1
Human Mobility and Spatial Disease Dynamics

Dirk Brockmann

1.1
Introduction and Motivation

The understanding of human mobility and the development of qualitative models as well as quantitative theories for it is of key importance in the research of human infectious disease dynamics on large geographical scales. Xia *et al.* state succintly [1]:

> "Spatial transmission of directly transmitted infectious diseases is ultimately tied to movement by the hosts. The network of spatial spread (the disease's spatial coupling) may therefore be expected to be related to the transportation network within the host metapopulation"

In our globalized world, mobility and traffic have reached a complexity and volume of unprecedented degree. More than 60 million people travel billions of miles on more than 2 million international flights each week as illustrated in Figure 1.1. Hundreds of millions of people commute on a complex web of highways and railroads, most of which operate at their maximum capacity. Despite this increasing connectivity and our ability to visit virtually every place on this planet in a matter of days, the magnitude and intensity of modern human traffic has made human society more susceptible to threats intimately connected to human travel. For instance, long-range human mobility is responsible for the geographical spread of emergent infectious diseases and plays a key role in human mediated bioinvasion, the dominant factor in the global biodiversity crisis. The prime example of modern epidemics is the severe acute respiratory syndrome (SARS). The SARS virus first appeared in a Chinese province from where it reached Hong Kong in 2003. It proliferated and spread around the world in a matter of weeks infecting nearly 10 000 individuals worldwide with a mortality of approximately 10%. Since then, epidemiologist have devoted an increasing amount of attention and modeling effort to understand in what way and to what extent modern traffic networks impact and determine the dynamics of emergent diseases, particularly facing

Reviews of Nonlinear Dynamics and Complexity. Edited by Heinz Georg Schuster

ISBN: 978-3-527-40850-4

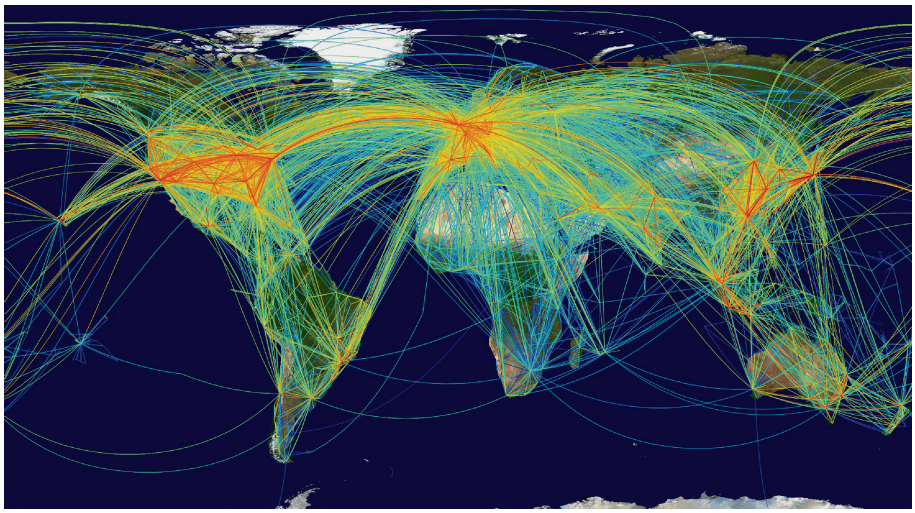

Figure 1.1 The worldwide air transportation network. More than three billion passengers travel on this network each year, on flights connecting approximately 4000 airports. The heterogeneity of the network is reflected by the flux of individuals between nodes, ranging from a few to more than 10 000 passengers per day between nodes.

an immanent H5N1 flu pandemic and the potential threat of the use of smallpox in bioterroist attacks [2, 3].

In a number of recent studies the statistical properties of particular human transportation networks were investigated in detail with a focus on air transportation and long-distance traffic [4–7]. However, human mobility occurs on many length scales, ranging from commuting traffic on short distances to long-range travel by air, and involves diverse methods of transportation (public transportation, roads, highways, trains, and air transportation). No comprehensive study exists that incorporates traffic on all spatial scales. This would require the collection and compilation of data for various transportation networks into a multi-component dataset; a difficult, if not impossible, task particularly on an international scale. Whereas central statistical features of air transportation networks have been studied in detail, it remains unclear whether these properties remain unchanged in traffic networks that comprise all other means of transportation and spatial scales. How do these properties depend on the length scale? Are they universal? In what way do they change as a function of length scale? What are the national and regional differences and similarities? In order to understand human mobility in the 21st century and the dynamics of associated phenomena, particularly the geographic spread of modern diseases, it is of fundamental importance to answer these questions.

Once a more comprehensive understanding of human mobility exists, the next step in the context of spatial epidemics is the translation of traffic information and topological features of complex traffic networks into dynamical components of models that can account for the spatial spread of infectious diseases. These type of models have been devised in the past on a wide range of complexity levels. On one end of the spectrum are reaction diffusion models in which local nonlinear infection dynamics is coupled with diffusive dispersal. Spatial heterogeneity in the host population is generally neglected in these models [8]. The type of questions which these models address are, for example; Under what circumstance does a propagating epidemic wave develop? How does the speed of the wave depend on the parameters of the model? What impact does spatial heterogeneity have on the disease dynamics, and what are the statistical regularities in spatial patterns?

On the other end of the spectrum are sophisticated models that are constructed with a high degree of detail [2, 3, 9, 10]. Examples of these models are agent-based simulation frameworks in which social, spatial and temporal heterogeneity are taken into account. Frequently these models contain entire global transportation networks and extrapolations where empirical data is lacking based on known statistics.

This chapter contains two parts. In the first part I will discuss recent progess in the study of multi-length scale transportation networks. I will show that, despite their complexity, these networks exhibit a set of scaling relations and statistical regularities. In the second part I will review how the topological features of traffic networks can be incorporated in models for disease dynamics and show that the way topology is translated into dynamics can have a profound impact on the overall disease dynamics.

1.2
Quantitative Assessments of Human Mobility

1.2.1
Preliminary Considerations

Formally we can address the issue of mobile individuals by the collection of individual trajectories of each of $\mathcal{N}$ individuals of a population, that is the collection $\{\vec{x}_i(t)\}_{i=1,\ldots,\mathcal{N}}$ where each individual is labeled i. Clearly the measurement and the prediction of each individual's location $\vec{x}_i(t)$ as a function of time is beyond a researcher's grasp. Some very recent experiments, however, employing high-precision measurements based on GPS (global positioning via satellite) or using cell phone location as a proxy for $\vec{x}_i(t)$ have made it possible at least to measure, individual trajectories with unexpected accuracy [11].

The next best approach to human mobility is based on population averages. To this end it is useful to define the microscopic time dependent density of individuals

$$u(\vec{x},t) = \frac{1}{A}\sum_i^{\mathcal{N}} \delta(\vec{x} - \vec{x}_i(t)) \tag{1.1}$$

where A is the spatial area under consideration. The global density of individuals in A is given by the integral of u, that is

$$u_0 = \frac{\mathcal{N}}{A} = \int \mathrm{d}\vec{x} u(\vec{x},t)\ . \tag{1.2}$$

The expecation value $\langle u(\vec{x},t)\rangle$ of the microscopic density is related to the probability $p_i(\vec{x},t)$ of individual i being located at $\vec{x}$ by

$$\begin{aligned}\langle u(\vec{x},t)\rangle &= \frac{1}{A}\sum_i^{\mathcal{N}} \langle \delta(\vec{x} - \vec{x}_i(t))\rangle \\ &= \frac{1}{A}\sum_i^{\mathcal{N}} p_i(\vec{x},t)\end{aligned} \tag{1.3}$$

Because for each i even the quantitiy $p_i(\vec{x},t)$ is usually inaccessible to measurement, a widespread assumption made in models is that individuals are indistinguishable and that although $\vec{x}_i(t) \neq \vec{x}_j(t)$ one assumes $p_i(\vec{x},t) = p_j(\vec{x},t)$ and thus

$$\langle u(\vec{x},t)\rangle = \frac{1}{A} p(\vec{x},t)\ . \tag{1.4}$$

Despite its simplicity, this equation is fundamental to the probabilistic interpretation of models that are based on the time-evolution of concentrations. It connects the probabilistic quantity $p(\vec{x},t)$ to the measurable density of individuals. The second assumption in the conceptual setup of analyzing human mobility is an ergodicity assumption, that is given by

$$\frac{1}{\Delta A}\int \mathrm{d}A u(\vec{x},t) \approx \langle u(\vec{x},t)\rangle\ , \tag{1.5}$$

in which $\Delta A \ll A$ is an area small in comparison to the spatial size of the entire system but large enough such that sufficient individuals reside in it at all times such that the spatial average (left-hand side of (1.5)) is approximately equal to the expected density. The degree to which these assumptions are fulfilled determines the right choice of model. Two structurally different models reflect a range of possibilities.

