TESIS DE LA UNIVERSIDAD DE ZARAGOZA
2023
13

Adriana Reyna Lara

Control of diffusion of infectious diseases in complex networks

Director/es
Gómez Gardeñes, Jesús

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

ISSN 2254-7606

Universidad Zaragoza
1542

Tesis Doctoral

CONTROL OF DIFFUSION OF INFECTIOUS DISEASES IN COMPLEX NETWORKS

Autor

Adriana Reyna Lara

Director/es

Gómez Gardeñes, Jesús

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

Programa de Doctorado en Física

2022

Control of diffusion of infectious diseases in complex networks

Adriana Reyna Lara

September, 2022

A mis padres,
Rubén & Gaby.

Acknowledgments

Primero y antes que nada quiero agradecer a mi supervisor, Gracias Jesús! Gracias por ser el líder que eres, por tu critica constructiva y tu paciencia. Gracias por toda la guía durante todo este tiempo y por compartir tu sabiduría y tu pasión por la ciencia. Sin duda, he aprendido mucho de tí, te admiro muchísimo. He tenido mucha suerte de tener un supervisor como tu! Nunca olvidaré que me invitaste aquella cena de navidad con tu familia, desde siempre me hiciste sentir como en casa.

David, no tengo palabras para agradecer y decir todo lo que has hecho por mi. Lo que me has hecho crecer y el apoyo que has significado para mi en todos los aspectos. El haber sido parte de la dirección de mi doctorado sin estar en el papel. Las largas discusiones científicas y no tan científicas que tuvimos por años. Gracias. Francho, mi estancia en España y mi doctorado no hubieran sido lo mismo sin tu amistad. Los jueves mágicos de pelis y cenas, las fiestas, los conciertos, he ir a escalar juntos. Las charlas larguísimas de ciencia y otras cosas. Me llevo mil aventuras contigo. Gracias por siempre estar. Sergio, esas tardes de mole y otras delicias culinarias, las tartas veganas y las limonadas nos dieron el tiempo para charlar de neuronas y de mil cosas más. No tengo más que decir que los quiero a todos por ser mi familia científico-española. Me los llevo para siempre como colegas y amigos. Gracias toatales!

I also want to acknowledge Wesley and Mozhgan for those long discussions about our works and the fun we had.

A los integrantes de GOTHAM, Mario, Alessio, Pablo y Hugo, por todas esas reuniones que nos enriquecieron a todos; y por supuesto a Santiago por todo lo que compartimos discutimos y aprendimos juntos. También quiero agradecer a todos mis colaboradores, especialmente a álex Arenas que a pesar de que no has sido mi supervisor como lo habíamos planeado has formado parte fundamental en el desarrollo de nuestro trabajo.

Lyere por todas esas tardes de foto, charlas y cervezas en el lugar de siempre, por crecer juntas. Al grupo de becarios por todas las mañanas de café.

A las amistades que he creado en España fuera de la universidad. María, gracias por todas esas cenas mexico-españolas y todos tus consejos que me han hecho mejor persona. María E. y Susana, nuestro crew fue el mejor, las quiero mucho. Manu, gracias por en-

señarme la España profunda. Caro y Sara, gracias por siempre escucharme.

A mi familia, mis personas favoritas en el mundo, por siempre apoyarme en todo y estar ahí para mi, siempre, siempre. A mis amigos que desde México siempre estuvieron al pendiente. Ulises y Naye M. han sido un apoyo emocional fundamental en este tiempo, me encanta compartir y viajar con ustedes. Iris por acompañarme siempre, por tu apoyo y el viajecito que casi nos hace residentes en Brujas. Ari e Ingrid, por siempre tener un espacio para mi, las adoro. Héctor, por todas esas discusiones sobre Física y Filosofía. Naye, por siempre ser positiva sobre mi, eres mi física y compañera preferida. Diego y Zuga, por mis escuchar potcast.

Todos ustedes han hecho una mejor versión de mi. Gracias!

Abstract

Our increasingly interconnected world allows communicable diseases to spread rapidly on a global scale with unprecedented ease. As populations grow, urbanization accelerates, climate change intensifies, and mobility increases, a new and complex battlefield is created in the fight against the emergence of new and old zoonotic pathogens. To face, or even anticipate new epidemic outbreaks, we must answer how, when, where, and which prevention and control strategies should be implemented.

The ever-increasing amount of data on human interactions, from daily human mobility patterns to face-to-face interactions, gives us the possibility to understand the pathways of contagious disease spread. It is possible to create data-driven models that allow us to study and predict statistical scenarios of the spread of epidemics under different circumstances. In this thesis, based on the physics of complex systems, we improve our understanding of epidemic modeling by integrating control systems into the dynamics. We study the interdependencies between the structure of human interaction networks and their functionality to spread infectious diseases while integrating control strategies into the dynamics.

Resumen

Nuestro mundo cada vez más interconectado permite que las enfermedades transmisibles se propaguen rápidamente a escala mundial con una facilidad sin precedentes. A medida que crecen las poblaciones, se acelera la urbanización, se intensifica el cambio climático y aumenta la movilidad, se crea un nuevo y complejo campo de batalla en la lucha contra la aparición de nuevos y viejos patógenos zoonóticos. Para enfrentar, o incluso anticipar, nuevos brotes epidémicos, debemos responder cómo, cuándo, dónde y qué estrategias de prevención y control se deben implementar.

La cantidad cada vez mayor de datos sobre las interacciones humanas, desde los patrones de movilidad humana diaria hasta las interacciones cara a cara, nos brinda la posibilidad de comprender las vías de propagación de enfermedades contagiosas. Es posible crear modelos basados en datos que nos permitan estudiar y predecir escenarios estadísticos de propagación de epidemias en diferentes circunstancias. En esta tesis, basados en la física de los sistemas complejos, mejoramos nuestra comprensión de la modelización de epidemias integrando sistemas de control en la dinámica. Estudiamos las interdependencias entre la estructura de las redes de interacción humana y su funcionalidad para propagar enfermedades infecciosas mientras integramos estrategias de control en la dinámica.

List of publications

- A. Reyna-Lara, D. Soriano-Paños, A. Arenas and J. Gómez-Gardeñes. Chaos, Solitons & Fractals **158**, 112012 (2022).
- A. Reyna-Lara, D. Soriano-Paños, J. H. Arias-Castro, H. J. Martínez, and J. Gómez-Gardeñes. Chaos: An Interdisciplinary Journal of Nonlinear Science **32 (4)**, 041105 (2022).
- A. Reyna-Lara, D. Soriano-Paños, S. Gómez, C. Granell, J. T. Matamalas, B. Steinegger, A. Arenas and J. Gómez-Gardeñes. Phys. Rev. Research **3**, 013163 (2021).
- D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni and J. Gómez-Gardeñes. Phys. Rev. Research **2**, 013312 (2020).

Contents

List of figures

List of tables

Chapter 1

Introduction

"...the totality is not, as it were, a mere heap, but the whole is something besides the parts ...",

Aristotle.

This thesis aims to understand how patterns of behavior that characterize human activity at different scales affect the spread of contagious diseases in order to heuristically find a way to control epidemic outbreaks. This analysis is carried out, in general, through the use of *Complex Networks* in which the patterns of human behavior are encoded and, in turn, the contagion dynamics that characterize the spread of an infectious pathogen takes place. Our goal, therefore, is the formulation and analysis of mathematical frameworks that capture the complex interplay between the structure of our interactions and propagation dynamics. However, before starting this thesis, I would like to address the question: why we physicists are interested in contagion phenomena? To answer this question we must understand how physics and, more specifically, the physics of *Complex Systems* has been developed.

Historically, physics evolved under reductionist ideas. Reductionism, although based on the atomic philosophy[1], was introduced by Renè Descartes in 1637, in his work *"Discourse on Method"* saying:

> *"The world is like a machine, its pieces like clockwork mechanisms, and that the machine could be understood taking its pieces apart, studying them and then putting them back together to see the larger picture".*

This idea was understood as the unification theory of all sciences attempting to be a *Theory of everything.* As the Nobel Laureate in Physics, P.W. Anderson mentions in its celebrated

[1]Perhaps introduced by Leucippus of Miletus (5th century BCE) before Democritus of Abdera (430 BCE)

article "*More is different*" [1] published 50 years ago, if reductionist ideas were right, we could, in principle, deconstruct the sociology to psychology, psychology to neurology, neurology to cell biology, cell biology to molecular biology, molecular biology to chemistry, chemistry to many-body physics, and many-body physics to particle physics.

Reductionist thinking has been very useful to answer many important questions in nature in many fields. However, although necessary, nowadays we understand that it is no longer sufficient. The limits of reductionist thinking in physics were 'rapidly' reached when Newtonian physics aimed to explain many-body systems, and, in general, any phenomena arising from the interactions among the constituents of an ensemble(*e.g.* phase transitions, pattern formation, and many others). In other scientific contexts, we find similar examples: cell biology cannot explain consciousness by the activity of a single neuron, the analysis of individual markets cannot predict the global economy, and the behavioral study of single bees does not allow us to understand how cooperation emerges in hives. Then, the approximation of a perfect linear world in which superposition principles apply is not true at all and, therefore, the mere addition of components does not allow us to reconstruct the phenomena we observe around us.

In light of this, at the end of the 19th century and the beginning of the 20th century, a revolution in physics was occurring. Physicists and mathematicians made great advances in creating scientific tools to understand natural phenomena beyond linearity. Poincaré portrayed the foundations of nonlinear physics and chaos theory when attempting to solve the three-body problem. Meanwhile, many-body physics made enormous advances in the characterization of critical phenomena, addressing ubiquitous phase transitions in different contexts: between states of matter (from liquids to solids to gases to plasma), magnetization, superconductivity, superfluidity, etc. In the light of these advances and thanks, above all, to the formulation of the renormalization group theory, it became apparent that the emergence of collective phenomena depends very little on the particular details of the components of the systems and very much on how the interaction between them is organized (e.g. the dimension of the system). In fact, as discussed by Anderson in [1], the breaking of internal symmetries taking place in the phase transitions of many-body systems can be viewed as a generalized phenomenon. In the macroscopic scale of a large system, matter will experience a phase transition to a new state in which microscopic symmetries of the system would be "violated". Anderson, thus, postulates that the emergent collective behaviors observed in nature are precisely the consequence of a symmetry breaking process that impedes to infer the properties of large aggregates of interacting elements by the sole characterization of the symmetries of their constituents. Thus, collective natural phenomena correspond to a macroscopic image that can only be understood through the lens of a holistic description. This description, in turn, requires a different conceptual structure to address each of the levels of organization that appear at different scales in nature.

The various advances in the theory of dynamical systems and statistical physics in the first half of the 20th century preceded the rise of the first computers for performing numerical calculus and simulation in the second half of the century. This meant a vertiginous advance in the field of Statistical and Nonlinear Physics in the characterization of natural phenomena at multiple scales. Some example of these advances are chaos theory [2], self-similar characterization (fractality) of strange attractors [3] or self-organized criticality [4, 5]. These theories focused on the activity of the field in the last decades of the 20th century and its theoretical value was coupled with their practical application in fields as diverse as economics [6], ecology [7], earthquakes [8], forest fires [9], or epidemiology [10] among others.

Despite the interest aroused by the aforementioned theories and their interdisciplinary applications, one of the major problems faced by Statistical and Nonlinear Physics was the scarcity (even absence) of reliable data to validate and allow the calibration of the models developed. However, after the popularization of the Internet and the World-Wide-Web at the beginning of the 21st century and, subsequently, the massive use of digital devices and the explosion of data science, the field begins to obtain data from both: the dynamics of social, biological, and technological systems; and the structure that governs the interactions between the constituents of these systems.

The availability of real data about the interaction backbone of real social, biological, and technological systems led to an unexpected result. Instead of homogeneous architectures (such as random graphs), the networks formed by the interactions between pairs of constituents were (in most of the cases) scale-free [11]. Thus, considering that most of the dynamical models developed in the last part of the 20th century assumed homogeneous interactions (such as mean-field approximations), the finding of largely heterogeneous and even scale-invariant interaction networks demanded a total revision of these models. Moreover, this finding suggested that the system's connectivity structure should be considered as another emergent property that joins to that observed from the dynamical behavior of these systems. Thus, if, as Anderson stated, emergent phenomena must be approached holistically, structure and dynamics must be integrated as two sides of the same coin in any model of reality. Indeed, the problem of structure *vs* functionality has been the most exciting line of research that has confronted Statistical and Nonlinear Physics in the present century and, more generally, has finally propelled the field of Complex Systems science towards the practical purpose for which it was created.

Now that we have motivated, from the perspective of Statistical and Nonlinear Physics, the study of Complex Systems and its current applied spirit, let us focus on the particular field of application that has been addressed in this thesis: the application of complex systems to the epidemiology of communicable diseases.

Complex Systems applied to the control of Infectious Diseases

During these years of the COVID-19 pandemic, it has become clear that the spread and control of communicable diseases depend both on the infectious potential of the particular pathogen and on human-to-human interaction patterns. Indeed, during the years 2020 and 2021 in the absence (or shortage) of effective vaccines and/or drugs to alleviate the health impact caused by COVID-19, health authorities in the vast majority of countries have perforce resorted to our behavioral patterns to cut the chains of SARS-CoV-2 virus transmission. In turn, once the vaccines have become available, their acceptance by the population has not been as expected in many countries, creating a social dilemma in which individual interest is in opposition to collective interest.

Therefore, the challenge of understanding the spread of infectious diseases can only be addressed from the perspective of a complex system in which our behavioral patterns are integrated with the dynamics of pathogen spread in a single formal framework.

In this thesis, we do not intend to provide a definitive framework for integrating all the social and biological dynamics that coexist and interact, but rather to advance the inclusion of two types of human behavior that play a very relevant role in the design of non-pharmacological containment policies: recurrent mobility and interpersonal interactions. To this end, we will make use of real data on recurrent mobility (round-trip) and face-to-face interactions. Both scales of interactions, provide us with a reliable source of information to perform the modeling tasks and to be able to apply the results in real contexts (populations) in which we will evaluate the efficacy of epidemiological control interventions. Thus, it is important to remark that the aim of this thesis is not to precisely forecast epidemics but to have an integrated system combining the relevant social and epidemiological information that enables us to make the right decisions about how to prevent and control epidemics.

To round off this introductory chapter, we anticipate some of the main problems that will be addressed in this thesis, each of them tackling the interplay of spreading phenomena and human behavior in different ways. Namely;

> **Virus spreading *vs* contact tracing** (Chapter 3). In this chapter, we propose a mathematical framework to analyze the effectiveness of contact tracing policies. To this end, we unify person-to-person transmission and quarantine of infectious contacts of positive individuals by means of a model in which both propagation processes compete over the same network of interpersonal contacts. In this model, we can also assess the effectiveness of this strategy when there are covert (asymptomatic) infectious agents and the available means are limited (both in terms of response times and the number of tests available).

Vulnerability and selective immunization driven by human mobility (Chapter 4). In this chapter, we present a mathematical framework for the treatment of vector-borne diseases in urban areas. To do so, we combine demographic information and human mobility data with the relevant contagion dynamics of this type of diseases. Our goal is to detect the most vulnerable geographic areas through mathematical means so that we can identify how to employ limited immunization resources most effectively.

Adaptive control by mobility restrictions and social cooperation (Chapter 5). Finally, in this chapter, we present a multiscale mathematical framework involving reactive control policies: global restrictions on travel to high-incidence areas, mesoscopic restrictions at the local level that reduce the mobility and activity of the residents of these areas, and also the social reaction (acceptance) to the local restrictions. This control system is activated or deactivated when the extent of the epidemic outbreak reaches a threshold. Therefore, we can analyze the appropriateness of reactive actions and also assess the success of these policies according to the architecture of human flows in each region.

Before explaining in detail the results obtained in the former scenarios, we first make a brief introduction in the next Chapter 2 to the main tools and models that will be used along the thesis.

Chapter 2

Epidemics in complex networks

I simply wish that, in a matter which so closely concerns the wellbeing of the human race, no decision shall be made without all the knowledge which a little analysis and calculation can provide.

Daniel Bernoulli.

Communicable diseases have always played a significant part in human history. Major pandemics and epidemics such as Bubonic plague, Small Pox, Spanish flu, severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV) and more recently the novel coronavirus (SARS-CoV-2) have scourged humanity [12] (see Table 2.1). Each of these historical events is accompanied by a multilayered crisis, that not only threatens human life but also causes profound changes in our socioeconomic and social structure [13, 14].

In this regard, historically, scientists from multiple disciplines have attempted to describe patterns of contagious diseases in human populations. They have investigated the characteristics of those diseases to nowcast and forecast with the final aim to control them [15]. Despite the great efforts to understand infectious diseases, the constant evolution of human interactions and mobility patterns that profoundly impact the spreading dynamics exacerbates the ongoing challenge. For instance, demographic changes, rapid urbanization, climate change and the globalization of travel and trade contribute to the emergence of zoonotic pathogens and their rapid spread on a global scale [16, 17].

