithmetic average of the local ones, definition which is also quite common in the literature

$$\langle C \rangle = \frac{1}{N} \sum_i \frac{L_i}{k_i(k_i - 1)} \tag{2.1}$$

where k_i is the degree of the i-th node, and L_i is the number of links between its neighbours. More general definitions of global clustering coefficient account for the fraction of triadic closures, but these technicalities go beyond the scope of this work.

2.1.3 Lattices: a regular network from solid state physics

Lattices are widely studied in the field of solid-state physics as models to represent the crystalline structures of matter. They can be seen as the most simple network structure, where atoms embody the nodes of the network and the chemical interactions among them embody the links. The defining properties of a lattice are the size, the structure layout and the dimensionality, but in some cases also the order of nearest neighbours that are interacting (usually only first nearest neighbours) and the lattice spacing (usually normalized to a unit). Network science refers to lattices as *regular networks*, as they present a translationally invariant pattern for which all nodes share the same individual metrics, e.g. the degree $\langle k \rangle = k_i \ \forall i$. The degree distribution of a lattice is simply $P(k) = \delta(k - \langle k \rangle)$, where for a first nearest neighbours hypercubic layout in d-dimension $\langle k \rangle = 2d$.

A lattice model has the advantage of naturally including the dimensionality of the space in its structure and naturally imposing a discretized representation of the physical world. On the contrary, most of the metrics that characterize other types of networks are less meaningful on lattices. Considering a first nearest neighbours cubic lattice, the connectivity is extremely low $c = \frac{2dN}{2N(N-1)} \approx 0$ for large networks, suggesting for a disconnected network, instead the connected component is only one and includes the totality of the nodes. Also the clustering coefficient becomes difficult to interpret when referred to a lattice: for a first nearest neighbours cubic lattice in 2 dimensions, the local clustering coefficient $C_i = 0 \ \forall i$, but considering second nearest neighbours it rises to $C_i \approx 0.45 \ \forall i$ and it quickly converges to 1 as more neighbours are linked to each node [22, 23]. The definition of clustering coefficient on a lattice is seldom used, as the spatial locality given by the structure layout favourites connections between neighbours of neighbours.

Another interesting metric to inspect is the diameter of the lattice, and in general all the distances between nodes. These scale with the system size, specifically as

$N^{1/d}$, where d is the dimensionality of the lattice. Intuitively, this is due to the number of steps that have to be taken when travelling from one point to another in a spatial setting. This aspect, together with the extreme rigidity of the network structure, are the major reasons of criticism to the description of social networks through lattice models.

2.1.4 Other networks: some real aspects captured

Independently from the studies on solid-state physics, mathematicians developed interest for graphs and, consequently, for generative models to build them. Erdős and Rényi pioneered this field by proposing the first generative model and opening the path to *random graphs* [24]. The generation relies on a Bernoulli process that builds a network of N nodes and connects each possible pair with probability p. The network so produced is an undirected unweighted graph, that has a low variability between nodes, as it is only due to the stochastic fluctuation in the Bernoulli process. Indeed, the degree distribution is the binomial distribution $P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$, so the average degree is $\langle k \rangle = p(N-1)$ and its variance is $\text{var}(k) = (N-1)p(1-p)$, which indeed suggests low variability in the individual connectivity between nodes. Erdős-Rényi networks show a small diameter compared to lattices, since random connections very likely shorten the paths between nodes creating "shortcuts". The characteristic of having small distances compared to the system size is known as *small-world*, and it is a salient characteristic of real networks [25]. Both the heterogeneity of node degrees and the under-scaling of the distances may suggest Erdős-Rényi networks to be more realistic models, compared to lattices, to describe real scenarios. Though, when looking at the clustering coefficient, Erdős-Rényi networks perform low scores, as the randomness in the link creation does not help in creating closures between neighbours of neighbours. Such feature does not resemble real networks, where clustering happens indeed more often than randomly.

The last aspect was the main driver for the development of the Watts-Strogatz generative model [26]. This model is a form of "interpolation" between the Erdős-Rényi and the lattice models, as the starting point is a one-dimensional k-th nearest neighbours lattice with periodic boundary conditions and it proceeds rewiring at random only a fraction of the existing links. In this way, the high clustering of the k-th nearest neighbours lattice model is preserved, but the rewiring creates shortcuts in the structure that contain the increase of the distances between nodes. Given the idea behind this model, the degree distribution of the generated network is an interpolation between a delta distribution (given by the lattice) and a binomial distribution (given by the Erdős-Rényi). The result is a distribution which has a less-than-binomial variability on the node degrees, so a small heterogeneity.

Observing real networks in very different settings, Barabási and Albert found a completely different scenario [20]: the node degrees follow a power-law tail distribution. A comparison between binomial and power-law tail distributions is

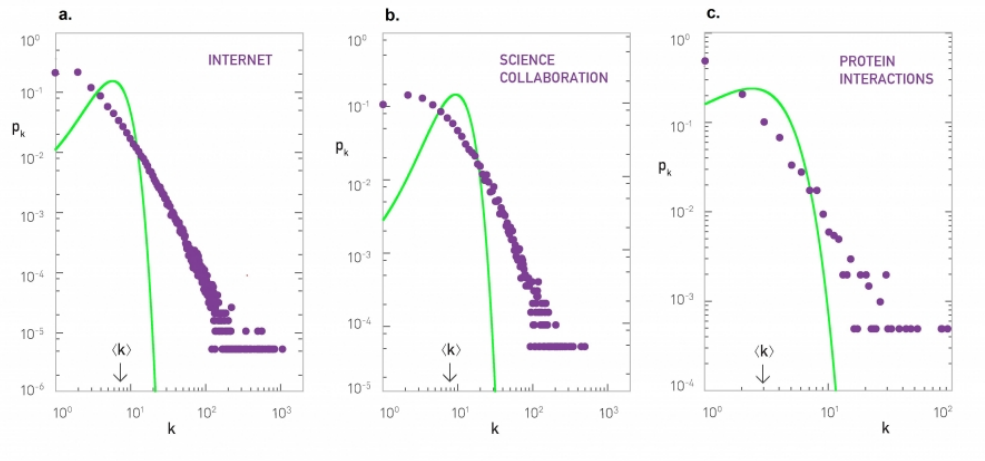


Figure 2.2: **The degree distributions of real networks have a different shape compared to Erdős-Rényi networks.** *The plots show the degree distributions of Erdős-Rényi networks (in green) compared to real networks (in purple). The real networks are observed in different contexts, but all present the same power-law tail: straight lines in the log-log plot. Figure from [19].*

shown in Figure 2.2. The authors proposed a new generative model, based on the principle of *preferential attachment*, to build a network showing high variability of node degrees [27]. The generative algorithm starts from a fully connected core and adds nodes one by one, connecting them to the already existing nodes with probability proportional to the degree; in this fashion, high degree nodes tend to grow more and end up being *hubs* of the network. The discovery of structures with many small nodes and few hubs in real networks from very different contexts boosted the belief in an emergent behaviour. The networks with long-tailed degree distributions are addressed as *scale-free*, meaning that the variance of the degree is constrained only by the size of the network and it scales up as the network grows: it is not an intrinsic property of the construction logic of the network. These type of network has also small diameter, high clustering and high variability, so it has been thought for long that they would well reproduce the statistics of the majority of real cases, and several research works have used this model for the network generation.

2.1.5 Temporal networks

Time is a fundamental dimension in all physical processes, especially because it constrains *causal* relations. Causality can only apply forward in time, so a temporal dimension in a model increases the information about the system. This is especially true for system under continuous evolution, such as human activities, for which some interesting phenomenon appear to be cyclic and bursty in time [28–30]. The networks introduced in subsection 2.1.1 and the following ones are static and completely lack the temporal dimension. So, temporal networks have

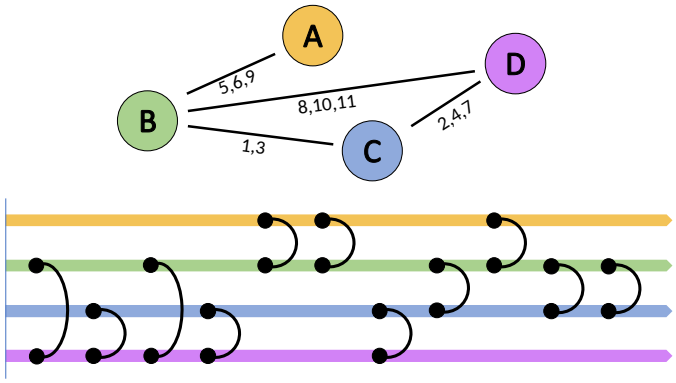


Figure 2.3: **Graphical representation of a temporal network.** *Two graphical representations of temporal networks. The lower representation shows how information can travel from node D to node A through B or C, but it is impossible vice-versa from A to D, at least within the shown time-window. On a static network, both directions would be allowed.*

been introduced to properly describe dynamical systems [31, 32]. The presence of time adds one dimension both in the graphical representation, as shown in Figure 2.3, and in the analytical representation. Indeed, the adjacency matrix becomes a $N \times N \times T$ matrix, with T the number of time-steps at which any link realizes; the edge lists adds the time component (usually in the first place) to each entry, assuming the structure (t, a, b, w), where t represents the time of occurrence of the link.

Temporal networks can be represented as collections of static networks, named *graphlets* or *snapshots*, each one that collects all the edges that share the same time coordinate t. Usually the sampling timescale of raw data is much lower than the timescale at which the phenomena of interest evolve, so it can be convenient to coarse grain the temporal resolution and *aggregate* all the graphlets that lie in the same time-window of length Δt. The choice of Δt is a delicate trade-off between the loss of information given by the aggregation and the analysis difficulty of high complexity datasets.

The introduction of the temporal dimension forces to rethink the metrics discussed in subsection 2.1.2 valid for static networks [33]. Some of them, e.g. the degree distribution, the clustering coefficient and the connectivity, may lose completely their meaning on highly sparse graphlets, so aggregation is the way to proceed if these metric are of some interest. Other metrics, e.g. the distance and the size of the connected components, acquire second meaningful interpretations as temporal quantities. The temporal distance between two nodes i and j, the *latency*, is the duration of the path that starts in i and ends up in j, respecting the order of links, that minimizes the arrival time. By this definition, the latency represents the smallest time for information to propagate from node i to node j. Though, this definition implies that the latency depends both on the starting and ending

nodes but also on the starting time of the path, and this dependence has a saw-tooth pattern difficult to keep under control [31]. The temporal analogous of the connected component size is the temporal reachability ratio, defined as the fraction of nodes that can be reached within time t by some information starting from node i at t_0. Similarly to the connected component sizes of static networks which undergo a sharp transition around the percolation threshold, the reachability ratio presents a transition in time.

2.2 Statistical mechanics for individual mobility

Naturally, when people meet face-to-face this happens in a spatial context. Moreover, individuals are rarely static in the environment for a long time, they move in their surrounding. The advantage of mobility models over the generative network models described in section 2.1 is that they naturally incorporate these mechanistic effects into the network, and do not just recreate the statistics [34, 35]. Mobility can be relevant both on large and small scales, depending on the subject of interest, specifically within populations or among populations. At the population level, meta-population models [36] and radiation models [37] took inspiration from theoretical ecology to describe the effects of human commuting and migration. Often, models for individual mobility find their roots in the realm of statistical mechanics, where moving agents are intensively studied to describe the behaviour of biological entities, as bacteria or animals [38]. I will begin this section reporting some works that have shown mobility to be strictly related to the contact network in real systems, and then I will describe some technical frameworks for the creation of contacts and the motion in a spatial setting.

2.2.1 Network of real human contacts

Real networks of social contacts are extremely complex as composed by the overlap of several sub-structures (private life, work, sports, school) and a strong dependence on the individual characteristics of people. Analysing real world data, Alessandretti et al. found a correlation between the strategies adopted by people in the spatial and social domains [39]. The strategies are individual dependent and persistent in time, and they define the balance between two opposite types of behaviour: the exploration of new places and new social contacts, and the exploitation of known and familiar locations and relationships.

Given the relation between spatial movement and social contacts, it results that human mobility plays an important role in determining the features of social networks. Several scientific works [40, 41] have observed power-law distributions in the travel distances of human mobility, suggesting some scale-free phenomenon in the underlying dynamics. Considering that human mobility is limited by distance and time constraints, it should have a properly defined scale, so the scale-free behaviour could be hard to interpret. For this reason, some recent works [42–46] have proposed

a new interpretation based on theoretical evidences matured in the field of Complex Systems [47, 48]: the aggregation of non-power-law distributions with ranges sufficiently close one to each other results in a power-law-like distribution. In the context of human mobility, it means that human mobility indeed has scales, one for each type of mobility that, after aggregation, induce a scale-free effect on the ranges. This may be true also for social contacts, considered the close relation discussed before. Indeed, analyzing the data from the Copenhagen Network Study [49] in Figure 2.4, the degree distribution that represents the number of daily contacts is not a power-law.

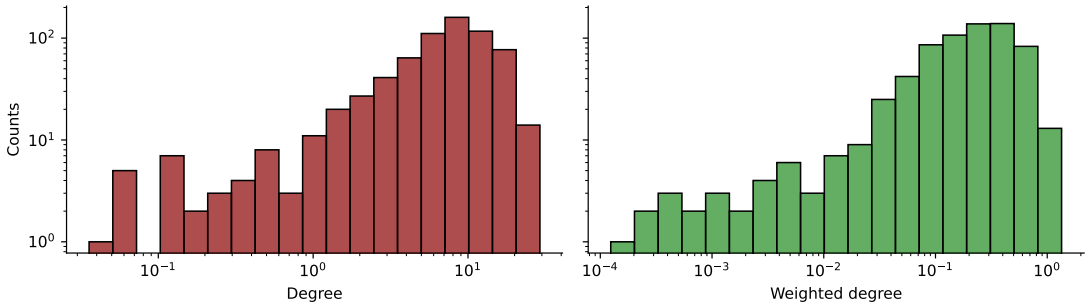


Figure 2.4: **The Copenhagen Network Study data present a daily-degree distribution that is not power-law shaped.** *The individual degree is computed aggregating the temporal network over each day of sampling and then averaging over all the* 28 d *of data availability. The degree in the left panel is the number of unique contacts in* 24 h*; the weighted degree in the right panel is the fraction of time spent being in contact with someone.*

2.2.2 Geometric graphs

In subsection 2.2.1 and in several works that analyze real scenarios [50, 51], the Euclidean spatial nature of direct human contacts as been proven to be an important point to consider. The generative models for networks discussed in subsection 2.1.4 are able to recreate some of the features of real networks, but do not directly embed any spatial causality in the network generation process. The only spatial model treated so far is the lattice; though, also the mobility plays an important role for the contact network, and lattices are static and inflexible frameworks. To work-around both restrictions it is possible to build geometric graphs [52], which are spatial networks in which the link between two nodes i and j exists if $\|\vec{x}_i - \vec{x}_j\| < r$. This is especially meaningful in the case of airborne infectious diseases, as the spreading happens mostly due to spatial proximity of an infectious and an healthy individual.

A generative model for a geometric graphs proceeds by randomly sampling the d-dimensional positions of N nodes according to some probability density function (uniform, unless differently specified), and then computing the mutual distances between them. This last step may be computationally expensive $\mathcal{O}(N^2)$ with a

brute-force approach, but using a domain decomposition algorithm it gets reduced to $\mathcal{O}(N^{1.3})$ [53]. This proceeds by dividing the d-dimensional space in smaller sub-cubes of size l, and mapping each node to the sub-cube it belongs; then, the distances $\|\vec{x}_i - \vec{x}_j\|$ are computed only between nodes i and j which belong to adjacent sub-boxes.

Random geometric graphs have been widely studied in percolation theory as continuous non-rigid versions of lattice models [54]. The authors of [53] studied some important metrics, useful also for this work. The framework is a d-dimensional system of size $L = 1$ where N points are uniformly randomly located; the *contact distance* r is the other free parameter. The excluded volume V_{ex} is the volume to be excluded from the system where to place a second node so that it lies at distance $\geq r$ from the firstly placed one. For the sake of brevity I will present the results only in $d = 2$, as it is the dimensionality used for this work, where $V_{\text{ex}} = \pi r^2$ is the "volume of influence" of each node. The average node degree (in [53] noted with α) is related to the excluded volume as

$$\langle k \rangle = \frac{N V_{\text{ex}}}{L^2} \tag{2.2}$$

To understand the reason behind this formula it is convenient to assume the point of view of one node in the random geometric graph. Once its position is assigned, there is a probability $p = \frac{V_{\text{ex}}}{L^2}$ that one other node is placed in its interaction range. This is a Bernoulli trial which is repeated for all the N nodes of the graphs, implying that the average degree is computed as in Equation 2.2 and that the degree probability distribution is a binomial. The reasoning on the degree is the same discussed in subsection 2.1.4 for the Erdős-Rényi model, but for the fact that in this case the link probability p is not just a fundamental parameter: it has a spatial interpretation. The most important difference between these two models is the fact that random geometric graphs present a form of "spatial locality", that can lead to the creation of sub-structures if by chance several nodes get placed close to each other (which is likely to happen at high densities). Indeed, clustering is a remarkable effect in random geometric graphs, and interestingly it does not depend on any free parameter but the dimensionality. In 2D the global clustering coefficient is computed as $\langle C \rangle = \frac{3\sqrt{3}}{4\pi}$, which is indeed not negligible. The average degree in Equation 2.2 is also related to the percolation threshold. From numerical simulation, it has been found that for a 2D system the critical value above which percolation happens is $\langle k \rangle_c \approx 4.52$.

2.2.3 Brownian motion

Individual human mobility is a complex phenomenon, and first-principles models only provide a rough approximation of some of its features. Often, the theoretical models used to emulate mobility are random walks, a general term that refers to a class of motions which present some stochastic term in the equation of motion. This class included a variety of models, like the Gaussian random walk [55–57] to

the run-and-tumble [58]. I will here focus on the description of one specific model, which is the one used in this work. The *Gaussian* random walk was introduced in statistical physics to describe the *Brownian motion*, i.e. the motion of a particle suspended in a static fluid at thermal equilibrium that undergoes random collisions with the other particles. It is a stochastic process that formalizes the equation of motion with a *Langevin equation* in the form $m\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = -\lambda\vec{v} + \vec{R}$, where m is the particle mass, λ is the friction coefficient and $\vec{R}$ is a Gaussian random variable with correlation $\langle R_i(t), R_j(t')\rangle = 2\lambda\kappa_{\mathrm{B}}T\delta_{ij}\delta(t-t')$. If the particle has negligible mass compared to the involved forces, the left-hand-side term is ≈ 0, and the equation simplifies to

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = \sqrt{2D}\vec{\eta} \tag{2.3}$$

where $D \coloneqq \frac{\kappa_B T}{\lambda}$ is the diffusion coefficient and $\vec{\eta} = \frac{\vec{R}}{\sqrt{2\lambda\kappa_B T}}$ is a vector of Gaussian random variables $\eta_i \sim \mathcal{N}(0,1)$.

A free random walk has a diffusive nature which tends to spread indefinitely the particle in the space. To prevent this dispersion, it is possible to confine the particle under the action of a potential V. The particle is then randomly diffusing with a *bias* given by the force exerted by the potential. The equation of motion for a Browian motion in a potential becomes

$$\frac{\mathrm{d}\vec{x}}{\mathrm{d}t} = -\nabla V + \sqrt{2D}\vec{\eta} \tag{2.4}$$

which has the analytical form of the stochastic process named as *Ornstein-Uhlenbeck*. This equation can be rewritten for convenience in its differential form, introducing the concept of *Wiener process*. This is a discrete stochastic process $\vec{W}_t$ such that the increments are stationary, independent and normally distributed as $\vec{W}_{t+\tau} - \vec{W}_t \sim \mathcal{N}(0,\tau)$. Consequently, $\frac{\mathrm{d}\vec{W}_t}{\mathrm{d}t} = \vec{\eta}(t)$ and $\mathrm{d}\vec{W}_t = \vec{\eta}(t)\mathrm{d}t$. Thus, Equation 2.3 can be rewritten in the differential form $\mathrm{d}\vec{x} = \sqrt{2D}\mathrm{d}\vec{W}$, and analogously Equation 2.4 becomes

$$\mathrm{d}\vec{x} = -\nabla V \mathrm{d}t + \sqrt{2D}\mathrm{d}\vec{W} \tag{2.5}$$

2.2.4 Fokker-Planck equation

The mobility of an agent regards the full trajectory, which is obtained integrating the equation of motion, using the tools of stochastic calculus. When dealing with stochastic differential equation, it is meaningful to look at the probability distribution of the random variable $\vec{x}$ instead of a specific realization of the process. The probability distribution $P(\vec{x},t)$ of the d-dimensional position of the particle at time t can be obtained by solving the *Fokker-Planck equation*. For an equation of motion in the form $\mathrm{d}\vec{x} = \vec{A}(\vec{x},t)\mathrm{d}t + \mathbf{b}(\vec{x},t)\mathrm{d}\vec{W}_t$, the associated Fokker-Planck equation reads [59]

$$\partial_t P(\vec{x},t) = -\sum_{i=1}^{d}\partial_{x_i}\left[A_i(\vec{x},t)P(\vec{x},t)\right] + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d}\partial_{x_i}\partial_{x_j}\left[B_{ij}(\vec{x},t)P(\vec{x},t)\right] \tag{2.6}$$

where $\mathbf{B} = \mathbf{b}\mathbf{b}^T$ represents the diffusion tensor, specifically a matrix noted with the bold sign.