On one hand, if $p(\vec{x},t)$ varies little in magnitude and the global density $\mathcal{N}/A$ is large enough, one can find a microscopic scale ΔA such that a sufficient amount of individuals are always contained in each microscopic unit

area for (1.5) to be valid. One a large scale one can then consider

$$n(\vec{x}, t) = \Delta A \langle u(\vec{x}, t) \rangle \tag{1.6}$$

a spatially continuous deterministic quantity and introduce dynamical equations for it.

Humans, however, are typically clustered in urban areas, cities, towns and villages in which the density of individuals is high as opposed to areas in between where is it negligible. In this case a metapopulation approach is more suitable. In this approach communities are defined by $p(\vec{x}, t)$ exceeding some threshold in some spatially compact area Ω_n and one labels these regions by a discrete index n. The size of each community n is given by

$$N_n(t) = \Omega_n \langle u(\vec{x}, t) \rangle \ . \tag{1.7}$$

In these models mobility of individuals is equivalent to exchange of them between the discrete set of communities. In metapopulation models $N_n(t)$ is typically considered a deterministic quantity for which (1.5) holds. The coupling of these communities is conveyed by mobility networks that quantify the exchange of individuals between them. Usually these traffic networks are quantified by a matrix $W_{nm} \geq 0$ whose elements reflect the traffic flux between communities.

1.2.2 The Lack of Scale in Human Mobility

By far the most studied human mobility system, particularly in the context of human infectious disease dynamics is the worldwide air transportation system, see Figure 1.1. The network is defined by a passenger flux matrix each element W_{nm} of which quantifies the number of passengers that travel between airport m and n. In a series of studies, air transportation networks were investigated using methods of complex network theory [4, 7, 12] and have been employed as the backbone in a set of models that attempt to account for the global spread of emergent human infectious diseases [5, 6, 13].

However, one of the central drawbacks of focusing on air transportation alone is that only long-range traffic is covered by it. If, for instance, one sets out to develop a model for disease dynamics on small to intermediate length scales, for example in countries such as Germany or the UK, air transportation does play a role, but an insignificant one compared to traffic on the network of highways and railways. Confronted with the difficulty of compiling a comprehensive dataset of human mobility covering all length scales, the idea was recently developed to employ proxies of human travel that indirectly provide information on mobility patterns of individuals. In [14] this idea was employed for the first time by analyzing the geographical circulation of bank

notes. In the study, data was analyzed which had been collected at the online bill tracker www.wheresgeorge.com founded by Hank Eskin in 1998. The idea of the game is simple. Individual dollars bills are marked and enter circulation. When new users come into possession of a marked bill, they can register at the site and report the current location of the bill by entering the zip code. Successive reports of a bill yield a spatio-temporal trajectory with a very high resolution. Since 1998 wheresgeorge.com has become the largest bill-tracking website worldwide with more than three million registered users and more than 140 million registered bills. Approximately 10% of all bills have had hits yielding a total of more than 14 million single trajectories consisting of origin $\vec{X}_1$ (initial entry location) and destination $\vec{X}_2$ (hit location). Figure 1.2 illustrates a sample of trajectories of bills with initial entries in five US cities. Shown are journeys of bills that lasted a week or less. Clearly, the majority of bills remains in the vicinity of their initial entry, yet a small but significant number of bills traversed distances of the order of the size of the US, consistent with the intuitive notion that short trips occur more frequently than long ones. One of the key results of the 2006 study was the first quantitative estimate of the probability $p(r)$ of a bill traversing a distance r in a short period of time, a direct estimate of the probability of humans performing journeys of this distance in a short period of time. This quantity is shown in Figure 1.2. This estimate was based on a dataset of 464 670 individual bills. On a range of distances between 10 and 3500 km, this probability follows an inverse power law, that is

$$p(r) \sim \frac{1}{r^{1+\mu}} \tag{1.8}$$

with an exponent $\mu \approx 0.6$. Despite the multitude of means of transportation involved, the underlying complexity of human travel behavior and the strong spatial heterogeneity of the United States, the probability follows this simple mathematical law indicating that human mobility is governed by underlying universal rules. Moreover, the specific functional form has important consequences. If one assumes that individual bills perform a spatial random walk with an arbitrary probability distribution $p(r)$ for distances at every step, one can ask: What is the typical distance $|\vec{X}(t)|$ from the initial starting point as a function of time? For ordinary random walks (Brownian motion) which are ubiquitous in the natural sciences, the behavior of $|\vec{X}(t)|$ is determined by the standard deviation $\sigma = \sqrt{\langle r^2 \rangle - \langle r \rangle^2}$ of the single steps and irrespective of the particular shape of the distance scales according to the "square-root law", that is $|\vec{X}(t)| \sim \sqrt{t}$, a direct consequence of the central limit theorem [15]. However, for a power law of the type observed in the dispersal of bank notes the variance diverges for exponents $\mu < 2$ and the situation is more complex. It implies that the dispersal of bank notes lacks a typical length scale, is fractal and the trajectories of bills are reminiscent of a particular class of random