In this thesis, we focus our attention on the dissemination process through interactions. Our main goal is to understand the paths of spreading phenomena to efficiently cut or reorganize interactions to keep under control the disease dissemination. Before we introduce our work, in this chapter, we aim to historically present the evolution of the mathematical frameworks of the spreading phenomena as we better understand human interactions. The

Epidemic	Time period	Type/pre-human host	Death toll
Antonine plague	165-180	Belived to be either smallpox or measles	5M
Plague of Justinian	541-542	Yersinia pestis bacteria / Rats, fleas	30-50M
Japannese smallpox	735-737	Variola major virus	1M
Black death	1347-1351	Yersinia pestis bacteria / Rats, fleas	200M
New world smallpox	1520-onwards	Variola major virus	56M
Great plague of London	1665	Yersinia pestis bacteria / Rats, fleas	100,000
Italian plague	1629-1631	Yersinia pestis bacteria / Rats, fleas	1M
Cholera pandemic	1817-1923	V. cholerae bacteria	1M+
Third plague	1885	Yersinia pestis bacteria / Rats, fleas	12M (China & India)
Yellow fever	late 1800s	Virus / Mosquitoes	100,000-150,000 (USA)
Russian flu	1889-1890	Believed to be H2N2 (avian origin)	1M
Spanish flu	1918-1919	H1N1 / Pigs	40-50M
Asian Flu	1957-1958	H2N2 virus	1.1M
Hong Kong flu	1868-1970	H3N2 virus	1M
HIV/AIDS	1981-present	virus / Chimpanzees	25-35M
Swine flu	2009-2010	H1N1 / Pigs	200,000
SARS	2002-2003	Coronavirus / Bats, civets	770
Ebola	2014-2016	Ebolavirus / Wild animals	11,000
MERS	2015-present	Coronavirus / Bats, camels	850
COVID-19	2019-preseent	Coronavirus / unknown (possibly pangolins)	6.3M (July, 2022)

Table 2.1: Major Pandemics & epidemics.
Adapted from [18].

basis of our scientific development and some essential aspects of epidemic modeling are presented. Starting from the building blocks, we will explain how the first simple compartmental models have evolved, from the mean-field models to structured approaches where detailed information about human interactions are added to the system under study. In Fig. (2.1), it is schematized the evolution of epidemic models. We focus on describing those models into the gray area. Those models are the most helpful to identify general

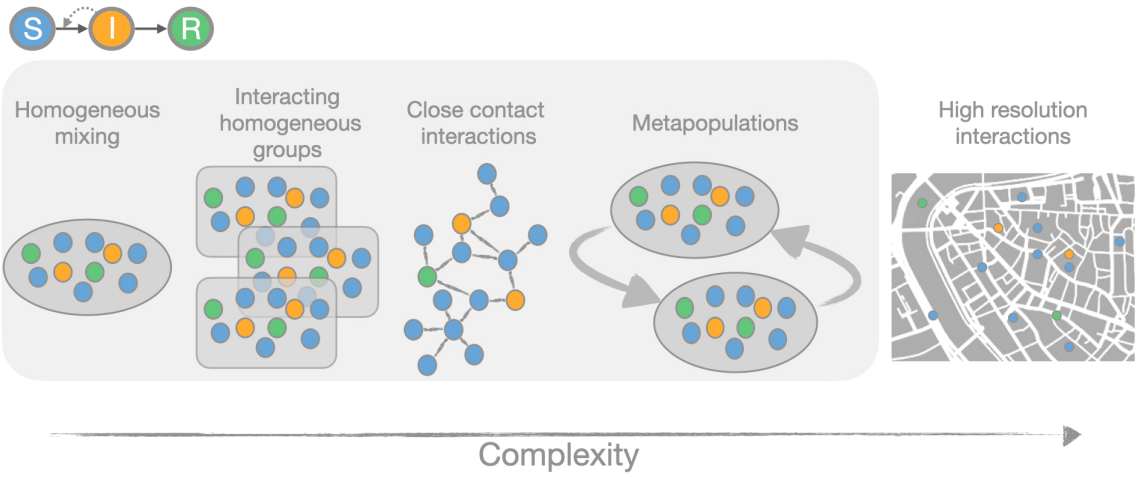


Figure 2.1: **Schematization of different approaches to epidemic modeling regarding interactions**. Circles represent individuals and the colors the stage of the disease (Blue represent susceptible individuals, orange infected individuals and green recovered/removed individuals). From left to right: *homogeneous mixing* hypothesis, where individuals are assumed to homogeneously interact with each other at random with k contacts per unit of time; *social structures* of interacting homogeneous groups, with people classified by sectors or groups according to demographic information (age, gender, etc.); *contact-network models*, where social interactions are depicted in between pairs in a complex network, thus opening the possibility to draw the propagation pathways of an infectious disease; *multi-scale models* in metapopulations, which divide a geographical region into sub-populations that are coupled by movements of individuals, while homogeneous mixing is assumed on the lower scale; *agent-based models*, which takes into account characteristic movements and interactions of any single individual on a very detailed scale (a schematic representation of a city is shown). *Adapted from:* [19].

principles that characterize spreading phenomena yet they integrate social complexity and capture the essence of human dynamics.

We aim to give the reader a general background to understand the theoretical frameworks presented in this thesis. In Section 2.1, we comment on the origins of the epidemiology before real interactions were added to their mathematical modeling, till the formulation of compartmental models. Then, in Section 2.2 we introduce the compartmental models by focusing our attention on SIR and SIS models as they are the building blocks of more sophisticated epidemiological models. In Section 2.3, we present the use of compartmental models beyond mean-field assumptions, i.e., incorporating the heterogeneous nature of human interactions by coarse-graining the system in classes, assuming that the dynamical properties within each class are the same. Then, we apply this method to specific interaction topologies. In Section 2.3.3, we break the epidemic spreading dialog to formally introduce heterogeneous interactions drawn in a complex network. Going back to the spreading dynamics, in Sections 2.3.4 and 2.3.5, we introduce the Microscopic Markov Chain Approach (MMCA) that is used along this thesis to describe spreading phenomena

in close contact networks. Finally, In Section 2.4 we discuss the description of spatial epidemics in metapopulations, extending the MMCA to account for human mobility patterns and large-scale interactions.

Let us note that this chapter is intended to provide an introduction with a broad and rough perspective on the dynamics of infectious diseases in humans, from the mean-field approach to the addition of human interaction and mobility patterns into the spreading dynamics. The reader could find more extended work in [10, 20] or in the review [21] for the epidemic processes in complex networks [22, 23] and metapopulations.

2.1 The origins of mathematical epidemiology

Long before accurate mathematical models for the spread of a communicable disease were developed, it was necessary to understand the means of spread of infection. Already Aristotle (384-322 BC) put forward the idea that invisible creatures were agents of disease. It was not until the 1670s, however, that Antonie van Leeuwenhoek first observed and experimented with *dierkens*, microbes. Later, in 1840, Jakob Henle published his essay *'On Miasma and Contagia'*. This text is considered to be the first argument for the germ theory, which states that pathogens or germs are the agents that cause diseases transmissible between humans, animals and plants. Subsequently, this theory was developed and experimentally confirmed by several scientists of the caliber of Louis Pasteur, Robert Koch and Joseph Lister at the end of the 19th and beginning of the 20th century. Among all the advances made during those years, John Snow's 1855 breakthrough in studying the spatial and temporal patterns of cholera cases and pinpointing the Broad Street (London) water pump as the source of the infection in 1855 [24] stands out. This study perhaps shed the most light on the germ theory, not only supporting the hypothesis of germs as disease-causing agents but also their ability to be transmitted between humans.

Based on the fundamentals of germ theory, the mathematical approaches to the spread of infectious diseases were progressively developed. The first known mathematical approach to epidemiology was developed by Daniel Bernoulli in 1760 after studying the effect of smallpox inoculation (variolation) on life expectancy [25]. Likewise, the first mathematical abstraction of a disease spreading across a population was developed in 1906, when W.H. Hamer suggested that the spread depend on the number of susceptible and infected individuals [26]. This idea was taken by Ronald Ross in 1908 to build the first dynamical model for the spread of Malaria, establishing the foundations of infectious disease dynamics. Then in his consecutive work, Ross deduced that the reduction of mosquito population below a critical level would be sufficient to control malaria in a particular region [27, 28]. After the seminal works by Ross, threshold concepts have been central in infectious disease models as they are closely related to control policies and, in theoretical terms, to bifurca-

tions separating disease-free and epidemic states. In particular, the most general threshold quantity in epidemic models is the basic reproduction number, $\mathcal{R}_0$, defined as: *the average number of secondary infections caused by a single infectious agent introduced in a fully susceptible population.* When this number is less than one the pathogen cannot provoke an outbreak, while, on the contrary, if it exceeds one, the initial infectious seed will trigger a cascade of infections leading to an epidemic. Thus, $\mathcal{R}_0 = 1$ pinpoints the threshold condition for epidemic models and its calculation for different models and epidemiological contexts has focused the activity of the field of mathematical epidemiology for decades.

2.2 Compartmental models

The pioneering works by Ross were also the first to implement the concept of compartments, *i.e.*, the assumption that the population is divided into a finite set of compartments that, usually, represent the possible stages of the course of infection. Ross ideas were extended by Kermack and McKendrick [29–31] to construct the well known Susceptible-Infected-Recovered (SIR) model, a paradigmatic epidemic framework that paved the way to the formulation of more refined compartmental models as the ones used in this thesis.

First, let us distinguish compartmental models' meaning as stages of infection from their intuitive resemblance to the clinical stages of disease caused by a pathogen. The evolution of infectious diseases can be defined qualitatively in terms of the invasion of the pathogen within the host; it is determined by a combination of the pathogen's growth rate and the host's immune response. Fig. 2.2 shows the timeline of a simplified course of infection. The grey area corresponds to the pathogen's growth and the black line to the immune response.

According to the figure the evolution of the (compartmental) state is as follows. The individual (host) remains in the Susceptible (S) compartment until, through a hazardous contact, he/she contracts the pathogen. At that moment, the host becomes Exposed (E), which means that the host is not yet able to transmit the pathogen because it is in a period of replication and its amount is not sufficient to produce infections. Once the pathogen population is sufficiently large to infect (regardless of the host immune response) the individual becomes Infectious (I). Finally, once the immune response is sufficiently activated, the host is able to reduce the abundance of the pathogen so that it is no longer infectious. At this point, the individual can be considered Recovered or Removed (R). It is important to note that the medical status does not correspond to the categorization of the infection status described above. In this sense, the infectious period of a disease is determined mainly by the abundance of the pathogen in the host, while, clinically, the individual is considered diseased when the immune response is large enough to show the symptomatology associated with the disease.

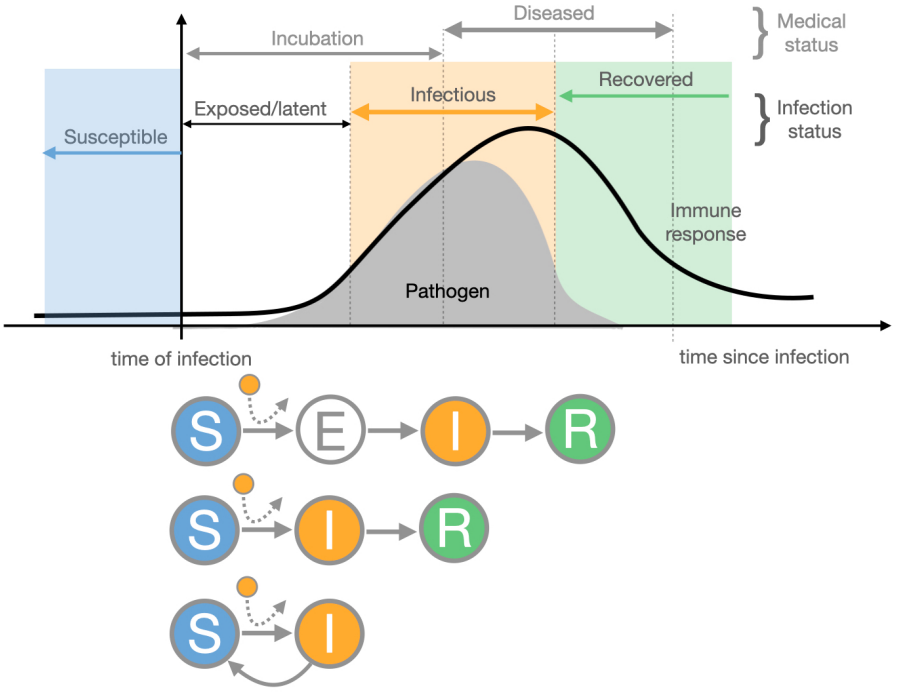


Figure 2.2: **Schematic representation of the timeline of infection and flow diagrams of SEIR, SIR & SIS compartmental models**. Top panel. Timeline of a generic infectious disease. The grey area represents the dynamics of the pathogen and the host immune response represented by the black line. The infection status can be classified as: susceptible, exposed, infectious, and recovered, different from the medical status that is classified as: the incubation period of the pathogen and diseased *i.e.* when symptoms are experienced. Bottom. Flow diagrams of SEIR, SIR and SIS compartmental models, as a simplification of the infection status. The dark grey arrows represent the movement between compartments. The curved dotted arrows represent the infectees influence over the rate at which a susceptible (potential host) moves into the infected class.

In the literature, a compartmental model can be interpreted as a stochastic reaction process. Individuals belonging to different compartments are interpreted as a kind of particles or chemicals. The reactions occur according to a set of mutual interaction rules that allow the individuals evolve to a new compartment. For instance, in the Susceptible-Exposed-Infected-Recovered (SEIR) model occur the following reactions,

$$S + I \xrightarrow{\beta} E + I \tag{2.1}$$

$$E \xrightarrow{\alpha} I \tag{2.2}$$

$$I \xrightarrow{\mu} R \tag{2.3}$$

where the transition rates between compartments are β, α and μ corresponding to a infection, latent and recovery rates respectively. In Fig. 2.2 we show three compartmental models. Although in Chapters 3 and 5, we will make use of more refined compartmental models in order to incorporate different stages of the infection course, in this introductory

chapter we will focus on the description of the most simplified versions of the compartmental models, the Susceptible-Infected-Recovered (SIR) and Susceptible-Infected-Susceptible (SIS) models, which flow diagrams are as well shown in Fig. 2.2.

2.2.1 The SIR model

The SIR model characterize those diseases that reach long lasting immunity after they recover from the infection, as occurs with many pathogens, like most strains of influenza. To construct the SIR model, let us consider a well mixed closed population (*i.e.*, without demographics, no births, deaths, or migration) with N individuals. Thus, during the course of an outbreak a population of N individuals can be divided into S susceptible, I infectious and $R = N - S - I$ recovered agents.

Given this compartmentalization, we focus on the microscopic contagion processes that lead to the infection of a susceptible agent. Let us suppose that a susceptible individual has $\langle k \rangle$ contacts per unit of time. Thus, the number of infectious contacts during a small period of time $(t + \delta t)$ that a susceptible individual has is $(\langle k \rangle I/N)(\delta t)$. Now let us define π as the the probability of successful contagion following an hazardous contact. Then, $1 - \pi$ corresponds to the probability of not getting the disease. If we redefine $1 - \delta q$ as the probability of escaping the infection after $(\langle k \rangle I/N)(\delta t)$ potentially infectious contacts, then,

$$1 - \delta q = (1 - \pi)^{(\langle k \rangle I/N)(\delta t)}. \tag{2.4}$$

Thus, the probability of getting the infection following those contacts is δq. In turn, if we define $\beta = -log(1 - \pi)$, we can rewrite Eq. (2.4) as,

$$\delta q = 1 - e^{-(\langle k \rangle \beta I/N)(\delta t)}. \tag{2.5}$$

Hence, we can translate this probability into a rate at which transmission occurs by expanding the exponential term ($e^x = \sum_{n=0}^{n=\infty} \frac{x^n}{n!}$), dividing both sides by δt and taking the limit of $\delta q/\delta t$ as $\delta t \to 0$. Then, the resulting transmission rate per susceptible individual evolve as,

$$\frac{dq}{dt} = \langle k \rangle \beta \frac{I}{N}. \tag{2.6}$$

Considering the evolution of the *force of the infection* per susceptible individual, it is easy to write the decreasing evolution of the number of agents in the susceptible compartment as they get infected, which reads as,

$$\frac{dS}{dt} = -\langle k \rangle \beta S \frac{I}{N}. \tag{2.7}$$

Once we have the Susceptible population dynamics as a function of the infectious contacts, we need to define the dynamics for the Infectious and Recovered/Removed population. Before we continue, let us redefine the variables of the SIR model by densities, so that,

s=S/N, i=I/N and r=R/N. Thus, a set of ordinary differential equations (ODEs) that denotes the temporal evolution of the compartments of the SIR model can be written as,

$$\dot{s} = -\langle k\rangle \beta s i, \tag{2.8}$$

$$\dot{i} = \langle k\rangle \beta s i - \mu i, \tag{2.9}$$

$$\dot{r} = \mu i, \tag{2.10}$$

where μ is the recovery rate, which is often related to the average infectious period τ as $\tau = 1/\mu$.

The characterization of the equilibrium solutions or the steady state of the SIR model depend on the infectious density i as shown in the set of ODEs. The obvious solution is when no infections have occurred, $i = 0$, then $\dot{s} = \dot{i} = \dot{r} = 0$. As we will further analyze, there is always another fixed point of the dynamics with an initial condition $i_0 \neq 0$. This fixed point is reached at the temporal extinction of the disease and is represented as $(s^\infty, 0, r^\infty)$, where r^∞ represent the extent of the epidemic outcome (the so-called 'attack rate') and $s^\infty = 1 - r^\infty$ the fraction of uninfected individuals.