For the Brownian motion (Equation 2.3) $\vec{A} = 0$ and $b_{ij} = \sqrt{2D}\delta_{ij}$, and all dimensions are independent from the others, so the Fokker-Planck equation simplifies into d identical 1D equations $\partial_t P(x,t) = D\partial_x^2 P(x,t)$. The solution to this equation is well known to be

$$P(x,t) = \frac{1}{\sqrt{4\pi Dt}} \exp\left(-\frac{x^2}{4Dt}\right) \tag{2.7}$$

2.3 Epidemiology

The other fundamental ingredient for this work is the epidemic spreading. Epidemiology is the quantitative study of the occurrence of diseases in a population. A population is generally a collection of individuals (not necessarily humans) at various scales, from a small community of few people to the whole world-wide population. Also *diseases* is a general term that includes several sub-categories, which can be distinguished on the basis of their cause: genetic, environmental or pathogenic. In the first two cases, the disease is non-infectious as the contraction does not depend on any contact with other individuals, but just on some causes that are external to the population. In this work, I will focus specifically on infectious diseases, for which the pathogen (of whatever kind) is *transmitted* from an infectious individual to a healthy one through a *contact* (that will be defined later on). When a mathematical framework was first applied by Ross in 1911 to understand the spreading of Malaria, the seeds of theoretical epidemiology were planted [60]. This work aims to extend the understanding of the physics of disease spreading, and to proceed I will present some important theoretical aspects that will be later referenced in the description of this work.

2.3.1 Mean-field compartmental models

The simplest models for epidemic spreading are *compartmental models* [1]. They are based on mutually exclusive *compartments*, that represent the status of the individual with respect to the disease, i.e. the stage of the disease course at which the individual is. Each individual is assigned to one compartment, and it can transition from that to another according to some well defined rules. Disease stages have been introduced with compartmental models, but they can be used in a wider range of models to represent the core characteristics of each phase of the disease. The most commonly used are:

1. susceptible S. This stage is occupied by healthy individuals who can be infected by infectious ones during an *infection*.

2. exposed E. This stage is occupied by individuals who already got infected but are still not able to infect others. It lasts on average for a *latent period*

τ_{lat}.

3. infectious I. This stage is occupied by infected individuals who are already able to infect others at a rate λ and spread the pathogen. It lasts on average for an *infectious period* τ_{inf}.

4. removed R. This stage is occupied by recovered (or dead) individuals who are immune to any infection and are not infectious anymore. It lasts on average for an *immunity period* τ_{imm}.

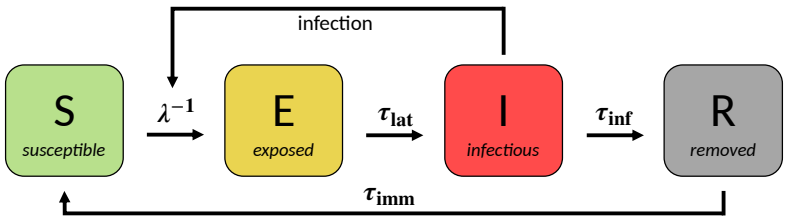


Figure 2.5: **Scheme of a SEIRS compartmental model**

Compartmental models model the evolution of the system using a set of coupled differential equations that rule the transitions of individuals between compartments. The system dynamics is mathematically described as a chemical solution in which reactions occur between substances (the compartments). This analogy has two important implications: all the molecules are indistinguishable, meaning that no individual characteristic is relevant for the system dynamics, i.e. a *mean-field* approximation is applied; the system itself is *well-mixed*, meaning that there are no sub-domains of prevalence of one substance and the whole solution is homogeneous. This approach carries the great advantage to be extremely fast, as the differential equations can be numerically integrated to obtain the evolution curves, and very simple, as there are few parameters that have to be tuned to reproduce in first approximation real data [61–67]. On the other side, it hardly captures phenomena that are due to individuality or sub-structures.

Despite the limits just discussed, compartmental models show some interesting spreading phenomena. In a pioneering work [68, 69], Kermack and McKendric discovered that not all the pathogens are able to invade a fully-susceptible population: there must be a sufficiently large infectious capability to prevent the epidemic from dying out beforehand. The *epidemic critical threshold* is the value that determines the transition between the absorbing phase (disease going extinct before spreading) and the active phase (disease spreading). An epidemic is in the absorbing phase if the infectiousness is so low that recoveries occur faster than infections, implying that the average number of infectious individuals decreases over time. An important parameter, related to the critical threshold, is the *basic reproductive number* R_0, that represents the average number of infections caused by a single individual in the infectious stage in a *fully susceptible population*. When this value $R_0 > 1$ there is a sufficient number of secondary cases to sustain the epidemic and the disease is in the active phase.

An important aspect to analyze when dealing with disease spreading is the growth of the *prevalence* $I(t)$, i.e. the number of infected individuals at time t. In this framework, R_0 new cases are generated from one infectious individual, and each of them will generate other R_0 new ones, until a sensible fraction of non-susceptible individuals occupies the system. This implies that the prevalence growth in a compartmental model is exponential $I(t) \propto e^{R_0 t}$ in the early stages. This fact is validated by several empirical observations [70, 71], but exponential growth is very delicate to identify [72] and it can lead to very unstable forecasts [73, 74] and fits.

2.3.2 Spatially diffusing diseases: lattice models

Mean-field compartmental models completely neglect any structure in the population, including the spatial domain. Most real diseases need a form of spatial proximity between the infecting agent and the infected, and lattice models allow to embed the spreading into a spatial setting, as already discussed in subsection 2.1.3. As widely studied models of solid-state physics, lattices have been at the core of the seminal works on epidemics diffusion [75, 76]. The first, great difference between the spreading dynamics on lattice and the mean-field compartmental models is the growth of the prevalence. In spite of being exponential in the early stages of the epidemic, the prevalence increases as $I(t) \propto t^{\alpha}$, where α depends on the dimensionality of the system and the type of disease.

Consider a 2D lattice with one infectious node in a SI model, in which infectious nodes never recover. The transmission rate determines the "velocity" of diffusion along every direction, so the spreading behaves as a "front wave" that travels at constant speed, as sketched in Figure 2.6. All the nodes within the front wave are infectious, so $I(t)$ is proportional to the area of the bubble which radius grows

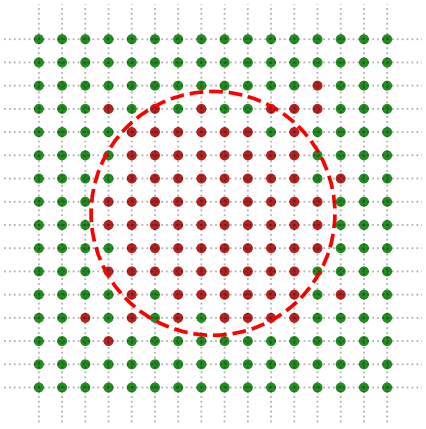


Figure 2.6: **Spatial spreading of a SI model on a 2D lattice.** *The image sketches the diffusion of a SI-disease on a 2D lattice. The nodes in green are in the susceptible stage and the red ones are in the infectious stage. The dashed red line is a representation of the wavefront of infection.*

linearly, meaning $\alpha_{\mathrm{SI}} = 2$. Generally, under the same assumptions, $\alpha_{\mathrm{SI}} = d$ when neglecting finite size effects and stochastic fluctuations. When considering an SIS disease, the exponent does not change $\alpha_{\mathrm{SIS}} = \alpha_{\mathrm{SI}}$ even if not all nodes within the front wave will be infectious. Indeed, the "infection bubble" converges to the equilibrium state with a fraction of S and I nodes that satisfies detailed balance between new infections and recoveries, implying that $I(t)$ differs only for a multiplicative constant. The same reasoning does not apply to SIR diseases, as equilibrium cannot be reached. In this case, the infectious nodes will be only *on* the wavefront, which is in general a hypersurface with one less dimension than the embedding space, implying that $\alpha_{\mathrm{SIR}} = \alpha_{\mathrm{SI}} - 1$. I point out that only the final stage is relevant for the growth exponent α, as all intermediate stages are absorbed by multiplicative constants. The difference between the power-law growth on lattice and the exponential in compartmental models has important implications for forecasts, due to the sensible difference in the prevalence growth at large times.

Diffusive processes on lattices have been intensively studied along with the birth of the physics branch of complex systems. In particular, it was discovered that disease spreading on lattices is a dynamical process that belongs to the universality class of directed percolation [77, 78]. Universality classes are groups of mathematical models that share the same scaling behaviour at criticality. The term "universality" stresses that the the system behaviour is independent of the contextual details, and it only depends on the symmetry properties of the mathematical description close to a phase transition. In particular, directed percolation describes the percolation of a fluid in a d-dimensional porous medium along a preferred direction. Epidemic spreading on a d-dimensional lattice show the same scaling behaviour of a percolating fluid in a $(d+1)$-dimensional medium, where the $+1$ refers to the time dimension that happens to be preferred direction of percolation[1]. This result implies that the scaling behaviour of several quantities are known providing that the critical threshold mentioned in subsection 2.3.1 identifies the phase transition. Some of these scaling laws are $I^{\infty}(h) \sim h^{\beta}$ and $p_{\mathrm{surv}}(h) \sim h^{\beta'}$, where $I^{\infty} \coloneqq I(t = \infty)$ is the prevalence at equilibrium, p_{surv} is the probability that an epidemic does not go extinct, $h \coloneqq p - p_c$ is the control parameter that determines the distance from the phase transition, and β and β' are the critical exponents, which for $(2+1)$D directed percolation are evaluated $\beta = \beta' \approx 0.5834$ [79].

2.3.3 Epidemics on complex networks

The complexity of network models allow to capture more different aspects of real networks compared to lattices, as discussed in subsection 2.1.4. Specifically, some structural characteristics observed in empirical networks can have positive or

[1] I want to point out that directed percolation in $(d+1)$-dimensions is sometimes referred to as *non-equilibrium directed percolation in d-dimensions*, as non-equilibrium already includes time in the assumptions. This use of similar definitions may lead to misunderstandings, so in this work I will stick to the first one, considering the dimensionality augmented for time.

negative impacts on the spreading velocity of the epidemic [80–82]. A network with a small diameter, or generally small average distances between nodes, is a good setting for a fast-spreading disease, because few infections are sufficient to largely extend the area of infection. This setting is therefore similar to the well-mixed system assumed by compartmental models, thus Erdős-Rényi, Watts-Strogatz and Barabási-Albert networks all present an exponential growth of the prevalence in the early stages [6]. Moreover, Barabási-Albert networks also have large hubs among their nodes, which can act as *super-spreaders* for the epidemic thanks to their large number of contacts. When these nodes get infected, they strongly enhance the spreading speed overcoming the absorbing effects of the low-connected nodes, and the super-spreading effect results to be dominating [83].

A very interesting aspect that emerges from the analysis of epidemics on networks is the implication of the presence of clusters. Clusters form sub-structures where the internal connectivity is above the average on the entire network. The high strength of the links within the sub-structures implies that the sub-structures are more loosely connected one another, for a given global connectivity. Under a spreading point of view, this configuration enhances infection speed within the sub-structures, but slows down and makes less stable the infection between sub-structures that happens through less and more fragile links. The overall effect is to slow-down the epidemic spreading, turning the exponential growth into a power-law [84, 85], at least for the networks in which super-spreading and short-distance effects are less relevant than clustering.

2.3.4 Epidemic diffusion on individual mobility models

As discussed in section 2.2, mobility models better incorporate spatial effects into the contact network generation than network models, while providing more flexibility and realism than lattice models. The authors in [55, 56] build a mobility model that joins local random walks with stochastic long-distance jumps. They observe that even with small jumping probability, the epidemic growth changes from polynomial to exponential, resembling the growth of mean-field models. This effect is due to the jumps that mix the system shuffling the positions of (some) walkers, so the locality of the spreading gets more and more dispersed as the jumping probability increases. Furthermore, being this a spatial model, the clustering is high, so the local spreading of the disease is boosted, and the slow-down of the spreading between clusters (discussed in subsection 2.3.3) is suppressed by the jumps. In another work, Rodríguez et al. have studied the response of a similar model, which approaches the mobility with the ”run-and-tumble” motion, to the variation of the diffusion length of the agents [58]. Also in this case, the authors obtain that the epidemic growth is similar to the one in a lattice for low mobility, and tends to a mean-field compartmental model when the scales of mobility and system size become comparable. They analyze the critical exponents to conclude that the universality class corresponds to the one of mean-field models when

mobility is sufficiently large, and obtain that the scaling laws are $I^{\infty}(h) \sim h^{\beta}$ and $p_{\text{surv}}(h) \sim h^{\beta'}$, where $I^{\infty} \coloneqq I(t = \infty)$ is the prevalence at equilibrium, p_{surv} is the probability that an epidemic does not go extinct, $h \coloneqq p - p_c$ is the control parameter that determines the distance from the phase transition, and β and β' are the critical exponents, which for the mean field are evaluated $\beta = \beta' = 1$

2.3.5 The importance of time for epidemics

In subsection 2.1.5 I mentioned some features of human behaviour that only emerge when the temporal dimension is taken into consideration [86]. The interesting question to address is their impact on the epidemic spreading. It is known that human activities are not uniformly distributed in time, indeed they tend to concentrate in short periods of very intense activity separated by long periods of very low level [28, 30]. This feature is known as *burstiness*, and it appears an important factor in the slowing-down of the spreading of an epidemic [87, 88]. The authors hypothesize that, when an individual gets infected, its chances to re-transmit the pathogen decrease when there is a high probability of a long waiting time before the successive contact. This effect would win over the boost due to easy transmission in the bursty phases. Such phenomenon would be the temporal analogy of the slow-down due to spatial clustering effects discussed in subsection 2.3.3.

Another prominent temporal feature of human behaviour is the cyclical pattern of the activities. These cyclic activities may regard the daily sleep routine, the weekly weekend-trip routine, the yearly organization of the summer holidays, and so on. All these cycles occur on specific timescales. Here I will focus on the circadian cycle, the most evident alternation between two completely different attitudes: highly active during the light hours, almost inactive during the night hours [29]. Zierenberg et al. [16] have studied the impact of the cyclic behaviour on the epidemic spreading, finding out that the spreading can be boosted or suppressed depending on the specific temporal characteristics of the disease. The idea is that most contacts happen during the daylight hours, so an individual is more likely infected in this time; after the infection, if it becomes infectious during the night hours and recovers before the next day, there is a very low chance that secondary cases are generated; on the contrary, if the infectiousness grows during the next daylight hours, there is high chance to infect other individuals. In figure Figure 2.7 is shown that the cyclic contact activity can suppress down to 50% or boost up to 150% the infectiousness of diseases that differ only by the duration of the latency stage and the infectious stage. An analogous result has been found by Kivela et al. [88], confirming that temporal features are as important as spatial ones in the understanding of spreading processes on networks.

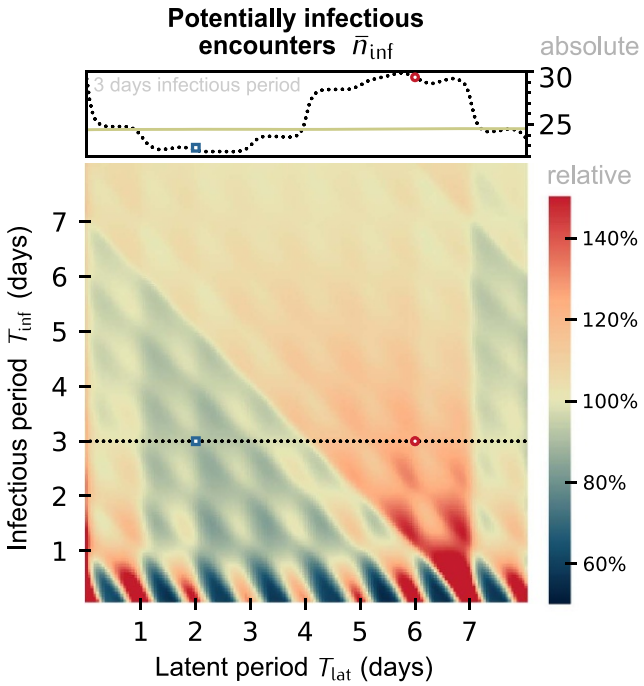


Figure 2.7: **Effective infectiousness with circadian cyclic contacts.** *The figure shows the variation of the effective infectiousness for a disease with fixed transmissibility and variable duration of the latency and the infectious stages. The variation is normalized to a randomized model that shuffles the temporal coordinate of the contacts. It is worth to notice that for infectious and latent periods that correspond to multiples of the cycle period (*$1\,\mathrm{d}$*), the variation is very small, and it is maximum in correspondence of half-cycle values. Figure provided by the authors of [16].*

Chapter 3

Model

In chapter 2 I provided insights on the scientific context in which this work sets. With this preamble in mind, I will now shift the focus on the actual project of this Master thesis, starting from a detailed discussion on the driving motivation. Later on, I will explain the assumptions and technical solutions that I implemented to tackle my research problem, both concerning the model for the agents' mobility and the model for the disease.

This project relies on numerical simulations. The coding framework consists of two different languages that interact with each other. The core routines that perform the extensive computations are written in `C++`, allowing for a very fast execution and efficient memory management. These routines are compiled and provided as `Python` modules to the user, enhancing the usability of the library and favouring the integration with the most common `Python` libraries for data analysis. The binding code that allows for interoperability between `C++` and `Python` is created making use of the open library `pybind11` [89].

3.1 Motivation

The importance of understanding disease spreading has been highlighted by recent human history, but in general spreading processes are worth the attention because they describe the mechanics at the core of a huge variety of real systems. Humans are extremely complex agents to be modelled, motivating the reason why human epidemics are one of the less understood spreading process. In chapter 1 I already mentioned the ethical and technical difficulties of data collection on human behaviour, due to privacy protection and data security. This has motivated the choice to build a *generative* model for the contact network, instead of applying other analysis on existing data.

Network of contacts are often created as random networks generated through the models presented in subsection 2.1.4, and they well recreate some statistical

features of real networks. Though, when narrowing the landscape to human contact networks, some of these features are either absent or due to a superposition of several undergoing processes, as discussed in subsection 2.2.1. There, I also present some works which highlight the close relationship between the mobility in the spatial domain and contact network. These considerations led me to avoid common models for network generation, and, instead, resort to spatial models. In subsection 2.1.3 I mentioned that lattice models have been largely used in the 19th century, but they provide no flexibility and, especially, cannot incorporate the so important mobility. Therefore, I decided to join a mobility framework with a geometrically generated network. Mobility is too complex to be fully-captured by first principle models, so I decided to focus only on a simple representation of the lowest level scale discussed in section 2.2. The geometric network generation (subsection 2.2.2) provides the flexibility needed to account for the mobility and allows to mechanistically incorporate the spatial dimension in the network.

The choice to include the spatial domain in the model has consequences on the spreading processes, as extensively discussed in section 2.3. When fitting the growth of the prevalence curve in the early stages of the epidemics, it can be hard to distinguish between the exponential increment, given by homogeneous mixing or random network models, and the polynomial growth, given by the spatial ones, and this can lead to imprecise predictions. The mobility takes the role of the hyper-parameter that allows to interpolate between these two extremes: for very low mobility the physical behaviour of the system resembles the pure spatial models (lattices), while for very high mobility it resembles the well-mixed systems, as discussed in subsection 2.3.4. Though, those mobility models do not consider any temporal aspect, which have been proven to be important for the spreading processes in real scenarios (subsection 2.3.5). This work aims to fill this research gap, embedding the temporal dimension through a circadian cyclic pattern for the spatial mobility. The goal is to understand the consequences of the interplay between space and time on the physics of epidemic spreading, regarding both the growth of the prevalence and the universality class of the absorbing-active phase transition.

3.2 Mobility

The setting of the model is a 2 dimensional squared *environment* of area L^2 with periodic boundary conditions that contains N agents. The choice of periodicity has the aim to make spatial boundaries effects vanish also for small system sizes, allowing for more scalability freedom. Moreover, the environment is translationally invariant, and there is no intrinsic difference in the behaviour of the agents due to their position in the environment. Each agent is assigned an initial position in the environment, representing its resting place (or its *home position*), and this is the only source of individuality in the model. The initial positions are uniformly

drawn at random $\vec{x}_0^i \sim \mathcal{U}(0, L) \times \mathcal{U}(0, L)$, where the index $i \in [1, N]$ represent the i-th agent.

Each agent (later addressed also as *walker* or *node* or *individual*) moves in the system according to the differential equation of motion

$$\mathrm{d}\vec{x}^i = -\nabla V^i(\vec{x}, t)\mathrm{d}t + \sqrt{2D}\mathrm{d}\vec{W} \tag{3.1}$$

analogous to Equation 2.4. The stochastic term on the right is independent from i, as all the agents undergo the same Gaussian random perturbation. The choice of a Gaussian random walk is motivated by the difficulty to find in the literature a more realistic model for individual mobility that would account for the spatial domain; in this respect, Gaussian random walks are used in the literature for both animal and human mobility (see section 3.2), and are very well studied in statistical physics.

The motion of each agent is solved independently through numerical simulation using the Euler-Maruyama method for stochastic differential equations [90]. Equation 3.1 is solved iteratively by computing a Markov-chain that follows the rule

$$\vec{x}_{t+\Delta t} = \vec{x}_t - \nabla V(x_t, t)\Delta t + \sqrt{2D}\Delta\vec{W}_t \tag{3.2}$$

where Δt is the time discretization and the subscript represents the time in the simulation. I note that by definition of Wiener process, each component $\Delta W_t \coloneqq W_{t+\Delta t} - W_t \sim \mathcal{N}(0, \Delta t)$ is a Gaussian random variable which variance scales as the time discretization. This is indeed correct, as the dispersion due to noise over a large period $\vec{x}_{t+T} - \vec{x}_t$ should be only proportional to the diffusion parameter D and the length of the period T and not to the discretization Δt taken. Indeed, assuming $\nabla V = 0$ and $x_0 = 0$ for simplicity, the spatial dispersion over the period T is $\vec{x}_T = \sum_{t=0}^{T} \sqrt{2D}\Delta\vec{W}_t$ is a sum of rescaled Gaussian random variables with variance $\sigma^2_{\Delta W} = \Delta t$. The sum of $\frac{T}{\Delta t}$ rescaled Gaussian random variables is again a Gaussian random variable with variance equal to the rescaled sum of the variances, thus

$$\sigma_x^2 = (2D)\frac{T}{\Delta t}\sigma^2_{\Delta W} = 2DT \tag{3.3}$$

which correctly proves to be independent from any choice of the time-discretization unit Δt.