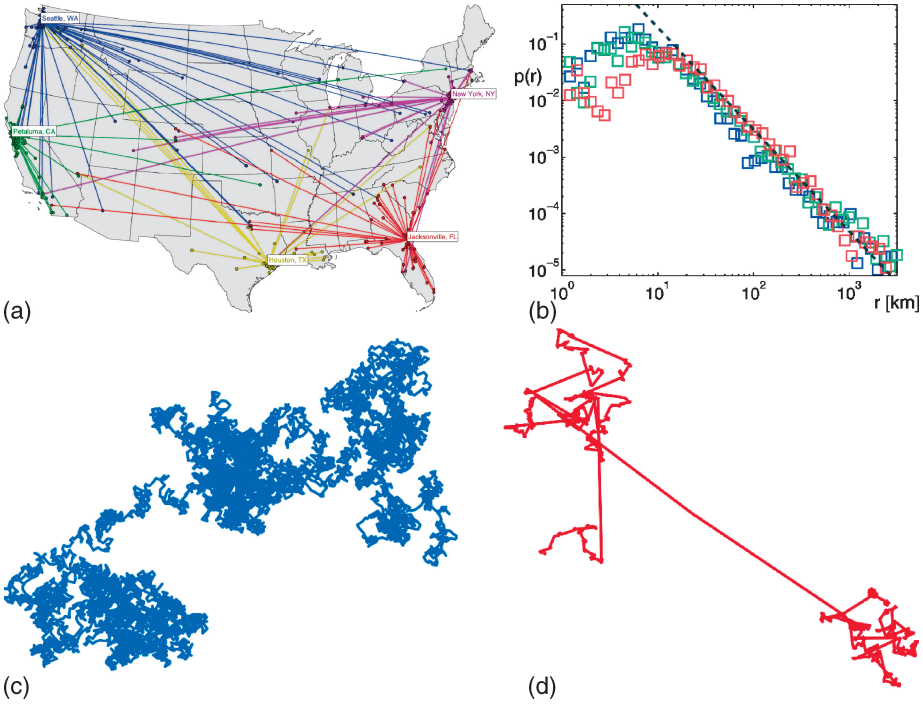

Figure 1.2 Short time trajectories of dollar bills in the United States. (a) Lines connect origin and destination locations of bills that traveled for less than a week. The majority of bills remain in the vicinity of their starting point, yet a small but significant fraction of bills travel long distances. (b) The probability $p(r)$ of traveling a distance r in a short period of time of T less than a week. The dashed line indicates the inverse power law of Equation (1.8) in the text. The colors encode the subsets of trajectories that started in large cities (blue), intermediate cities (green) and small towns (red). Despite systematic deviations for small distances, the asymptotic power law behavior is the same for all subsets indicating the universality of dispersal. (c) Two-dimensional trajectory of and ordinary random walk or Brownian motion. (d) Trajectory of a superdiffusive Lévy flight. The Lévy flight geometry consists of small clusters interconnected by long leaps. The dispersal of bank notes is reminiscent of Lévy flight trajectories such as the one depicted.

walks known as Lévy flights [16, 17]. Lévy flights, as opposed to ordinary random walks are anomalously diffusive, they exhibit a scaling relation that depends on the exponent:

$$|\vec{X}(t)| \sim t^{1/\mu} \,. \tag{1.9}$$

Because Lévy flights are superdiffusive, they disperse faster than ordinary random walks, and their geometrical structure differs considerable from ordinary random walks, see Figure 1.2. The discovery that the dispersal of bank notes and therefore human travel behavior lacks a scale and is related to Lévy flights was a major breakthrough in understanding human mobility on global

scales. This result is particularly intriguing because power laws of the type above and Lévy flight dispersal have been observed in foraging animals such as the albatross, deer and marine predators as well [18–20] and have since then been validated by a recent study on mobile phone dynamics [11], indicating that emergent mobility patterns are determined by similar underlying rules.

1.3 Statistical Properties and Scaling Laws in Multi-Scale Mobility Networks

Figure 1.3 illustrates a proxy network obtained from the flux of dollars in the United States, including all spatial scales. This network is defined by 3109 nodes (counties in the United States excluding Alaska and Hawaii) connected by weights W_{nm} that represent the flux rate of bills from county m to n in units of bills per day. The entire network structure is thus encoded in the 3109×3109 flux matrix $\vec{W}$. As each location has a well-defined geographical position, this multi-scale US traffic network can be visualized as a geographically embedded network as shown in the figure. Qualitatively, one can see that prominent East coast–West coast fluxes exist in the network. Yet the strongest connections are the short to intermediate length scale connections, as opposed to the air transportation network that serves long distance only. Although every day 2.35 million passengers travel on the US air transportation network, this represents only a small subset of the multi-scale traffic network. The histogram in Figure 1.3 illustrates these properties more quantitatively, comparing the relative frequency of distances in the multi-scale wheresgeorge network to the air transportation network. Clearly, the majority of distances served by air transportation, peaks around 1000 km, whereas distances in the multi-scale network are broadly distributed across a wide range from a few to a few thousand kilometers.

In order to understand human mobility on all spatial scales it is therefore essential to include all means of transportation indirectly involved in the wheresgeorge money circulation network. The bill circulation network quantified by the flux matrix can give important insight into the statistical features of human mobility across the United States. In order to quantify the statistical features of the network we will concentrate on the flux of bills in and out of a node given by

$$F_n^{\text{in}} = \sum_m W_{nm} \quad F_n^{\text{out}} = \sum_m W_{mn} \tag{1.10}$$

respectively. These flux measures are a direct proxy for the overall traffic capacity of a node in the network. Furthermore, we will investigate the in- and

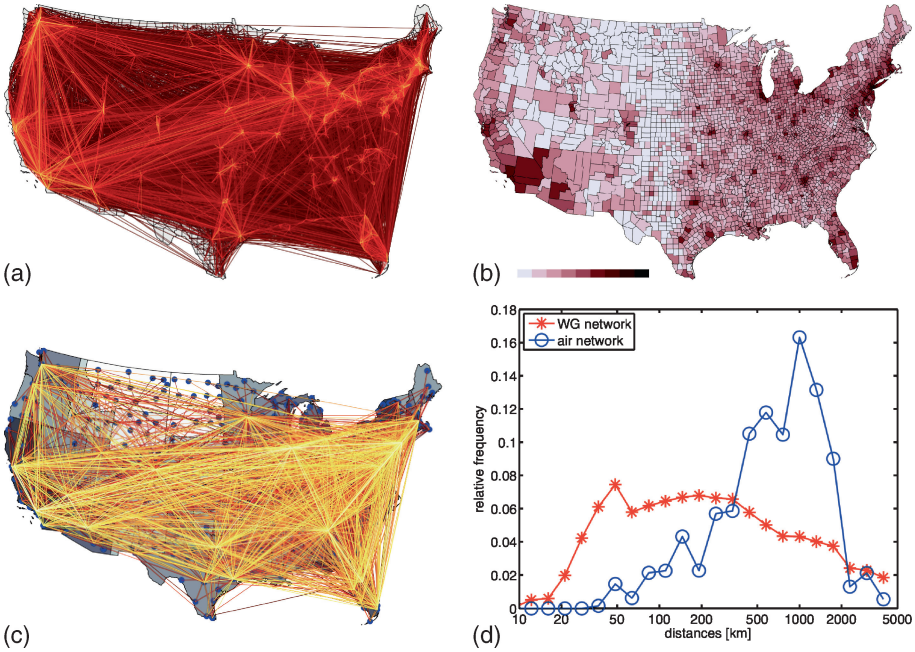

Figure 1.3 (a) The flux of dollar bills in the United States. Each line represents the flux of bills between the counties it connects. The color encodes the magnitude of the flux, bright lines indicate heavy flux, dark lines weak flux. The figure illustrates the strong heterogeneity of money dispersal, short distance connections typically exhibit strong fluxes, long distance connections are weaker but significant. (b) The population density of the United States spatially resolved and colored on a logscale. (c) The US air transportation network. The lines indicate connections between the 413 major airports in the US. The color encodes the magnitude of connections in passengers per day. (d) Relative frequency of distances in the multi-scale traffic network obtained from the wheresgeorge dataset (red) compared to the air-transportation network (blue). Air transportation mainly serves long distance whereas multi-scale traffic exhibits a broad distribution ranging from a few to a few thousand kilometers.

out-degree of a node defined according to

$$k_n^{\text{in}} = \sum_m A_{nm} \quad k_n^{\text{out}} = \sum_m A_{mn} \tag{1.11}$$

where the elements A_{nm} are entries of the adjacency matrix $\mathbf{A}$. These elements are either one or zero depending on whether or not nodes are connected. The degree of a node quantifies the connectivity of a node, that is to how many other nodes a given node is connected. A first important but expected feature of the multi-scale mobility network is its degree of symmetry. Figure 1.4 depicts the correlation of the flux of bills in and out of each node and a correlogram of the in- and out-degrees. These quantities exhibit a linear relationship