To analyze the possible epidemiological outcomes, let us introduce a small perturbation to the system. A small fraction i_0 of infectious agents is placed into the susceptible population approximated as $s_0 \simeq 1$ thus, $r_0 = 0$. Then the Eq. (2.9) can be rewritten as,

$$\dot{i} = i(\langle k\rangle \beta s - \mu) \simeq i_0(\langle k\rangle \beta - \mu)\ , \tag{2.11}$$

the perturbation of infectious agents will grow or decay as a function of the constants k, β and μ since the solution of the linear initial evolution dictated by Eq. (2.11) is $i(t) \propto e^{(\langle k\rangle \beta - \mu)t}$. It is easy to see that if the $\mu > \langle k\rangle \beta$ then the infection will die out. On the contrary, when $\mu < \langle k\rangle \beta$ the fraction of infected grows exponentially yielding into an epidemic state. This threshold condition is closely related to the basic reproductive number or ratio, $\mathcal{R}_0$, that, as previously mentioned in Section 2.1 represents *the average number of secondary cases arising from an average primary case in an fully susceptible population.* The basic reproductive number can be derived from the threshold condition as,

$$\mathcal{R}_0 = \frac{\langle k\rangle \beta}{\mu} = \langle k\rangle \beta \tau. \tag{2.12}$$

that depends on the number of contacts an infected individual makes per unit of time $\langle k\rangle$, the transmisibility rate per contact β, and the average infectious period $1/\mu$. The former expression implies that introducing an infected individual to a fully susceptible population will trigger a cascade of infectious cases if $\mathcal{R}_0 > 1$, otherwise the infection will die out. If $\mathcal{R}_0 = 1$ we can obtain the critical infectivity, $\beta_c = \mu/\langle k\rangle$, better known as *epidemic threshold.*

Leaving aside the initial stage of an epidemic, let us now focus on the steady state, i.e., when the infectious pathogen ceases its propagation. As anticipated above, the relevant

question to answer is: what is the ultimate extent of the outbreak r^{∞}? To answer this question, we can divide Eq.(2.8) by Eq.(2.10) obtaining,

$$\frac{ds}{dr} = -\mathcal{R}_0 s, \tag{2.13}$$

integrating respect to r one easily obtains,

$$s(t) = s_0 e^{-r(t)\mathcal{R}_0}. \tag{2.14}$$

In this equation, the susceptible pool is not explicitly described, but there is an implicit dependence with the evolution of the recovered/removed fraction. Assuming that at the beginning of the epidemic outbreak we have $s_0 = (1 - i_0)$, and keeping in mind that in a SIR model transmission will eventually die out, i.e., $r^{\infty} = 1 - s^{\infty}$, then, we can obtain the proportion of recovered individuals at the steady state as,

$$r^{\infty} = 1 - (1 - i_0)e^{-r^{\infty}\mathcal{R}_0}. \tag{2.15}$$

Again, we do not obtain a closed expression of for r^{∞} although for SIR epidemics with small attack rates we can rewrite and approximate Eq. (2.15) as,

$$r^{\infty}\mathcal{R}_0 = -\log(1 - r(\infty)) \simeq r^{\infty} + \frac{(r^{\infty})^2}{2}\ , \tag{2.16}$$

so that we can approximate the attack rate as $r^{\infty} \simeq 2(\mathcal{R}_0 - 1)$.

Apart from the attack rate, Eq. (2.14) also allows to obtain other interesting observables such as quantities related to the epidemic peak, i.e., the situation in which the fraction of infectious agents reaches its maximum and the epidemic wave starts its decline due to the depletion of the susceptible pool. To this aim, we consider $r(t) = 1 - s(t) - i(t)$ and write Eq. (2.14) as,

$$i(t) = 1 - s(t) + \frac{1}{\mathcal{R}_0}\log\left(\frac{s(t)}{s_0}\right)\ . \tag{2.17}$$

This equation allows to derive the value of the susceptible pool at the epidemic peak. In particular, by solving the condition $\dfrac{di}{ds} = 0$ we obtain $s(i^{max}) = \mathcal{R}_0^{-1}$, and by substituting this value into Eq. (2.17) yields,

$$i_{max} = 1 - \frac{1}{\mathcal{R}_0}\left(1 + \log(\mathcal{R}_0 s_0)\right)\ . \tag{2.18}$$

The former information about the epidemic peak is of utmost importance when dealing with infectious diseases whose impact in the health system is large, since it is related to the maximum needs of hospitalizations and clinical means that a population will require during an outbreak.

To confirm and expand the former theoretical derivations it is useful to solve numerically Eqs. (2.8-2.10) and obtain the time evolution of infections in the SIR model. In Fig. 2.3-(a) we show the attack rate r^{∞}_{SIR} as a function of the basic reproductive number $\mathcal{R}_0$.

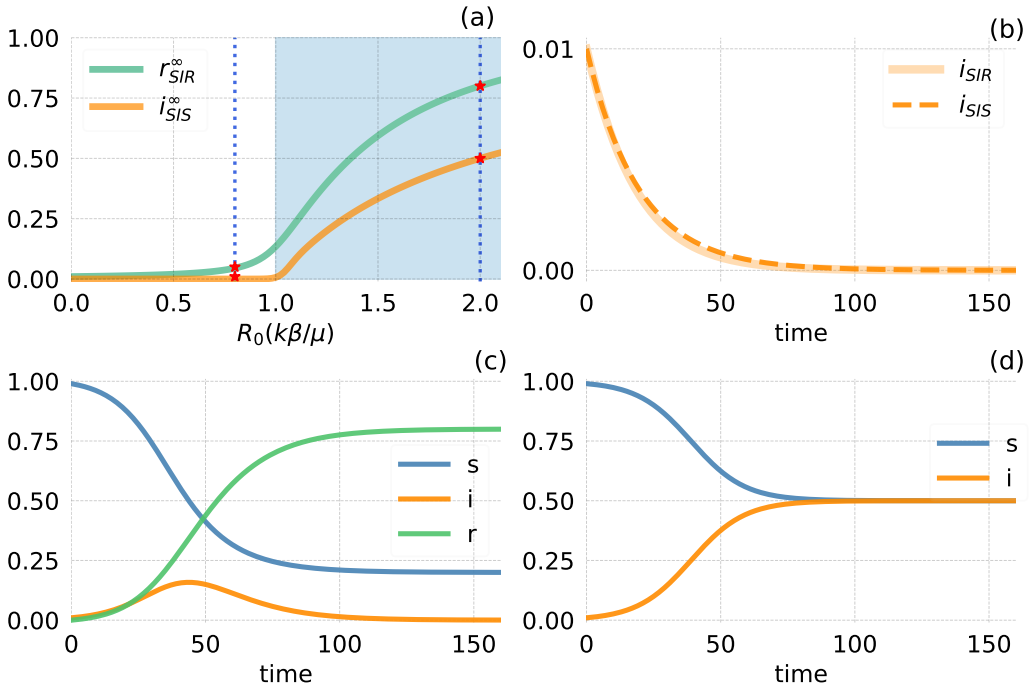


Figure 2.3: **Dynamics & stationary state of SIS & SIR models**. (a). Stationary state of the R_{SIR} and I_{SIS} compartments as a function of $\mathcal{R}_0$ (b) Temporal evolution of the infected (I) compartment for the mean-field approach to epidemics of SIR and SIS models in the subcritical regime $\mathcal{R}_0 = 0.8$ (red markers in the white area in (a)). (c) and (d) Temporal evolution of SIR (c) and SIS (d) compartmental models in the supercritical regime $\mathcal{R}_0 = 2$(corresponding to the red dots in the blue area in (a)). We simulate the dynamics over a population with 10^4 individuals, with an initial condition of $I_0 = 50$ infectees, and with $k = 2$, $\mu = 0.1$.

Note that the theoretical value $\mathcal{R}_0 = 1$ does not exactly correspond to a transition in the r^{∞}_{SIR} curve. However, as shown in 2.3-(b), if we use $\mathcal{R} = 0.8$ we observe numerically that the dynamics corresponds to an exponential decrease and the eventual extinction of the infections. In contrast, if $\mathcal{R}_0 > 1$ we observe an exponential growth until the pool of susceptible individuals decreases up to $s(t) = \mathcal{R}_0^{-1}$. At this point, the fraction of infectious agents reaches its maximum and starts to decrease exponentially. This last effect can be observed in Fig. 2.3-(c), where we plot the temporal evolution for each compartment for $\mathcal{R}_0 = 2$.

2.2.2 The SIS model

There are several infectious diseases that confer non-lasting immunity, such as rotaviruses and sexually transmitted diseases, like syphilis or gonorrhea, in which individuals after recovering become susceptible again. Consequently, hosts can contract the same infectious disease multiple times during their lifetime, with no apparent immunity. These diseases can be mathematically described by the SIS model where the recovery rate return the

infectious to the susceptible pool. Then, for a close population with N individuals, we can write the corresponding ODEs analogously to the previous section as,

$$\dot{s} = -\langle k\rangle\beta si + \mu i, \tag{2.19}$$

$$\dot{i} = \langle k\rangle\beta si - \mu i. \tag{2.20}$$

Note that Eqs. (2.20) and (2.9) are the same, thus the basic reproductive number $\mathcal{R}_0 = \langle k\rangle\beta/\mu$ coinciding with the SIR compartmental model. However, at variance with the SIR model, the SIS model allows for an analytical solution of the time evolution of infected agents. In fact, taking Eq. (2.20) and substituting $s = 1 - i$, integrating we obtain,

$$i(t) = \left(1 - \frac{\mu}{\langle k\rangle\beta}\right)\frac{Ce^{(\langle k\rangle\beta-\mu)t}}{1 + Ce^{(\langle k\rangle\beta-\mu)t}}, \tag{2.21}$$

where $C = i_0/(1 - i_0 - \mu/\langle k\rangle\beta)$.

By inspecting the asymptotic behavior of Eq. (2.21) we obtain two possible outcomes. The called *disease-free state* occurs when $\mu > \langle k\rangle\beta$ or $\mathcal{R}_0 < 1$, since, in this scenario, the number of cured individuals per unit of time exceeds the number of new infections. As a result, the pathogen disappears from the population. On the contrary, the *endemic state* shows up when $\mu < \langle k\rangle\beta$ or $\mathcal{R}_0 > 1$. In this case, a small initial infectious seed i_0 produces an exponential increase of the epidemic size reaching, after some transient time, a constant steady value i^*. The precise value of i^* can be derived by inspection of Eq. (2.21) for $t \to \infty$, which yields,

$$i^* = 1 - \frac{\mu}{\langle k\rangle\beta} \tag{2.22}$$

These two outcomes imply that the SIS model predicts that some pathogens will persist over time while others will die out. In Fig. 2.3-(a), we plot the prevalence of the disease i^∞ as a function of the basic reproductive number $\mathcal{R}_0$. The epidemic threshold signalizes the transcritical bifurcation point $\mathcal{R}_0 = 1$, which separates the absorbing phase (white area) where the epidemic outbreak its unsuccessful and the epidemic active phase (blue area). To dynamically show the absorbing phase, the Fig. 2.3-(b) shows the temporal evolution of the infectious compartment $\mathcal{R}_0 < 1$ where the outbreak vanishes. In Fig. 2.3-(d) we show the temporal evolution on the epidemic active phase ($\mathcal{R}_0 > 1$) for each compartment in the SIS model, note that the infection is persistent over time at a baseline level, in epidemiology this is known as endemic state.

Due to its simplicity, in the following sections of this chapter, we will focus on the dynamics of the SIS model to analyze the results when different aspects of human behavior are incorporated.

2.3 Epidemic modeling and heterogeneous interactions

The assumption of the homogeneous interactions presented in the previous section, although very useful to qualitatively understand epidemic outbreaks, is far from describing real-world scenarios. While in the homogeneous mixing hypothesis it is assumed that any individual interacts with any other and the average number of contacts of all individuals is $\langle k \rangle$, in reality, human interactions are more complex than that. Historically, the heterogeneity of interactions in epidemic modeling was first tackled statistically, separating individuals in groups according to their connectivity (number of contacts) and considering that all the individuals belonging to the same class share the same dynamical evolution.

2.3.1 Heterogeneous mean-field model

Heterogeneity in epidemic models was introduced by Pastor-Satorras and Vespignani in 2001 [32]. They supposed a closed population of N individuals, classifying them by the number of interactions each individual makes every time step referred to as *degree* k. Under this framework, they considered a mean-field approach with the addition of a compartmentalization of the population in groups according to the degree of the agents. For each group associated with a particular connectivity k, they supposed that all the members are statistically equivalent, so that their evolution can be monitored by a single equation. In addition, the equations associated to the groups differ from each other provided their different force of infection. Thus, for the standard SIS model, instead of having one single equation as in the mean-field formulation, now we have a set of equations (one for each degree present in the system). Each of these equations takes the following form,

$$\dot{i}_k = -\mu i_k + (1 - i_k) k \beta \sum_{k'} P(k'|k) i_{k'}, \tag{2.23}$$

representing the time evolution of the probability of being infected given that the individual has k contacts. The first term in the right hand of the equation accounts for those already infected that may recover with rate μ. The second term in the right hand side of the equation represents the new infections and, accordingly, take into consideration all possible ways a node of degree k can be infected by members of any other degree-class k'. Thus, in the sum it appears the conditional probability $P(k'|k)$ which weights each term as the probability that a node with degree k is connected with another one with degree k'.

Assuming that there is not degree correlations (*i.e.* $P(k'|k)$ does not depend on k) then $P(k'|k) = k'P(k')/\langle k \rangle$. Thus, the sum in the right hand side now accounts for the fraction of infected neighbors of any kind of node, Θ, and it can be written as,

$$\Theta = \sum_{k'} \frac{k' P(k')}{\langle k \rangle} i_{k'}. \tag{2.24}$$

Inserting this expression into Eq. (2.24) and imposing the steady-state condition ($\dot{i}_k = 0$), we obtain the solution for the stationary fraction of infected nodes of degree k,

$$i_k = \frac{k\beta\Theta}{\mu + k\beta\Theta}. \tag{2.25}$$

This equation provides an implicit solution to the steady state. In order to be able to analyze under which conditions this state corresponds to an endemic state, we take Eq. (2.25) and insert it into Eq. (2.24). This way, the steady-state satisfies,

$$\Theta = \frac{1}{\langle k \rangle} \sum_k \frac{k^2 P(k)\beta\Theta}{\mu + k\beta\Theta}. \tag{2.26}$$

Eq. (2.25) has a trivial solution $\Theta = 0$ that (from the expression in Eq. (2.24)) corresponds to the epidemic-free state. However, this is not the only possible solution to Eq. (2.25). By inspecting the intersections between the function contained in left hand side and the right hand side of Eq. (2.25), it is easy to check that when the derivative of the right hand side with respect to Θ evaluated at $\Theta = 0$ is larger that the corresponding to the left hand side (which value 1) then a second solution (intersection of the two curves) appears at $\Theta > 0$. This second solution corresponds to the endemic solution and, thus, the condition for this solution to exist can be obtained by the derivative of the Eq. (2.26) evaluated at $\Theta = 0$,

$$1 = \frac{\beta^{HMF}}{\mu\langle k \rangle} \sum_k k^2 P(k). \tag{2.27}$$

From this expression, we can obtain the minimum infectivity needed to trigger an endemic state, the so-called epidemic threshold,

$$\beta_c^{HMF} = \mu \frac{\langle k \rangle}{\langle k^2 \rangle}. \tag{2.28}$$

The epidemic threshold plays an analogous role as the basic reproduction number of the network. In fact, for a particular network and fixed value of μ, the value β^{HMF} is the infectivity value that makes $R_0 = 1$.

The expression of the epidemic threshold contains the ratio between the average degree and the average square degree of the nodes of a network. These two quantities represent, respectively, the first and the second moment of the degree distribution of the network, $P(k)$, defined as the probability that a randomly chosen node has degree k. The expression for the m-th moment of this distribution reads,

$$\langle k^m \rangle = \sum_0^\infty P(k) k^m dk. \tag{2.29}$$

2.3.2 HMF for synthetic contact networks

Although we have not defined a network formally yet, we have introduced above the degree distribution, which contains the histogram of interactions of a population. Before we

mathematically introduce those concepts, let us continue with the intuition of what a contact network is. We will soon clarify these concepts in the next section.

Random networks. Paul Erdös and Alfréd Rényi, in 1960 [33], created a mathematical model to generate random graphs. The model proposes that any two elements or *nodes* chosen from a closed system with N elements, are connected with probability p. The probability distribution of the connectivity results in a binomial distribution,

$$P(k) = \binom{N-1}{k} p^k (1-p)^{1-p^k}. \tag{2.30}$$

The epidemic threshold for a random network in a SIS heterogeneous model can be found if the first and the second moments of the distribution are substituted on the Eq. (2.28), $\langle k \rangle = p(N-1)$ and $\langle k^2 \rangle = \langle k \rangle (1-p) + \langle k \rangle^2$. Thus, the epidemic threshold is,

$$\beta_c^{HMF} = \frac{\mu}{1 - p + \langle k \rangle}. \tag{2.31}$$

Scale-free networks. Real-world networks have highly heterogeneous properties, such as the existence of some nodes that are highly connected to others, known in the literature as *hubs*. Laszlo Barábasi and Reka Albert, in 1999 [11] proposed a mechanism called *preferential attachment* obtaining long-tailed distribution observed in many real networks [34]. The preferential attachment is applied in a newborn network. It states that the likelihood of a new connection towards one of the already existing nodes in the network is proportional to their degree of already existing nodes in the network. So, hubs tend to accumulate interactions as new nodes are added to the network. They found a degree distribution $P(k) \sim k^{-3}$.