The individual attraction potential $V^i(t)$ in Equation 3.1 is introduced to constrain each agent in a region close by its home (the random initial position). This is a braking point with respect to the models described in subsection 2.3.4, where the random walk is boundless and spans through the entire environment. The aim is to allow for mobility while preserving the individuality of the agents, as the home position is the only features that distinguishes them. The potential must fulfil three main requirements: it must have a bell-shaped global minimum centered on the home position, it must be differentiable so that the gradient exists, and its gradient must be reasonably bounded within the domain of the environment

so that the attractive force does not overwhelms the dynamics of motion. The individual potentials have the same shape and only differ in their center, in formula

$$V^i(\vec{x},t) = A(t)\frac{1}{\alpha}\left(-\delta^\alpha + \left(\delta^\gamma + \left(\frac{\left\|\vec{x}-\vec{x}_0^i\right\|}{\beta}\right)^\gamma\right)^{\frac{\alpha}{\gamma}}\right) \tag{3.4}$$

where $A(t) > 0$ is the amplitude, $\alpha > 0$ sets the long-range behaviour, $\beta > 0$ sets the spatial scale, $\gamma \in 2\mathbb{N}$ sets the short-range behaviour and $\delta > 0$ sets the scale to distinguish short versus long range. I point out that the constraint imposed on the parameter γ shapes the potential to be *rotationally invariant* with center on the home position. In Figure 3.1 the effect of the variation of each parameter (excepts β that is trivial) is shown for an easier intuition about their meaning. The choice

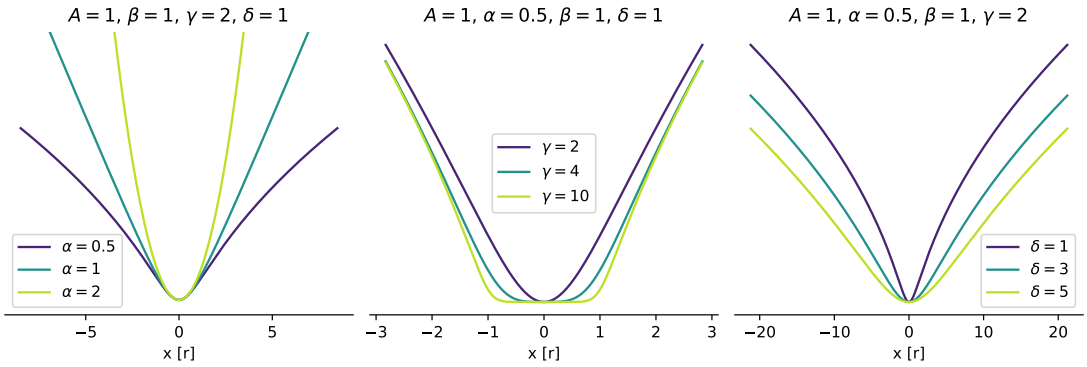


Figure 3.1: **The potential is shaped by the value of its parameters.** *The potential in Equation 3.4 is shown. In the left panel, the potential shows different behaviours in the tails, where $V(x) \sim x^\alpha$. In the central panel, the potential shows different behaviours in the well, where $V(x) \sim x^\gamma$. In the right panel, the potentials plotted have the same behaviour in the tails and in the central well, but the well has different amplitude.*

of the functional form of the potential is motivated by the extreme flexibility it provides, which extends further the needs of this work, as discussed in Appendix C. The explicit form of the gradient is

$$\nabla V^i(\vec{x},t) = A(t)\beta^{-\gamma}\left\|\vec{x}-\vec{x}_0^i\right\|^{\gamma-1}\left(\delta^\gamma + \left(\frac{\left\|\vec{x}-\vec{x}_0^i\right\|}{\beta}\right)^\gamma\right)^{\frac{\alpha}{\gamma}-1} \tag{3.5}$$

The temporal heterogeneity is introduced in the mobility model through the amplitude parameter of the potential $A(t)$, which is the only parameter with a time dependence. The basic idea is to create a circadian pattern that alternates an exploratory phase and a confinement phase: the former to represent the active daily routine during the day hours, the latter to represent the passive stay at home

during the night hours. This translates into an attractive potential which is much stronger during the night hours than during light hours. The functional profile chosen to reproduce this effect is a rectangular function in the form

$$A(t) = \begin{cases} A_{\text{day}} & \frac{1}{2} - \frac{e}{2} + k \leq \frac{t}{1\,\text{d}} \leq \frac{1}{2} + \frac{e}{2} + k \\ A_{\text{night}} & \text{elsewhere} \end{cases} \qquad \forall k \in \mathbb{Z} \tag{3.6}$$

where $A_{\text{day}} < A_{\text{night}}$ are the values constraining, respectively, the exploratory and confinement phases, and e represents the fraction of time spent in the exploratory phase in a day.

3.3 Epidemic

The relevance of the motion of the agents in the environment is the contact network that emerges among them. Since the motion of the agents evolves in time, the most proper representation of the contact network is a temporal network, with sampling rate that matches the discretization of the agents' motion. At each timestamp, if two walkers i and j are in the condition such that $\left\|\vec{x}_t^{\,i} - \vec{x}_t^{\,j}\right\| < r$, they are defined to be *in contact*. In other words, at each timestamp the *snapshot* of the temporal network of contacts is a geometric graph modelled as in subsection 2.2.2. By this definition, the duration of each contact equals the time-discretization unit Δt chosen for the random walk. Since a geometric graph is generally very sparse, especially at low densities, I choose to represent the temporal network of contacts as a timed edge list.

The contact network is the topology underlying the epidemic spreading. In this work, the epidemic is generally modelled as a SEIS disease, unless differently specified and motivated. This model has been chosen so that an equilibrium solution is reached after the epidemic spreads across the entire environment, and this is only possible if the course of the disease forms a closed loop over the compartments. The equilibrium is useful, if not necessary, for the analysis of the universality class and the physical properties of the system. Any immunity is ignored in this model, as it would augment the complexity of the model without adding any foreseeable interesting effect. The exposed stage has been proven to have an important effect on the spreading, as discussed in subsection 2.3.5, but it also slows down the growth of the epidemic because of the delayed infections. This is a useful effect when dealing when finite size systems, for which an early saturation would make impossible to observe the growth of the spreading. Though, the most important consequence of the existence of a latency between the moment of infection and the onset of infectiousness is that it decouples the location of infection from the generation of secondary cases: suppose that agent i infects agent j at position $\vec{x}_c$ and there is no latent period; right after this infection, both i and j are still in the neighbourhood of $\vec{x}_c$, so all the other agents in that neighbourhood are likely infected by j other than by i; on the contrary, if there is a latent period that lasts for $1 - e$d at least

(the duration of the day hours), then the node j will return in the neighbourhood of his home position $\vec{x}_0^j$ before becoming infectious, so the secondary infections generated by j will be only dependent on $\vec{x}_0^j$. This coupling may or may not be desirable in the model, but a deeper understanding would be required to perform a conscious choice. In order to avoid uncontrolled effects, I choose to insert the exposed stage in the model.

The transition from one stage to the other undergoes the following rules:

- $S \to E$. The infection can happen only between two agents in contact, one infectious and one susceptible. In such situation, the infection has probability p_{inf} to actually realize. The contacts have duration equal to Δt, so the rate of infection is $\lambda_{\text{inf}} = \frac{p_{\text{inf}}}{\Delta t}$. Since the dynamics of an epidemic is related to λ_{inf}, if one wants to change Δt while preserving the same force of infection, then p_{inf} must be tuned accordingly.
- $E \to I$. The onset of the infectiousness happens after a time t_{lat} since the infection, and this time is randomly drawn from the distribution $P(t_{\text{lat}})$ when the infection takes place.
- $I \to S$. The recovery happens after a time t_{inf} since the onset of the infectiousness, and this time is randomly drawn from the distribution $P(t_{\text{inf}})$ when the onset takes place.

In all the runs presented in this work, the generality previously described is not exploited. All simulations are performed with the same time-discretization, and the sampling distributions are $P(t_{\text{lat}}) = \delta(t_{\text{lat}} - \tau_{\text{lat}})$ and $P(t_{\text{inf}}) = \delta(t_{\text{inf}} - \tau_{\text{inf}})$. The choice of imposing single-valued transition rates to all the agents is operated to avoid stochastic effects that would have hidden the interesting effects without changing the average behaviour of the system. This is a first step that might be relaxed in a follow-up work.

The evolution of the agent' motion and the creation of the contact network happen concurrently to the spreading of the disease. The simulations start by placing the home positions of the agents uniformly at random in the environment, and creating an empty queue of events. Each event is the transition of an agent from one compartment to another, and it happens at a certain time t_e. The second step of the algorithm is the seeding of the epidemic: one agent, drawn at random, gets infected at a random time during the first day of simulation, and its transition is added to the queue. Then, the algorithm proceeds looping over the time-steps, indexed by t, as follows:

1. all the entries in the event queue that happen at a time $t_e < t$ are removed from the queue and the corresponding transitions are applied. If the event agent is in the exposed stage, than $E \to E - 1$ and $I \to I + 1$, and an event is appended in the queue for the same agent at time $t_e = t + t_{\text{inf}}$. If the event agent is in the infectious stage, than $I \to I - 1$ and $S \to S + 1$, and nothing

is add to the event queue.

2. the positions of all the agents are updated according to Equation 3.1.
3. the contact list is computed onto the updated positions of the agents using the algorithm for geometric graph discussed in subsection 2.2.2 and shown in Figure 3.2. The sub-domains size is chosen to be $l = r$, so that each agent can be in contact with other individuals that lie only in the adjacent sub-domains. This solution allows to have a computational complexity $\mathcal{O}(N \log N)$.
4. the contact list is parsed; for each contact, if one is in the infectious stage and the other is in the susceptible stage, the infection happens with probability p_{inf}.
5. if the infection actually realizes, $S \to S - 1$ and $E \to E + 1$, and an event is appended in the queue for the infected agent at time $t_e = t + t_{\text{lat}}$.
6. the occupancy of each pool is recorded.
7. the algorithm advances by one time-step, and proceeds to the next iteration only if the event queue is not empty and the maximum duration of the simulation is not yet reached $t < t_{\max}$.

In the case a SEIR model is needed, the transition from $I \to S$ becomes $I \to R$ and nothing in the algorithm has to be changed; in the case of a SEI model, the transition $I \to S$ has no effect and again nothing has to be changed.

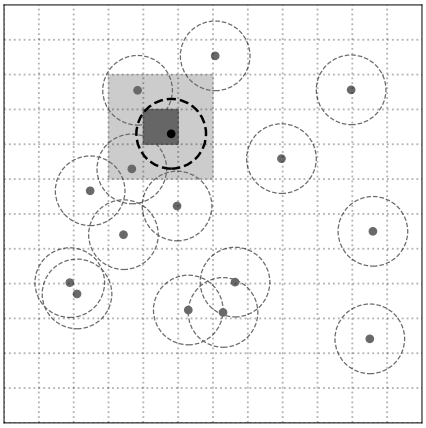


Figure 3.2: **Graphical representation of the algorithm for geometric graphs.** *The outer box in the figure represents the 2D environment, the dotted lines the boundaries of the sub-domains, and the gray points the agents. The circumferences represent the contact range (of size r). The agent in black is the subject, and its sub-domain is dark coloured, while its adjacent sub-domains are light gray. When computing the distances between agents, for the agent in black only the ones from the agents within the gray area are computed. For this purpose, the size of each sub-domain equals the interaction range r.*

Chapter 4

Tuning the model

I explained in chapter 3 the ideas behind the model used in this work, mentioning the relevant parameters and their meaning. This chapter is dedicated to the discussion of the values to assign at those parameters, some of which require extensive consideration. The analysis of the behaviour of the model will come along with the first results obtained, with the aim to keep all the effects under control. Some of the parameters have been kept fixed throughout all the simulations carried out during this work; others have been varied to inspect their effects on the system's behaviour. Table 4.1 summarizes the values used in the simulations, to facilitate the reader's understanding of the results presented in chapter 5.

4.1 Epidemic parameters

The three epidemic parameters of the model are the probability of infection per contact p_{inf}, the latent period τ_{lat} and the infectious period τ_{inf}. The first one is the subject of interest of most results, as it is the one that represents the force of infection of the disease. In particular, it is the parameter which determines the transition between the active phase (spreading) and the absorbing phase (extinction). On the contrary, the other two parameters are fixed throughout the whole project, as the implication of their variation is out of the scope of this work. As discussed in section 3.3, the latent period is chosen to be larger than the period of the cyclic behaviour of interest, i.e. $\tau_{\text{lat}} \geq 1\,\text{d}$, so that the diffusion of the disease only depends on the home positions of the agents. Since it is desirable to have a fast epidemic to lower the computation power needed for the simulations, I set $\tau_{\text{lat}} = 1\,\text{d}$. For what concerns the infectious period, the evolution of a SEIS epidemic presents damped oscillations at equilibrium if $\tau_{\text{lat}} \geq \tau_{\text{inf}}$. Since it is an undesirable effect for this study, I set $\tau_{\text{inf}} = 2\,\text{d}$ to disrupt the intrinsic oscillatory behaviour of the system, so that shorter simulations are needed to ensure the equilibrium regime.

Parameter	Symbol	Standard Value
agents N°	N	$*10^5$
agents density	ρ	$*0.025\,\mathrm{r}^{-2}$
environment size	L	$*\sqrt{N\rho}$
discretization time	Δt	1 min
simulation time	$t_{\max}$	*
contact range	r	1
avg. displacement on exploration	σ_{free}	$*10\,\mathrm{r}^{-2}$
diffusion coefficient	D	*Equation 4.5
potential amplitude on exploration	A_{day}	0
daily fraction of exploration phase	e	*0.5
avg. displacement on confinement	σ_{conf}	$1\,\mathrm{r}^{-2}$
potential coefficients	$\alpha, \beta, \gamma, \delta$	$1, 1, 4, 0.5$
potential amplitude on confinement	A_{night}	*Equation 4.6
probability of infection	p_{inf}	*
latent period	τ_{lat}	1
infectious period	τ_{inf}	2
initial N° of infected	I_0	1

Table 4.1: *The meaning, symbol and value of the parameters are here collected. All non-starred values are kept fixed in all the simulations. The starred values have been used as the* standard condition*, and in the text it is explicitly stated when other values are being used. The values of $t_{\max}$ and p_{inf} are missing because they change in the majority of the simulations.*

4.2 Environmental parameters

The environmental parameters of the model are the contact range r, the number of agents in the system N and the size of the environment L. The contact range is chosen to be the spatial unit of measure, so that all the other quantities are expressed in terms of interaction radii, meaning that $r = 1$. In real settings with air-transmitted pathogens, the spatial proximity needed to define a contact (thus a possible infection) is usually between 1 m and 2 m; this gives a sense of the magnitude of the distances in my model. The number of agents is again a parameter that is not fixed for all the simulations, as some cases require very large systems and others only perform fast runs. Generally $N = 10^4$–10^5 unless differently specified, so that system size is large enough to observe the interesting phenomena while preserving a reasonable computation time.

The environment size L is not a directly meaningful parameter, its role is relevant only because the agents' density varies. Thus, I select this density $\rho \coloneqq \frac{N}{L^2}$ as a model parameter, as it has a dominant effect on the system's behaviour, implying that $L(N, \rho)$ turns into a derived quantity. The system must be dense enough so

that the epidemic spreading is not arrested by the absence of contacts, but also sparse enough so that the mobility (and not the over-density) plays the major role as the spreading driver. In other words, if the agents were fixed into their home positions, there should exist no relevant spreading, i.e. the system should be *non-percolating*. For a random geometric graph, the density threshold for percolation is computed inverting Equation 2.2. Inserting in the equation the definition of $\rho := \frac{N}{L^2}$, the excluded volume in 2D $V = \pi r^2$ and the percolation threshold $\langle k \rangle_c = 4.52$, the system results being non-percolating if $\rho \ll \frac{4.52}{\pi r^2} \approx 1.43\,\mathrm{r}^{-2}$. The lower-bound on the agents' density can only be checked a-posteriori: first a density value is set, then the system is evolved and the contact traced, so to check if the temporal network percolates. I arbitrarily set $\rho = 0.025\,\mathrm{r}^{-2}$, and evolved the system with the lowest mobility to capture the worst-case scenario. Then the temporal network obtained is aggregated over the smallest time-windows ΔT that are relevant for the spreading, i.e. the infectious period. Figure 4.1 shows that during $\Delta T = \tau_{\mathrm{inf}}$ most of the nodes remain isolated, implying that the disease has a high chance to get extinct because of the absence of contacts; but already during a $\Delta T = 2\tau_{\mathrm{inf}}$ the fraction of isolated nodes significantly drops, implying that the spreading is not prevented by the absence of contacts and it mostly depends on the epidemic parameters, namely the force of infection. Hence, the proposed value $\rho = 0.025$ is indeed a good trade-off

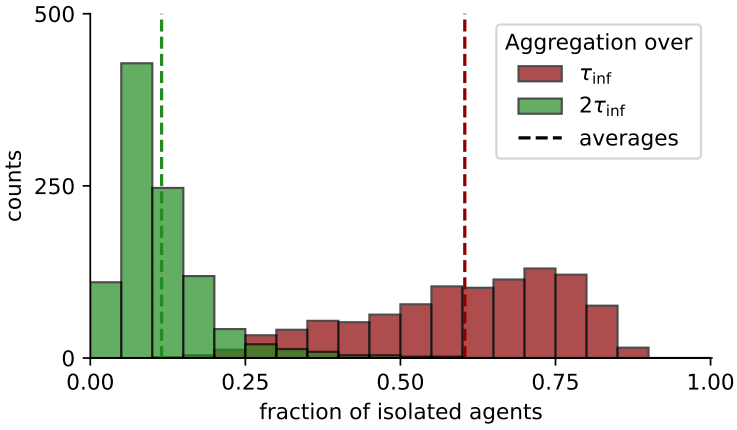


Figure 4.1: **Fraction of isolated agents.** *The plot shows the distributions of the fraction of isolated agents over* 1000 *runs. The colour coding represents the length of the time-window over which the temporal network is aggregated. The aggregation over* $2\tau_{\mathrm{inf}}$ *is already sufficient to have a percolating network.*

and it is the default values for all the simulations run in this work. This observation also suggests that the mobility is indeed crucial in this model, as it drives the instantaneously-non-percolating geometric graph towards a percolating aggregated network within timescales that are faster the epidemic spreading. I point out that the unit of measure of the density is $[\rho] = \frac{1}{r^2}$, so considering $r = 2\,\mathrm{m}$ (as previously discussed) the density value in real units becomes $\rho = 6250\,\mathrm{people/km}^2$, which is of

the order of the population density of large crowded cities (Wikipedia data report $\rho_{\text{Milan}} \approx 7500\,\text{people/km}^2$ and $\rho_{\text{London}} \approx 5600\,\text{people/km}^2$). This is not an attempt to pursue realism, as this model is meant to be purely theoretical, but it gives an intuition about the scales of the quantities involved.

4.3 Analytical approaches

Solving analytically the equation of motion of each walker allows to reduce the number of simulations needed to explore the mobility parameter space in search for suitable values. Since all the walkers have identical behaviour and only differ by their home position, all solutions are equal but translated. The equation of motion is much easier to solve in the assumption of null attractive potential (here addressed as *free motion*), meaning that the motion is purely noise-driven. In this case, the noise is Gaussian and the probability density is given by Equation 2.7 that becomes for agent i

$$P^i_{\text{free}}(\vec{x}, t) = \frac{1}{Z(t)} \exp\left(-\frac{\left\| \vec{x} - \vec{x}^i_0 \right\|^2}{4Dt} \right) \tag{4.1}$$

where $Z(t)$ is the normalization constant. This solution is exactly Gaussian only assuming a boundless environment, but it is approximately Gaussian if the amplitude is much smaller than the environment size. The goodness of the approximation can be checked computing $Z(t)$ on the finite domain and estimating the approximation needed for it to converge to the Gaussian normalization constant $Z_G = 2\pi\sigma^2$. Using the periodicity of the boundary conditions I simplify $\vec{x} - \vec{x}^i_0 \to \vec{x}$ and defining $\sigma(t) = \sqrt{2Dt}$, the normalization constant becomes

$$\begin{aligned} Z(t) &= \iint_{-L/2}^{L/2} \mathrm{d}x\mathrm{d}y \exp\left(\frac{x^2 + y^2}{2\sigma^2(t)} \right) = 4 \left(\int_0^{L/2} \mathrm{d}x \exp\left(\frac{x^2}{2\sigma^2(t)} \right) \right)^2 \\ &= 2\pi\sigma^2(t) \left(\operatorname{erf}\left(\frac{L}{2\sqrt{2}\sigma(t)} \right) \right)^2 \end{aligned} \tag{4.2}$$

where at the first step I have used the independence between x and y and the Gaussian being an even function, and at the second step the definition of error function $\int_0^x \exp(-ax'^2)\mathrm{d}x' = \frac{\sqrt{\pi}}{2\sqrt{a}} \operatorname{erf}(\sqrt{a}x)$ and $\operatorname{erf}(x) = -\operatorname{erf}(-x)$. Equation 4.2 shows that when $L \ll \sigma \propto \sqrt{Dt}$, then $\operatorname{erf}(x) \to 1$ and $Z(t) \to 2\pi\sigma^2$, meaning that the profile of P^i_{free} is well described by the Gaussian approximation. Considering that $\operatorname{erf}(2)^2 \approx 0.991$, if $\sqrt{Dt} < \frac{L}{8}$ the relative error is $\frac{|Z(t) - Z_G|}{Z_G} < 1\%$. Figure 4.2 shows the Gaussian functions for different values of D at fixed t and vice versa. Considering the values of N and ρ discussed in section 4.2, the environment size is always $L > 500$, meaning that the constraint on D and t to preserve a good Gaussian approximation is not stringent.