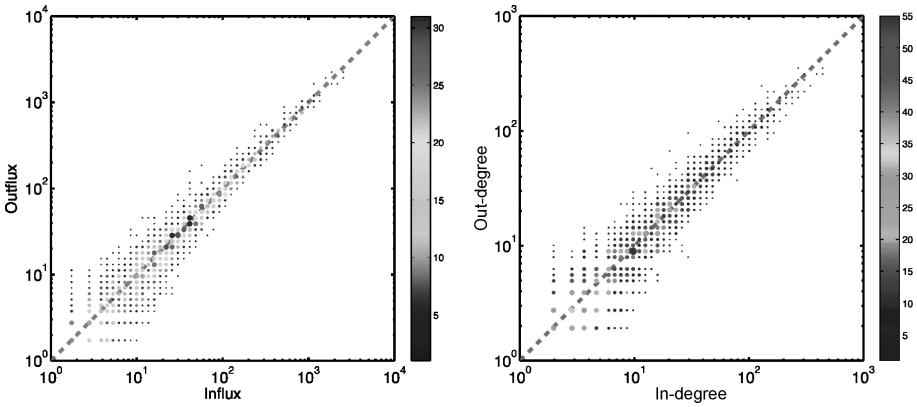

Figure 1.4 Symmetry of the money circulation network. The figures depict the correlation F_n^{in} and F_n^{out} of flux of bill in and out and the in- and out-degree k_n^{in} and k_n^{out} of a node n for all 3109 nodes in the network. The dashed lines represent the linear relationships.

subject to fluctuations,

$$F_n^{\text{in}} \approx F_n^{\text{out}} \quad k_n^{\text{in}} \approx k_n^{\text{out}} \tag{1.12}$$

indicated by the dashed lines in the figure. Note also that the magnitude of the flux values ranges over nearly four orders of magnitude, a first indication of the strong heterogeneity of the network.

This high degree of heterogeneity is further illustrated by the cumulative distributions of the weights, the fluxes and the degrees of all the nodes in the network as depicted in Figure 1.5. All quantities are broadly distributed across a wide range of scales. Very similar broad distributions have been observed in studies of the air transportion networks [4, 7, 12]. A very important issue in transportation theory is the development of a plausible evolutionary mechanism that can account for the emergence of these distributions; a task that has not been accomplished so far. There is no plausible "theory" for human traffic networks, as of today, that predicts the precise functional form of the distributions shown in Figure 1.5.

1.3.1 Scaling Laws in the Topological Features of Multi-Scale Transportation Networks

In order to reveal additional structure in multi-scale human mobility networks we investigated the functional relation of the quantities defined above; that is, what is the functional relation of fluxes and degrees with respect to the population size of a node? Figure 1.6 illustrates the statistical relationship between the population size of a node and the flux of bills into a node. The dashed line

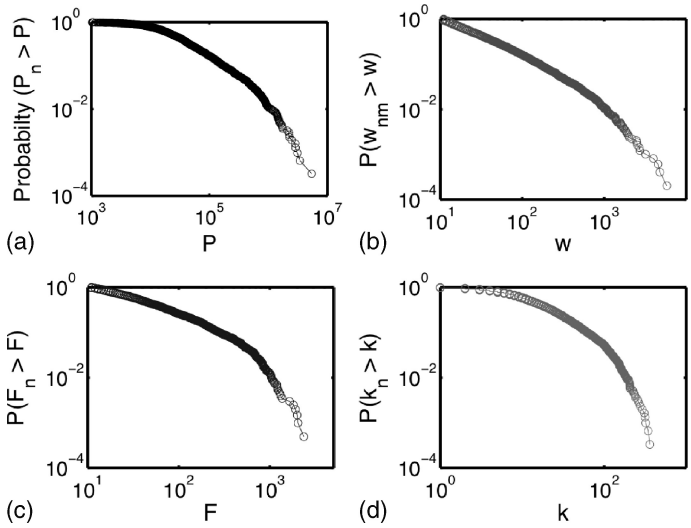

Figure 1.5 Heterogeneity of multi-scale human mobility networks. Cumulative probability distributions of the population size of the nodes (a), the weight matrix elements W_{nm} (b), the flux of bills F_n in and out of nodes, see (c) and the degree k_n of the nodes (d). The broadness of these distributions is a consequence of the strong heterogeneity of the network.

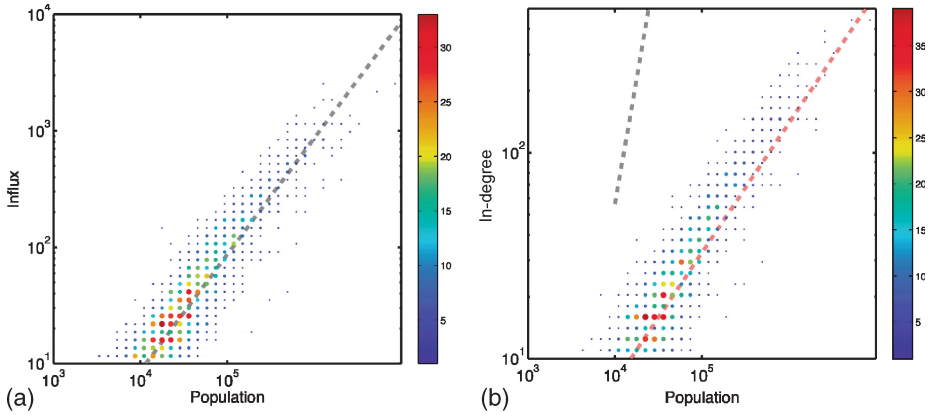

Figure 1.6 The functional dependence of influx F^{in} (a) and in-degree k^{in} (b) on the population size P of a node. The flux of bill depends linearly on the population size (gray dashed line), whereas the degree exhibits a sublinear dependence (pink dashed line).

in the figure represents a linear relationship with slope one, indicating that traffic through a node grows linearly with the population size.

$$F(P) \sim P \tag{1.13}$$

Intuitively, this is expected, as the larger the population of a node the more traffic flows in and out of it. However, correlating the degree of a node against the population size indicates a sublinear relationship:

$$k(P) \sim P^{\xi} \tag{1.14}$$

with an exponent $\xi \approx 0.7$, contrasting the intuitive notion that the connectivity of a node also grows linearly with population size. From the scaling relations (1.13) and (1.14) we can determine an important property of multi-scale mobility networks. The typical strength of a connection is given by the ratio of flux and degree and one obtains heuristically

$$W \sim P^{1-\xi} \tag{1.15}$$

This implies that larger counties are not only connected to a larger number of other counties but also that the typical strength of every connection is stronger. Both relations are determined by the universal exponent $\xi = 0.7$ and these relations hold over nearly four orders of magnitude, a surprising regularity exhibited by the multi-scale mobility network. Again, no theory exists that can account for these scaling relations and the value of the exponent.

1.4 Spatially Extended Epidemic Models

In summary, two prominent features of multi-scale human mobility networks emerged in the analysis above. (1) Networks exhibit a strong heterogeneity, the distribution of weights, traffic fluxes and populations sizes of community range over many orders of magnitude. (2) Although the interaction magnitude in terms of traffic intensities decreases with distance, the observed power laws indicate that long-range interactions play a significant role in spatial disease dynamics. In the models to be discussed below, we will introduce a class of spatially extended models in which the impact and interplay of both spatial heterogeneity and long-range spatial interactions can be investigated in a systematic fashion. It will also become clear that another key issue in spatial disease dynamics is the translation of topological features of transportation networks, that is the flux matrix **W** into dynamical entities which generate the dispersal in space. At first glance, this may seem a straightforward process. However, as we will see, this is a nontrivial task, and the behavior of a spatially extended epidemic model depends sensitively on the precise choice of

method of translating the topology of a transportation network into dynamics. To understand this, we first review some of the paradigmatic models for disease dynamics in a single population.