The networks generated by this model belong to a class of networks called scale-free and characterized by a potential degree distribution $P(k) \sim k^{-\gamma}$. The model is particularly interesting since most of real networks display this power-law distribution and, when $\gamma \leq 3$, the second moment of the degree distribution diverges in the $N \to \infty$ limit, thus the epidemic threshold in the HMF tends to zero. The absence of an epidemic threshold in scale-free networks indicates that infections could proliferate despite whatever spreading rates they may have, and it is a direct consequence of the existence of nodes that act as super-spreaders (the hubs) in the network. This suggests that the existence of heterogeneous social structures can have a great impact in the spread of a disease, enhancing its spread potential in comparison with that expected in a random graph with the same average connectivity.

2.3.3 Heterogeneous interactions as complex networks (a parenthesis)

Up to now, we have described the spreading phenomena in heterogeneous interactions statistically through a degree distribution $P(k)$. In reality, an individual can only transmit

a pathogen to those with whom is in contact. Therefore, if one desires to specifically study the paths of an infectious disease, across a particular network, one should include the explicit interaction backbone that dictates the possible contagion pathways in the system. In this regard, it is worth making a parenthesis to present how close contact interactions can be formally translated into a complex network. These concepts might help to clarify what we just have and are about to introduce.

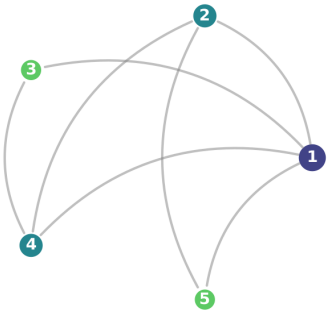
Close contact interactions can be naturally depicted into a complex network. A complex network is formally defined as a catalogue of any set of interacting elements that belong to a system. The elements of the system are called nodes or vertices and their interactions links or edges. Mathematically a network or graph $\mathcal{G}$ can be written as $\mathcal{G} = (N, L)$, for all set of N nodes and L links. The nodes can be labeled with an index $i = \{1, 2, ..., N\}$, while the information of the L interactions can be stored in a matrix $\mathbf{A}_{N\times N}$, known as the adjacency matrix, in which each entry a_{ij} accounts for the interaction between the node i with the node j. If the interaction between those nodes exists, then $a_{ij} = 1$, otherwise $a_{ij} = 0$. Interactions between nodes can be directional or bidirectional. The direction of the links is reflected in the adjacency matrix: if $\mathbf{A}$ is symmetric then the network is undirected or with all bidirectional links; if $\mathbf{A}$ is asymmetric then the network is directed. Network directionality has an important role in determining diffusion processes, since it limits the paths that may be followed. The interactions between nodes can also be weighted, as interactions are stronger or weaker. This can be reflected in the elements of the adjacency matrix that is renamed as $\mathbf{W}$ since each element w_{ij} would reflect the weight of the connection (see Fig. 2.4).

Once encoded the network in an adjacency matrix, it is possible to characterize its architecture. In the last two decades, network science has provided with many metrics that capture network properties across different scales, from global to local structures. In Sections 2.3.1 and 2.3.2, we have already used the simplest metric that characterize nodes, their degree that, given the adjacency matrix of a unweighted network $\mathbf{A}$, can be easily computed as,

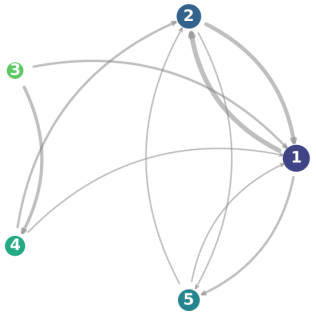
$$k_i = \sum_{j=1}^{N} a_{ij}. \tag{2.32}$$

As anticipated in Section 2.3.2, once computed the degrees of the nodes, a useful global property of the network is to analyze how these degrees are distributed in the graph. This can be easily analyzed through the degree distribution, $P(k)$, whose first and second moments play a crucial role in determining the spreading capacity of a network as we have derived through the HMF. In Fig. 2.5 we show three examples of degrees distributions corresponding to synthetic graphs.

Degree is a local property that can be considered as a measure of how central (important) is a node in a network. In turn, the centrality of a node can be considered as



$$\mathbf{A}_{5\times5} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$



$$\mathbf{W} = \begin{pmatrix} 0 & 5 & 0 & 0 & 2 \\ 4 & 0 & 0 & 0 & 1 \\ 2 & 0 & 0 & 3 & 0 \\ 1 & 2 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 2.4: **Cartoon of an undirected network and a directed and weighted network with their adjacency matrix respectively**. Size and colors (from small/green to big/dark blue) are proportional to the degree of the node. The thickness of the links are proportional to the weights appearing in the adjacency matrix.

its relevance for the spreading processes. Thus, it is convenient to explore other ways for ranking the centrality of nodes in a more global way than degree does. To this aim, one can rely on this statement: the importance of a node is proportional to the importance of the nodes it is connected to. This definition captures the fact that even nodes that are poorly connected can be crucial for the network if the small set of neighbors is composed by important nodes. By writing $\mathbf{x} = (x_1, x_2, ..., x_N)$ the set of the centralities of the nodes (where x_i is the centrality value of node i) we can translate the former definition into the following mathematical equation,

$$\mathbf{A}\mathbf{x} = \lambda\mathbf{x}, \tag{2.33}$$

where λ is the proportionality parameter. From the former equation it is clear that the set of centralities corresponds to an eigenvector of $\mathbf{A}$ and, hence, this centrality measure is usually termed as eigenvector centrality. In particular, applying the Perron-Frobenius theorem[1] the set of nodes' centralities corresponds to the set of entries of the eigenvector

[1]Perron-Frobenius theorem. Let $\mathbf{A}$ be a matrix with entries $\in \mathbb{N}$, then:

i. There exist a unique eigenvalue $\in \mathbb{R} > 0$ of $\mathbf{A}$ whose absolute value is bigger than those of other eigenvalues. Leading eigenvalue.

ii. There is a unique eigenvector whose entries $\in \mathbb{R} > 0$ corresponding to the leading eigenvalue. Leading eigenvector.

iii. The only eigenvectors with the same absolute eigenvalue as the leading one are on the same circle

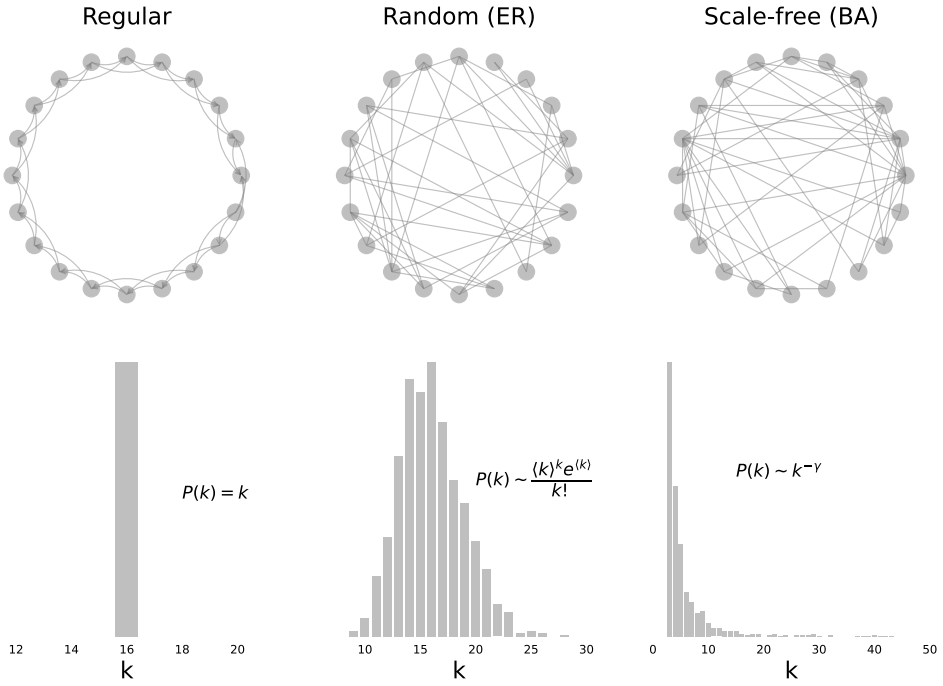


Figure 2.5: **Complex networks**. At the top complex networks with 20 nodes and $\langle k \rangle = 4$. From left to right: the regular network, every node has the same linked neighbors; the random network is obtained by rewiring each edge with probability $p = 1$, also known as Erdös-Rényi network (ER); the scale-free network is generated by the preferential attachment model of Barábasi-Albert (BA) with $m = 3$ edges to attach from a new node to existing nodes. At the bottom, the degree distributions of alike networks with 1000 nodes and $\langle k \rangle = 16$.

corresponding to the largest eigenvalue (λ) of a matrix $\mathbf{A}$. As we will see later, this set of centralities plays a crucial role in the spread of diseases close to the epidemic threshold.

Another indicator of the structural importance of nodes at the global level is the betweenness centrality b_i. Betweenness aims at capturing the importance of nodes as connectors of the network, highlighting those acting as important bridges between parts of the network that, and sometimes without their presence, would be disconnected.

The computation of betweenness relies on the characterization of the shortest paths between all the pairs of nodes. A path between two nodes, i and j, is a sequence of n connected edges to hop from the node i to node j. Obviously, there are many different paths between two nodes and, therefore, the shortest path is that having the minimum number of edges. Note that, for a given pair of nodes i and j, there can be more than one shortest path.

Once characterized all the shortest paths between all the pairs of nodes, the betweenness of a node i accounts for shortest paths between the rest of nodes of the network that pass

in the complex plain (having at least a negative or non real component). Symmetry of eigenvalues.

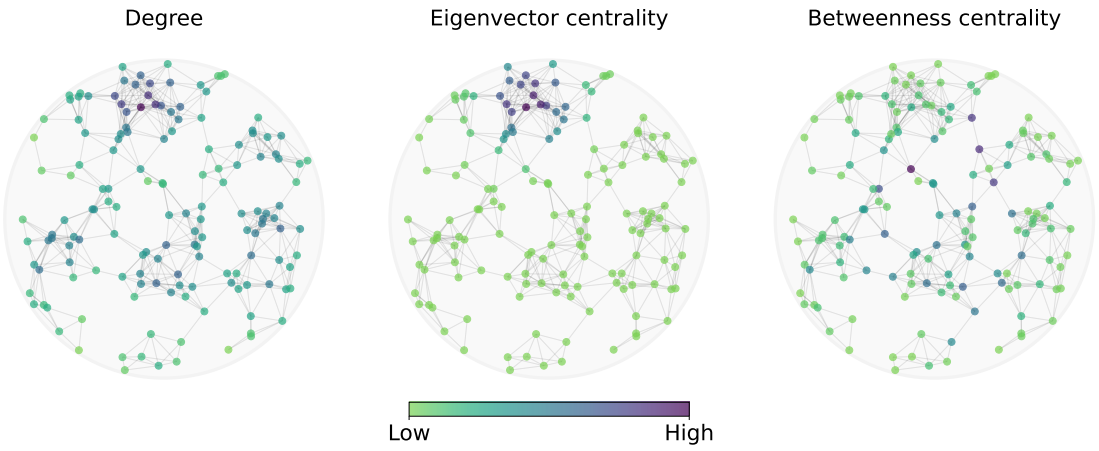


Figure 2.6: **Visualization of centrality measures in a random geometric graph**. Characterization of the nodes of an unweighted, undirected network by their degree, eigenvector centrality, and the betweenness centrality.

through node i and is computed in the following way,

$$b_i = \sum_{j \neq l \neq i} \frac{\sigma_{j,i,l}}{\sigma_{j,l}}, \tag{2.34}$$

where $\sigma_{j,l}$ is all the shortest paths between j and l, and $\sigma_{j,i,l}$ all the shortest paths between j and l that contains i as an intermediary node.

In Fig. 2.6, we show a visualization of the three centrality measures described above in a graph. For more information on centrality measures we recall the works [35, 36].

2.3.4 Microscopic Markov Chain Approach

Once introduced the basic features of networked architectures beyond their degree distribution, we can present now the formulation of epidemic models on networks beyond heterogeneous mean-field assumptions. Let us recall that this latter approximation neglected all the individual particularities of nodes and considered all the nodes with the same degree being identical from the point of view of the dynamics. This degree-block aggregation of individuals, cuts off other structural features of the contact network where the particular role of each of these individuals in the spreading dynamics is lost.

In light of this, several frameworks have attempted to study spreading phenomena using the specific information of the close contact interactions drawn in a complex network, where is possible to separate the dynamical contribution of each individual (*e.g.* pair quenched mean-field theory, recurrent dynamical message passing to mention some, etc.) [37–42]. Statistically, the microscopic contact reactions between infectious and susceptible individuals in real contact networks can be studied with the Microscopic Markov Chain Approach (MMCA) formulation [43]. This mathematical framework is based on probabilis-

tic discrete-time Markov chains that can be applied to weighted and unweighted complex networks. Most of the work of this thesis is based on this approach, since it allows for calculating the probability that every node of a network is in a particular compartment at each step of the computation.

2.3.5 MMCA for the SIS model

Let us consider an undirected and unweighted network $\mathcal{G}$ with N nodes whose interactions are encoded in the adjacency matrix $\mathbf{A}$. Each node represents an individual (or any entity, *e.g.* edge groups as statistical patterns, city or airport) and a link represents the interaction where the infection spreads. Regarding the dynamics of the node, we keep considering a SIS model but now follow a time-discrete framework for characterizing its dynamical evolution. This way, at each time step t, an infected node i can transmit the disease with probability β to each of its close neighbors, while it can recover with probability μ. Under this SIS conditions, we consider that the dynamical state of each node i of the network is characterized by the probability of being infected at time t, $\rho_i(t)$, so that, the whole network is characterized by a set of N equations (one per node). The discrete time evolution of the probability of the node i to be infected, ρ_i^I at time $t+1$ is given by,

$$\rho_i^I(t+1) = \rho_i^I(t)(1-\mu) + (1-\rho_i^I(t))\Pi_i(t), \tag{2.35}$$

where $\Pi_i(t)$ is the probability of node i being infected,

$$\Pi_i(t) = 1 - \prod_{j=1}^{N}(1-\beta A_{ij}\rho_j^I(t)). \tag{2.36}$$

The structure of the right hand side of the master equation, Eq. (2.35), contains both the lost of infectiousness by recovering their susceptibility (first term) and the gain of infectiousness by being having hazardous contacts (second term). These kind of master equations can be generalized to other compartmental models (*e.g.* SEIR see Chapter 3) where the number of equations to be solved are $N(\Omega_C - 1)$, being Ω_C the number of compartments. In general terms, MMCA equations represent the average probability that an individual i is in compartment m at time $t+1$ computed over an ensemble of stochastic epidemic trajectories generated with the same parameters and initial conditions.

As usual in the SIS model, the order parameter characterizing the overall state of the system is the fraction of infectious individuals in the stationary state. To compute such quantity, one can numerically iterate the equations reaching, for large t, the steady state in which the individual probabilities of being infected are also stationary and the order parameter reads,

$$i^{\infty} = \frac{1}{N}\sum_{i}^{N}\rho_i^{\infty}. \tag{2.37}$$

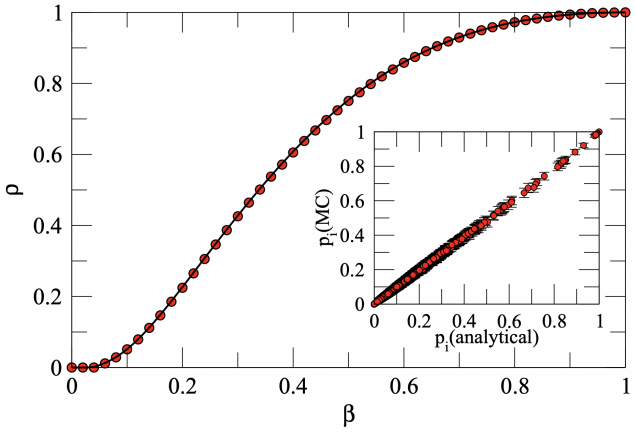


Figure 2.7: **Epidemic prevalence ρ of a SIS model as a function of β for a Scale-free network with 10^4 nodes evaluated with the MMCA**. The red dots correspond to MC simulations and the continuous line to the steady state of the MMCA. The scale-free network with $\gamma = 2.7$ is considered undirected and unweighted. The inset represents the probabilities of each node p_i being infected in the steady state for MC simulations in the y-axis and for MMCA in the x-axis ($\beta = 0.1$). Both results have been obtained with $\mu = 1$. *Figure taken from* [43].

The Fig. 2.7, taken from [43], shows the epidemic prevalence ρ as a function of the infectivity β obtained by Monte-Carlo (MC) stochastic simulations in a scale-free network (red points) and compared with the MMCA formulation (black line). The agreement between both results is flawless. The inset of the same figure shows the microscopic effect of the epidemic outcome p_i in the steady state for each node with the simulations and the theoretical approach.

Given the accuracy of Eqs. (2.35-2.36), we now analyze them to derive the epidemic threshold under this formulation. To this aim, let us to assume that we are in the steady-state regime and close (but above) to the epidemic threshold. Under these two conditions, we can consider that the infection probabilities of the nodes are small enough, such that $\rho_i = \epsilon_i$, to neglect terms of order higher than linear ($> \mathcal{O}(\epsilon^2)$) in Eqs. (2.35-2.36), so that they read,

$$\frac{\mu}{\beta}\epsilon_i = \sum_{j=1}^{N} A_{ij}\epsilon_j. \tag{2.38}$$

As usual after linearizating, the former equation is an eigenvalue problem in which the solution $\vec{\epsilon}$ is an eigenvector of the adjacency matrix with eigenvalue μ/β. However, since we are interested in the minimum infectivity that fulfills the former relation (*i.e.* the epidemic threshold). This implies that the eigenvalue of interest is the maximum one, and this way, we obtain the following expression for the epidemic threshold under the MMCA,

$$\beta_c = \frac{\mu}{\Lambda_{max}\left(\mathbf{A}\right)}. \tag{2.39}$$

This equation, at variance with Eq. (2.39), gives us the threshold for a particular net-

work architecture given by **A**, instead of (as in the HMF approach) the average epidemic threshold for the family of networks with the same degree distribution.