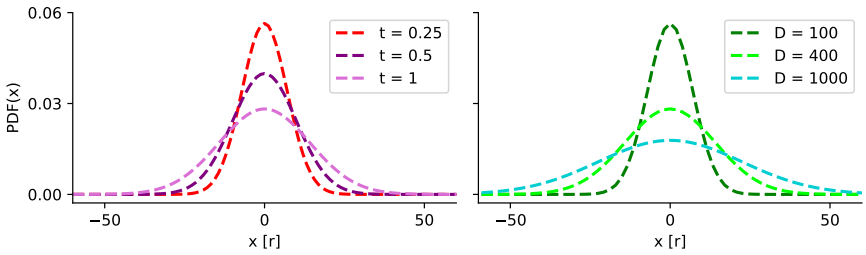


Figure 4.2: **Free Gaussian diffusion in 1D.** *Marginalized probability density function in Equation 4.1, parameterized with:* $D = 100\mathrm{r}^2/\mathrm{d}$ *in the left panel;* $t = 0.25\mathrm{d}$ *in the right panel;* $x_0 = 0$ *in both panels.*

When the motion gets constrained by the attractive potential, analytical computations become more involved, indeed the Fokker-Planck equation can be hardly solved analytically for arbitrary potential at arbitrary times. Though, the existence of the potential enables the possibility for the motion to reach an equilibrium state, instead of diffusing indefinitely in the whole environment. The equilibrium state is usually convenient to study, because the probability density $P(\vec{x}, t) \to P_{\mathrm{eq}}(\vec{x})$ becomes stationary and the left-hand-side of the Fokker-Planck equation vanishes. For the equilibrium state to be realized, the potential must fulfil two constraints. First, the tails must not be exponentially suppressed, so that the containing force is strong enough to counteract the diffusive tendency. Second, the potential itself must have an equilibrium state to converge to. The potential in Equation 3.4 cannot fulfil the second constraint, as the amplitude parameter depends on time in a cyclic fashion. Though, for this analytical calculation it is convenient to impose a fixed value $A(t) = \mathrm{const}$ and study the equilibrium state. With these premises, the equilibrium solution for the agent's location probability density takes the form

$$P^i_{\mathrm{eq}}(\vec{x}) = \frac{1}{Z} \exp\left(-\frac{V^i(\vec{x})}{D}\right) \tag{4.3}$$

where Z is the normalization factor that can be analytically computed only for particular choices of V, while in general is numerically obtained as $Z = \iint_{\mathcal{D}} P^i_{\mathrm{eq}}(\vec{x})\mathrm{d}\vec{x}$. The details of this calculation are shown in Appendix A. Now, if one wants to reintroduce the periodic temporal dependency of the potential, specifically through the Equation 3.6, then the equilibrium cannot be reached. Indeed, two different equilibrium solutions could be separately computed $P^i_{\mathrm{eq,night}}$, and $P^i_{\mathrm{eq,day}}$, and the probability density $P(\vec{x}, t)$ oscillates between these two extremes. Hence, the analytical calculation can be useful to constrain the parameters of motion, even if it cannot fully capture its temporal evolution.

The analytical calculation carried out above allows also to open another work branch. When all the probability densities $P^i(\vec{x}, t)$ are computed for all the agents in the system (and I recall they only differ for the center), it is possible to compute the contact probability between every pair of agents $P^c_{ij}(t)$. This can be obtained

with the spatial convolution

$$P_{ij}^{c}(t) = \iint_{-L/2}^{L/2} \int_{x_1-r}^{x_1+r} \int_{x_2-\sqrt{r^2-(x_1'-x_1)^2}}^{x_2+\sqrt{r^2-(x_1'-x_1)^2}} \mathrm{d}\vec{x}\mathrm{d}\vec{x}' P^i(\vec{x},t)P^j(\vec{x}',t) \tag{4.4}$$

This integral can only be solved numerically and the computation becomes rather expensive because it must be carried out for all the $N(N-1)$ pairs $\{i,j\}$ in each realization of the system. It is much more convenient to introduce a second variable h representing the distances between the centers of the distributions, i.e. the home positions, and numerically compute $P^c(h,t) = \iint_{\mathcal{D}} \iint_{\|\vec{x}'-\vec{x}\|<r} \mathrm{d}\vec{x}\mathrm{d}\vec{x}' P(\vec{x},t)P^*(\vec{x}',t)$, with $P^*(\vec{x},t) \coloneqq P(\vec{x}-h\hat{e}_1,t)$. The bivariate quantity $P^c(h,t)$ is very expensive to compute but it can later be used for all the pairs $\{i,j\}$ in all the realizations of the system, interpolating the numerical estimation of $P^c(h,t)$ with $h = \left\|x_0^i - x_0^j\right\|$. Unfortunately, this approach can only be useful when the full $P(\vec{x},t)$ is known (at least numerically), which does not happen to be the case for a potential shaped as in Equations 3.4 and 3.6.

4.4 Mobility parameters

The parameters of mobility are the ones that determine the shape of the potential and its amplitude, together with the diffusion coefficient that determines the stochastic contribute. The exploratory phase of the motion is set to be a free Gaussian diffusion, completely unconstrained $A_{\mathrm{day}} = 0$, as shown in Figure 4.3. The only free parameter in the exploratory phase is the explored area per unit time, i.e. the diffusion parameter D. Inverting Equation 3.3 so that

$$D = \frac{\sigma_{\mathrm{free}}^2}{2T} \tag{4.5}$$

the diffusion parameter can be constrained by imposing an average range σ_{free} explored within a given time window T. This consideration becomes handy because the scales of distances and time are easier to understand than the scale of the diffusion constant. The value used for most of the simulations is $D = 100\,\mathrm{r}^2/\mathrm{d}$, obtained fixing a $\sigma_{\mathrm{free}} = 10\,\mathrm{r}$ in $T = 0.5\,\mathrm{d}$. I point out that σ_{free} represents the average projection along one axis (coordinate) of the agent's position with respect to a reference frame centered on its home position, and *not the average distance* from home. The average distance from home is computed as $\langle r_{\mathrm{free}}\rangle(t) = \iint_{\mathcal{D}} \|\vec{x}-\vec{x}_0\| P(\vec{x},t)\mathrm{d}\vec{x}$, which is conveniently rewritten in polar coordinates with respect to the home position thanks to the rotational symmetry of the probability density in Equation 4.1, so to obtain $\langle r_{\mathrm{free}}\rangle(t) = \sqrt{\frac{\pi}{2}}\sigma_{\mathrm{free}}(t)$.

All the other mobility parameters are related to the attraction potential in Equation 3.4, specifically for what concerns the confinement phase (the night). The coefficients $\alpha, \beta, \gamma, \delta$ have been introduced for the sake of generality, but all the

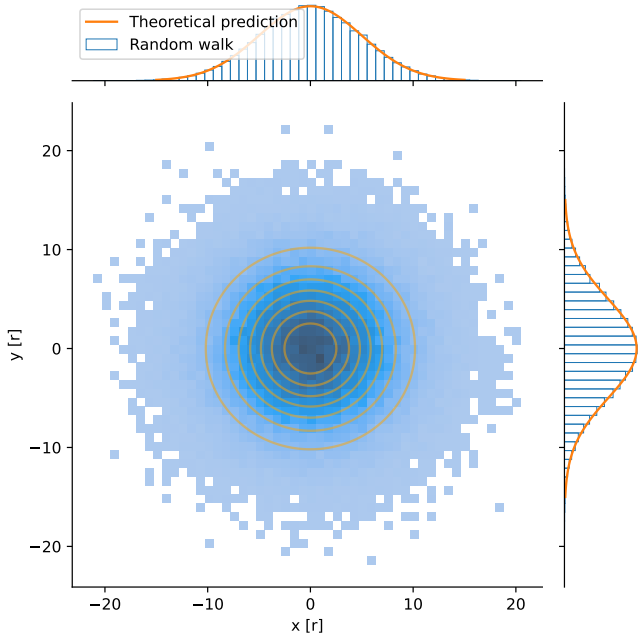


Figure 4.3: **Simulation of the motion of an agent free from confinements.** *The simulation of* 10^5 *unconstrained random walks with* $D = 25\,\mathrm{r}^2/\mathrm{d}$ *is performed for a time* $t = 0.5\,\mathrm{d}$. *The histogram of the final positions is compared to the analytical solution in Equation 4.1. On the sides, the marginal distributions are shown.*

results shown in this work have been obtained with the same set of values. For the most trivial one, β, the value is set to be $\beta = 1$ in order to not alter the spatial scale imposed by the contact range r. $\gamma = 4$ has been set so that the potential well is flatter than for $\gamma = 2$. $\delta = 0.5$ has been chosen so that the potential well is narrow around the home position, allowing for a stringent confinement during the night without needing the potential amplitude to explode. Finally, $\alpha = 1$ is set so that the agent undergoes a constant force when far away from home; values $\alpha > 1$ would have led to a divergence of the gradient and, eventually, meaningless oscillations around the home position; values $\alpha < 1$ would have led to a decreasing force of attraction at large distances, which is somehow opposite to the realistic scenario where people run back home faster when farther away.

The potential amplitude A_{night} is the most delicate parameter to set. The idea is to proceed in a similar way as discussed above for D, i.e. to derive the value from other more intuitive quantities. Assuming to wait long enough so that the equilibrium probability density in Equation 4.3 is reached, the average confinement range σ_{conf} can be fixed as constraint. The spreading σ_{conf}, as in the case of σ_{free}, is the standard deviation of the 1D projection of the position of the agent with respect to the home position. It is then computed as $\sigma^2_{\mathrm{conf}} = \int_{\mathcal{D}_{x_1}} x_1^2 P_{\mathrm{eq}}(x_1)\mathrm{d}x_1$, where $P_{\mathrm{eq}}(x_1)$ is the equilibrium probability density of Equation 4.3 marginalized

over x_2. This leads to the integral equation

$$\begin{aligned}\sigma_{\text{conf}}^2 &= \iint_{\mathcal{D}} x_1^2 \exp\left(-\frac{V(\vec{x})}{D}\right)\mathrm{d}\vec{x} \\ &= \iint_{\mathcal{D}} x_1^2 \exp\left(-\frac{A_{\text{night}}}{D}\left(-0.5 + \left(0.5^4 + \left\|\vec{x} - \vec{x}_0^i\right\|^4\right)^{\frac{1}{4}}\right)\right)\mathrm{d}\vec{x}\end{aligned} \tag{4.6}$$

where in the potential $V(\vec{x})$ the standard values (Table 4.1) of the coefficients $\alpha, \beta, \gamma, \delta$ have been inserted. Thus, the value of A_{night} can be constrained by solving numerically this implicit integral equation. I point out that the choice to marginalize over x_2 and integrate over x_1 could be switched, as the potential is rotationally invariant. The equation above highlights that for fixed σ_{conf} one gets $A_{\text{night}}(D \mid \sigma_{\text{conf}}) \propto D$, whereas there is no analytical closed-form relation for $A_{\text{night}}(\sigma_{\text{conf}} \mid D)$. The numerical solution to Equation 4.6 is shown in Figure 4.4 for different values of the diffusion parameter. The plots shows a profile that

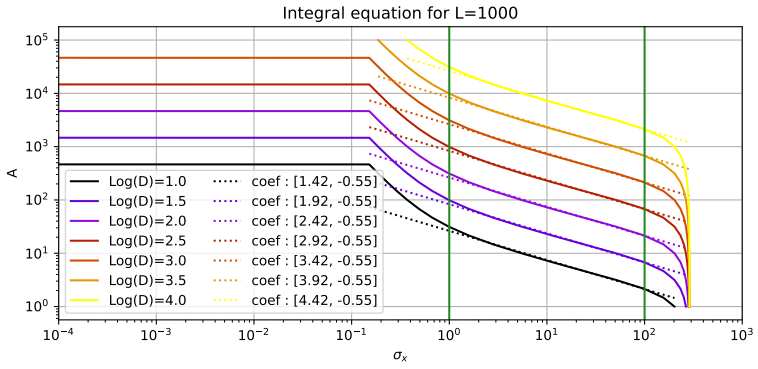


Figure 4.4: **Numerical solution of the integral equation for the amplitude parameter.** *The numerical solutions to Equation 4.6 are shown for various D. The green vertical lines delimit the range of linear behaviour of the solutions. The dotted lines are the linear fits, which coefficients are shown in the legend.*

is approximately linear in the central range, which luckily happens to be the range of interest, so a linear fit is performed to look for the relation $\log(A_{\text{night}}) = c_0(D) + c_1(D)\log(\sigma_{\text{conf}})$. Interestingly, $c_1 \approx -0.55$ independently on D, and c_0 appears to grow as $c_0 \approx \log D + 0.42$. Exponentiating both sides of the equation, one gets

$$A_{\text{night}} \approx 10^{0.42} D\, \sigma_{\text{conf}}^{-0.55} \tag{4.7}$$

which is consistent to the former observation $A_{\text{night}}(D \mid \sigma_{\text{conf}}) \propto D$.

4.5 Discretized displacement

The Euler-Maruyama method relies on discretization to numerically solve the stochastic differential equation of motion. Considerations about the size of time-step may be important to avoid discretization issues and other uncontrolled effects.

I start considering the unconstrained motion of the exploratory phase, imposing $V(\vec{x}, t) = 0$ in Equation 3.2. First of all, I point out that the noise term ΔW_t depends on time only when a specific realization of the stochastic process is considered. The Wiener process is not time-dependent per se, it is stationary and depends only on the time-discretization step, so the subscript can be dropped when referring to the random variable. With a reasonable abuse of notation, here I will refer to this random variable as $\Delta W_i \sim \mathcal{N}(0, \Delta t)$, where the subscript i refers to the i-th component of the vector. With no potential, the spatial displacement $\Delta\vec{x} = \sqrt{2D}\Delta\vec{W}$ is also random variable that does not depend on time, and it is isotropically distributed so that $\Delta x_i \sim \mathcal{N}(0, 2D\Delta t)$. The interesting quantity here are not the single components of the displacement, rather its modulus, i.e. the spatial step-size of the walk. The two components of the displacement are independent and normally distributed with same variance, so the squared norm is distributed according to a scaled chi-square distribution with two degree of freedom $\|\Delta\vec{x}\|^2 \frac{1}{2D\Delta t} \sim \chi^2(2)$. The modulus $\|\Delta\vec{x}\|\sqrt{\frac{1}{2D\Delta t}} \sim \chi(2)$ is distributed according to the Rayleigh distribution, a particular case of the chi distribution with two degrees of freedom. For this distribution, the mean is known to be $\langle\chi(2)\rangle = \sigma\sqrt{\frac{\pi}{2}}$ and the median to be $\mathrm{Med}(\chi(2)) = \sigma\sqrt{2\log(2)}$, where $\sigma = \sqrt{2D\Delta t}$ is the scale parameter. Thus, the mean and median displacements are

$$\langle\|\Delta\vec{x}\|\rangle = \sqrt{\pi D\Delta t} \qquad , \qquad \mathrm{Med}\left(\|\Delta\vec{x}\|\right) = \sqrt{4\log(2)D\Delta t} \tag{4.8}$$

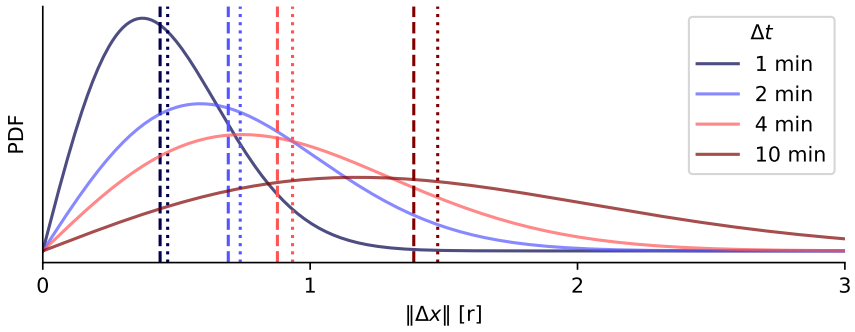


Figure 4.5: **Distribution of the displacement modulus for an unconstrained walk.** *The probability distributions $P(\|\Delta\vec{x}\|)$ are shown for different values of the discretization time, with the D fixed to the standard value. Mean and median values of the distributions are computed as in Equation 4.8.*

Figure 4.5 shows the distribution of the displacement magnitude for different values of the diffusion coefficient. The goal of this analysis is to set the Δt so that the median displacement has the same order of magnitude of the contact range r, for two main reasons. The first emerges when considering two agents that have

intercepting trajectories: if the discretization is too large, the agents would likely be too far apart to make a contact either at the step before and at the step after the intersection, implying that no contact would be recorded. The smaller is the discretization, the harder it becomes to end up in this situation. The second reason is that it is impossible to confine the agents around their home positions at distances smaller than the stochastic jumps. To contain these problems, the time-discretization step is set to $\Delta t = 1\,\text{min}$, balancing the trade-off with the increasing computation time needed for the simulations.

Performing the same analysis becomes much harder when dealing with the confinement phase, i.e. when the agent undergoes the attraction of the potential. The displacement is the vectorial sum of the stochastic and the deterministic terms in Equation 3.2, so its square modulus is

$$\|\Delta\vec{x}\|^2 = (\|\nabla V\|\Delta t)^2 + 2D\left\|\Delta\vec{W}\right\|^2 + 2\sqrt{2D}\Delta t\|\nabla V\|\left\|\Delta\vec{W}\right\|\cos\theta$$

where θ is the angle formed by ∇V and $\Delta\vec{W}$. It is evident that $\|\Delta\vec{x}\|$ is a random variable that cannot follow a known distribution, as neither $\|\Delta\vec{x}\|^2$ does. Analyzing the right-hand-side in the equation, the first term is the pure deterministic contribution, which is position dependent; the second term is the pure stochastic contribution, and it is the same random variable discussed for the free walk distributed as $\sim 2D\chi^2(2)$; the third term is the mixed contribution, stochastic and deterministic, and it is a random variable distributed as $\sim \mathcal{N}\left(0, 8D(\Delta t\|\nabla V\|)^2\right)$. This last claim is motivated observing that a chi-squared distributed random variable multiplied by the cosine of a uniformly distributed one represents the 1D projection of a bivariate uncorrelated Gaussian random variable. The relative contributions of the three terms is spatially dependent, because the strenght of the gradient modulates them differently. Given this comment, $\|\Delta\vec{x}\|^2$ is a random variable which distribution is complex to compute (even numerically) and hard to plot because dependent on all the variables and parameters of the potential. Though, considering that in my mobility model the potential plays a role only in the transition phase between the exploration and the confinement phase, meaning when the walker is returning home, the uncontrolled effects on the contact network are considered negligible.

A last interesting insight comes when dividing Equation 4.8 by Δt, i.e. when the average *velocity* of the agents is computed. Indeed, the velocity appears to be *dependent* on the time-discretization step Δt. Despite this fact might appear non-trivial, it is due to setting the diffusion as a constant of motion. To better explain it, I resort to an analogy with a thermodynamical system, such as the diffusion in an ideal gas. According to the kinetic gas theory, the diffusion coefficient $D = \frac{1}{3}\ell v_T$ is related to the thermal velocity v_T and the mean free path ℓ. The mean free path is the average displacement without change of direction, meaning $\ell = \langle\Delta\vec{x}\rangle \propto \sqrt{\Delta t}$. Since the diffusion coefficient is a constant of motion, $v_T \propto \ell^{-1}$, so the thermal velocity scales inversely with respect to the squared discretization

time, in agreement with our results. The physical intuition may come recalling that in kinetic gas theory $\ell \propto \frac{1}{n}$, with n the number density of the gas: a smaller time-discretization physically translates into a denser gas, where collisions are more frequent. To ensure the same diffusion length within a denser gas, the particles in the system must have a higher mobility, i.e. a higher thermal velocity, which consequently cannot be a constant of motion.

Chapter 5

Results and discussion

The two previous chapters presented the model and the meaning of its key parameters. In this chapter I will illustrate the most interesting results obtained applying the model through numerical simulations. In section 5.1, I will start by presenting the characteristics of the contact network produced with the mobility model. Then, in section 5.2, I will show the growth of the prevalence of the infectious disease, which turns out to be unexpectedly super-quadratic. Hypothesis on the source of this behaviour will be inspected and discussed with the help of the network metrics introduced in section 5.1. The key role of the mobility with its spatial and temporal features will be highlighted in section 5.3. Finally, in section 5.4, I will discuss the absorbing phase transition of the disease spreading, observing that the spatial setting together with the temporal cycle are sufficient for the model to break the universality classes of directed percolation and mean-field.

5.1 Contact network characterization

The model used in this work for individual mobility has been largely discussed in chapter 4 with its essential ingredients: the circadian alternation between an exploratory phase and a confinement phase. When applied to all the agents in the environment, it translates into a collective cyclic behaviour which has implications on the contacts established between agents. I will not resort to any comparison to real datasets because this work does not aim to reproduce reality in its full complexity. This analysis has the aim to clearly understand the consequences of the mobility principles onto the contact network, which are good to discuss before tackling the epidemic spreading.