1.4.1 Disease Dynamics in a Single Population

One of the simplest models for an epidemic in a single population is the SIR model [21]. In this model a population of N individuals is classified according to infectious state, that is a person can be susceptible (S) to the disease, infected (I) by the disease, and recovered (R) from the disease. Recovered individuals are assumed to have aquired immunity to the disease and can no longer be infected. Each individual in a population may undergo the transition

$$S \to I \to R \tag{1.16}$$

during the time course of an infection. The dynamics of an epidemic is governed by only two reactions:

$$S + I \xrightarrow{\alpha} 2I \tag{1.17}$$

$$I \xrightarrow{\beta} R\,, \tag{1.18}$$

a contact-initiated disease transmission and the recovery from disease, respectively. Models of this type are known as compartmental models, because a population is divided into different compartments defining the state and individuals of various compartments which interact by a set of reactions. A key assumption in the SIR model and, in fact, most single population compartmental models is the mixing assumption. It means that: (1) all individuals of a given class are identical in their behavior; (2) independent of one another; and (3) reactions between a given pair of individuals occurs with the same likelihood as a reaction of any other pair.

The structure of compartmental models is very similar in nature to chemical reactions, in fact one usually employs the mass-action principle to derive ordinary differential equations for the dynamics of the number of susceptibles, infected and recovereds. At any point t in time the probability that an infected individual recovers in $[t, t + \Delta t]$ is assumed to be constant and proportional to Δt. The change in infecteds and recovereds is thus

$$\Delta I = -\Delta R \approx -\beta \Delta t\,. \tag{1.19}$$

The probability that an infected succesfully transmits the disease to a susceptible in Δt is given by

$$P = \Delta t \times \sigma \times T \times \frac{S}{N}\,, \tag{1.20}$$

where σ is the contact rate between individuals, T the transmission probability and S/N the probability that the contact made by the infected is with a susceptible individual. This yields

$$-\Delta S = \Delta I \approx \alpha \frac{S\,I}{N} \Delta t \tag{1.21}$$

where $\alpha = \sigma T$ is the force of infection, that is the effective transmission rate. For the SIR model this yield the following system of nonlinear ordinary differential equations (ODEs):

$$\begin{aligned} \partial_t S &= -\alpha \frac{S\,I}{N} \\ \partial_t I &= \alpha \frac{S\,I}{N} - \beta I \\ \partial_t R &= \beta I\,. \end{aligned} \tag{1.22}$$

We can define fractions $s = S/N$, $j = I/N$ and $r = R/N$ and noting that $S(t) + I(t) + R(t) = N$ (i.e. the population size is conserved) we obtain the SIR model in its canonical form [22]:

$$\begin{aligned} \partial_t s &= -\alpha s j \\ \partial_t j &= \alpha s j - \beta j \\ r &= 1 - s - j\,. \end{aligned} \tag{1.23}$$

The key parameter in the SIR model is the basic reproduction number

$$R_0 = \frac{\alpha}{\beta} = \frac{T_{\text{recovery}}}{T_{\text{contacts}}}$$

the ratio of the force of infection and recovery rate. It is the average number of secondary infections caused by one infected individual in the time that individual is infected, on average. When $R_0 > 1$, a population with an initially small fraction of infecteds will be subject to an epidemic: a fast exponential increase and a subsequent decay of $j(t)$, see Figure 1.7. When $R_0 < 1$ no epidemic occurs. The basic reproduction number is thus a threshold parameter.

1.4.1.1 The SIS Model

In the SIS model the second reaction scheme (1.18) is replaced by $I \to S$, infected individuals do not aquire immunity but rather recover from the disease to become susceptible again. This model lacks the R class and is governed by only one ODE for the infecteds

$$\partial_t j = \alpha j(1 - j) - \beta j \tag{1.24}$$

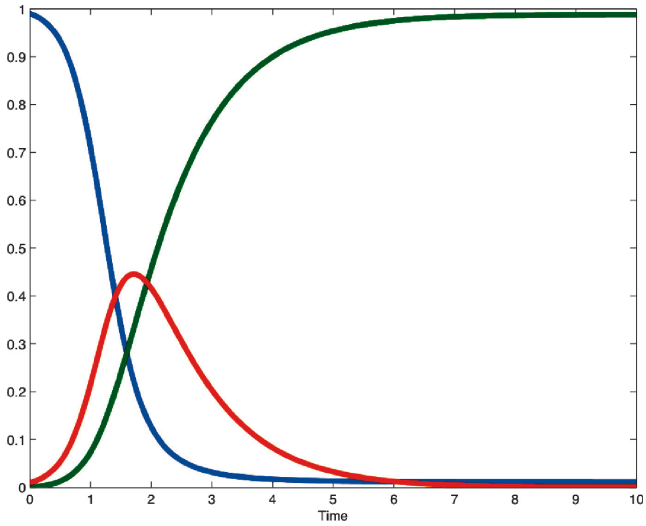

Figure 1.7 Time evolution of the SIR model as defined by (1.23). Parameters are $\beta = 1$ and $R_0 = 4.5$. The time course of the fraction of infecteds, susceptibles and recovereds are shown in red, blue and green, respectively. The initial condition was $j(0) = 0.01$, $s(0) = 1 - j(0)$ and $r(0) = 0$.

where the conservation of individuals $s = 1 - j$ is assumed. For $R_0 = \alpha/\beta > 1$ the SIS model evolves to a stable stationary state given by

$$j^{\mathrm{s}} = 1 - \frac{1}{R_0} ,$$

in which a fraction j^{s} of the population is infected, the disease is endemic. The SIS model is a useful system for investigating the impact of space on disease dynamics and we will discuss the spatially extended SIS model in the next section.

1.5 Spatial Models

In the heart of all spatial models is the motivation to forsake the assumption of homogenous mixing of individuals and incorporate the fact that individuals belonging to different populations exhibit different interaction probabilities and that they are mobile in space. The conceptual tool underlying the development of spatial models is that of a metapopulation. A metapopulation is a set $m = 1, \ldots, M$ of populations of size N_n. The total number of individuals

of the metapopulation is

$$N = \sum_{n=1}^{M} N_n \,. \tag{1.25}$$

It is usually assumed that the dynamics in each population is governed by dynamics that adhere to homogeneous mixing but interaction of individuals between populations are governed by additional laws. The most important of these interactions for disease dynamics is the random exchange of individuals between populations. The most straightforward generalization of the SIS model, including metapopulations, is given by:

$$\begin{aligned} S_n + I_n &\xrightarrow{\alpha} 2I_n \\ I_n &\xrightarrow{\beta} S_n \\ S_n &\xrightarrow{w_{mn}} S_m \\ I_n &\xrightarrow{w_{mn}} I_m \end{aligned} \tag{1.26}$$

In addition to the first two reactions, that is ordinary SIS dynamics in each population n, susceptibles and infecteds can randomly move between population m and n, the rate of which is governed by the probability rate w_{mn}. The assumption in this model is that individuals of all types randomly travel between populations in the same fashion. The set of ODEs governing disease dynamics is then given by a set of $2M$ coupled ODEs:

$$\begin{aligned} \partial_t S_n &= -\alpha \frac{S_n I_n}{N_n} + \beta I_n + \sum_{m \neq n} \left[w_{nm} S_m - w_{mn} S_n \right] \\ \partial_t I_n &= \alpha \frac{S_n I_n}{N_n} - \beta I_n + \sum_{m \neq n} \left[w_{nm} I_m - w_{mn} I_n \right] \,. \end{aligned} \tag{1.27}$$

The total rate of leaving a node n is given by $\sum_{m \neq n} w_{mn}$ and the expected time an individual remains in a population n is

$$\langle T_n \rangle = \frac{1}{\sum_{m \neq n} w_{mn}} \,. \tag{1.28}$$

Note, that in the metapopulation system, the number $N_n(t) = S_n(t) + I_n(t)$ of individuals in each subpopulation is generally time-dependent, in fact, adding the ODEs pairwise we obtain

$$\partial_t N_n = \sum_{m \neq n} \left[w_{nm} N_m - w_{mn} N_n \right] \,. \tag{1.29}$$