Finally, it is also worth the connection between Eq. (2.38) and the definition of eigenvector centralities, Eq. (2.33), implying that, close to epidemic threshold, those nodes with higher infection probability (and thus those nodes participating the most in the epidemic) are those nodes with larger eigenvector centrality.

2.4 Metapopulations models

Upon this point, in this chapter, we have studied spreading phenomena in close contact networks, whose nodes represent individuals and links encode the nontrivial interaction patterns existing among them. Despite explaining the social factors driving the reaction phase of the disease, contact networks on large scale are difficult to obtain. In this section, we aim to present the extension of the MMCA to a metapopulation model, which includes another social behavior aspect that influences spreading as much as the structure of contact does: human mobility patterns.

The development of theoretical frameworks for spreading phenomena has evolved in parallel with our understanding of human behavior patterns and the characteristics of human populations. To accurately predict the spread of infectious diseases, it is necessary to integrate the interaction of large-scale social systems containing millions of people, whose dynamics can be characterized spatially and temporally. However, these high resolution interactions are hard to obtain in large-scale, and using them to model spreading of infectious diseases would require a large computational resources and the implementation of agent-based models. Here, instead, we present metapopulation models in which human mobility and other characteristics of social interactions are considered in a statistical way, enabling its analysis while keeping close to real data about actual populations.

2.4.1 Human mobility patterns

The current explosion of extensive high spatio-temporal resolution data-sets related to human movement has enabled scientists to analyze and quantify individual and collective mobility patterns. Furthermore, inspired in these data, scientists have been able to develop theoretical models that capture many of the observed spatio-temporal patterns of human trajectories in cities and countries, or even the flows of people traveling around the world [44].

As we strive to understand the diffusion of diseases on larger scales based on complex networks, we are interested in human mobility data-sets or models that describe the mobil-

ity patterns of large aggregates of individuals [2]. To this aim, an useful approach is the use of metapopulations in which the geographical areas under study are divided into patches or supopulations, each having a different number of residents as dictated by demographical data. These patches can correspond to urban districts within a city, or zip codes within regions, regions with countries, etc.

Through the lens of network theory, the patches are associated to nodes of a graph whereas the links correspond to the mobility relationship between these subpopulations, i.e. edges account for the ability of residents to hop between pairs of patches. To capture the architecture of the links in a metapopulation, it is convenient to make use of a weighted and directed network, usually encoded in the so-called Origin-Destination (OD) matrix, $\mathbf{W}$. This network estimates the number of individuals traveling between pairs of patches in which a given geographical area is divided over a given period of time. Mathematically, the OD matrix is an $n \times m$ matrix with n locations as *origin* and m locations as *destination*. In general, the origin and destination points are the same $n = m$ and corresponds to the administrative divisions of the studied geographical area $n = m = N_p$.

Once the OD matrix is constructed and the demographic partition into patches of the entire population is set, the metapopulation is formulated. Thus, this mathematical object encapsulates the human mobility flows among of the N_p subpopulations (or patches), recorded in the OD matrix, $\mathbf{W}$, where each patch $i \in 1, 2, ..., N_p$ has n_i inhabitants such that the overall human population size N is: $N = \sum_{i=1}^{N_p} n_i$. Equipped with this formalism, any type of dynamical model involving the interaction of reaction and diffusion processes can be studied. Typically, these models assume that reaction stages occur within each subpopulation while diffusive stages take place across the metapopulation's edges. In our case, reaction processes will correspond to the spread of pathogens, while diffusion will account for human mobility.

Usual ways to construct OD matrices

Before modelling the interplay between contagion processes and human mobility by means of metapopulations, it is worth making a brief parenthesis to review some of the approaches followed to construct OD matrices. In this regard, in some situations, empirical data can be used to quantify flows between subpopulations. Such empirical data can come from a variety of sources, such as transportation surveys, traffic counts, or geolocated data. However, in the absence of this information, it is also possible to determine interdependencies between locations and mobility flows. In this case, the flows will be determined by a set of intrinsic attributes of the locations (nodes) that allow predictions to be made about the number of trips between pairs of subpopulations. The two most common techniques or models to estimate the OD matrix $\mathbf{W}$ are briefly presented below.

[2] The reader can find more information in [45] an extended review on human mobility.

Gravity model. The migratory flows between two communities based on the Newton Gravity law was first proposed by George K. Zipf in 1946 [46]. Zipf showed that the number of persons moving between two areas, i and j, is proportional to their number of residents, n_i and n_j respectively, and inversely proportional to their shortest transportation distance r_{ij}. With this premises one can write,

$$W_{ij} = K\frac{n_i n_j}{r_{ij}}, \tag{2.40}$$

where K is a constant of proportionality. A generalization of the former equation can be made as it follows,

$$W_{ij} = K m_i m_j f(r_{ij}), \tag{2.41}$$

where m_i and m_j represent the masses of the respective zones (that can relate with any socio-economical variable of the patches under study (such as the population or GDP-per-capita) and $f(r_{ij})$ is considered as the deterrence function.

Intervening opportunities model (radiation model). In 1940, Stouffer suggested migration is determined by the number of intervening opportunities between the origin and the destination [47]. This model was the inspiration for what is know as the radiation model [48]. The radiation model was constructed taking into account the jobs opportunities by zones and a bias from the benefits, so that a traveler chooses the destination with the aim of maximizing the former benefits among the set of closest patches. This way, the former ingredients can be captured into the following explicit formula for each entry of the OD matrix as,

$$\langle W_{ij} \rangle = W_i \frac{m_i m_j}{(m_i + s_{ij})(m_i + m_j + s_{ij})}, \tag{2.42}$$

where $W_i = \sum_{j \neq i} W_{ij}$ is the total number of travelers from location i, which is proportional to the population of the source location; s_{ij} is the total population in the circle of radius r_{ij} centered at i and m_i and m_j are the respective populations.

2.4.2 Movement-Interaction-Return model

Once the metapopulation formalism is established (by means of empirical data or based on synthetic models) we are ready to study the reaction-diffusion processes. As we anticipated earlier, in our case the reactions will correspond to contagions within each patch. For diffusive processes, however, we can choose between different approaches. For example, we can consider that agents follow (under the rules imposed by the flows described in W) random walks, navigating the network in a stochastic way without retaining any memory about their residence patch. In counterpart to this memoryless approach, we can model diffusive processes as back-and-forth events (from residence patch to destination and back), thus capturing the (mostly) recurrent form of flows in large cities. In this thesis, we will take the latter approach and, in particular, we will follow the one proposed by the Movement-Interaction-Return (MIR) model whose main ingredients are presented in this section.

Model formulation

Before discussing the main premises of the MIR model, it is convenient to transform the OD matrix into a different object that will allow us to formulate MIR. Let us consider our metapopulation divided in N_p patches, with n_i the population in the patch i, and whose OD matrix is give by $\mathbf{W}$. Then, we define a new matrix $\mathbf{R}$ whose elements R_{ij} are defined as,

$$R_{ij} = \frac{W_{ij}}{\sum_{l=1}^{N_p} W_{il}} \, . \tag{2.43}$$

From this definition, it is clear that $\mathbf{R}$ is a row-stochastic matrix, i.e., the sum across rows are equal to 1 and have positive entries. This matrix can be interpreted as the probability that a resident in patch i moves to patch j.

Equipped with matrix $\mathbf{R}$, we can define the tree basic stages of the MIR model. These stages are defined for every time step of the dynamics so, in this sense, the MIR model is of time-discrete nature, being the length of the time step defined by the time resolution associated to the mobility flows captured in the original OD matrix. In our case, each time step of the MIR model corresponds to a single day.

Movement. At the beginning of each time step, each agent in the population moves from their residence with probability p. This probability accounts for the sector of the population that moves to other places to accomplish their usual activities (work, school, etc.). Mathematically p is a control parameter that activates or deactivates mobility flows at convenience.

Each of the agents that decides to move at this time step makes it so according to matrix $\mathbf{R}$. This way, an agent with residence i selects patch j as the destination node with probability R_{ij}.

Interaction. Once the population has been relocated among all the patches, individuals interact with others accordingly to their epidemiological state and the epidemiological model at work. The number of interactions that each agents makes usually depends on the density of agents in the patch where the agent is placed at, capturing the effect that more saturated areas give rise to more mixing between agents. At the end of this stage, i.e. after interaction, agents update their epidemiological status.

Return. Finally, to incorporate the recurrent nature of commuting flows, all the individuals are forced to return to their residence and start the process again. At this stage, close contacts interactions also take place at home, however, they are neglected. There are more refined models that incorporate them [49] as we do in Chapter 5.

Note that this model is thought for recurrent mobility patterns, that draw daily activities. It has been measured that this mobility patterns represent from 30%-50% in daily activities [50]. Thus, it represent a reasonable approach to study the influence of this

commuting flows on the spread of communicable diseases.

Model equations

The three stages described above can be seen as the steps of a mechanistic simulation in which one monitors the evolution (according to location and epidemiological status) of each agent. This computationally costly procedure can be skipped by formulating a set of effective equations in which we follow the average epidemiological state of the residents in each patch. To this aim, and given the time-discrete nature of the MIR model we can follow a Microscopic Markovian Chain approach.

To show how the MMCA can be used in the context of the MIR model, in the following, we show its implementation when the SIS compartmental model is the epidemiological framework under study. In this case, we can assign a single dynamical variable to each patch, $\rho_i(t)$, accounting for the probability that the agents whose residence is patch i are infectious at time t. This way, the time-discrete evolution equations read as,

$$\rho_i(t+1) \quad = \quad \rho_i(t)(1-\mu) + (1-\rho_i(t))\Pi_i(t), \tag{2.44}$$

where the first term of right-hand side of the Eq. (2.44) represents the probability that those infected residents of node i do not recover at time t, while the second term represents the probability that those susceptible residents of i at time t get infected. In this latter probability, we have written $\Pi_i(t)$, which is the probability that a susceptible resident from i contracts the disease because of the interaction with other agents. This probability reads,

$$\Pi_i(t) = (1-p)P_i(t) + p\sum_{j=1}^{N_p} R_{ij}P_j(t)\ , \tag{2.45}$$

where the first term encapsulates those contagions happening at the residential patch and the second one captures the probability of contagion when traveling to other patches. Therefore, $P_i(t)$ is the probability that a susceptible individual, regardless of their residence, acquires the infection inside patch i. Then, the evolution of this probability in accordance with the well-mixing hypothesis is,

$$P_i(t) = 1 - \left(1 - \beta\frac{\tilde{n}_i^I(t)}{\tilde{n}_i(t)}\right)^{f_i(t)}\ , \tag{2.46}$$

where $f_i(t)$ is the average number of contacts made in patch i at time t, $\tilde{n}_i(t)$ represents the effective population at patch i at time t, that is, the number of individuals placed but not necessarily residents at patch i after the movement stage, and $\tilde{n}_i^I(t)$ is the effective number of infected individuals inside patch i at time t.

To round off and close the set of effective equations that describe the MIR model under

SIS dynamics, we quantify the effective population at patch i at time t,

$$\tilde{n}_i = \sum_{j=1}^{N_P} n^I_{j\to i} = \sum_{j=1}^{N_P} n_j[(1-p)\delta_{ij} + pR_{ji}(t)], \tag{2.47}$$

and the effective number of infected individuals inside patch i at time t,

$$\tilde{n}^I_i = \sum_{j=1}^{N_P} n_{j\to i} = \sum_{j=1}^{N_P} n_j\rho^I_i(t)[(1-p)\delta_{ij} + pR_{ji}(t)], \tag{2.48}$$

where $n_{j\to i}$ and $n^I_{j\to i}$ represent the total number of residents and the number of infected individuals moving from patch j to patch i respectively.

Epidemic threshold

The Markovian formulation presented above allows, to obtain fast and reliable numerical estimations about both the stationary properties of endemic diseases and the transient dynamics during epidemic outbreaks. This reliability can be observed in Fig. 2.8.a) in which the stationary prevalence obtained with the MMCA is compared with that obtained in mechanistic simulations. However, despite this computational advantage in comparison with mechanistic simulations, the most salient feature of this approach is the possibility of obtaining analytical results.

As already commented in the former sections, one of the most important questions when analyzing epidemic models is to determine the epidemic threshold, i.e., the smallest value of β that enables the epidemic/endemic state to occur. To do so in the MIR model, let us consider that the dynamics given by Eq. (2.44) has reached the steady state, i.e., $\rho_i(t+1) = \rho_i(t) = \rho^*_i$, then,

$$\mu\rho^*_i = (1-\rho^*_i)\Pi^*_i. \tag{2.49}$$

Now we consider that, in addition to the steady-state condition, the values of the stationary prevalence of each patch is negligible, i.e., $\rho^*_i \ll 1 \implies \rho^*_i \sim \epsilon$. This allows us to linearize the rest of equations by neglecting terms with $\mathcal{O}(\epsilon^2)$ or higher. After this linearization we can rewrite Eq. (2.46) as,

$$P_i(t) = 1 - \left(1 - \beta\frac{\tilde{n}^I_i(t)}{\tilde{n}_i(t)}\right)^{f_i} \approx \frac{\beta f_i}{\tilde{n}_i}\sum_{j=1}^{N_P}[(1-p)\delta_{ij} + pR_{ji}(t)]\epsilon_j n_j. \tag{2.50}$$

Substituting this expression in Eq. (2.45) we obtain the linearized contagion probability for susceptible residents of patch i as,

$$\Pi^*_i = \sum_{j=1}^{N_P}\epsilon_j n_j\left[\delta_{ij}(1-p)^2\frac{f_i}{\tilde{n}_j} + p(1-p)\left(\frac{R_{ij}f_j}{\tilde{n}_j} + \frac{R_{ji}f_i}{\tilde{n}_i}\right) + p^2\sum_{l=1}^{N_P}\frac{R_{il}R_{jl}f_l}{\tilde{n}_l}\right]. \tag{2.51}$$

Note that, from the form of the linearized contagion probability Π_i^*, what is on the square parenthesis can be easily interpreted the expected probability of interactions that one individual from patch i has with residents in patch j. In fact, we can write a matrix $\mathbf{M}$ whose entries,

$$M_{ij} = n_j \left[\delta_{ij}(1-p)^2 \frac{f_i}{\tilde{n}_j} + p(1-p) \left(\frac{R_{ij} f_j}{\tilde{n}_j} + \frac{R_{ji} f_i}{\tilde{n}_i} \right) + p^2 \sum_{l=1}^{N_P} \frac{R_{il} R_{jl} f_l}{\tilde{n}_l} \right], \tag{2.52}$$

capture the expected number of interactions that a resident in patch i makes with residents in patch j.

Interestingly, by inspecting the entries of matrix $\mathbf{M}$ we identify the three basic mechanisms of interaction under the MIR framework. Namely, the first term represents those contacts that a resident in i makes with other residents from the same patch. The second term corresponds to contacts made by the resident of patch i with residents of j when either visiting their patch (j) or when the residents of j visit patch i. Finally, the third term contains the interaction that a resident in patch i makes with residents of patch j when visiting a different patch l (different from the patches associated with the agents in contact). Taken these contributions together, matrix $\mathbf{M}$ is called the *Mixing matrix* of the metapopulation.

To follow the derivation of the epidemic threshold, let us assume that the interactions take place all-to-all within each patch *i.e.* $f_i = \tilde{n}_i$. Under this assumption the element of the matrix $\mathbf{M}$ can be rewritten as,

$$M_{ij} = n_j \left[\delta_{ij}(1-p)^2 + p(1-p) \left(R_{ij} + R_{ji} \right) + p^2 \left(\mathbf{RR} \right)_{ij}^T \right]. \tag{2.53}$$

Finally, considering the contribution of the mixing matrix $\mathbf{M}$ in Eq. (2.49), we obtain the following equation,

$$\frac{\mu}{\beta} \epsilon_i = \sum_{j=1}^{N_P} M_{ij} \epsilon_j. \tag{2.54}$$

It is clear that Eq. (2.54) defines an eigenvalue problem and, as it happened for the SIS model on contact networks, since our interest relies on the minimum infectivity value (the epidemic threshold) that satisfies the problem, we should calculate the maximum eigenvalue of the involved matrix. This way, we obtain that the epidemic threshold of the MIR model (under SIS dynamics) reads,

$$\beta_c = \frac{\mu}{\Lambda_{max}\left(\mathbf{M} \right)}. \tag{2.55}$$

Note that, unlike the SIS model on contact networks, here the maximum eigenvalue is computed in a matrix that does not simply capture the links between nodes (which in a metapopulation would be given by $\mathbf{W}$ or $\mathbf{R}$), but in a matrix (the mixing matrix $\mathbf{M}$) in which information about the demographic partition of the population ($\{n_i\}$), the architecture of recurrent mobility flows ($\mathbf{R}$) and the degree of mobility (p) interplay.