First of all, at each time-step the contact network is generated as a pure random geometric graph, so, as explained in subsection 2.2.2, the degree distribution is a binomial. Though, accounting for the standard values of the parameters used in the simulations (Table 4.1), the average degree at each time-step is $\langle k \rangle \approx 8 \times 10^{-2}$,

computed according to Equation 2.2. The extremely low connectivity makes it useless to reason on single snapshots of the temporal network, and rather proceed through temporal aggregations to look for interesting observables. To do so, the temporal coordinate in the timed edge list is binned, and all the repeated contacts between the same two nodes in the same bin are discarded. I point out that, on one side, this procedure destroys the information about repeated contacts, at least at time-scales smaller than the binning, but on the other side, it highlights the information about the number of different agents that have been in contact. In Figure 5.1 the degree distributions for different aggregation windows are shown. The temporal aggregation has the obvious consequence to increase the degrees

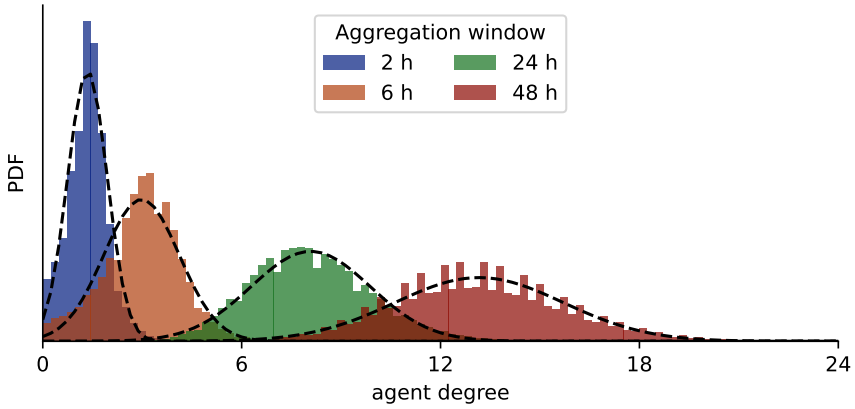


Figure 5.1: **Degree distributions for different aggregation windows.** *The distributions are obtained aggregating the temporal contact networks in time-windows of specified amplitude. The distributions run over each agent, over each window, over each realization of a standard system with* $N = 10^4$*. The dashed lines represent the Gaussian distribution fitted on the data using the maximum likelihood estimation.*

of the agents, because it sums all the unique contacts that each agent has. The agent degree in each snapshot is a Poisson random variable, so, for the central limit theorem, summing over different snapshots the degree should become a random variable distributed in a Gaussian fashion. This was true if all contacts were unique, which does not happen to be the case, as agents that get in contact are also close to each other, and are likely to meet again. The result is a degree distribution that is close to a Gaussian, but not exactly Gaussian, especially for small aggregation windows. This might appear illogical, considering that the smaller is the aggregation window, the less duplicate contacts are removed, but the real reason is due to the cyclic behaviour of the agents' mobility. Since the cycle period is 1 d, all aggregations on larger time-windows smooth out the difference between the exploration and the confinement phase. The net effect is that more duplicate contacts are discarded, but in a more uniform way, so that the distribution preserves its shape. For smaller aggregation windows, the distributions are distorted because averaged over different phases.

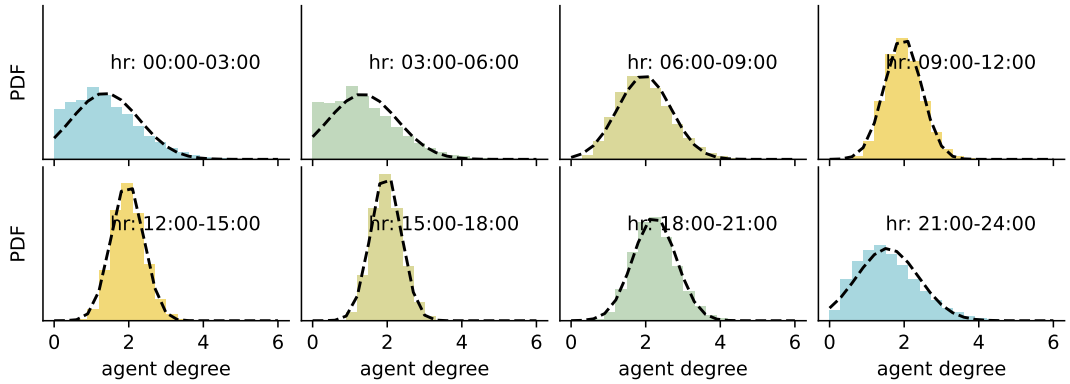
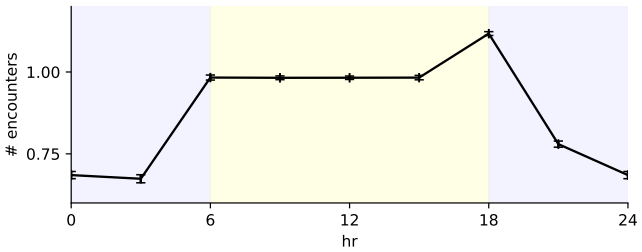


Figure 5.2: **Degree distributions for different clock-times.** *The distributions are obtained aggregating the temporal contact network in* 3 h *windows. The distributions run over each agent, over each realization of a standard system with* $N = 10^4$*. The dashed lines represent the Gaussian distribution fitted on the data using the maximum likelihood estimation. The color coding refers to the daily phases, associated to the night in blue and the day in yellow.*

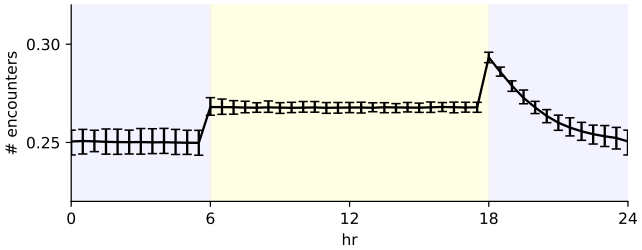
To understand the difference between the degree distributions obtained during the exploratory and during the confinement phases, I plot them for different clock-times in Figure 5.2. The degree distributions are here aggregated over a time window of the same amplitude, but are averaged between different days only for corresponding clock-times. The linear combination of these distributions would reproduce the ones in Figure 5.1, so the non-Gaussian shape can be ascribed to the agents' activity during the confinement phase. Indeed, the good matching of the fit with the exploration distributions suggests that during the day the majority of the contacts that each agent has is unique, or at least they are homogeneously repeated. On the contrary, at night there is a dominance of duplicated contacts, that alters the shape of the distribution for low degrees. This is indeed a wanted effect, that reproduces the higher social activity during the light hours compared to the night ones, when people tend to isolate or preserve already existing contacts.

Another point of view of the same phenomenon is provided by the plot in Figure 5.3a. The points represent the number of unique encounters per capita, which corresponds to half the average degree. The profile clearly shows the difference in social activity between the night hours and the day hours. It also shows an interesting pattern, that is already visible in Figure 5.2 but less evident: an increase in the later stage of the exploration phase and a slow decay in the early stage of the confinement phase. To better inspect this effect, I resort to Figure 5.3b that presents a temporally fine-grained version of the same plot.

The smaller aggregation window allows to observe that the increase in the number of encounters happens exactly at the turning point between the day and the night. This is probably due to the fact that the transition between the two phases happens



(a) *Time window amplitude* 3 h.



(b) *Time window amplitude* 30 min.

Figure 5.3: **Average number of encounters during a full day.** *Each point represents the mean number of encounters per agent that are happening in the system within a time-window. The average is computed over each day of simulation and over each realization of the standard system with* $N = 10^4$*. The error bars describe the central* 95% *confidence interval. The background color represents the mobility phases, confinement in blue and exploration in yellow.*

with a sharp activation of the attraction potential, that drives the agents back home. The potential adds the deterministic contribution to the velocity of the agents, which turn their motion from a random exploration to a directed walk. This does not change the average number of contacts in each time-step, which I recall remain a Poisson process, but it decreases the chance that the contacts are repeated, as the agents quickly leave the meeting location right after the contact. The slow decrease of the number of encounters is due to the coexistence of some agents that are still returning home, contacting several different agents on the way, and some agents that have already reached their home position, lowering their social activity.

At night one might expect to see no encounters between agents, since they are confined at their home position. Though, the random placement of the home positions in the environment creates heterogeneous density regions, for which the home positions can be closer to each other than the contact range. I will refer to the kind of structures so formed as *households*, as they recall what happens in real scenarios where different people share the same house or building. This configuration has consequences on the epidemic spreading, as the probability that an infectious agent infects a susceptible housemate is much higher than for any

other agent, because of the extremely repeated contact happening at night.

Another interesting characteristic of the contact network is the temporal distance between the agents, because it represents the minimum time needed by any kind of information (in this case, the pathogen) to travel from one agents to the other. The latency would be the perfect metric to measure this quantity, but it is expensive to compute and very unstable with respect to stochastic effects, as mentioned in subsection 2.1.5. For this reason, I resort to another metric to understand the spreading velocity of the information on the network: the temporal reachability. It is computed seeding the information on one agent and spreading it through all the contacts that involved agents who already possess the information. It is possible to introduce a *waiting time* between the instant in which the node is reached by the information and the instant in which it becomes able to spread it to other agents. The fraction of agents reached by the information at a certain time t after the seeding is plotted in Figure 5.4. I point out that this metric exactly reproduces the

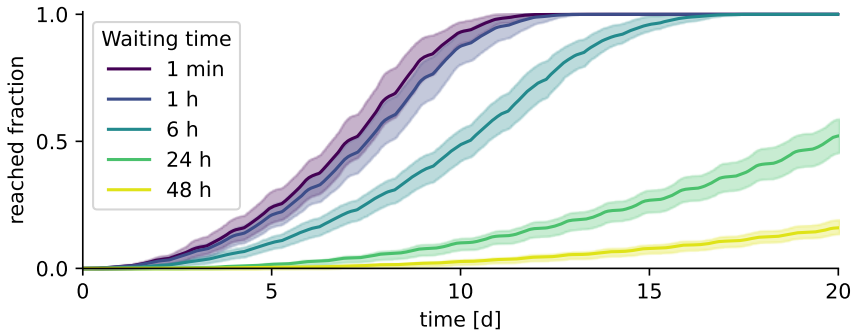


Figure 5.4: **Reachability fraction of some information spread on the contact network.** *The curves show the fraction of nodes reached by some information spreading onto the contact network for different waiting times. The solid lines are obtained as the average over different realizations of the standard system with* $N = 10^4$*, and the colored stripes represent the central* 95% *confidence interval.*

epidemic spreading of a SEI model in which the waiting time corresponds to the latent period τ_{lat} and each contact between a susceptible and an infectious agents becomes an infection, meaning that $p_{\text{inf}} = 1$. In other words, the reachability curve represents the prevalence growth of the fastest possible epidemic with fixed τ_{lat} on the contact network. This metric is useful as it highlights the spreading speed limit imposed by the structure of the contact network. Observing the curves in Figure 5.4, the first aspect that emerges is the oscillatory trend. The period of the oscillation coincides with the period of the mobility cycle 1 d. The interpretation is that the reachability grows faster during the exploration phase, because more distinct contacts are established by the agents, while at night the growth slows down because the contacts are mainly repeated. This observation has been already presented in the previous paragraphs, but Figure 5.4 highlights the consequences on the spreading process.

5.2 Epidemic grows super-quadratically

Once the contact network is characterized and its relevant features are under control, the focus can be shifted towards the real goal of this project: studying the effects of the mobility on the epidemic spreading. In particular, one fundamental quantity to observe is the prevalence, i.e. the number of agents in the infectious state $I(t)$, as it provides information about the amount of agents that are currently able to spread the disease. The epidemic parameters p_{inf}, τ_{lat} and τ_{inf} have an effect on the scales of the $I(t)$ curve, specifically, they all impact the steepness, but only p_{inf} and τ_{inf} have effect on the saturation value. At the beginning, the seeds of the epidemic are possessed only by few agents, so the early stages are the moment in which the epidemic is most vulnerable to stochastic fluctuations. If the epidemic survives the early stages, then it unlikely gets extinct at later times. A survived epidemic then converges towards the equilibrium condition, that is reached when the rate of recoveries equals, on average, the rate of infections. Figure 5.5 shows some epidemic curves for different infectiousness parameter p_{inf}. The plot immediately shows that

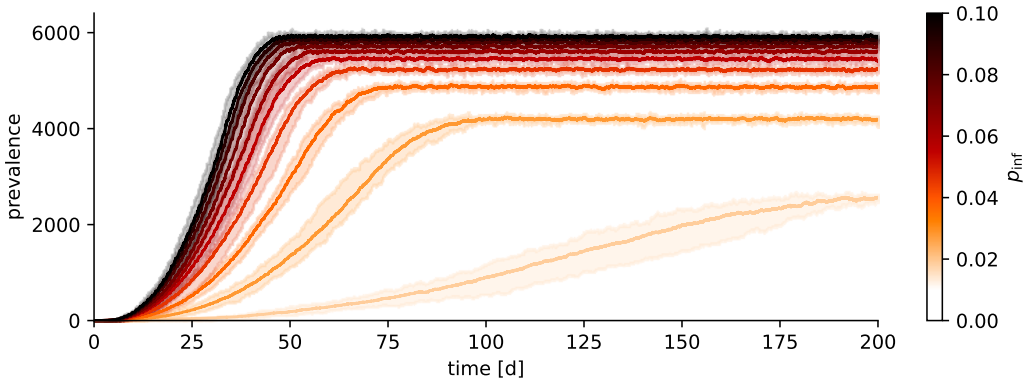


Figure 5.5: **Prevalence curves for different probabilities of infection.** *Each solid line represents the time-wise average $\langle I \rangle(t)$ of different realizations of a standard system with $N = 10^4$, and the colored stripes the 95% credibility interval.*

the higher the infectiousness, the easier is the transmission from agent to agent, so the spreading grows faster and reaches higher prevalence values at equilibrium. Moreover, epidemics that are closer to the critical absorbing threshold are affected by larger stochastic fluctuations, because the "spreading force" is more effectively counteracted by stochastic sparsity in the contact network.

The temporal and prevalence scales of the epidemic curve are not the only important quantities to observe. When facing an epidemic spreading, it is important to understand what is the functional form of growth that one has to expect. The functional form is not dependent on the epidemic parameters, it is only dependent on the characteristics of the topology onto which the spreading happens, i.e. the contact network. Indeed, as discussed in section 2.3, the prevalence growth is

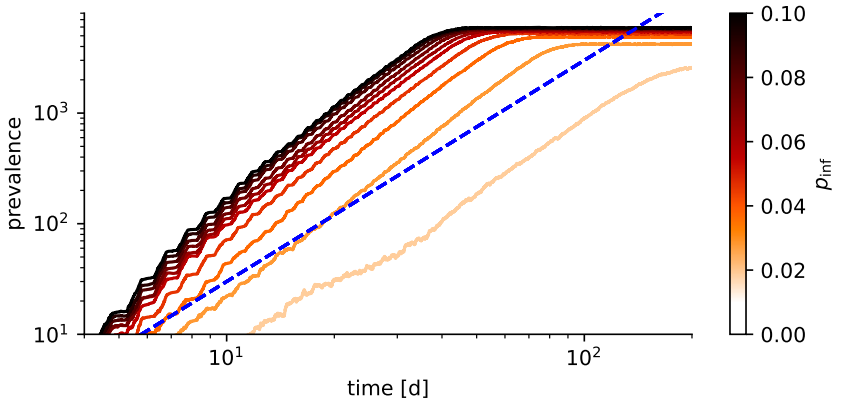


Figure 5.6: **Prevalence curves for different probabilities of infection in a log-log plot.** *Each solid line represents the time-wise average $\langle I \rangle(t)$ of different realizations of a standard system with $N = 10^4$. The dashed blue line represents the quadratic growth in the log-log plane.*

exponential in an Erdős-Rényi network and polynomial on a lattice (quadratic in 2D), and this difference is due to the discrete Euclidean topology induced on the lattice by the metric of the Euclidean space. Since the contact network is generated in a spatial setting, one might expect the prevalence growth to be similar to the lattice case, but since contacts are built between mobile agents that wander randomly in the environment, one might expect it to be similar to the mean-field case. Spotting the functional form of the prevalence curve from Figure 5.5 is rather hard, but luckily log-log plots come in handy when looking for power-laws, which appear as straight lines with angular coefficient equal to the exponent of the power-law. Figure 5.6 presents the same epidemic curves shown in Figure 5.5 in the log-log visualization. The first observation is that the prevalence present the oscillatory trend already noted in Figure 5.4 that is caused by the day/night alternation. Secondly, all curves but the one with lowest p_{inf} (which is the closest to the critical threshold) appear to be straight and parallel, implying power-law growth with similar exponents. The third, most important, observation comes from the comparison between the prevalence curves and the dashed line, which represents a quadratic power-law, i.e. the growth observed on a lattice model. It is quite evident how different is the slope of the quadratic power-law with respect to all the simulations performed, suggesting that the prevalence growth observed in these systems is for some reason super-quadratic. To measure the growth exponent I resort to a linear least-square fit in the log-log plane, which produces the values shown in Table 5.1. The fitting ranges are chosen looking at the log-log plots and choosing the largest range in which the curve appears linear. The fitted values are quite sensible to the choice of the fitting range, which must be chosen wisely. Even though a more solid framework to compute the growth exponents and their relative uncertainties would improve the reliability of this result, Figure 5.6 suggests that

p_{inf}	exponent	fit range
0.01	2.32	$[50, 110]$
0.02	2.46	$[25, 55]$
0.03	2.49	$[20, 40]$
0.04	2.36	$[15, 40]$
0.05	2.40	$[12, 35]$
0.06	2.47	$[11, 32]$
0.07	2.48	$[10, 31]$
0.08	2.37	$[9, 30]$
0.09	2.37	$[9, 30]$
0.1	2.38	$[9, 30]$

Table 5.1: **The fit results show super-quadratic exponents.** *The values obtained from the fitting routine are rounded to three significant figures and the uncertainties are omitted because extremely small. The fitting ranges are arbitrarily chosen observing the linear regime in the log-log plots.*

the discrepancy between the simulated systems and the theory about 2D lattices is too large to be only due to the poorness of the fitting procedure.

To cross-check the values obtained through the least-square fit, I follow also another pathway to compute the growth exponents. The idea is to find the slope $\alpha(t)$ of the tangent to the prevalence curve in the log-log plot, and letting it vary in time, so to get rid of the problem to choose the fitting ranges. I call this quantity the *instantaneous exponent*, which, considering $\mathcal{I} \coloneqq \log(I)$ and $\tau \coloneqq \log(t)$, is defined as

$$\alpha(t) \coloneqq \frac{\mathrm{d}\mathcal{I}}{\mathrm{d}\tau} = \frac{t}{I}\frac{\mathrm{d}I}{\mathrm{d}t} \tag{5.1}$$

To better understand the meaning of this quantity, one can focus on the specific value $\alpha^* \coloneqq \alpha(t^*)$ computed from data. Rewriting Equation 5.1 in its differential form in the neighbourhood of t^* $\alpha^* \frac{\mathrm{d}t}{t^*} = \frac{\mathrm{d}I}{I^*}$ and integrating both sides, one obtains $\log(I^*) = \alpha^* \log(t^*) + c$, where c is the integration constant. Exponentiating both terms, one gets $I^* = ct^{*\alpha^*}$, which holds in the neighbourhood of t^*. This means that computing the instantaneous exponent corresponds to finding the exponent of a power-law in the form $f(t) = ct^{\alpha}$ which is locally fitting the actual prevalence curve $I(t)$. The advantage of this procedure is that is parameter free, and no fitting range should be arbitrarily provided; it is only necessary to filter the data, so to reduce the stochastic noise and improve the readability of the results. In Appendix B I provide some plots that confirm the stability of this method, proven with the reconstruction of synthetic data that well match the original ones. In Figure 5.7 I plot the values, numerically computed after averaging the prevalence curve over a time window of 1 d. Looking at all the different simulations, one can notice that the instantaneous exponent are definitely > 2 before the epidemic reaches the equilibrium, and then suddenly drop to 0. The profile of $\alpha(t)$ seems to grow even

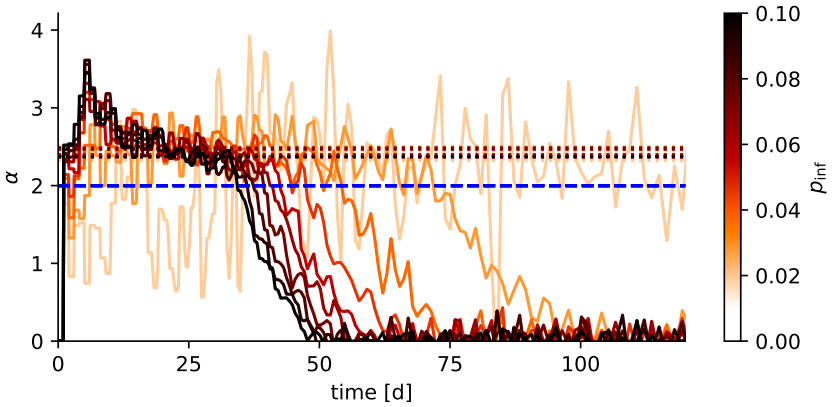


Figure 5.7: **Instantaneous exponents of standard simulations** *Each solid line represents the instantaneous exponent numerically computed as in Equation 5.1 from the data $\langle I \rangle(t)$, obtained averaging over different realizations of a standard system with $N = 10^4$. The dashed blue line represents the instantaneous exponent of a perfectly quadratic curve. The dotted lines represent the values of the exponents obtained with the linear fit.*

above 3 in the early stages, at least for the most infectious diseases, and then flats out towards a stable value. To understand what is the $\alpha(t)$ profile associated to each functional form, it is possible to recall Equation 5.1. For an exponential curve $I(t) = c\, e^{kt}$, the instantaneous exponent becomes $\alpha(t) = \frac{t}{c\, e^{kt}}\, c\, k\, e^{kt} = kt$, i.e. $\alpha(t)$ grows linearly in time with some angular coefficient that depends on the epidemic parameters. While for a power-law curve $I(t) = c\, t^k$, the instantaneous exponent becomes $\alpha(t) = \frac{t}{c\, t^k}\, c\, k\, t^{k-1} = k$, i.e. $\alpha(t)$ is constant. With this interpretation, we can explain the very initial peak as an initial exponential growth proper of the mean-field models, but then the growth slows down to a power-law behaviour that is not purely quadratic as expected from the spatial models.