In most models it is usually assumed that the system is equilibrated with respect to dispersal, that is N_n does not change over time and is therefore equal

to the fixed point of (1.29), that is

$$N_n(t) = N_n^s = C_n = \text{const.} \tag{1.30}$$

In the following we will refer to the stationary population size of node n as the capacity C_n. In equilibrium the flux of individuals from n to m balances that of m to n (detailed balance condition):

$$w_{nm}C_m = w_{mn}C_n \ . \tag{1.31}$$

In this case the spatial SIS model (1.27) reduces to a set of M coupled ODEs for the fraction of infecteds in each population:

$$\partial_t j_n = \alpha j_n(1 - j_n) - \beta j_n + \sum_{m \neq n} [w_{nm}j_m - w_{mn}j_n] \tag{1.32}$$

with $j_n = I_n/C_n$. The system defined by (1.32) is an example of an infectious disease dynamical system extended to the metapopulation level. A large class of contemporary models for spatial disease dynamics are related to it in structure [6, 9, 23, 24]. One of the key difficulties in theoretical epidemiology are: (1) the identification of effective communities of populations that make up a metapopulation; and (2) the quantitative assessment of traveling rates w_{nm} between these populations. Note that the introduction of populations n making up the metapopulation did not specify spatial locations. In the dynamical system (1.32), the relation between communities is solely defined by the dynamical coupling ω_{nm}. In most models, however, all communities are typically embedded in space such that each population n has a well defined geographical location $\vec{x}_n$. One can then use the geographical information to make and test assumptions as to how the exchange rates w_{nm} depend on geography. One of the most popular assumptions in this context is that the flux of individuals between two communities depends on their size and their distance. The total flux of individuals in equilibrium from community m to n and vice versa is given by the left- and the right-hand side of the detailed balance condition (1.31), respectively. In the majority of models it is assumed that the flux F_{nm} increases with the capacities (i.e. the stationary size of the populations) C_m and C_n and decreases monotonically with the geographical distance between them, that is

$$F_{nm} = \omega_0(C_mC_n)^{\xi}G(|\vec{x}_n - \vec{x}_m|) = F_{mn} \tag{1.33}$$

with $0 \leq \xi \leq 1$. The function G takes care of the dependence on distance. Depending on the type of metapopulation and dynamical context, this kernel can be exponential, gaussian or show algebraic deay with x. Using the relation $F_{nm} = w_{nm}C_m$ between absolute flux and probability rates in equilibrium, (1.33) implies for the hopping rate

$$w_{nm} = w_0C_n^{\xi} \times G(|\vec{x}_n - \vec{x}_m|) \times C_m^{\xi-1} \ . \tag{1.34}$$

Inserted into the rate equation (1.29) one can check that C_n is the equilibrium community size. In epidemiological contexts, spatial communities often reflect cities, towns and villages. The specific choice of $G(x)$ put forth by (1.8) or human mobility which is frequently used is the power-law decay

$$G(x) \sim x^{-1-\mu} \tag{1.35}$$

when this is inserted into (1.33) and (1.34) it gives

$$\omega_{nm} = \omega_0 \frac{C_n^{\xi} \times C_m^{\xi-1}}{|\vec{x}_n - \vec{x}_m|^{D+\mu}} \tag{1.36}$$

where $D = 2$ is the spatial dimension. The parameter ξ quantifies the impact of origin and destination in the travelling event $m \to n$.

- When $\xi = 1$ we have

  $$w_{nm} \propto C_n \quad \text{and} \quad F_{nm} \propto C_n C_m \tag{1.37}$$

  This implies that the rate is independent of properties of the origin and the flux is proportional to the size of both communities.

- When $\xi = 0$ we have

  $$w_{nm} \propto \frac{1}{C_m} \quad \text{and} \quad F_{nm} \propto 1 \tag{1.38}$$

  that is the rate of traveling to destination n is independent of properties of the destination and the flux is independent of community sizes of both places.

- An interesting system is the symmetric case when $\xi = 1/2$. This implies that

  $$w_{nm} \propto \sqrt{C_n/C_m} \quad \text{and} \quad F_{nm} \propto \sqrt{C_n C_m}\,. \tag{1.39}$$

  In this situation, the rate w_{nm} is independent of scaling the entire metapopulation size uniformly by some factor and the flux is the geometric mean of the community sizes of origin and destination. That implies, for example, that if we scale the entire population size by $C_n \to 2C_n$ this also scales the flux by a factor of two.

1.5.1 Continuity Limit and Fractional Transport

With the definition of the rate according to (1.36) the dispersal of individuals is given by

$$\partial_t N_n = w_0 \sum_{m \neq m} \left[\frac{C_n^{\xi} C_m^{\xi-1}}{|\vec{x}_n - \vec{x}_m|^{1+\mu}} N_m - \frac{C_m^{\xi} C_n^{\xi-1}}{|\vec{x}_m - \vec{x}_n|^{1+\mu}} N_n \right]$$

with $0 < \mu$. A useful insight into the properties of this master equation can be gained by performing a continuity limit. Letting $\vec{x}_n$ be points on a grid of microscopic areas ΔA and $N_n(t) = n(\vec{x}_n, t)\Delta A$, $C_n = c(\vec{x}_n)\Delta A$, the above equation becomes

$$\partial_t n(\vec{x}, t) = w_0 \lim_{\Delta A \to 0} \int_{\vec{y} \notin \Delta A} \mathrm{d}\vec{y} \frac{c^\xi(\vec{x}) c^{\xi-1}(\vec{y}) n(\vec{y}, t) - c^\xi(\vec{y}) c^{\xi-1}(\vec{x}) n(\vec{x}, t)}{|\vec{x} - \vec{y}|^{2+\mu}} \quad (1.40)$$

The integral is over all points outside of an area centered at $\vec{x}$. One has to be careful when carrying out this limit, because of the divergent denominator. In fact, the rate $m \to n$ was originally only defined for interacting communities $n \neq m$ and it is meaningless for $n = m$. One can, however, carry out the limit $\Delta A \to 0$ and interpret the integral as a Cauchy integral. The limit of the rhs of (1.40) then depends sensitively on the value of the exponent μ. For $\mu > 2$ one obtains [25, 26]

$$\partial_t n = D_0 \left[c^\xi \Delta c^{\xi-1} n - c^{\xi-1} n \Delta n^\xi \right] , \quad (1.41)$$

with $n = n(\vec{x}, t)$ and $c = c(\vec{x})$ and $\Delta = \partial_x^2$. This implies that when the exponent μ exceeds the critical value $\mu_c = 2$ the process becomes a diffusion process in the limit above. However, this diffusion process evolves in a heterogeneous environment determined by the function $c(\vec{x})$.

If $\mu < 2$ as, for example, observed in the dispersal of bank notes (in that case $\mu \approx 0.6$) the limit yields

$$\partial_t n = D_0 [c^\xi \Delta^{\mu/2} c^{\xi-1} n - c^{\xi-1} n \Delta^{\mu/2} n^\xi] \quad (1.42)$$

where the operator $\Delta^{\mu/2}$ is known as the fractional Laplacian, a non local singular operator defined by

$$(\Delta^{\mu/2} f)(\vec{x}) = C_\mu \int \mathrm{d}\vec{y} \frac{f(\vec{y}) - f(\vec{x})}{|\vec{x} - \vec{y}|^{D+\mu}} \quad (1.43)$$

where C_μ is a constant and D is the spatial dimension [27, 28]. The reason why $\Delta^{\mu/2}$ is refered to as a fractional derivative is that in Fourier space it exhibits a particularly simple form, a multiplication by $-|\vec{k}|^\mu$. Equations of the type (1.42) are known as fractional diffusion equations and have been employed in a number of physical, biological and chemical systems [29–32]. Ranging from anomalous diffusion of protein motion on folded polymers to human eye movements [27, 28, 33]. The derivation above relates dispersal of individuals in metapopulations for the first time to fractional diffusion equations, an approach that may well prove to be valuable in the future.