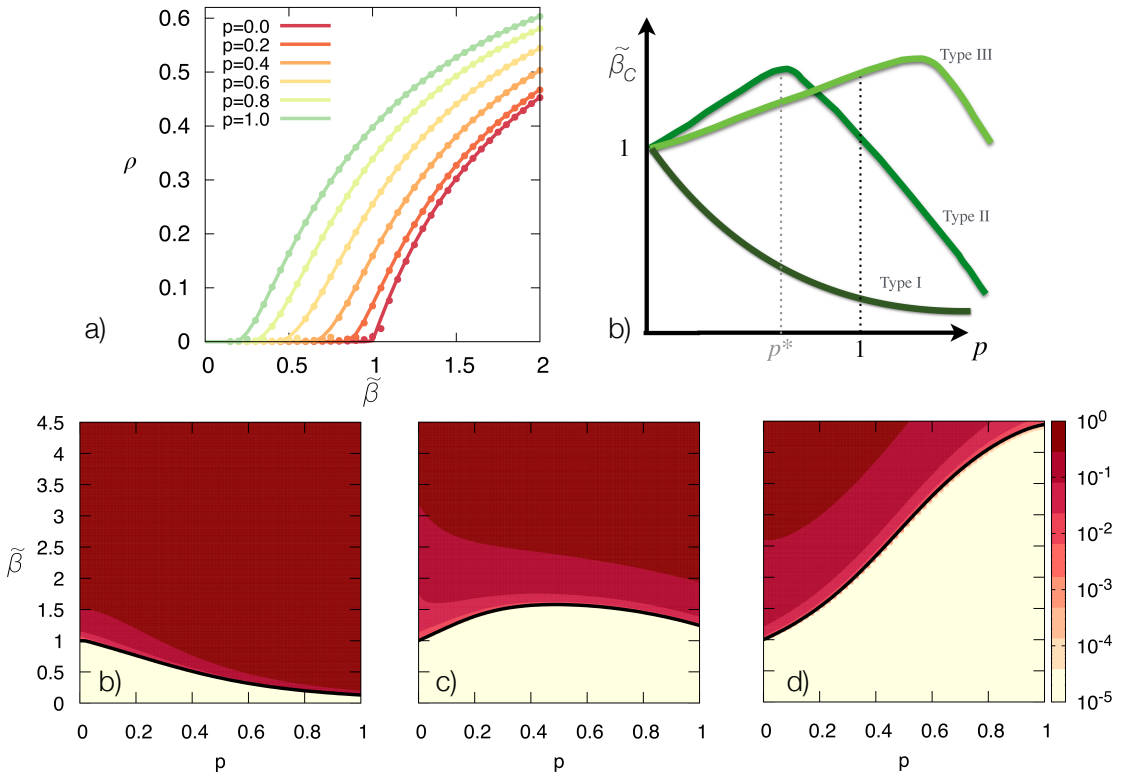


Figure 2.8: **MIR model validation, epidemic threshold & detriment**. a) In this panel the stationary overall prevalence $\rho = \sum_{i=1}^{N_p} \rho_i^*$ is represented as a function of β and different values of the mobility parameter p. The curves correspond to the solution of the MMCA while dots account for the result of implementing mechanistic simulations of the MIR model. b) Qualitative behavior of the epidemic threshold β_c (re-scaled to its value at $p = 0$) as a function of the mobility p for the three different qualitative behaviors observed. The value of p^*, corresponding to the change of the qualitative impact of the mobility on $\beta_c(p)$ in type II metapopulations, is highlighted. In panels c-e) we show the stationary prevalence ρ (color code) as a function of the infectivity β_c (re-scaled by its value at $p = 0$) and the mobility p. The solid black line depicts the analytical estimation for the re-scaled epidemic threshold provided by Eq. (2.55). *Figure taken from* [23].

Epidemic detriment

To round off the presentation of the MIR model, we now briefly present an interesting phenomenon found when analyzing its dependence with the degree of mobility (or fraction of active population) p. In [51] the authors found a beneficial effect on protection against outbreaks as people become more mobile. A close inspection of this detrimental effect of mobility on epidemic processes revealed the existence of three different types of metapopulations:

Type I. Those metapopulations for which, as mobility increases p, the epidemic threshold is reduced monotonically. Usually, these typologies are synthetic metapopulations with

uniformly distributed populations in their patches.

Type II. In this metapopulaton, the epidemic threshold shows a non-monotonous behavior with respect to mobility. In particular, for small values of p, the epidemic threshold increases with p from $p = 0$. This increasing behavior (signalling the aforementioned epidemic detriment) holds up to a value p^*, in which the epidemic threshold turns into a decreasing function of p, thus recovering the behavior of Type I metapopulations.

Type III. In this type of metapopulation the epidemic threshold always increases with p, so that epidemic detriment appears consistently with the increase of mobility.

In Fig. 2.8.b it is shown the behavior of the epidemic threshold as a function of p, $\beta_c(p)$, for the three types of metapopulations. Also, in the panels c-e in the same figure the whole epidemic diagrams $\rho(p, \beta)$ are shown for three star-like metapopulations [51] corresponding to each of the types.

Finally, let us note that the MIR model can be adapted to more realistic scenarios (as we will show in chapters 4 and 5) considering different aspects of both human behavior and the epidemiological characteristics of real diseases. We recall to [23] for more details on this metapopulation framework.

Chapter 3

Epidemic control on contact networks: contact tracing

The highest form of generalship is to thwart the plans of the enemy

Sun Tzu (The Art of War)

As we have introduced in Chapter 2, contact patterns between individuals play a crucial role in shaping the spreading paths of infectious diseases in human populations. Thus, it is clear that a good characterization of the laws governing these interactions is key to advance our understanding of the spread and control of infectious diseases. In recent years, the development of this knowledge has been greatly advanced by the increased spatio-temporal resolution of empirical data on social interactions, enhanced by the persuasive use of mobile and wireless technologies in our daily lives. However, despite the improvements made in mapping our social cartography, we are still far from having comprehensive access to real-time interactions. Therefore, our efforts remain focused on extrapolating the characteristics of the already collected data in real-world environments. Gaining a greater understanding of the transmission of infectious diseases through the paths that follows in human interactions will allow us to enhance our ability to create and implement containment measures.

The research presented in this chapter addresses the possibility of cutting the statistical paths of transmission between close contacts. Specifically, we leverage our experience on incorporating the interplay between social interactions and propagation phenomena, presented in Chapter 2, to analyze the outcome of contact tracing as a control strategy. The key idea of the work presented in this chapter is to consider that, in the same way that a virus spreads through the links of a social network, the identification of contacts of infectious individuals can be seen as a counter-propagation through the same network

of a warning that aims to anticipate the, often covert, transmission of the pathogen by isolating as many infectious agents as possible. Based on this approach, we evaluate the roots of digital contact tracing and study how well contact tracing apps, coupled with the quarantine of identified contacts, can mitigate the spread in real environments. We study the spreading phenomena in real-world face-to-face interactions, where the cascade of detections is triggered in the virtual network constructed by the application's users.

The chapter is structured as follows. After a brief introduction (section 3.1) about the benefits of contact tracing on the mitigation of epidemic outbreaks, in Section 3.2 we structure a mathematical framework that allows us to describe the competition dynamics between contagion processes and contact tracing on real-world data sets. We modify a SEIR compartmental model by taking into account some epidemiological aspects present in the transmission of SARS-CoV-2 and that play a key role in the outcome of contact-tracing policies. Once the model is introduced and the effective Markovian equations are derived, we validate the latter by performing stochastic mechanistic simulations and evaluate the efficacy of tracing protocols. We show that contact tracing is a very effective policy that can reduce a harmful epidemic outbreak by more than 50% of the original outcome, while the isolation of the symptomatic infections cases can reduce the epidemic outbreak by less than 25% in the best scenario (depending on the contact network topology). Third, in Section 3.3, we study the microscopic roots of the symptomatic and contact tracing detection. We found the efficiency of contact tracing relies on the detection of highly connected nodes, highlighting the effects of the competition process between contagion and contact tracing in contact networks. Fourth, in Section 3.4, we investigate the dynamical effects of contact tracing. We found that while symptomatic detection flattens the curve (the maximum of the curve is lowered), contact tracing bends it (there is fast change in the derivative of the temporal incidence). In this same section, we analytically study the effects of contact tracing in the effective reproductive number that keeps the outbreak under better control (*i.e* $\mathcal{R}(t) < 1$). Finally, in Section 3.5, we analyze the effectiveness of contact tracing strategies when resources are limited. We show that the scarce resources to detect infectious cases could lead to a collapse of the control strategy *i.e.* any detection control measure taken might have been in vain to control the infectious disease.

3.1 The effect of contact tracing into the effective reproductive number

We start by exposing the effects that contact tracing and the implementation of Test-Trace-Isolate (TTI) policies have on the spread of communicable diseases. In the Chapter 2, we introduced the basic reproductive number $\mathcal{R}_0$ which is the average number of new cases, or secondary infections, caused by an infected individual in a totally susceptible population.

However, this measure only accounts for the infectious potential at the very beginning of an outbreak for it assumes that all the population is susceptible. Thus, in order to monitor how the infectiousness of pathogen evolves in time, one should rely on the so called effective reproductive number, $\mathcal{R}(t)$, which is the average number of contagions that an infectious individual, infected at time t, makes during their infectious period [52, 53]. This way, the effective reproductive number quantifies the stage of an ongoing outbreak rather than the potential infections made by an index case in a fully susceptible population as $\mathcal{R}_0$ does. The general expression of the effective reproductive number resembles the one introduced in Chapter 2 for $\mathcal{R}_0$ with the addition of the fraction of susceptible individuals in the population at time t, hereafter denoted by $\rho_S(t)$. Then, $\mathcal{R}(t)$ can be expressed as the combination of four contributions in the following way:

$$\mathcal{R}(t) \simeq \beta \cdot k \cdot \tau \cdot \rho_S(t) \ , \tag{3.1}$$

where β is the probability of infection per contact, k is the average number of contacts per unit time, and τ is the average duration of the infectious period. When this number takes values above 1, the number of new infections will grow exponentially in time, whereas when $\mathcal{R}(t) < 1$ the disease is on decline and, if this trend continues over time, the number of new infections will decrease until it disappears. Therefore, one of the main objectives when facing a running outbreak is to reduce the effective reproductive number to values below the epidemic threshold $\mathcal{R}(t) < 1$ through pharmaceutical or non-pharmaceutical interventions.

Pharmaceutical interventions, such as vaccines, might reduce the fraction of the susceptible population. In fact, from Eq. (3.1), it is easy to obtain that, in order to achieve herd immunity, the fraction of susceptible people individuals should decrease up to:

$$\rho_S(t) = (\tau \cdot k \cdot \beta)^{-1} = \mathcal{R}_0^{-1} \ . \tag{3.2}$$

However, in the lack of pharmaceutical interventions, such as vaccines, non-pharmaceutical interventions must come into play in order to decrease either of the three factors (β, k and τ) that interplay in Eq. (3.1) aiming $\mathcal{R}(t) \lesssim 1$. This way, prophylactic measures, such as physical protective means and hygiene, might reduce the disease transmissibility β, whereas social distancing help decreasing the number of interactions, thus affecting k.

The implementation of strict social distancing (such as lockdowns) might be the best and fastest solution to flatten and bend the epidemiological curve as has been the case of COVID-19 pandemic [54–57]. However, this might harm the socioeconomic activities (as we will discuss in Chapter 5). After taking such actions and having some control of the epidemic outbreak, recovering the usual socioeconomic activity involves a delicate trade-off between increasing social interactions, *i.e.* k, while keeping $\mathcal{R}(t) \lesssim 1$. In this scenario. the target focuses on trying to decrease the infectious period τ as much as possible.

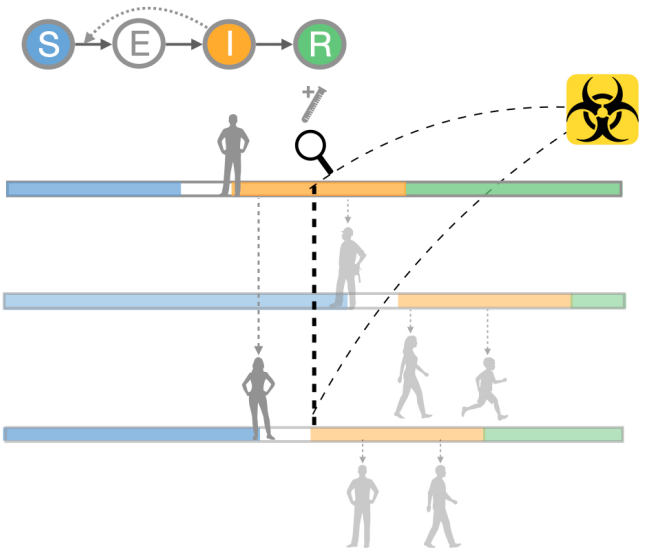


Figure 3.1: **Contact tracing and the infectious period**. The colored lines represent the time-line of observation during an outbreak, colors signalize the compartment at which individuals are according to a SEIR model. Individuals, when infectious (I), may transmit the disease to others during a close contact. Considering $\mathcal{R}(t) = 2$ the tree of infections observed in the image can be generated. This tree of infections can be cut off by 1) isolation when presenting symptoms and seeking medical assistance (testing and isolating, first line) and 2) contact tracing (testing and isolating, last time-line). Note that contact tracing shortens the infectious period of time of infected individuals even though they may be presymptomatic or asymptomatic, making the control more efficient.

Diagnostic testing, treatment and surveillance, have been proven to be effective to in achieve the former goal and avoid secondary waves [58, 59]. The strategy, known as Test-Treat-Isolate (TTI), relies on the detection of symptomatic individuals, isolating them, and, more importantly, trace the contacts that these individuals have had in the recent past as they represent potential infections before the index case has been detected. Thus, a successful contact tracing stops the spread of the virus caused by identifying and isolating secondary cases, thus reducing their average infectious period, τ, in Eq. (3.1) as sketched in Fig. 3.1.

A successful contact tracing requires a personalized and exhaustive search of the contacts of each detected case, taking into account the complex and heterogeneous nature of human relationships [60–63]. This arduous task, however, becomes critical when, as in the case of SARS-CoV-2, presymptomatic and asymptomatic infections are abundant [64–70]. Under these conditions, the symptomatic cases, that are detected, probably have already infected some of their contacts. Moreover, it is possible that a large fraction of their known infectees do not present symptoms during the entire infectious period. Therefore, the contact tracing would remove those many 'silent' potential spreaders.

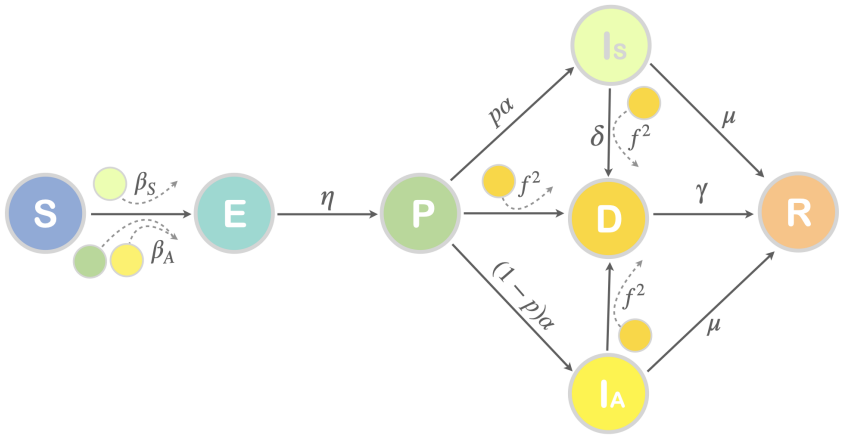


Figure 3.2: **Compartmental epidemic model for CT strategies**. The epidemiological compartments are: susceptible (S), exposed (E), presymptomatic (P), infected asymptomatic (I_A), infected symptomatic (I_S), detected (D) and recovered (R). Arrows represent the possible transitions between the different states. Note that both contagion and detection events are driven by pairwise interactions (curved arrows) whereas the rest of transitions are governed by constant probabilities encoding the typical epidemiological times.

3.2 The competition: contagion dynamics *vs* contact-tracing

The essential attributes of contact tracing (hereafter CT) can be captured by modeling it as a contagion process, in which detected infectious agents spread a warning to those with whom they have been in contact. The propagation of the awareness is a process that competes with the spread of the pathogen itself and aims to suppress the transmission of the virus by removing the active spreaders whose infection is related to the identified case. It is important to note that, unlike other competing spreading processes in which different pathogens spread simultaneously in a population and interfere with the transmission of others [71–78], here the spread of the warning via the cascade of CT identifications is activated only by the presence of the communicable pathogen.

To incorporate the CT dynamics into a framework capturing the transmission characteristics of SARS-CoV-2, we construct an epidemic model including 7 compartments (shown in Fig. 3.2): Susceptible (S), Exposed (E), Presymptomatic (P), Infectious asymptomatic (I_A), Infectious symptomatic (I_S), Detected (D), and Removed (R). Let us start by explaining the main features of the transitions between these states when CT is absent, *i.e.* in the absence of detection of symptomatic individuals:

(i) Susceptible (S) agents are healthy individuals who can be infected by direct interaction with Presymptomatic, Asymptomatic and Symptomatic Infectious agents. The infection probability of a S individual depends on the type of infectious interaction, due to possible viral loads in the different stages of the contagious disease. In general,

we considered, that susceptible agents have the probability of being infected β_A for the interactions with an individual in the states of P and I_A and β_S for the contact with a symptomatic infected individual.

(ii) When an S agent is infected, she converts into Exposed (E) in which the individual displays no symptoms and is not contagious. This state lasts for an average period of η^{-1} days.