The causes of the observation of the super-quadratic growth deserve a further discussion. There are three main hypothesis that might explain this phenomenon: (i) the smallness of the system, (ii) the existence of disease-reservoirs, (iii) the existence of mobility. The first hypothesis to address is that the growth would be exponential as for mean-field models, but that the system considered is too small and saturation effects kick in before the exponential pattern is able to emerge. To evaluate this hypothesis, I increase the number of agents from 10^4 to 10^5 (value that becomes the standard for simulations) while preserving all the other parameter values. The second hypothesis is that the growth would be quadratic as for lattice models, but the existence of households creates a "disease-reservoir" effect that further boosts the epidemic growth. The "disease-reservoir" effect is given when two or more agents that share the same household go through multiple alternated reinfections: consider agent that gets infected during the day and returns home; after τ_{lat} it become infectious and infects its housemates with extremely high

probability, since they make lots of contacts at night; after τ_{inf}, it recovers and becomes susceptible again, but, in the meantime, the housemates have become infectious and reinfect the agent. The net effect on the system is that the household becomes a "disease reservoir" that enhances the spreading of the disease in the neighbourhood. To evaluate this hypothesis, I simulate a system with different epidemic models, an SEI and an SEIR, which, through the elimination of the recovery and the introduction of the immunity, completely neutralize any "disease reservoir" effect. The third hypothesis to address is that the growth would be quadratic as for lattice models, but the mobility of the agents introduces an effective fractal dimension to the diffusion of the wavefront of infection. To evaluate this hypothesis, I simulate a system in which all agents are static in their home positions, removing any effect due to the mobility. This type of simulation requires some further attention, because the contact network so produced is completely different from the others obtained in this work, as it is a static random geometric graph built on the home positions of the agents. To ensure the possibility for an epidemic to spread in the system, such static network must have a much larger density than the value used for the other simulations. Indeed, if the random geometric graph is undesirable to be connected when mobility is present, it becomes essential when dealing with this static network. Recalling Equation 2.2, the percolation threshold for a random geometric graph is at $\rho_c \approx 1.43\,\mathrm{r}^{-2}$, and I set the agents' density value at $\rho = 2\,\mathrm{r}^{-2}$ so that the giant component of the static random geometric graphs includes almost all the agents in the system.

Figure 5.8 shows the comparison of the prevalence curves and the instantaneous exponents of all the different models mentioned in the previous paragraph, with the aim to understand what are the actual causes of the observed super-quadratic growth. The system's smallness hypothesis can be checked by comparing the "SEIS" and the "SEIS-small" curves. It is straightforward to notice that the two almost perfectly overlap, but for the fact that the epidemic grows for longer time in the larger system. Both the prevalences and the instantaneous exponents present the same behaviour, so it can be excluded that saturation effects are preventing the spreading to show its exponential increment. Indeed, the growth can be assumed to be a power-law, so this model does not present mean-field characteristics. To check for the disease-reservoir hypothesis, one can compare the "SEIS", "SEI" and "SEIR" curves. I recall, as discussed in subsection 2.3.2, that the growth on spatial networks is quadratic for SEIS and SEI and linear for SEIR. For this reason, SEIR is not directly comparable to SEIS and SEI, but all of them are comparable to their reference exponent for a spatial model. SEIS and SEI show a negligible difference for what regards the exponent growth, except for the early stages where SEI seems to grow exponentially. In general, all these three epidemic models present a steeper prevalence and higher instantaneous exponents compared to their reference, implying that "disease reservoir" effects do not alter the growth exponent. The mobility hypothesis can be checked by comparing the "static" curve with the reference of the spatial model and with the "SEIS" and "SEI" curves.

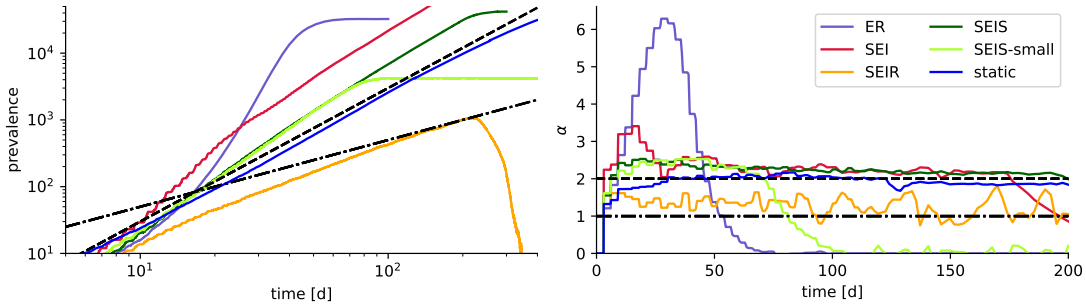


Figure 5.8: **Growth comparison of different network and epidemic models.** *In the left panel, the prevalence curves $\langle I \rangle(t)$ averaged over all the realizations are plotted in the log-log plane to highlight power-law trends. In the right panel, the respective instantaneous exponents are numerically computed as in Equation 5.1. The "SEIS-small" curve is obtained with standard parameters, $N = 10^4$ and $p_{\text{inf}} = 0.02$, and it is one of the curves in Figure 5.5. All the other ones are standard systems with $p_{\text{inf}} = 0.02$. The "static" curve is obtained changing the density parameter to $\rho = 2\,\text{r}^{-2}$. The "ER" curve is obtained simulating the spreading of Erdős-Rényi network. The dashed line emulates a purely quadratic growth, and the dash-dotted line emulates a purely linear one.*

The growth on the static network is almost perfectly quadratic, which implies that mobility is the real cause of the change of the growth exponent.

5.3 Mobility scales and resetting

In section 5.2 I have shown that when the agents move in the environment, the epidemic spreading grows with a larger exponent. It is now interesting to inspect the consequences on the spreading of *how* the agents move. Some research works mentioned in subsection 2.3.4 have shown that a system composed of highly mobile agents presents an exponentially growing prevalence before saturation is reached, meaning that the system is *well-mixed* and behaves as a mean-field model. This exponential increment has not been observed in the simulations presented in section 5.2, possibly because the mobility is too short-ranged and the agents do not mix. So, I perform other simulations extending the range of the exploration of the agents, letting them travel farther from their home position during the day. The mobility during night is kept constrained in a close neighbourhood around the home position, to check if the well-mixing condition can be obtained just extending the exploration range. In Figure 5.9 are shown the prevalence curves and the instantaneous exponents of simulations run with different values of σ_{free}. The prevalence curves exhibit the well discussed super-quadratic growth for small exploration ranges, but it quickly becomes exponential as the range is extended. Indeed, already for $\sigma_{\text{free}} = 25\,\text{r}$, it is possible to observe an exponential behaviour in the early stages of the epidemic, when the instantaneous exponent

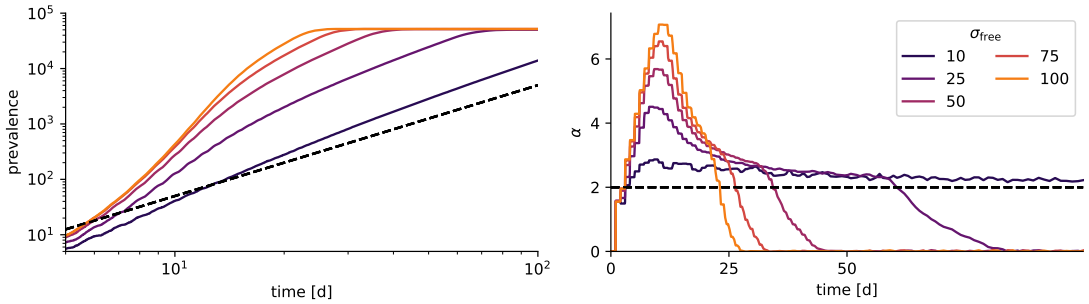


Figure 5.9: **The exploration range affects the epidemic growth.** *In the left panel, the prevalence curves $\langle I \rangle(t)$ are plotted in the log-log plane to highlight the power-law trends. In the right panel, the respective instantaneous coefficients are computed according to Equation 5.1. The simulations are run on standard systems with $p_{\text{inf}} = 0.02$. The dashed line is the quadratic growth of reference for a spatial model.*

grows linearly. After the early stage, the instantaneous exponent reaches a maximum and decreases towards the super-quadratic value. On a much longer timescale, the growth asymptotically decays towards the quadratic regime, which is not reached because equilibrium is established suppressing the exponent down to 0. Three regimes are so observed: exponential in the early stage, super-quadratic in the mature stage and null at equilibrium. For large exploration ranges, the intermediate regime is not even observed because the epidemic spreads so fast that the equilibrium is reached right after the early stage. Looking at the scales of mobility, the average displacement from the home position at the end of the day is $\langle \|\vec{x} - \vec{x}_0\| \rangle = \sqrt{\frac{\pi}{2}\sigma_{\text{free}}^2}$ (from the considerations in section 4.5). This means that for the lowest values in the plot $\langle \|\vec{x} - \vec{x}_0\| \rangle \approx 12.5\,\text{r}$ while for the highest value $\langle \|\vec{x} - \vec{x}_0\| \rangle \approx 125\,\text{r}$. Comparing these values with the environment size, which in these simulation is $L = 2000\,\text{r}$, one understands that agents that explore on average $\frac{1}{16}$ of the environment are mobile enough to mix well the system. This observation can be useful to apply in real scenarios, as it provides a scale above which the spreading of an epidemic can be well represented by a mean-field model. Analogously, if the exploration range is $\approx \frac{1}{160}L$ or below, then spatial models would better fit the spreading process.

The three regimes observed in Figure 5.9 deserve some further attention, particularly the first one that shows the exponential growth. Recalling that exponential is associated to well-mixing, the interesting question is why the system appears to be well-mixed for large mobility ranges. Consider a portion of the environment with size l. If $l \ll \sigma_{\text{free}}$, the agents travel across the whole portion within the exploration phase, so the contacts occur at random locations between random agents. The epidemic spreads randomly in the portion, without any spatial reference, so the portion happens to be well-mixed and the epidemic grows exponentially. Coarse-

graining the size of the portion, when $l \gg \sigma_{\text{free}}$ the agents are not able to mix well in the portion and remain confined in a small area, implying that contacts are more localized and the epidemic spreads spatially. If the mobility range is comparable to the size of the whole environment, this coarse-graining cannot be large enough to appreciate the spatial regime. Hence the second regime in the curves with larger mobility range is absent in Figure 5.9. If a larger system was simulated, this second regime would have been observed also for $\sigma_{\text{free}} = 100\,\text{r}$. This explanation implies that for all exploration ranges there should be an initial exponential spreading, and one might wonder why that is not seen in all the curves. The answer is that exponential growths starts with an instantaneous exponent that is low, and grows linearly in time up to the moment when the epidemic has spread at the spatial scale of exploration. If this scale is reached before infecting enough agents, the exponential growth is still sub-quadratic, so the exponential regime is not observed even if present. Indeed, if one boosts the infection rate, i.e. increases p_{inf}, then all growths should show the first regime of exponential increment. At the same time, for fixed p_{inf}, the larger is the exploration range, the longer lasts the exponential regime, as Figure 5.9 confirms.

Provided that the "level of mixing" of the system is determinant for the functional form of the growth, it is useful to have some metrics that are capable of quantifying this level. One possible choice is the temporal reachability of the contact network, already discussed in section 5.1. Figure 5.10 shows the reachability curves of the simulations considered in Figure 5.9. From the plot, it appears evident how the

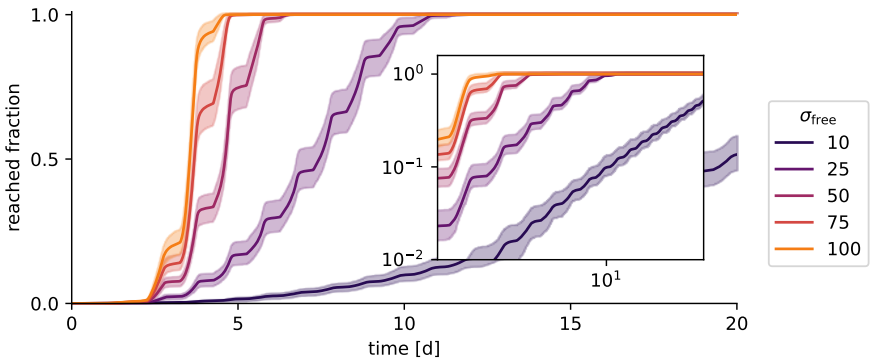


Figure 5.10: **Temporal reachability for different exploration range.** *The solid lines represent the temporal reachability for standard simulations with different exploration ranges. The colored stripes represent the central* 95% *confidence interval over multiple realization of the system. The waiting time for the reachability curves presented is* 1 d*, so to match the latent period of the epidemic* τ_{lat}*.*

timescale needed for the temporal percolation through the whole network change when the exploration range changes. In particular, the increment in reachability becomes extremely steep during the day for systems with high mobility, meaning that in each day the exchange of information (i.e. possibly infections) reaches

the majority of the agents. The effect is that the whole network is reached by the information in less than 5 d, for the highest σ_{free} in the plot. Comparing this timescale to the one proper of the epidemic, it comes that in less than two generations almost any agent in the system has some probability to get infected, almost regardless of the distance of its home position from the source of infection. This metric can be useful to provide information on the level of mixing of the system, but it requires the full knowledge about the contact network, which is impractical in real scenarios.

The previous paragraphs were devoted to the discussion and understanding of the impact on the epidemic spreading of the mobility scale, precisely of the exploration phase. This is the period of the day when most of the unique contacts between agents are established, so the period when the epidemic can mostly diffuse in the space. The interesting question that emerges is what happens if the duration of the exploration phase changes, so I simulate a set of systems with standard parameters except for e, which is the daily fraction of duration of the exploratory phase. The first aspect to understand is how this change affects the characteristics of the contact network. In Figure 5.11 are shown the number of encounters in the systems for different values of e. The extreme values $e = 0\%$ and $e = 100\%$

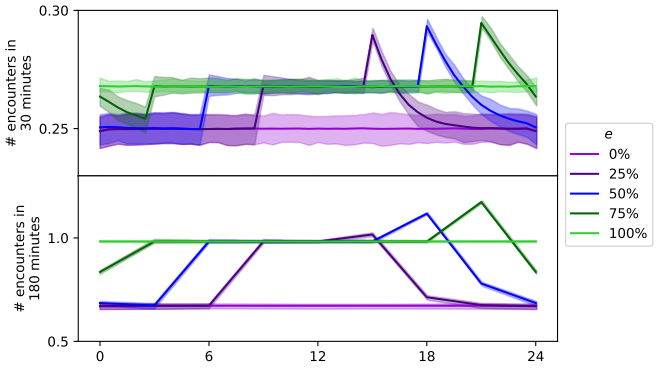


Figure 5.11: **Number of encounters for different duration of the exploration phase.** *Each solid line shows the average number of encounters between agents in the whole system over different realizations. The colored stripes represent the central* 95% *confidence interval. The simulations are obtained with different e and standard parameters.*

correspond to agents that are constantly in the confinement and in the exploration phase respectively. In these settings, no temporal cycles are present and the system becomes translationally invariant in time. The intermediate values present a trivial linear interpolation in between the two extremes. Looking at these curves, one can draw the same observations already discussed for Figure 5.3b. The only noticeable difference between the curves is in the velocity of transition between the exploration regime and the confinement one. Indeed, for $e = 75\%$ the transition is slower than for $e = 25\%$, but this is simply explained by considering that the longer the

exploration phase, the farther is the agent from its home position, and the longer it will take to return there. For $e = 75\%$, the return home is so slow that the encounter curve does not complete its transition towards the confinement level. This can be interpreted as the fact that some of the nodes may still be on the way before the night finishes, and they start the following exploration phase without having reset their position.

The other interesting metric that one could look at is the temporal reachability, shown in Figure 5.12. Comparing these curves with the ones plotted in Figure 5.10,

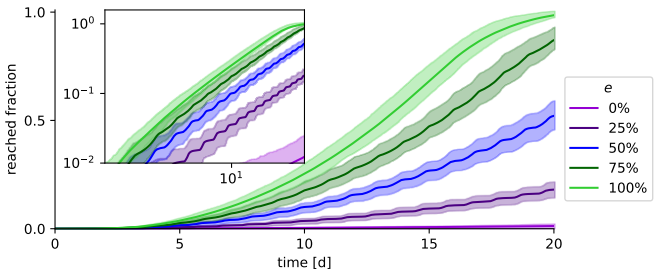


Figure 5.12: **Temporal reachability for different duration of the exploration phase.** *The solid lines represent the temporal reachability for standard simulations with different exploration ranges. The colored stripes represent the central* 95% *confidence interval over multiple realization of the system. The waiting time for the reachability curves presented is* 1 d*, so to match the latent period of the epidemic* τ_{lat}*.*

one observes that the difference between different values of e does not affect much the temporal reachability. The information takes 20 d to spread across a system in which the agents are free to explore for the entire day; if the agents are confined at home for half of the day, in the same time the information can spread in just half of the system. Looking at the curves for $e = 25\%$ and $e = 75\%$ it is evident that the relation between e and the fraction of agents reached by the information is not exactly linear, but the approximating with a linear interpolation does not lead to huge errors. This metric shows smaller differences between the simulations than one might have expected from the previous observations.

Having checked the characteristics of the contact network, it is interesting to inspect what are the effects on the epidemic spreading. Figure 5.13 shows the prevalence and instantaneous exponents for the simulations with different exploration duration. Surprisingly, the simulations obtained with $e = 75\%$ and especially $e = 100\%$ definitely show the exponential regime in the early stages of the spreading. Despite the reachability does not speed up much, the epidemic spreading has a great initial boost compared the simulation with $e = 50\%$, similarly to the growth observed in the simulation with $\sigma_{\text{free}} = 25\,\text{r}$ in Figure 5.9. In this case, the exponential growth is due to the fact that for $e = 100\%$ there is no resetting of the agents' position over time, so they are free to explore indefinitely the environment and the system ends up being well-mixed. Since the potential is never switched on, the agents perform

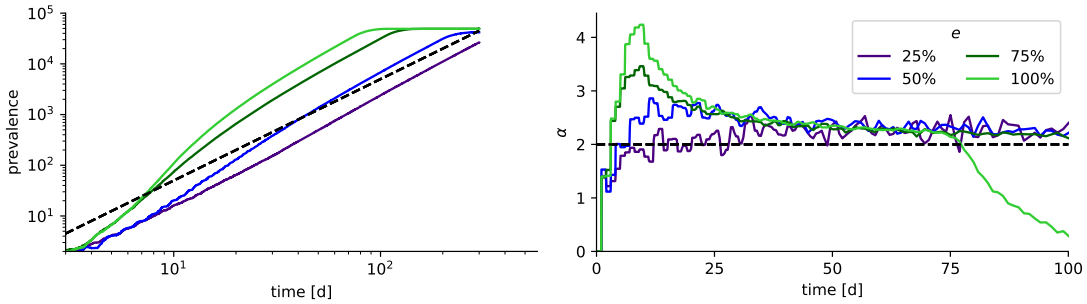


Figure 5.13: **The duration of the exploration phase affects the epidemic growth** *In the left panel, the prevalence curves* $\langle I \rangle(t)$ *are plotted in the log-log plane to highlight the power-law trends. In the right panel, the respective instantaneous coefficients are computed according to Equation 5.1. The simulations are run on standard systems with* $p_{\text{inf}} = 0.02$*. The dashed line is the quadratic growth of reference for a spatial model.*

a free Gaussian random walk, so the probability density to be at position $\vec{x}$ at time t is the Gaussian distribution reported in Equation 4.1, which standard deviation increases over time as $\sqrt{t}$. Since the motion is a Gaussian random walk and not a directed motion, the absence of the resetting is neutralized after some time. Indeed, the wavefront of infection grows linearly in t, and after some time it overruns the exploration range of the agents, slowing down the epidemic growth towards the super-quadratic regime. This effect is observed also for the curve obtained with $e = 75\%$ because the potential is too weak, or the confinement phase lasts for too short, to reset all the agents in their respective homes, as observed also in Figure 5.11. In this case, the temporal reachability has not been a good estimator to predict the regime of the epidemic growth. When increasing the exploration range, the area explored by agents per unit time increases quadratically, and the area explored is proportional to the number of unique contacts per agent. The reachability is extremely sensible to the number of new contacts, so it increases very steeply with σ_{free}. On the contrary, when increasing the exploration duration, the number of unique contacts per agent per unit time remains equal, so the reachability does not increase much. Since the total number of contacts is constant in time, having more unique contacts implies that less contacts are repeated, and the repeated contacts are good pathways for infections. To sum up, reachability is a good metric to look at when the epidemic spreading is mainly driven by unique contacts (as in high mobility scenarios), while it is not very accurate when the spreading is mainly driven by repeated contacts (as in low mobility scenarios). In other words, a fast reachability curve is a sufficient but not necessary condition to prelude exponential growth.

5.4 Universality class of the phase transition

In sections 5.2 and 5.3 I presented the effects of the mobility model onto the functional form of the prevalence. All the observations have been obtained on epidemics in their active phase, so to observe the spreading features of diseases that actually invade the system. One more interesting question to address is what is the effect of the mobility model onto the transition between the active phase and the absorbing phase. The absorbing phase is the region of the parameter space in which the epidemic gets extinct before being able to invade the system. As discussed in subsection 2.3.1, this happens when the rate of recoveries is faster than the rate of infections, and on a contact network it depends on the network structure, on p_{inf} and on τ_{inf}. To simplify the study of the phase transition, in this section all

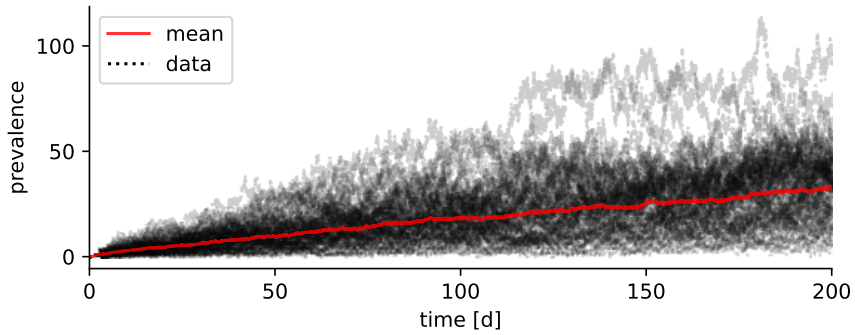


Figure 5.14: **Epidemic growth near the phase transition.** *The black curves represent the prevalence $I(t)$ of several realization of a standard system with $N = 10^4$, obtained with $p_{\text{inf}} = 0.005$. The red line is the time-wise average $\langle I \rangle(t)$ computed over the data in black.*

the parameters but p_{inf} will be fixed to their standard values, so to reduce the dimensionality of the parameter space to sample. The critical threshold for the control parameter is denoted as p_c, meaning that it represents the probability of infection at criticality.