1.5.2 Limiting Cases

Before re-inserting the dispersal component into the original spatial SIS model, it is worthwhile considering known marginal cases of the general fractional diffusion equation (1.42). For example, when $\mu = 2$ and $c(\vec{x}) = 1$, the dynamics equation reduces to

$$\partial_t n = D_0 \Delta n \,, \tag{1.44}$$

that is ordinary diffusion in a homogeneous environment. When $\mu = 2$ but $c(\vec{x})$ is a variable function of position, that is (1.42) is the same as (1.41), the dispersal is governed by a Fokker–Planck equation

$$\partial_t n = -\nabla F\, n + \tfrac{1}{2} \Delta D\, n \tag{1.45}$$

which is equivalent to (1.41) and force and the diffusion coefficients $F = F(\vec{x})$ and $D = D(\vec{x})$, respectively, are related to the heterogeneity function $c(\vec{x})$. This relation depends, of course, on the value of the parameter ξ. For example, when the system is origin-driven, that is when $\xi = 0$, then (1.41) reduces to

$$\partial_t n = D_0 \Delta n / c \,, \tag{1.46}$$

a Fokker–Planck equation with a space-dependent diffusion coefficient

$$D(\vec{x}) = \frac{D_0}{c(\vec{x})} \,, \tag{1.47}$$

which is inversely proportional to the stationary population density $c(\vec{x})$. This means that, in this system, diffusion is high in regions where the population is small and vice versa. In the destination-driven system $\xi = 1$, we obtain a Fokker–Planck equation with

$$D(\vec{x}) = 2D_0 c(\vec{x}) \quad \text{and} \quad F(\vec{x}) = 2D_0 \nabla c(\vec{x}) \,. \tag{1.48}$$

in which diffusion increases with population density but, more importantly, a nonzero drift towards regions with higher population density is introduced. When $\mu = 1/2$, that is the impact of the origin and destination are the same, the diffusion coefficient is contant and the force term is given by

$$F(\vec{x}) = D_0 \nabla \log c(\vec{x}) \,. \tag{1.49}$$

One can see that it is only in this situation that the dynamics does not change when the population density $c(\vec{x})$ is scaled uniformly by a factor. In this case $-\log c(\vec{x})$ can be considered a potential $V(\vec{x})$ of the system, with minima in densely populated areas and maxima in weakly populated ones.

The most interesting case, and certainly the one closest to reality, is the general case when the dynamical system is fractionally diffusive and spatially heterogeneous. The combination of the rhs of (1.42) with the spatial SIS model of (1.32) gives the spatially extended fractional SIS model.

$$\partial_t j = \alpha j(1-j) - \beta j + D_0 c^{\xi-1}[\Delta^{\mu/2} c^{\xi} j - j\Delta^{\mu/2} c^{\xi}] \,. \tag{1.50}$$

The spatial SIR model or related systems that differ in the local dynamics can be derived analogously; for instance, the spatial SIR model is given by

$$\begin{aligned} \partial_t s &= -\alpha j s + D_0 c^{\xi-1}[\Delta^{\mu/2} c^{\xi} s - s\Delta^{\mu/2} c^{\xi}] \,, \\ \partial_t j &= \alpha j s - \beta j + D_0 c^{\xi-1}[\Delta^{\mu/2} c^{\xi} j - j\Delta^{\mu/2} c^{\xi}] \,. \end{aligned} \tag{1.51}$$

The key question is: What are the general properties of the solutions to these reaction-fractional diffusion equations? How do the solutions depend on the parameters $0 \leq \xi \leq 1$ and $0 < \mu \leq 2$? And what are approximate choices for these parameters for real epidemics?

To address the first question: solutions of three variants of the spatial SIR model are depicted in Figure 1.8. One system is spatially homogeneous and dispersal is by ordinary diffusion. The solution exhibits traveling wavefronts that propagate at constant speeds, a fact known for similar systems such as the Fisher equation. In fact, a spatially homogeneous SIR variant was employed to estimate the speed of propagation of the black death in Europe in the 14th century. The second simulation is a system with some degree of spatial heterogeneity, that is $c(\vec{x})$ is variable but $\mu = 2$. As in the spatial homogeneous system, solutions to the spatial SIR model still exhibit traveling well-defined wavefronts that exhibit some irregularity imposed by the spatial heterogeneity. However, the key feature of a wave-front propagating with a constant speed remains unchanged.

If, however, one introduces nonlocal dispersal by choosing a value $\mu < 2$, the overall statistical features of the spreading pattern change drastically. Instead of a well-shaped wavefront, the pattern exhibits localized islands in the time course of the epidemic. This behavior is a direct consequence of the interplay of the spatial heterogeneity and the nonlocal superdiffusive nature of dispersal incorporated in the fractional SIR model (1.51).

The last questions can be answered by a comparison with the empirical results presented above. The fact that the flux of dollar bills into nodes is proportional to the population size suggests that human dispersal is destination driven; see, for example (1.37), and that $\xi = 1$. The power law in the short-time dispersal probability for the distance, that is $p(r)$ in (1.8) implies that $\mu \approx 0.6$. With these parameters, and the equilibrium distribution of individuals in a large geographical area, we can investigate the spreading pattern in

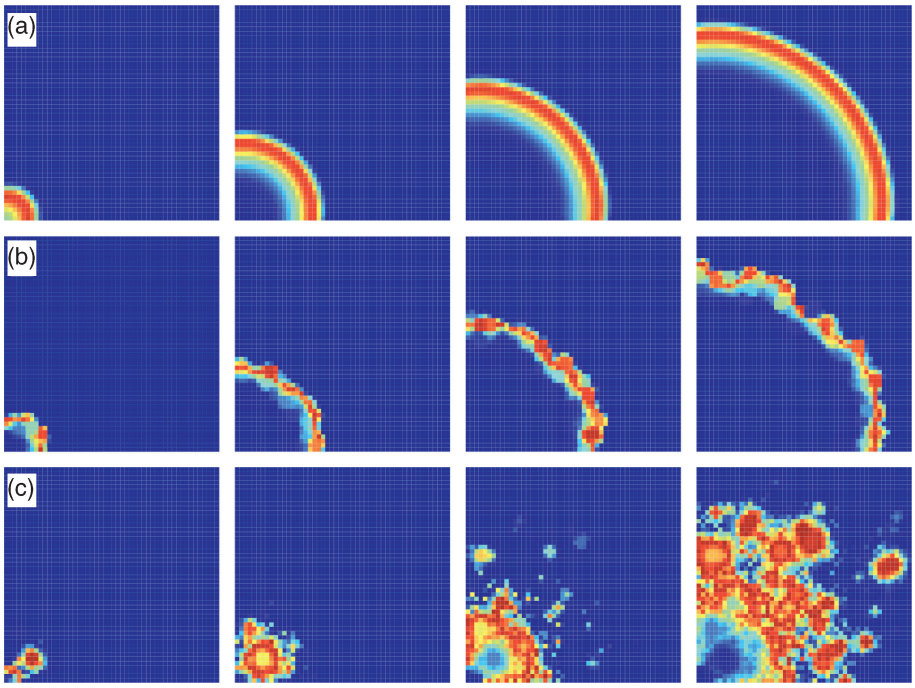

Figure 1.8 Snapshots of a two-dimensional spatial extended SIR model. (a) A spatially homogeneous system with $c(\vec{x}) = \text{const.}$ and ordinary diffusion in space. This system exhibits a propagating front at constant speed. (b) The same as above but with spatial heterogeneity. The heterogeneity induces randomness in the shape of the wavefront but introduces no qualitatively different patterns. (c) The fractional SIR model with heterogeneity. The combination of scale-free diffusion and heterogeneity introduces a novel type of spatiotemporal pattern with fractal properties.

a real geographical context. Results are shown in Figure 1.9 for a fractional SIS model with parameters $\mu = \mu_h \approx 0.6$ in the United States for an initial outbreak in Washington, DC. For $c(\vec{x})$ we chose the population density of the counties in the United States. In comparison to a system with only local dispersal, the fractional SIS systems shows a pattern similar in structure to the idealized system of a square grid (i.e. Figure 1.8). For instance, well before the bulk of the epidemic reaches the midwest, the disease has already almost reached its maximum on urban areas on the West Coast. Dispite its structural simplicity and the crude assumptions made on the course of deriving the fractional SIS model, these spreading patterns are strikingly similar to recently published large-scale, agent-based simulation studies on the most likely spread of new human influenza H5N1 subtype in the United States.