(iii) After being in E, agents pass to the Presymptomatic state (P). In this state, no symptoms are observed but the individual is already contagious, with contagion probability per contact of β_A. At this P stage, and without detection, the individual lasts an average of α^{-1} days.

(iv) After the P stage, individuals can continue being asymptomatic (I_A) with some probability $(1-p)$ that is given by the fraction of fully asymptomatic infections. Individuals in this compartment share the same characteristics regarding infectivity and detectability as P. Without detection, an individual lasts an average of μ^{-1} days in this compartment before entering the Removed state (R).

(v) The rest (a fraction p) of P agents pass to be Symptomatic Infectious (I_S). This compartment is characterized by an infectivity β_S when contacting an S agent and, as in the case of the I_A compartment, and (average) of μ^{-1} days before passing to R, if not detected before.

With the former ingredients, the model ($SEPI_SI_AR$) can be viewed as a refined variant of the $SEIR$ class in which compartment I is split in three infectious states, P, I_S and I_A to accommodate the specific contagion forms observed for the dynamics of SARS-CoV-2 [79, 80]. This model can be used as a framework for studying the spreading of SARS-CoV-2 and to assess the impact of contention measures such as social-distancing, prophylactic behavior or strict quarantines. However, to study the influence that detection of symptomatic cases has on the transmission dynamics and, more importantly, to incorporate the possibility of tracing those infectious contacts of symptomatic individuals, a further, and fundamental, compartment capturing those infectious agents detected (D) is needed.

As shown in Fig. 3.2, compartment D can be reached by agents in P, I_S and I_A states. First, those symptomatic infectious can be directly detected as they display symptoms; this happens with probability δ, which is related to the average time spanned from the onset of symptoms to the availability of the test (*e.g.* 2 days would correspond to $\delta = 0.5$). Once a symptomatic agent is detected, CT is activated. However, the possibility of tracing recent contacts is subject to the availability of information about the social activity of those detected. Here, we consider that a fraction f of subjects are equipped with an application that record those acquaintances that have installed it as well. Thus, those contacts of D

Parameter	Value	Description	Reference
η	$1/2.5$ day^{-1}	Probabilty $E \to P$	[81]
α	$1/2.5$ day^{-1}	Probabilty $P \to I_A\,, I_S$	[81]
p	0.65	Fraction of Symptomatic	[82]
μ	$1/7$ day^{-1}	Probability $I_A\,, I_S \to R$	[55, 83]

Table 3.1: Epidemiological parameters of the compartmental model.

individuals that are in the P, I_A, and I_S states can transit to state D by means of an infection-like process in which the infection probability is equal to f^2, *i.e.* the probability that both the detected individual (in D) and the corresponding infectious contact (either in P, or I_A or I_S) are equipped with the CT application. Finally, any individual entering in D transits to Removed (R) with a probability γ, *i.e.*, the CT contagion-like process has an *effective infectious period* of γ^{-1} days. In the following we will set $\gamma = 1$ considering that, once an agent is detected, the corresponding infectious contacts are immediately identified.

The complete model has 6 epidemiological parameters (those of the $SEPI_SI_AR$ model) and 3 additional ones, δ, f, and γ, that characterize the CT process triggered by the detection of symptomatic agents. The values of the epidemiological parameters are presented in Table 3.1 with the exception of $\beta_{\rm S}$ and $\beta_{\rm A}$ that are assumed to be equal, $\beta_{\rm S} = \beta_{\rm A}$, and whose value is taken so that the attack rate in the absence of detection, R^∞, is the same in all the networks analyzed. Having fixed the epidemiological parameters, those corresponding to detection are used to analyze the impact of CT on epidemics.

3.2.1 Microscopic Markovian chain approach

The dynamical evolution of the compartmental model can be studied under a microscopic Markovian time-discrete formulation [84–86]. In this framework, the dynamical state of a node i at time t is given by the probability of being in any of the 7 compartments denoted by $\rho_i^m(t)$, with $m = \{S, E, P, I_A, I_S, D, R\}$. The evolution of these probabilities reads,

$$\rho_i^{\rm S}(t+1) = \rho_i^{\rm S}(t)(1-\Pi_i^{\rm S\to E}(t)) \tag{3.3}$$

$$\rho_i^{\rm E}(t+1) = (1-\eta)\rho_i^{\rm E}(t) + \rho_i^{\rm S}(t)\Pi_i^{\rm S\to E}(t) \tag{3.4}$$

$$\rho_i^{\rm P}(t+1) = (1-\Pi_i^{\rm P\to D}(t))(1-\alpha)\rho_i^{\rm P}(t) + \eta\rho_i^{\rm E}(t) \tag{3.5}$$

$$\rho_i^{\rm I_A}(t+1) = (1-\Pi_i^{\rm I_A\to D}(t))(1-\mu)\rho_i^{\rm I_A}(t) + (1-\Pi_i^{\rm P\to D}(t))(1-p)\alpha\rho_i^{\rm P}(t) \tag{3.6}$$

$$\rho_i^{\rm I_S}(t+1) = (1-\Pi_i^{\rm I_S\to D}(t))(1-\mu)\rho_i^{\rm I_S}(t) + (1-\Pi_i^{\rm P\to D}(t))p\alpha\rho_i^{\rm P}(t) \tag{3.7}$$

$$\rho_i^{\rm D}(t+1) = (1-\gamma)\rho_i^{\rm D}(t) + \Pi_i^{\rm I_S\to D}\rho_i^{\rm I_S}(t) + \Pi_i^{\rm I_A\to D}\rho_i^{\rm I_A}(t) + \Pi_i^{\rm P\to D}(t)\rho_i^{\rm P}(t) \tag{3.8}$$

$$\rho_i^{\rm R}(t+1) = \rho_i^{\rm R}(t) + \gamma\rho_i^{\rm D}(t) + \mu(1-\Pi_i^{\rm I_S\to D}(t))\rho_i^{\rm I_S}(t) + \mu(1-\Pi_i^{\rm I_A\to D}(t))\rho_i^{\rm I_A}(t) \tag{3.9}$$

In the former equations, the quantities $\Pi_i^{\mathrm{S}\to\mathrm{E}}(t)$, $\Pi_i^{\mathrm{P}\to\mathrm{D}}(t)$, $\Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t)$ and $\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t)$ account for the probabilities that an individual passes from Susceptible to Exposed, from Presymptomatic to Detected, from Infectious asymptomatic to Detected, and from Infectious symptomatic to Detected respectively. Considering the adjacency matrix $\mathbf{A}$ capturing the contacts between the nodes, these probabilities read as,

$$\Pi_i^{\mathrm{S}\to\mathrm{E}}(t) = 1 - \prod_{j=1}^{N} \left\{ 1 - A_{ij} \left[\beta_{\mathrm{A}} \left(\rho_j^{\mathrm{P}}(t) + \rho_j^{\mathrm{I}_A}(t) \right) + \beta_{\mathrm{S}} \rho_j^{\mathrm{I}_S}(t) \right] \right\}, \tag{3.10}$$

$$\Pi_i^{\mathrm{P}\to\mathrm{D}}(t) = \Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t) = 1 - \prod_{j=1}^{N} \left(1 - A_{ij} f^2 \rho_j^{\mathrm{D}}(t) \right), \tag{3.11}$$

$$\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t) = 1 - (1-\delta) \prod_{j=1}^{N} \left(1 - A_{ij} f^2 \rho_j^{\mathrm{D}}(t) \right). \tag{3.12}$$

Note that in Eqs. (3.11)-(3.12), the probabilities of being detected, $\Pi_i^{\mathrm{P}\to\mathrm{D}}(t)$, $\Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t)$ and $\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t)$, are calculated as the complement of the probability of not being detected from each of the original compartments (P, I_A and I_S respectively). In the first two cases, $\Pi_i^{\mathrm{P}\to\mathrm{D}}(t)$ and $\Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t)$, the complement of the probability contains the product of the probabilities of not being traced from any detected neighbor j, $(1 - A_{ij} f^2 \rho_j^{\mathrm{D}}(t))$. In contrast, for symptomatic individuals, this probability also accounts for not being detected $\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t)$, through symptomatic detection $(1-\delta)$.

Although the equations above only give information about the probability that a node i is in the Detected compartment at each time, it is possible to construct the probability that a given node i is detected at time t, either after showing symptoms, $D_i^S(t)$, or via contact tracing, $D_i^{CT}(t)$. Thus, the expected number of symptomatic detections at time t is,

$$D^S(t) = \sum_{i=1}^{N} D_i^S(t) = \delta \sum_{i=1}^{N} \rho_i^{I_S}(t) \,, \tag{3.13}$$

and the expected number of detections via CT at time t is,

$$\begin{aligned} D^{CT}(t) = \sum_{i=1}^{N} D_i^{CT}(t) \;&=\; \sum_{i=1}^{N} \rho_i^{\mathrm{P}}(t) \Pi_i^{\mathrm{P}\to\mathrm{D}}(t) + \rho_i^{\mathrm{A}}(t) \Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t) + \rho_i^{I_S}(t) \\ &\times\; (1-\delta) \left(1 - \prod_{j=1}^{N} \left(1 - A_{ij} f^2 \rho_i^{\mathrm{D}}(t) \right) \right) . \end{aligned} \tag{3.14}$$

This last expression captures the effects of network topology on the success of CT strategies. Note that the expressions of $D^S(t)$ and $D^{CT}(t)$ are not used as inputs for the Markovian equations. On the contrary, they are obtained from the time evolution of the individual probabilities associated to compartments P, I_S, I_A and D.

Network	N	$< k >$	r	Reference
Hospital	75	30	-0.18076	[87]
Science Gallery	410	13	0.22575	[88]
School	784	60	0.22814	[89]

Table 3.2: **Characteristics of the three proximity networks**. For each network we show the number of nodes N, the average degree of the nodes $\langle k \rangle$, and the assortativity measured as the Pearson correlation between the degrees of adjacent nodes r. We also report the reference where these networks were presented and analyzed.

To confirm the validity of the Markovian equations, we have compared the results obtained from these equations with mechanistic stochastic simulations. We ran the simulations by assuming that the fraction f of people adopting the CT application is randomly distributed over the network. In addition, we set a small fraction (1%) of Exposed individuals as initial condition to trigger the epidemic outbreak over the rest of susceptible population. Then, at each time step, a stochastic process takes place, the transition rules described in Fig. 3.2. For each value of f and δ we perform 10^3 realizations to calculate the corresponding averages.

In Fig. 3.3 we compare the time evolution of the occupation of each compartment given by the iteration of the Markovian equations with the media of the evolution given by the different realizations of the stochastic dynamics of the compartmental model for $\delta = 0.1$ and $f = 0.4$. We also plot, for each evolution, the 95% CI obtained from the pool of stochastic simulations. In all the cases the Markovian equations reproduce well the trajectories obtained from mechanistic simulations, showing its accuracy and validating its use in what follows.

3.3 Detection: symptomatic *vs* contact-tracing

The solution of the former Markovian equations allows to explore the performance of CT on any particular social network characterized by its adjacency matrix $\mathbf{A}$ in a fast and accurate way. In the following, we will focus on three real proximity networks with different populations and social structures in which data were obtained by means of face-to-face sensors that capture interactions with a temporal resolution of 20 s. In Table 3.2 we report the main structural descriptors of these social networks. Although these networks can be represented as time-varying or weighted graphs, we created the static unweighted versions of the proximity graph in each case. In particular, for the school network, we set a temporal window of 5 minutes as the minimum interaction time to define a link between individuals and focus our analysis on its giant component.

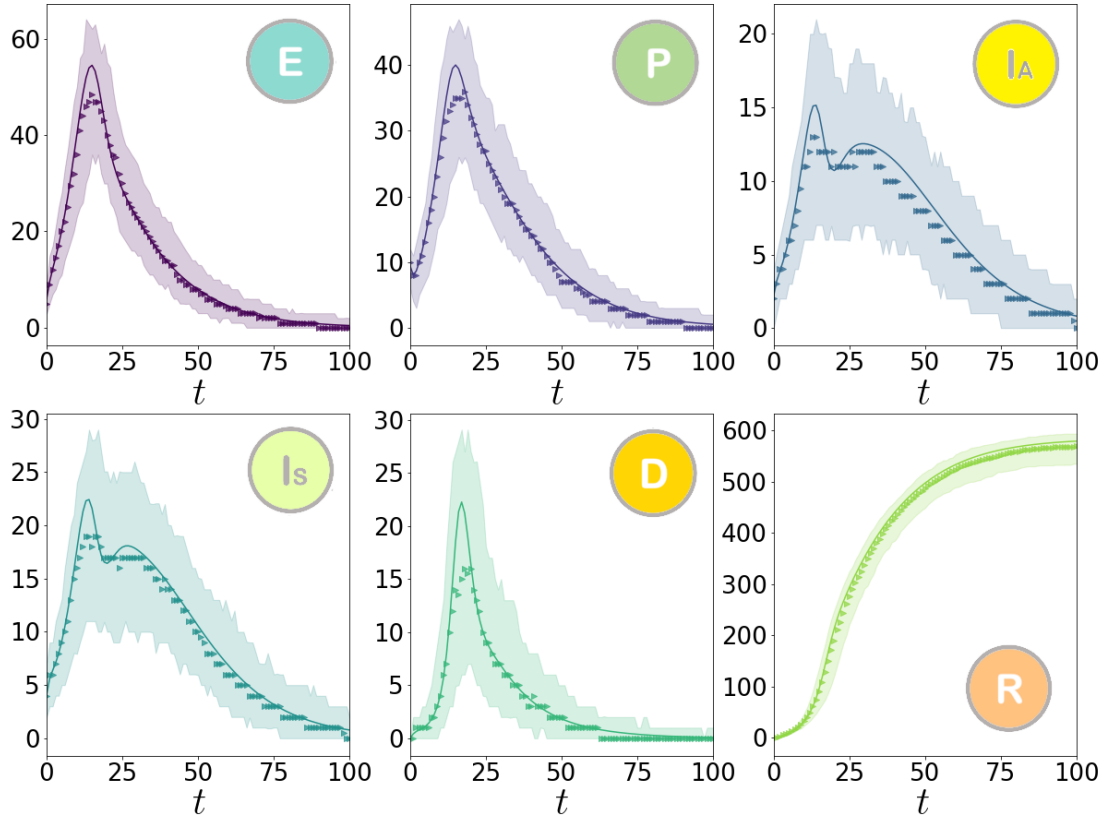


Figure 3.3: **Validation of Markovian dynamics by stochastic simulations**. Each graph shows the evolution of the occupation of each compartment as obtained from stochastic simulations (bands and points (median)) and the Markovian equations (lines). The network substrate is the School proximity network and the detection parameters are set to $\delta = 0.1$ for symptomatic detection, and $f = 0.4$ for CT.

Once the contact networks are constructed from proximity data, we implement the compartmental model equipped with the symptomatic detection and the subsequent CT process to unveil the effects of these two mechanisms on the spread of the disease. In Fig. 3.4.a-c we plot the epidemic diagrams $R^{\infty}(\delta, f)$ showing the attack rate of the disease as a function of the symptomatic detection probability δ and the degree of penetration f of the CT application for the three proximity networks analyzed. In these diagrams it becomes clear that the sole implementation of symptomatic detection ($f = 0$) does not lead to a dramatic decrease of the final attack rate. On the contrary, even with a poor symptomatic detection (*e.g.* $\delta = 0.2$), the addition of CT with a moderate penetration (*e.g.* $f = 0.5$) yields much lower attack rates than the case with perfect symptomatic detection and no CT (*i.e.* $\delta = 1$ and $f = 0$).