The first relevant observation regarding the phase transition is the epidemic growth close to criticality. A paragraph will be devoted further on to the explanation of the procedure to compute the critical threshold p_c, while I here assume that the value is known so to discuss the growth near criticality. Figure 5.14 shows the simulation of a system with $p_{\text{inf}} \approx p_c$. The first important observation is that the timescale of the growth is hugely increased compared to the simulations shown in Figure 5.5. In this case, only less than 1% of the system is reached by the infection in the $t_{\text{max}} \approx 200\,\text{d}$ of simulation. The other evidence is that these simulation are much more affected by stochastic noise compared to the ones farther away from criticality. Both effects are expected to be shown by a critical system, and the outcome is that the super-quadratic and the exponential regimes discussed in the previous sections are completely destroyed.

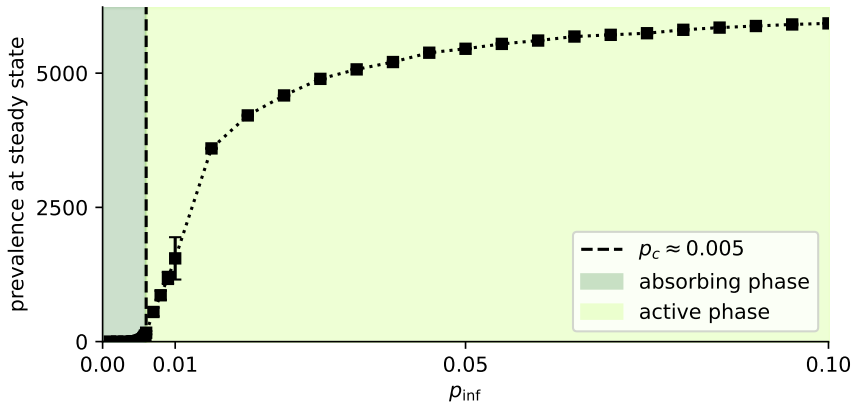


Figure 5.15: **The prevalence at equilibrium curve is a proxy to locate the phase transition.** *The black points identify the average prevalence at equilibrium* $\langle I^{\infty} \rangle$ *computed over different realizations of a standard system, varying* p_{inf}*. Error bars are obtained as the standard deviation of the data resampled with the bootstrap method.*

The critical threshold p_c can be numerically estimated by sampling the parameter space and plotting the prevalence at equilibrium, as done in Figure 5.15. The active and absorbing phases can be easily seen from the plot, and the approximate value of p_c is located in between the two. It can either be computed through fits or directly looking at the plots. For the purpose of this work, the value has been selected by looking at the plots.

The curve plotted in Figure 5.15 is the first interesting quantity that characterizes the universality class of the system. Universality classes are collections of mathematical models that share the same symmetries The universality class of the epidemic phase transition strongly depends on the symmetry of the topological space of spreading, i.e. on the characteristics of the contact network. Depending on the model used to construct the contact network, the universality class of the phase transition will be different. As mentioned in section 2.3, random network and homogeneous mixing models belong to the class of mean-field models, while lattice and random geometric graph models are in the class of directed percolation. The difference is strictly related to the spatial topology that is induced by the underlying Euclidean space, which is absent for random networks and homogeneous mixing. Mobility models are usually an ”interpolation” between these two classes, where the closeness to one or the other depends on the mobility parameters: in high-mobility settings, the universality class is reconstructed to be mean-field (see subsection 2.3.4 for references).

All the models belonging to the same universality class share the same scaling

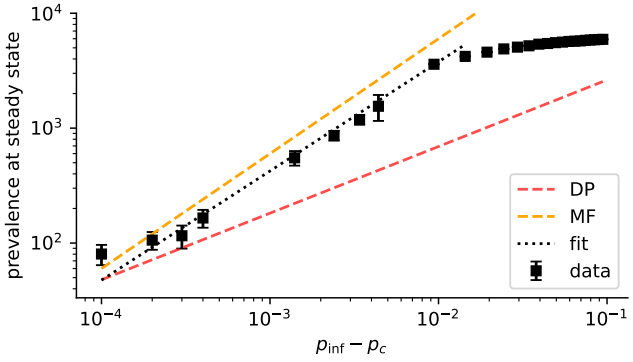


Figure 5.16: **The prevalence at equilibrium shows a similar behaviour to the mean-field model.** *The black points identify the average prevalence at equilibrium* $\langle I^\infty \rangle$ *computed over different realizations of a standard system, varying the distance from criticality* $p_{\text{inf}} - p_c$*. Error bars are obtained as the standard deviation of the data resampled with the bootstrap method. The fitting procedure provides a value* $\beta \approx 0.95$ *for the critical exponent in Equation 5.2.*

behaviour. Some canonical relations and the respective exponents are

$$I^\infty(h) \sim h^\beta \tag{5.2}$$

$$p_{\text{surv}}(h) \sim h^{\beta'} \tag{5.3}$$

where $h \coloneqq p_{\text{inf}} - p_c$. I start inspecting the quantities in Equation 5.2, already plotted in Figure 5.15. To measure the value of β, it is convenient to plot the log-log plot the curve $I^\infty(p_{\text{inf}} - p_c)$, so that β becomes the angular coefficient of the fitting line. This is represented in Figure 5.16. The value obtained is $\beta \approx 0.95$, very close to the value of the mean-field class $\beta_{\text{MF}} = 1$ and quite far from the value of the directed percolation class $\beta_{\text{DP}} \approx 0.58$. This result suggests that the mobility model adopted in this work drives the phase transition far from being related to the spatial topology of the directed percolation class. The mobility seems mix the system sufficiently well that the systems' behaviour at criticality is close to be of the mean-field type. Though, this observation is quite unexpected, as the epidemic growth discussed extensively in section 5.2 shows a behaviour that is definitely not mean-field like. There are two possible explanations to this apparent contradiction: either there is some hidden effect to be yet understood, or the fitting procedure is affected by some errors. The second explanation deserves some more discussion to point out the weaknesses of the procedure. First of all, the fitting is performed with a linear least-square method in the log-log plane, and the results is not stable with respect to the choice of the fitting range. Secondly, the data are clearly not linear for all p_{inf}, they saturate at some point, and the fitting range may be hard to choose correctly. Thirdly, the linear trend is strongly dependent on the choice of p_c, which then should be evaluated in a sophisticated manner. Finally, each data point requires long simulations (as equilibrium must be reached, and close to

criticality the timescale needed is extremely long) with several replications, so it is expensive to obtain very accurate and dense data for the fit. Despite all these comments, the aspect of the fit in Figure 5.16 seems to be reasonable, so I proceed assuming that the value obtained is approximately reliable and there is some other hidden cause underlying the contradiction.

To further inspect the scaling exponents related to the phase transition, I focus on the other relevant quantity, that is the survival probability p_{surv} of Equation 5.3, i.e. the probability for an epidemic to be able to invade the system and reach equilibrium. In order to obtain more reliable results, I want to perform the same analysis both on SEIS and on SEIR epidemic models. The intuition is that the scaling behaviour is only related to the symmetry properties of the contact network, which is the same for the two cases, and not on the epidemic model. The biggest difference is that the equilibrium condition is not defined for a SEIR model, so is not the survival probability, as any SEIR epidemic will eventually go extinct. Though, if one only considers *early* extinctions, it is possible to relax the definition of survival probability as the probability for an epidemic to only be able to invade the system, even without reaching equilibrium. This leads to the double advantage of saving some simulation time also in the case of SEIS epidemics, as the simulation can he halted before reaching the equilibrium. Attention must be paid in this case, because this solution may introduce some systematical errors that lead to overestimate the value of p_{surv}. To contain such errors, I performed a preliminary study to measure the extinction rate as a function of time, here omitted for the sake of brevity. All the simulations have been run for the minimum time needed to obtain a negligible extinction rate. The data are shown in Figure 5.17. SEIS

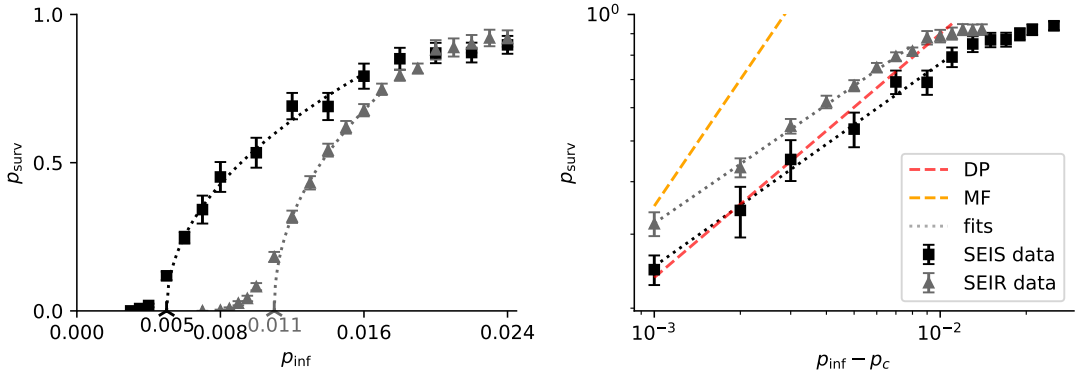


Figure 5.17: **The survival probability shows a different behaviour from mean-field and directed percolation models.** *The black points identify the survival probability of a standard system, while the gray points correspond to SEIR simulations. Error bars are obtained as the standard deviation of the data resampled with the bootstrap method. The fitting procedure provides a value* $\beta' \approx 0.48$ *for the critical exponent in Equation 5.3.*

and SEIR present two different critical thresholds. This difference is due to the presence of removed agents in the SEIR model, which counteract the epidemic spreading forming an immunity barrier and reducing the susceptible population. Despite the phase transition happens at larger p_{inf} for SEIR, when entering the active phase both SEIS and SEIR present the same scaling exponents, confirming the intuition about the independence of the epidemic model. More importantly, the values of both scaling exponents obtained with the fits are $\beta'_{\text{SEIS}} \approx \beta'_{\text{SEIR}} \approx 0.48$, and they are much below the values for the mean-field class $\beta'_{\text{MF}} = 1$ and the directed percolation class $\beta'_{\text{DP}} = 0.58$. Moreover, for both universality classes $\beta_{\text{MF}} = \beta'_{\text{MF}}$ and $\beta_{\text{DP}} = \beta'_{\text{DP}}$, while for the SEIS simulation $\beta \neq \beta'$. Differently from the previous comments about the measured β, the value of β' cannot be explained with any form of trivial interpolation. This mobility model imprints a topology on the contact network which breaks the requirements of both the universality classes that would describe its limiting cases. There are two main reasons that may explain this phenomenon: the periodic resetting of the spatial configuration and the change in the number of unique contacts between the exploratory and confinement phases. My interpretation is that the key role is played by the latter, because it breaks the temporal symmetry of the contact network system and so of the spreading capability of the epidemic. In this respect, the alternated mobility is a form of non-linear interpolation between a fully exploratory mobility and a fully confining one, that alters the scaling laws at criticality, that is when the system is most susceptible to stochastic effects. The check of this hypothesis could be implemented by two new set-ups, one in which the mobility phases alternates without resetting to the home positions, and the other in which the mobility is fully exploratory but with instantaneous periodic resets to the home positions. This implementation is deferred to a future work.

Chapter 6

Conclusion

In this thesis I have addressed the question of the effects of individual mobility on the epidemic dynamics. The mobility imprints the topological features of a spatio-temporal domain on the structure of the contact network, which in turn is the substratum of spreading for infectious diseases. Given the complexity of human behaviour, it is impractical to aim at theoretical models that reconstruct the full spectrum of its features, so here I have inspected the consequences of a few elements: the spatial extension of the exploration area, its cyclic temporal alternation with a confinement phase and the resetting of the spatial positions of the agents. The mobility range determines the amplitude of the neighbourhood that each agent often explores, determining how many different partners it gets in contact with. The properties of systems that assume the two limiting values of the mobility range are well studied: a null range corresponds to static geometric graphs, and an infinite one corresponds to well-mixed systems.

Through the analysis presented in section 5.2 and 5.3, I have shown that the scale of individual mobility can shape the functional form of the epidemic growth. Indeed, until the epidemic is locally spreading in a region that is smaller or comparable in size to the exploration range of the agents, the system behaves as well-mixed and the growth is exponential. As soon as the region of spreading exceeds the mobility scale, the growth quickly shifts towards a polynomial form with an effective fractal exponents. The fractal contribution is due to the effect of the mobility, that augments the spreading pathways with respect to the dimensionality of the wavefront of infection, i.e. the dimensionality of the spatial environment. In the infinite size limit, the ratio between the mobility scale and the environment size becomes negligible, and so epidemic growth tends asymptotically to become purely quadratic, as for static geometric models. Considering that all real systems are finite in size, this is an important effect to be taken into consideration when studying specific scenarios, especially if dealing with epidemic predictions. The very early growth of a real epidemic can be hardly determined to be polynomial or exponential, and the knowledge about the mobility scale could serve as a proxy to

improve the prediction accuracy.

In section 5.3, I have shown that also resetting is relevant to consider for epidemic growth. For agents that are never reset to their home positions, their neighbourhood is a spatial region that travels through the environment at a pace dictated by the mobility range; on the contrary, for agents that periodically return home, their neighbourhood is a region that is geographically static in the environment, and only its extension depends on the mobility range. Over several cycles, resetting is responsible for attributing a spatial locality to the contacts between agents, thus for preventing the system from becoming well-mixed and averting an exponential epidemic growth.

In section 5.4 I shifted the focus to the behaviour close to the absorbing phase transition. I have shown that the scaling exponents for the prevalence at the steady state and for the survival probability do not belong to neither the mean-field nor the directed percolation universality classes. These classes include, respectively, the epidemics spreading on well-mixed systems and on static geometric graphs. Despite the directed percolation class is rather flexible and inclusive of several models studied in literature [91–97], this model generates a contact network which topology breaks the symmetry requirements of this universality class. The potential causes of the asymmetries are either the spatial periodic resetting of the agents' position or to the cyclic pattern of the contact network, or to both simultaneously. The result is that the few elements considered in this work have been shown to alter the universality class of the epidemic dynamics close to the phase transition.

This project could be a starting point for some future scientific effort, as it opens up some intriguing research paths. The universality class deserves further analysis, especially concerning the two possible causes of the symmetry break mentioned above. To uncover this question, one could implement: on one side, a temporally homogeneous random walk with instantaneous periodic resets, similarly to what has been done in [98]; on the other side, a random walk with cyclically alternated exploration range. Furthermore, different sets of parameters could be testes with the aim to recreate the limiting cases, and check if the corresponding universality classes are reached with a smooth or a sudden transition.

One more feature of reality that would be nice to introduce is the directed motion. In this model, individual mobility is build on a Gaussian random walk, which turn useful to introduce non-deterministic effects but produces an unrealistically diffusive motion. Human mobility is mostly aimed towards specific locations, and this effect could be introduced through one (or multiple) meeting spot(s). Practically, this target could be reached by summing other properly shaped potentials, common for some agents, to the individual potential, as discussed in Appendix C. This setting would also recreate high density regions where several contacts would be established, that may represent working places or free-time activities. I expect that the creation of these meeting spots should drive the system towards a higher level of mixing. Alternatively, directed motion without meeting spots could be

obtained by reversing the sign of the potential, turning it from attractive during the confinement phase to repulsive during the exploration phase.

A third interesting extension to this work could be the addition of an attraction potential between groups of agents. The contacts modelled in this work are instantaneous, as they are established and broken randomly at each time-step. Real contacts have a temporal duration which is rather heterogeneous. To recreate this effect it could be possible to introduce an attraction potential that drives the agents one towards another, and keeps them in contact for subsequent time-steps. The attraction could also be weighted by some friendship coefficients, that would determine an heterogeneous probability for an agent to meet one or another agent. These elements would probably break even more the spatio-temporal symmetries of the contact network, possibly with interesting implications on the epidemic spreading.

To sum up, mobility is important to be adequately modelled when studying disease spreading because of the dominant effects it has on the spreading velocity and epidemic survival. Individual mobility is only one of the many aspects of the complex human world, but it surely deserves further research investment for the many open questions that are yet to be understood.

Appendix A

Fokker-Planck equilibrium solution

The equation of motion 3.1 is the differential form of a Langevin equation. Defining $\vec{A}(\vec{x}) \coloneqq -\nabla V(\vec{x})$ and $\mathbf{b} \coloneqq \sqrt{2D}\mathbf{I}$, where the bold indicates a matrix, the differential equation of motion assumes the shape $\mathrm{d}\vec{x} = \vec{A}(\vec{x},t)\mathrm{d}t + \mathbf{b}\mathrm{d}\vec{W}$. As discussed in subsection 2.2.4, the probability density of the agent's location evolves according to the Fokker-Planck equation

$$\partial_t P(\vec{x},t) = -\sum_{i=1}^{d} \partial_{x_i}\left[A_i(\vec{x},t)P(\vec{x},t)\right] + \frac{1}{2}\sum_{i=1}^{d}\sum_{j=1}^{d} \partial_{x_i}\partial_{x_j}\left[B_{ij}(\vec{x},t)P(\vec{x},t)\right]$$

First of all, for this motion $\mathbf{B} = 2D\mathbf{I}$ is independent on $\vec{x}$. Moreover, I am looking for the equilibrium solution, which is independent on time and stationary, so the left-hand-side $\partial_t P_{\mathrm{eq}}(\vec{x}) = 0$. Thus the equation becomes

$$0 = -\sum_{i=1}^{d} \partial_{x_i}\left[A_i(\vec{x})P_{\mathrm{eq}}(\vec{x})\right] + D\sum_{i=1}^{d}\sum_{j=1}^{d} \partial_{x_i}\partial_{x_j} P_{\mathrm{eq}}(\vec{x})$$

The equation in this form can be conveniently rewritten in vectorial notation, observing that the first term is a divergence and the second is a Laplacian. Thus

$$0 = -\nabla\cdot\left[\vec{A}(\vec{x})P_{\mathrm{eq}}(\vec{x})\right] + D\nabla^2 P_{\mathrm{eq}}(\vec{x})$$

$$0 = -\nabla\cdot\left[\vec{A}(\vec{x})P_{\mathrm{eq}}(\vec{x}) - D\nabla P_{\mathrm{eq}}(\vec{x})\right]$$

All the probability densities that form a solenoidal vector field $\vec{F} \coloneqq \vec{A}P_{\mathrm{eq}} + D\nabla P_{\mathrm{eq}}$ are solutions to the Fokker-Planck equation. Since I am not interested in all the solutions, rather in the simplest solution, then I want the particular solution P^*_{eq} that forms a *null* vector field $\vec{F} = 0\ \forall\vec{x}$. Thus, the equation becomes

$$\vec{A}(\vec{x})P^*_{\mathrm{eq}}(\vec{x}) - D\nabla P^*_{\mathrm{eq}}(\vec{x}) = 0$$

Writing it in components, this equation becomes is a system of two coupled partial differential equations, and substituting $\vec{A} = -\nabla V$ one gets

$$\begin{cases} \partial_{x_1} P^*_{\text{eq}}(\vec{x}) = -\dfrac{\partial_{x_1} V}{D} P^*_{\text{eq}}(\vec{x}) \\ \partial_{x_2} P^*_{\text{eq}}(\vec{x}) = -\dfrac{\partial_{x_2} V}{D} P^*_{\text{eq}}(\vec{x}) \end{cases} \tag{A.1}$$

The first equation can be solved by separation of variables and integrating over domain section $\mathcal{D}_{x_1}$

$$\int_{\mathcal{D}_{x_1}} \frac{1}{P^*_{\text{eq}}(\vec{x})} \frac{\partial P^*_{\text{eq}}(\vec{x})}{\partial x_1} \mathrm{d}x_1 = -\frac{1}{D} \int_{\mathcal{D}_{x_1}} \frac{\partial V(\vec{x})}{\partial x_1} \mathrm{d}x_1$$

$$\log\Big(P^*_{\text{eq}}(\vec{x})\Big) = -\frac{V(\vec{x})}{D} + c(x_2) \tag{A.2}$$

where $c(x_2)$ is an integration constant that might depend on x_2. Deriving now the equation for x_2 one gets

$$\frac{\partial}{\partial x_2}\Big[\log\Big(P^*_{\text{eq}}(\vec{x})\Big)\Big] = \frac{\partial}{\partial x_2}\left[-\frac{V(\vec{x})}{D} + c(x_2)\right]$$

$$\frac{1}{P^*_{\text{eq}}(\vec{x})} \frac{\partial P^*_{\text{eq}}(\vec{x})}{\partial x_2} = -\frac{1}{D} \frac{\partial V(\vec{x})}{\partial x_2} + \frac{\partial c(x_2)}{\partial x_2}$$

$$\partial_{x_2} P^*_{\text{eq}}(\vec{x}) = -\partial_{x_2} \frac{V(\vec{x})}{D} P^*_{\text{eq}}(\vec{x}) - \partial_{x_2} c(x_2) P^*_{\text{eq}}(\vec{x})$$

By comparison with the second equation in Equation A.1, one sets $\frac{\partial c}{\partial x_2} = 0$, meaning that $c = \text{const}$. Inserting this result into Equation A.2 and exponentiating both sides one gets

$$P^*_{\text{eq}}(\vec{x}) = e^c \exp\left(-\frac{V(\vec{x})}{D}\right) \tag{A.3}$$

Finally, one can define $Z \coloneqq e^{-c}$ so to obtain the expression in Equation 4.3.