Figure 1.9 Simulation of the fractional, spatially extended SIS model (b) in the United States compared to a system with ordinary diffusion (a). Each column represents a snapshop of the time evolution of both systems. The initial outbreak in the simulations was in Washington DC and parameters of the dispersal were $\xi = 1$ (destination driven) and $\mu = 0.6$ (superdiffusion, (b)) and $\mu = 2$ (ordinary diffusion, (a)). The patterns exhibit the qualitative behavior of the idealized system shown in Figure 1.8.

Although these results are promising, from a theoretical point of view little is known about the general properties of fractional and heterogeneous reaction diffusion equations such as (1.50) and (1.51). This is primarily due to the fact that these equations are difficult to solve numerically and the analytical tools for investigation are currently underdeveloped. The richness of the possible applications of this approach, not only in spatial epidemiology, leads us to believe that in the near future novel and interesting properties of fractional diffusion systems in heterogeneous environments will be discovered and will find their intentification in natural systems.

References

1 Xia, Y.C., Bjørnstad, O.N. and Grenfell, B.T. (2004) Measles metapopulation dynamics: A gravity model for epidemiological coupling and dynamics. *Am. Nat.*, **164**(2), 267–281.

2 Ferguson, N.M., Cummings, D.A.T., Cauchemez, S., Fraser, C., Riley, S., Meeyai, A., Iamsirithaworn, S. and Burke, D.S. (2005) Strategies for containing an emerging influenza pandemic in southeast Asia. *Nature*, **437**(7056), 209–214.

3 Ferguson, N.M., Cummings, D.A.T., Fraser, C., Cajka, J.C., Cooley, P.C. and Burke, D.S. (2006) Strategies for mitigating an influenza pandemic. *Nature*, **442**(7101), 448–452.

4 Dall'Asta, L., Barrat, A., Barthelemy, M. and Vespignani, A. (2006) Vulnerability of weighted networks. *J. Stat. Mech. Theory E*, P04006

5 Colizza, V., Pastor-Satorras, R. and Vespignani, A. (2007) Reaction-diffusion processes

and metapopulation models in heterogeneous networks. *Nat. Phys.*, **3**(4), 276–282.

6 Colizza, V., Barrat, A., Barthelemy, M. and Vespignani, A. (2006) The role of the airline transportation network in the prediction and predictability of global epidemics. *P. Natl. Acad. Sci. USA*, **103**(7), 2015–2020.

7 Barrat, A., Barthelemy, M. and Vespignani, A. (2005) The effects of spatial constraints on the evolution of weighted complex networks. *J. Stat. Mech. Theory E*, P05003.

8 Noble, J.V. (1974) Geographic and temporal development of plagues. *Nature*, **250**(5469), 726–728.

9 Grenfell, B.T., Bjørnstad, O.N. and Kappey, J. (2001) Travelling waves and spatial hierarchies in measles epidemics. *Nature*, **414**(6865), 716–723.

10 Grenfell, B.T., Bjørnstad, O.N. and Finkenstädt, B.F. (2002) Dynamics of measles epidemics: Scaling noise, determinism, and predictability with the TSIR model. *Ecol. Monogr.*, **72**(2), 185–202.

11 Gonzalez, M.C., Hidalgo, C.A. and Barabasi, A.-L. (2008) Understanding individual human mobility patterns. *Nature*, **453**(7196), 779–782.

12 Guimera, R. and Amaral, L.A.N. (2004) Modeling the world-wide airport network. *Eur. Phys. J. B*, **38**(2), 381–385.

13 Hufnagel, L., Brockmann, D. and Geisel, T. (2004) Forecast and control of epidemics in a globalized world. *P. Natl. Acad. Sci. USA*, **101**(42), 15124–15129.

14 Brockmann, D., Hufnagel, L. and Geisel, T. (2006) The scaling laws of human travel. *Nature*, **439**(7075), 462–465.

15 Gardiner, C.W. (1985) *Handbook of Stochastic Methods*, Springer Verlag, Berlin.

16 Metzler, R. and Klafter, J. (2000) The random walks guide to anomalous diffusion: A fractional dynamics approach. *Phys. Rep.*, **339**:1–77.

17 Shlesinger, M.F., Zaslavsky, G.M. and Frisch, U. (eds) (1995) *Lévy Flights and Related Topics in Physics*, Lecture Notes in Physics, Springer Verlag, Berlin.

18 Viswanathan, G.M., Afanasyev, V., Buldyrev, S.V., Murphy, E.J., Prince, P.A. and Stanley, H.E. (1996) Lévy flight search patterns of wandering albatrosses. *Nature*, **381**:413–415.

19 Heinrich, B. (1979) Resource heterogeneity and patterns of movement in foraging bumblebees. *Oecologia*, **40**:235–245.

20 Ramos-Fernandéz, G., et al. (2004) Lévy walk patterns in the foraging movements of spider monkeys (Ateles geoffroyi). *Behavioral Ecology and Sociobiology*, **55**(3):223–230.

21 Anderson, R.M. and May, R.M. (1979) Population Biology of Infectious Diseases. *Nature*, **280**(5721), 361–367.

22 Kermack, W.O. and McKendrick, A.G. (1932) Contributions to the mathematical theory of epidemics II – the problem of endemicity. *P. R. Soc. Lond. A-Conta*, **138**(834), 55–83.

23 Colizza, V. and Vespignani, A. (2008) Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *J Theor Biol*, **251**(3), 450–467.

24 Hufnagel, L., Brockmann, D. and Geisel, T. (2004) Forecast and control of epidemics in a globalized world. *PNAS*, **101**:15124–15129.

25 Belik, V.V. and Brockmann, D. (2007) Accelerating random walks by disorder. *New J. Phys.*, **9**:54.

26 Brockmann, D. and Sokolov, I.M. (2002) Lévy flights in external force fields: From models to equations. *Chem. Phys.*, **284**:409–421.

27 Brockmann, D. and Geisel. T (2003) Particle dispersion on rapidly folding random hetero-polymers. *Phys. Rev. Lett.*, **91**:048303.

28 Brockmann, D. and Geisel, T. (2003) Lévy flights in inhomogeneous media. *Phys. Rev. Lett.*, **90**(17), 170601.

29 Barkai, E., Metzler, R. and Klafter, J. (2000) From continuous time random walks to the fractional Fokker–Planck equation. *Phys. Rev. E*, **61**(1), 132–138.

30 Barkai, E. (2001) Fractional Fokker–Planck equation, solution, and application. *Phys. Rev. E*, **63**:46118.

31 Saichev, A.I. and Zaslavsky, G.M. (1997) Fractional kinetic equations: solutions and applications. *Chaos*, **7**(4), 753–764.

32 Metzler, R. and Klafter, J. (2000) Boundary value problems for fractional diffusion equations. *Physica A*, **278**, 107–125.

33 Brockmann, D. and Geisel, T. (2000) The ecology of gaze shifts. *Neurocomputing*, **32–33**, 643–650.