Appendix B

Instantaneous exponents

In chapter 5 several results are discussed through the instantaneous exponent defined as in Equation 5.1. Here I provide some proofs about the reliability of this analysis. On one side, this metric is more robust with respect to a standard fitting, because there is no fitting range or other parameters to be arbitrarily chosen. On the other side, it only seems to be completely parameter free, even if it is not. From the definition $\alpha \coloneqq \frac{\mathrm{d}\mathcal{I}}{\mathrm{d}\tau}$, the variables $\mathcal{I}$ and τ could in principle represent whatever quantities. It is when defining $\mathcal{I} \coloneqq \log(I)$ and $\tau \coloneqq \log(t)$ that the meaning of α becomes the instantaneous exponent of a locally fitting curve in the form $f(t) = c\, t^{\alpha(t)}$. This definition is implicitly imposing the constraint that $I(0) = 0$, which is in principle wrong, as the epidemic is seeded at $0\,\mathrm{d} \leq t_{\mathrm{seed}} \leq 1\,\mathrm{d}$ with one infectious node $I(t_{\mathrm{seed}}) = 1$. Moreover, the discrete nature of the system may introduce non-negligible discretization issues for low I in the early stages of the spreading. This implies that a more robust functional form for the locally fitting curve could be $g(t) = c(t-a)^{\alpha(t)} + b$, and it is not obvious that $a = b = 0$. If this was not true, then the computed values $\alpha(t)$ would suffer of strong systematic biases, thus would be not reliable.

To obtain the proof of reliability, I resort to the comparison of the measured data with new *synthetic* data generated with the following procedure. I first compute $\alpha(t)$ from the measured values I and t using the definition Equation 5.1. The, I rewrite the equation as

$$\frac{\alpha(t)}{t}\mathrm{d}t = \frac{1}{I}\mathrm{d}I$$

Integrating both sides I get

$$\int_{t_0}^{t} \frac{\alpha(t')}{t'}\mathrm{d}t' = \int_{I_0}^{I} \frac{1}{I'}\mathrm{d}I'$$

Solving the integral in the right-hand-side and exponentiating both sides, I obtain

$$I = c \exp\left(\int_0^t \frac{\alpha(t')}{t'}\mathrm{d}t'\right) \tag{B.1}$$

where c is a free scale parameter that comes from the integration constants t_0 and I_0. Computing the synthetic prevalence I_s with equation Equation B.1, if the synthetic data well approximate the measured data $I_s \approx I$, then the hidden assumptions discussed in the previous paragraph are well justified. Figure B.1 shows the plots of the comparison between the measured data and the synthetic data. From the plot, the match between the two appears to be very close, implying that the instantaneous exponents are well defined and provide a useful tool for the analysis of the results in chapter 5.

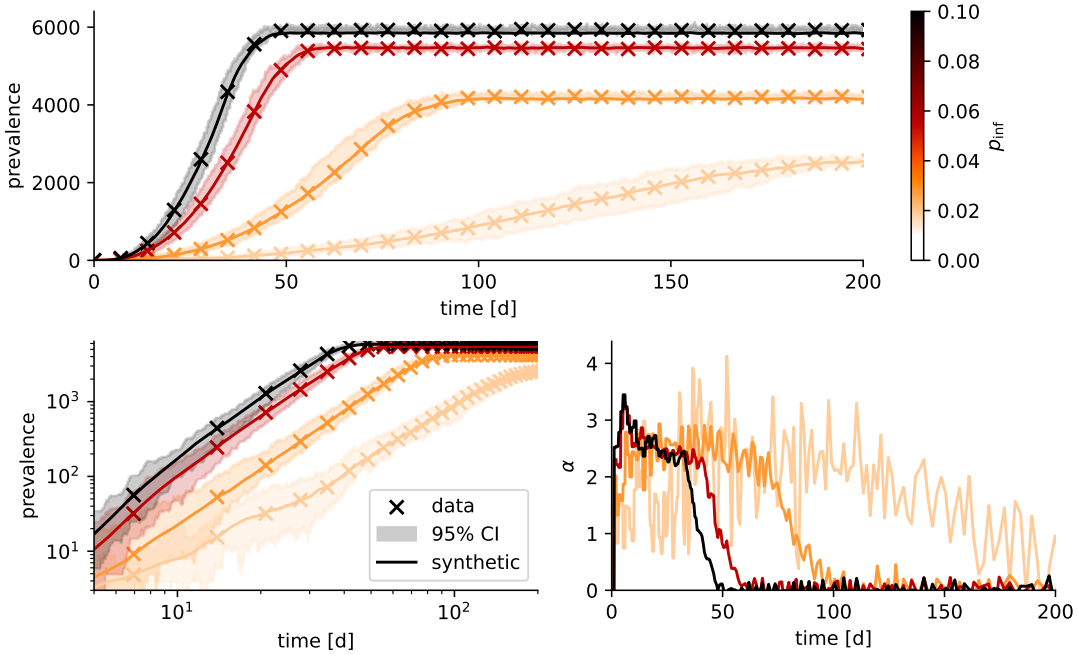


Figure B.1: **The instantaneous exponents are a reliable metric to describe the prevalence growth.** *In the upper and lower-left panels a sub-sample of the measured prevalence and its central* 95% *confidence interval is plotted along with the synthetic prevalence computed as in Equation B.1. In the lower-right panel, the instantaneous exponents of the respective data is reported for reference. The curves here shown are some of the ones presented in Figure 5.5.*

Appendix C

Design of the potential

The functional shape of the potential in Equation 3.4 is rather complicated because it is designed to be very versatile. If one desires to introduce a directed motion towards a specific location in space, another potential must be added to the one centered in the home position. This "composite" potential should present two (or more) local minima, so to allow the agent perform stochastic transitions from one to the other, and the depth of the minima may vary in time, so that the preferred location varies depending on the time of the day. To have a smooth transition from one minimum to the other, there are some constraints that must be imposed on the potential. The first is that the amplitude parameter $A(t)$ must be continuous in time. The second is the existence of two (or more) minima when two (or more) potentials of the same functional form and different center are added. This second constraint translates into the requirement for the potential to present an inflection point so that the gradient in the tails is lower than the gradient near the central well.

Mathematically this can be computed more easily by writing the potential in polar coordinates, which is easy possible if $\gamma \in 2\mathbb{N}$. The second derivative of the potential along the radial direction (the only interesting component, since the rotational invariance) is

$$\frac{\mathrm{d}^2 V}{\mathrm{d}r^2} = A(t)\beta^{-\gamma} r^{\gamma-2}\left(\delta^\gamma + \left(\frac{r}{\beta}\right)^\gamma\right)^{\frac{\alpha}{\gamma}-2}\left((\gamma-1)\delta^\gamma - (1-\alpha)\left(\frac{r}{\beta}\right)^\gamma\right) \tag{C.1}$$

where $r \coloneqq \|\vec{x} - \vec{x}_0\|$ is the radial distance from the center of the potential. This expression proves the presence of one inflection point at

$$r^* = \beta\delta\sqrt[\gamma]{\frac{\gamma-1}{1-\alpha}} \tag{C.2}$$

which exists only if $\alpha < 1$, i.e. if the increment in the tails of the potential is sub-linear. Considering for simplicity two potentials V_1 and V_2 with the same

amplitude and centers in $\vec{x}_0^1 = 0$ and $\vec{x}_0^2 = h\hat{e}_1$, their sum will present two local minima if $h > 2r^*$. In a more general case, the condition over h depends on the ratio between the amplitude parameters of the two potentials. Thus, if one wants to have a potential landscape that includes two local minima, provided appropriate amplitude coefficients, it is sufficient to choose $\alpha < 1$ and the centers of the minima that are sufficiently far one from the other. If these conditions are not fulfilled, then only one minima will be present and its center will be located in an intermediate point.

Bibliography

[1] M. J. Keeling and P. Rohani, *Modeling infectious diseases in humans and animals* (Princeton University Press, 2008).

[2] S. Peng, Q. Chen, and E. Liu, "The role of computational fluid dynamics tools on investigation of pathogen transmission: prevention and control", Science of The Total Environment **746**, 142090 (2020).

[3] N. C. Achaiah, S. B. Subbarajasetty, and R. M. Shetty, "R0 and re of COVID-19: can we predict when the pandemic outbreak will be contained?", en, Indian J. Crit. Care Med. **24**, 1125 (2020).

[4] P. Bajardi et al., "Human mobility networks, travel restrictions, and the global spread of 2009 h1n1 pandemic", PLOS ONE **6**, 1 (2011).

[5] I. Z. Kiss, J. C. Miller, P. L. Simon, et al., *Mathematics of epidemics on networks* (Springer, 2017).

[6] R. Pastor-Satorras et al., "Epidemic processes in complex networks", Rev. Mod. Phys. **87**, 925 (2015).

[7] A. Barrat et al., "Measuring contact patterns with wearable sensors: methods, data characteristics and applications to data-driven simulations of infectious diseases", Clinical Microbiology and Infection **20**, 10 (2014).

[8] P. Sapiezynski et al., "Interaction data from the copenhagen networks study", Scientific Data **6**, 315 (2019).

[9] SocioPatterns Collaboration.

[10] A. Stopczynski et al., "Privacy in sensor-driven human data collection: a guide for practitioners", 2014.

[11] J. Lane et al., *Privacy, big data, and the public good: frameworks for engagement* (Cambridge University Press, 2014).

[12] A. Barrat and C. Cattuto, "Temporal networks of face-to-face human interactions", in *Temporal networks*, edited by P. Holme and J. Saramäki (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013), pp. 191–216.

[13] B. Ivanovic et al., "Generative modeling of multimodal multi-human behavior", in 2018 ieee/rsj international conference on intelligent robots and systems (iros) (Oct. 2018), pp. 3088–3095.

[14] M. Starnini et al., "Robust modeling of human contact networks across different scales and proximity-sensing techniques", in Social informatics, edited by G. L. Ciampaglia, A. Mashhadi, and T. Yasseri (2017), pp. 536–551.

[15] V. Sekara, A. Stopczynski, and S. Lehmann, "Fundamental structures of dynamic social networks", Proceedings of the national academy of sciences **113**, 9977 (2016).
[16] J. Zierenberg et al., "How contact patterns destabilize and modulate epidemic outbreaks", New Journal of Physics **25**, 053033 (2023).
[17] S. Altizer et al., "Seasonality and the dynamics of infectious diseases", Ecology Letters **9**, 467 (2006).
[18] J. Dehning et al., "Impact of the euro 2020 championship on the spread of COVID-19", Nature Communications **14**, 122 (2023).
[19] A.-L. Barabási and M. Pósfai, *Network science* (Cambridge University Press, Cambridge, 2016).
[20] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks", Science **286**, 509 (1999).
[21] G. T. L. da F. Costa F. A. Rodrigues and P. R. V. Boas, "Characterization of complex networks: a survey of measurements", Advances in Physics **56**, 167 (2007).
[22] C. H. Sanabria-Montaña and R. Huerta-Quintanilla, "Generalization of clustering coefficient on lattice networks applied to criminal networks", International Journal of Computer and Information Engineering **11**, 868 (2017).
[23] G. Chalancon, K. Kruse, and M. M. Babu, "Clustering coefficient", in *Encyclopedia of systems biology*, edited by W. Dubitzky et al. (Springer New York, New York, NY, 2013), pp. 422–424.
[24] P. Erdős, A. Rényi, et al., "On the evolution of random graphs", Publ. math. inst. hung. acad. sci **5**, 17 (1960).
[25] B. Uzzi, L. A. Amaral, and F. Reed-Tsochas, "Small-world networks and management science research: a review", European Management Review **4**, 77 (2007).
[26] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks", Nature **393**, 440 (1998).
[27] A.-L. Barabási, E. Ravasz, and T. Vicsek, "Deterministic scale-free networks", Physica A: Statistical Mechanics and its Applications **299**, 559 (2001).
[28] J. Stehlé, A. Barrat, and G. Bianconi, "Dynamical and bursty interactions in social networks", Phys. Rev. E **81**, 035101 (2010).
[29] H.-H. Jo et al., "Circadian pattern and burstiness in mobile phone communication", New Journal of Physics **14**, 013055 (2012).
[30] M. Karsai, H.-H. Jo, and K. Kaski, "Empirical findings in human bursty dynamics", in *Bursty human dynamics* (Springer International Publishing, Cham, 2018), pp. 31–46.
[31] P. Holme and J. Saramäki, "Temporal networks", Physics Reports **519**, Temporal Networks, 97 (2012).
[32] V. Gelardi et al., "From temporal network data to the dynamics of social relationships", Proceedings of the Royal Society B: Biological Sciences **288**, 20211164 (2021).

[33] R. K. Pan and J. Saramäki, "Path lengths, correlations, and centrality in temporal networks", Phys. Rev. E **84**, 016105 (2011).
[34] R. Gallotti et al., "A stochastic model of randomly accelerated walkers for human mobility", Nature Communications **7**, 12600 (2016).
[35] Z. Toroczkai and H. Guclu, "Proximity networks and epidemics", Physica A: Statistical Mechanics and its Applications **378**, Social network analysis: Measuring tools, structures and dynamics, 68 (2007).
[36] I. Hanski and O. Gaggiotti, in *Ecology, genetics and evolution of metapopulations*, edited by I. Hanski and O. E. Gaggiotti (Academic Press, Burlington, 2004), pp. 3–22.
[37] F. Simini et al., "A universal model for mobility and migration patterns", Nature **484**, 96 (2012).
[38] A. Okubo and S. A. Levin, *Diffusion and ecological problems: modern perspectives*, Vol. 14 (Springer, 2001).
[39] Alessandretti, Laura, Lehmann, Sune, and Baronchelli, Andrea, "Understanding the interplay between social and spatial behaviour", EPJ Data Sci. **7**, 36 (2018).
[40] C. Song et al., "Modelling the scaling properties of human mobility", Nature Physics **6**, 818 (2010).
[41] M. C. González, C. A. Hidalgo, and A.-L. Barabási, "Understanding individual human mobility patterns", Nature **453**, 779 (2008).
[42] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data", SIAM Review **51**, 661 (2009).
[43] A. D. Broido and A. Clauset, "Scale-free networks are rare", Nature Communications **10**, 1017 (2019).
[44] L. Alessandretti, U. Aslak, and S. Lehmann, "The scales of human mobility", Nature **587**, 402 (2020).
[45] P. Holme, "Rare and everywhere: perspectives on scale-free networks", Nature Communications **10**, 1016 (2019).
[46] H. Barbosa et al., "The effect of recency to human mobility", EPJ Data Science **4**, 21 (2015).
[47] S. Gheorghiu and M.-O. Coppens, "Heterogeneity explains features of "anomalous" thermodynamics and statistics", Proceedings of the National Academy of Sciences **101**, 15852 (2004).
[48] K. Zhao et al., "Explaining the power-law distribution of human mobility through transportationmodality decomposition", Scientific Reports **5**, 9136 (2015).
[49] A. Stopczynski, A. 's. Pentland, and S. Lehmann, "How physical proximity shapes complex social networks", Scientific Reports **8**, 17722 (2018).
[50] S. Chang et al., "Mobility network models of COVID-19 explain inequities and inform reopening", Nature **589**, 82 (2021).
[51] A. Aleta et al., "Quantifying the importance and location of SARS-CoV-2 transmission events in large metropolitan areas", en, Proc Natl Acad Sci U S A **119**, e2112182119 (2022).

[52] E. N. Gilbert, "Random plane networks", Journal of the Society for Industrial and Applied Mathematics **9**, 533 (1961).
[53] J. Dall and M. Christensen, "Random geometric graphs", Physical review E **66**, 016121 (2002).
[54] M. Penrose, *Random Geometric Graphs* (Oxford University Press, May 2003).
[55] M. Frasca et al., "Dynamical network model of infective mobile agents", Phys. Rev. E **74**, 036110 (2006).
[56] A. Buscarino et al., "Disease spreading in populations of moving agents", Europhysics Letters **82**, 38002 (2008).
[57] M. Starnini, A. Baronchelli, and R. Pastor-Satorras, "Modeling human dynamics of face-to-face interaction networks", Phys. Rev. Lett. **110**, 168701 (2013).
[58] J. P. Rodríguez et al., "Epidemic processes on self-propelled particles: continuum and agent-based modeling", Phys. Rev. Res. **4**, 043160 (2022).
[59] H. Risken and H. Risken, *Fokker-planck equation* (Springer, 1996).
[60] N. Bacaër, *A short history of mathematical population dynamics*, Vol. 618 (Springer, 2011).
[61] M. P. Ward et al., "Estimation of the basic reproductive number (r0) for epidemic, highly pathogenic avian influenza subtype h5n1 spread", Epidemiology &; Infection **137**, 219 (2009).
[62] R. de Vries et al., "Systematic screening for chlamydia trachomatis: estimating cost-effectiveness using dynamic modeling and dutch data", Value in Health **9**, 1 (2006).
[63] G. P. Garnett et al., "The transmission dynamics of gonorrhoea: modelling the reported behaviour of infected patients from newark, new jersey", Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **354**, 787 (1999).
[64] J. O. Giraldo and D. H. Palacio, "Deterministic sir (susceptible–infected–removed) models applied to varicella outbreaks", Epidemiology & Infection **136**, 679 (2008).
[65] L. Zhou et al., "Global dynamics of a discrete age-structured sir epidemic model with applications to measles vaccination strategies", Mathematical Biosciences **308**, 27 (2019).
[66] T. Kuniya, "Global behavior of a multi-group sir epidemic model with age structure and an application to the chlamydia epidemic in japan", SIAM Journal on Applied Mathematics **79**, 321 (2019).
[67] P. Dönges et al., "Interplay between risk perception, behavior, and covid-19 spread", Frontiers in Physics **10**, `10.3389/fphy.2022.842180` (2022).
[68] W. O. Kermack and A. G. McKendrick, "Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity", Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character **138**, 55 (1932).
[69] R. M. Anderson, "Discussion: the kermack-mckendrick epidemic threshold theorem", Bulletin of Mathematical Biology **53**, 1 (1991).

[70] J. Dehning et al., "Inferring change points in the spread of covid-19 reveals the effectiveness of interventions", Science **369**, eabb9789 (2020).
[71] S. d. Picoli Junior et al., "Spreading patterns of the influenza a (h1n1) pandemic", PLOS ONE **6**, 1 (2011).
[72] G. Chowell and C. Viboud, "Is it growing exponentially fast? – impact of assuming exponential growth for characterizing and forecasting epidemics with initial near-exponential growth dynamics", Infectious Disease Modelling **1**, 71 (2016).
[73] E. Z. Martinez, D. C. Aragon, and A. A. Nunes, "Long-term forecasts of the covid-19 epidemic: a dangerous idea", Revista da Sociedade Brasileira de Medicina Tropical **53**, e20200481 (2020).
[74] C. Viboud, L. Simonsen, and G. Chowell, "A generalized-growth model to characterize the early ascending phase of infectious disease outbreaks", Epidemics **15**, 27 (2016).
[75] R. Dickman, "Universality and diffusion in nonequilibrium critical phenomena", Phys. Rev. B **40**, 7005 (1989).
[76] D. Mollison, "Spatial contact models for ecological and epidemic spread (with discussion)", J Roy Stat Soc **39**, 283 (1977).
[77] P. Grassberger, "Directed percolation in 2+1 dimensions", Journal of Physics A: Mathematical and General **22**, 3673 (1989).
[78] P. Grassberger, "Universality of critically pinned interfaces in two-dimensional isotropic random media", Phys. Rev. Lett. **120**, 200605 (2018).
[79] M. Henkel et al., *Non-equilibrium phase transitions*, Vol. 1 (Springer, 2008).
[80] J. Stehlé et al., "Simulation of an seir infectious disease model on the dynamic contact network of conference attendees", BMC Medicine **9**, 87 (2011).
[81] G. Chowell et al., "Mathematical models to characterize early epidemic growth: a review", Physics of Life Reviews **18**, 66 (2016).
[82] M. Kretzschmar and M. Morris, "Measures of concurrency in networks and the spread of infectious disease", Mathematical Biosciences **133**, 165 (1996).
[83] E. Massad et al., "Scale-free network of a dengue epidemic", Applied Mathematics and Computation **195**, 376 (2008).
[84] B. Szendrói and G. Csányi, "Polynomial epidemics and clustering in contact networks", Proceedings of the Royal Society of London. Series B: Biological Sciences **271**, S364 (2004).
[85] M. Salathé and J. H. Jones, "Dynamics and control of diseases in networks with community structure", PLOS Computational Biology **6**, 1 (2010).
[86] N. Masuda and P. Holme, "Predicting and controlling infectious disease epidemics using temporal networks", F1000prime reports **5** (2013).
[87] M. Karsai et al., "Small but slow world: how network topology and burstiness slow down spreading", Phys. Rev. E **83**, 025102 (2011).
[88] M. Kivelä et al., "Multiscale analysis of spreading in a large communication network", Journal of Statistical Mechanics: Theory and Experiment **2012**, P03005 (2012).

[89] W. Jakob, J. Rhinelander, and D. Moldovan, *Pybind11 – seamless operability between c++11 and python*, https://github.com/pybind/pybind11, 2017.

[90] K. Burrage, P. M. Burrage, and T. Tian, "Numerical methods for strong solutions of stochastic differential equations: an overview", Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences **460**, 373 (2004).

[91] A. Jiménez-Dalmaroni, "Directed percolation with incubation times", Phys. Rev. E **74**, 011123 (2006).

[92] A. Badie-Modiri et al., "Directed percolation in temporal networks", Phys. Rev. Res. **4**, L022047 (2022).

[93] A. Badie-Modiri et al., "Directed percolation in random temporal network models with heterogeneities", Phys. Rev. E **105**, 054313 (2022).

[94] H. Hinrichsen and M. Howard, "A model for anomalous directed percolation", The European Physical Journal B - Condensed Matter and Complex Systems **7**, 635 (1999).

[95] S.-W. Son et al., "Percolation theory on interdependent networks based on epidemic spreading", Europhysics Letters **97**, 16006 (2012).

[96] H.-K. Janssen and O. Stenull, "Field theory of directed percolation with long-range spreading", Phys. Rev. E **78**, 061117 (2008).

[97] P. Grassberger, "Two-dimensional sir epidemics with long range infection", Journal of statistical physics **153**, 289 (2013).

[98] M. R. Evans and S. N. Majumdar, "Diffusion with stochastic resetting", Phys. Rev. Lett. **106**, 160601 (2011).