We analyze the combined impact of CT and symptomatic detection in panels (d)-(f) of the same Fig. 3.4. There we plot the fraction of cases detected via CT with respect to

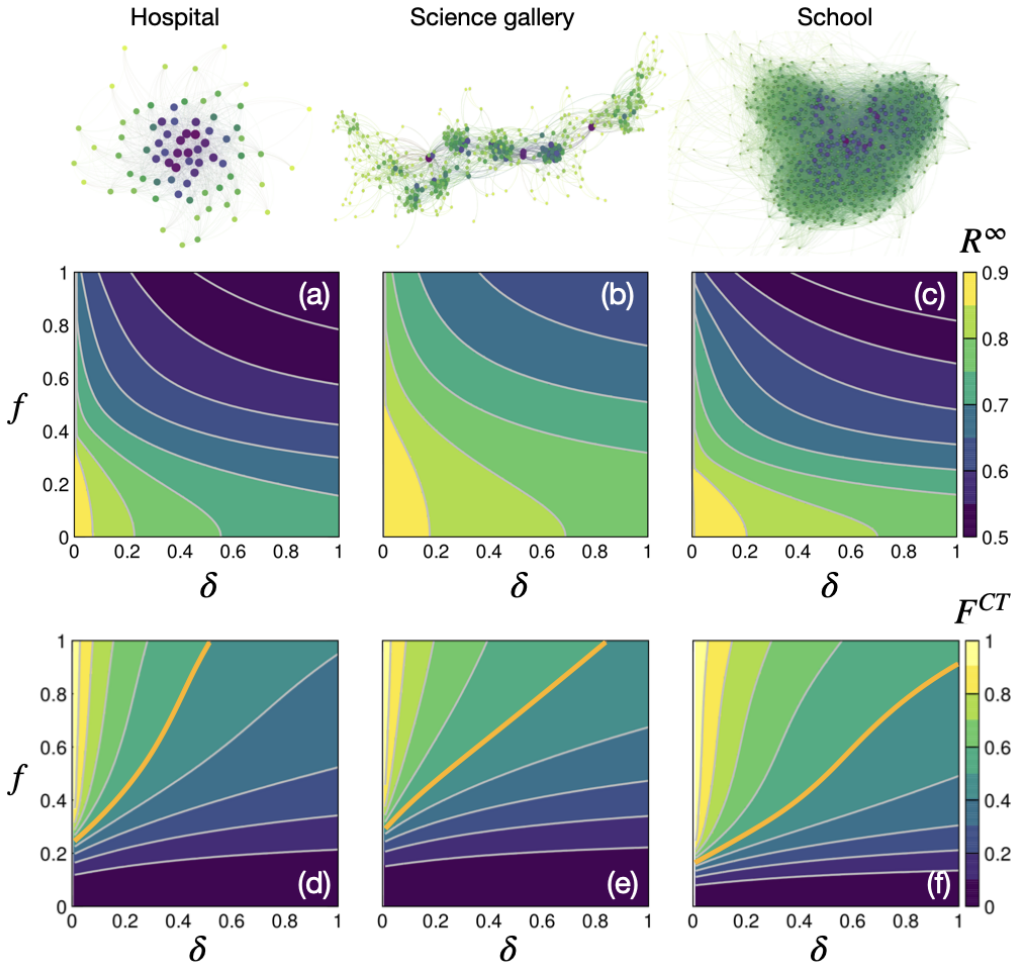


Figure 3.4: **CT *vs* symptomatic detection in three proximity networks**. The color and size of the nodes is proportional to their connectivity, being the small (large) and yellow (blue) ones those vertices with the smallest (largest) degree. Panels (a)-(c) show the attack rate, R^∞ as a function of the quality of symptomatic δ and CT detection f. Panels (d)-(f) show the fraction of detected cases by CT, F^{CT}. We have highlighted the case $F^{CT} = 0.5$ (orange line) signalling that symptomatic and CT detections identify the same number of cases. The infectivity probability per contact ($\beta_S = \beta_A = \beta$) is chosen so that the attack rate in the absence of any kind of detection ($\delta = f = 0$) is $R^\infty = 0.9$ for the three networks.

the total number of infectious cases identified,

$$F^{CT}(\delta, f) = \frac{\sum_{t=0}^{\infty} D^{CT}(t)}{\sum_{t=0}^{\infty} \left(D^{CT}(t) + D^{S}(t)\right)} \,. \tag{3.15}$$

In these plots we highlight the curve (orange) corresponding to those values of δ and f that yield $F^{CT} = 0.5$. Although the partition between CT and symptomatic detections depends on the precise network architecture, from the panels it becomes clear that CT alone is not responsible for the large decrease in the attack rate produced for large values of δ and f, but it is the combination of both mechanisms what allows the efficient suppression of transmission chains.

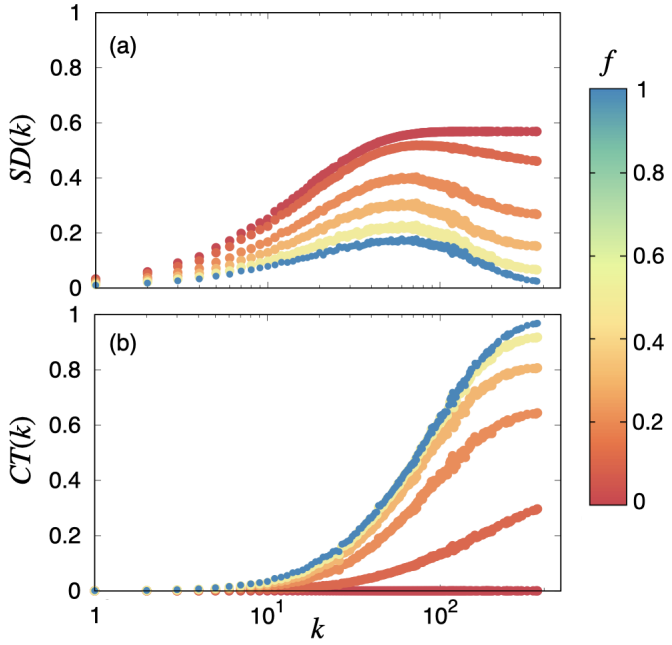


Figure 3.5: **Symptomatic detection and CT as a function of degree k of the nodes**. The panel shows the probability of being detected via symptoms (a) and CT (b) as a function of the degree of the nodes for the School proximity network.The curves correspond (from red to blue) to $f = 0$, 0.1, 0.2, 0.3, 0.5 and 1.

3.3.1 Microscopic differences between symptomatic and CT detections

The last result is quite expected since CT is triggered by symptomatic detections and, consequently, it cannot show up alone. However, from the panels (a)-(c) in Fig. 3.4 it is clear that symptomatic detection alone does not allow a significant decrease of the epidemic impact but it needs the addition of CT policies. Thus, as already mentioned, it is the combination of the two policies what makes detection effective. However, although the number of detections made with each of the two mechanisms is roughly similar, ($F^{CT} \sim 0.5$), when reaching the maximum decrease of the attack rate R^{∞}, not all the identified cases are equally useful to stop the advance of the disease as we show below.

To shed light on the mechanisms behind the effectiveness of CT we analyze the connectivity pattern of those cases detected by CT and symptomatic detection. To this aim, we construct the probability that a node of degree k has been detected during the course of an epidemic by symptomatic and CT detection:

$$SD(k) = \frac{1}{N_k} \sum_{i;|\, k_i=k} \sum_{t=0}^{\infty} D_i^S(t) \,, \tag{3.16}$$

$$CT(k) = \frac{1}{N_k} \sum_{i;|\, k_i=k} \sum_{t=0}^{\infty} D_i^{CT}(t) \,, \tag{3.17}$$

where N_k is the total number of nodes with degree k.

In Fig. 3.5 we plot the functions $SD(k)$ and $CT(k)$ when $\delta = 0.5$ and f varies in the range $f \in [0, 1]$, *i.e.*, from no CT to a situation in which CT is always possible. From panel (a) it is clear that, when no CT is at work ($f = 0$), the function $SD(k)$ is an increasing function of the degree, *i.e.* the largest the connectivity of a node the more probable that it is detected. This is clearly due to the high risk of infections of those nodes with a large connectivity that, consequently, have more probability of being at compartment I_S and hence being detected. However, as f increases the probability $SD(k)$ becomes a non-monotonous function of k and displays a maximum at some degree class $k^\star$. The reason behind this behavior is the action of CT, which shows [see panel (b)] an increasing pattern for $CT(k)$ for any value of $f > 0$. As f increases the identification of those infected nodes with the largest degrees, *i.e.* the super-spreaders, is progressively replaced by CT in detriment of symptomatic detection. In fact, by comparing with the function $SD(k)$ for $f = 0$, we notice that, for $f = 0.2$, CT already outperforms the ability of symptomatic detection in the identification of super-spreaders. Moreover, when $f > 0.5$ the probability that a super-spreader is detected via CT is close to 1, pinpointing that the effectiveness of CT is rooted on the identification and isolation of super-spreaders that have been in contact with those symptomatic cases previously detected.

3.4 Dynamical effects: flattening *vs* bending of the epidemic curve

An early identification of super-spreaders is key to achieve an effective control of an outbreak. Super-spreaders can be identified by symptomatic detection in the first stages of an epidemic as they are usually exposed to a number of potential infections due to their large connectivity. However, symptomatic detection restricts its identification to those that display symptoms and, moreover, their identification always happens once after they have transited the P compartment, thus provoking contagions in a number of neighbors prior to detection. On the contrary, CT allows catching super-spreaders at any infectious compartment, specially those in P, thus providing with an early suppression of super-spreading events. The earliness of CT with respect to symptomatic detection is manifested in the progressive replacement in the identification of large degree nodes as f increases observed in Fig. 3.5.

The early identification of super-spreaders provided by CT is more evident when analyzing the time evolution of the epidemic curve when subjected either to symptomatic detection or to CT. To monitor the effects that both detection mechanisms have on the epidemic trajectory, we monitor in panels (a) and (b) of Fig. 3.6 the time evolution of the number of new contagions when either symptomatic detection or CT is the only detection mechanism at work and tuning their corresponding strength by varying δ and f respec-

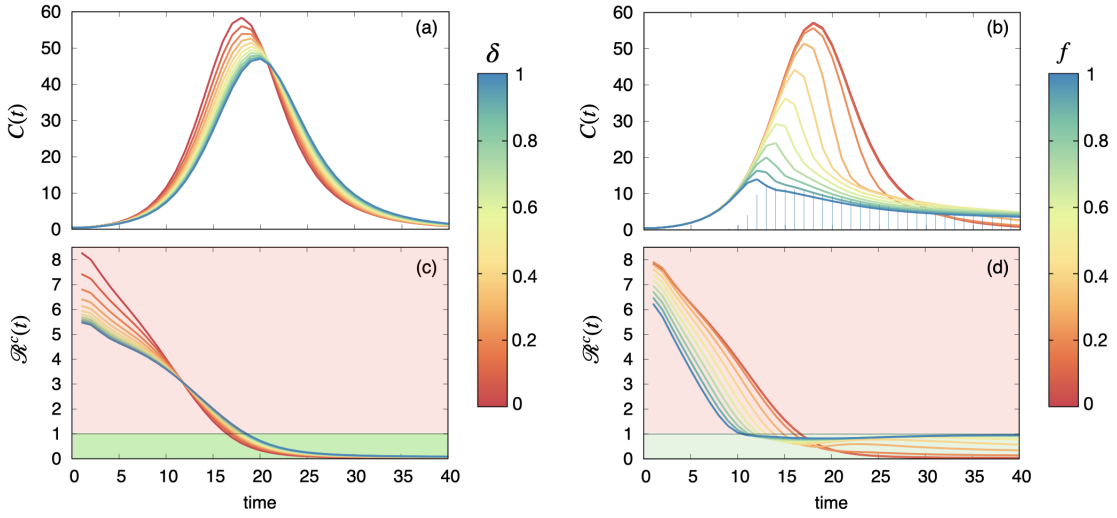


Figure 3.6: **Dynamical evolution of the epidemic trajectory under symptomatic detection and CT**. Panels (a) and (b) show the evolution of the new contagions when different degrees of symptomatic detection and CT are implemented, respectively. In panel (b) we also show (impulses) the evolution for number of detected cases via CT when $f = 1$. The bottom panels, (c) and (d), show the evolution of $\mathcal{R}(t)$ corresponding to the different epidemic curves shown above. The calculations are performed using the School proximity network.

tively. The number of new contagions that occur at a given time t, $C(t)$, can be readily computed from the Markovian dynamics as,

$$C(t) = \sum_{i=1}^{N} \rho_i^S(t) \Pi_i^{S\to E}(t) \, . \tag{3.18}$$

The first set of curves, panel (a), shows how symptomatic detection changes the epidemic curve as δ varies from 0 to 1. This plot shows that the sole action of symptomatic detection causes the so-called flattening of the epidemic curve in which the peak of the epidemic curve is delayed and decreased. This flattening becomes more pronounced as δ increases, thus reducing progressively the final attack rate. On the other hand, the effect of CT, see panel (b), yields a qualitative different scenario. In this case, we set a very small degree of symptomatic detection $\delta = 0.05$ to trigger the CT cascade, and vary f from 0 to 1. The result implies that the epidemic curve is no longer flattened but bent, *i.e.* CT is able to reverse the increasing tendency of the curve corresponding to $f = 0$. This bending is further anticipated as the fraction f of individuals adopting the CT application increases. To illustrate the bending action of CT, we show (blue bars) the evolution of the number of CT detections, $D^{CT}(t)$ [Eq. (3.14)], for the case $f = 1$. It is clear that, as soon as CT is triggered, the increasing trend of the epidemic curve is reversed leading to a successful mitigation.

3.4.1 Effective reproduction number $\mathcal{R}(t)$

To shed more light on the qualitative differences between CT and symptomatic detection, we can monitor the time evolution of effective reproduction number $\mathcal{R}(t)$ to analyze their respective impact on the evolution of the infective power of the nodes in the network during an epidemic outbreak.

As introduced in Section 3.1, $\mathcal{R}(t)$ is defined as the average number of secondary cases that a case infected at time step t will eventually infect during her infectious period [90]. In our particular compartmental dynamics, an agent can transit three infectious states, namely P and I_S and I_A, making a different average number of secondary infections during its permanence each state. Therefore, in principle, the effective reproduction should be written as,

$$\mathcal{R}(t) = \mathcal{R}_P(t) + p \cdot \mathcal{R}_{I_S}(t) + (1-p) \cdot \mathcal{R}_{I_A}(t) \,, \tag{3.19}$$

where $\mathcal{R}_\star(t)$ is the average number of infections made by an agent infected at time t when staying at compartment $\star$. Unfortunately, the former partition is not straightforward when CT is active, which is the central ingredient of our study. In the presence of CT the time window associated to each infectious compartment does depend in a non-trivial way on the instant state of the system. In this case, the calculation of $\mathcal{R}(t)$ should be performed starting from its general definition,

$$\mathcal{R}(t) = \frac{\sum_{i=1}^{N} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)\mathcal{I}_i(t)}{\sum_{i=1}^{N} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)} \tag{3.20}$$

where $\mathcal{I}_i(t)$ is the number of infections caused by agent i provided she has been infected at precise time t.

To calculate $\mathcal{I}_i(t)$ in a general way, we introduce the joint probabilities $\mathcal{P}_i(\tau_E, \tau_P, \tau_A|t)$ and $\mathcal{P}_i(\tau_E, \tau_P, \tau_S|t)$ that account for the probabilities that an agent i infected at time t stays a time τ_E in the exposed compartment, a time τ_P in the presymptomatic stage, and a time τ_A or τ_S in the infectious asymptomatic or symptomatic stages respectively. Note that these two probabilities does not depend on t and factorize,

$$\mathcal{P}_i(\tau_E, \tau_P, \tau_\star|t) = \mathcal{P}(\tau_E)\mathcal{P}(\tau_P)\mathcal{P}(\tau_\star) \,, \tag{3.21}$$

only when the time interval in each compartment does not depend on the state of the system, as it is the case when $f = 0$. The general form of these conditional probabilities for any value of δ and f, *i.e.* when symptomatic detection and CT are present, is derived in the Appendix A.

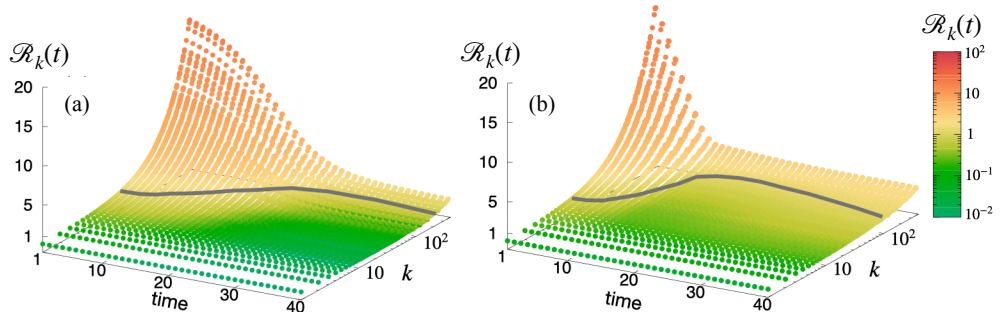


Figure 3.7: **Evolution of the reproduction number for the different degree classes**. The panels show the evolution of $\mathcal{R}_k(t)$ for (a) symptomatic detection ($\delta = 1$, $f = 0$) and (b) CT ($\delta = 0.05$, $f = 1$) for all the degree classes present in the School proximity network. The curve in grey account for the points where $\mathcal{R}_k(t) = 1$.

Once known the probabilities $\mathcal{P}_i(\tau_E, \tau_P, \tau_S|t)$ and $\mathcal{P}_i(\tau_E, \tau_P, \tau_A|t)$, the average infections made by an individual i can be written as,

$$\begin{aligned}\mathcal{I}_i(t) &= (1-p)\sum_{\tau_E=1}^{\infty}\sum_{\tau_P=1}^{\infty}\sum_{\tau_A=0}^{\infty}\mathcal{P}_i(\tau_E,\tau_P,\tau_A|t)\left\{\sum_{s=t+\tau_E+1}^{t+\tau_E+\tau_P+\tau_A}\sum_{j=1}^{N}A_{ij}\beta_A\rho_j^S(s)\right\}\\ &+ p\sum_{\tau_E=1}^{\infty}\sum_{\tau_P=1}^{\infty}\sum_{\tau_S=0}^{\infty}\mathcal{P}_i(\tau_E,\tau_P,\tau_S|t)\\ &\times\left\{\sum_{s=t+\tau_E+1}^{t+\tau_E+\tau_P}\sum_{j=1}^{N}A_{ij}\beta_A\rho_j^S(s)+\sum_{s=t+\tau_E+\tau_P+1}^{t+\tau_E+\tau_P+\tau_I}\sum_{j=1}^{N}A_{ij}\beta_I\rho_j^S(s)\right\}, \qquad (3.22)\end{aligned}$$

and the evolution of $\mathcal{R}(t)$ can be readily computed from the Markovian evolution to illustrate the qualitative differences between symptomatic detection and CT.

In the bottom panels, (c) and (d), of Fig. 3.6 we show the evolution of $\mathcal{R}(t)$ for the different epidemic curves shown in panels (a) and (b). The evolution of $\mathcal{R}(t)$ when symptomatic detection is at work shows the fingerprint of the flattening effect observed in panel (a), *i.e.*, the effective reproduction number, while being smaller in the beginning of the epidemic, slows down the decreasing trend as δ increases, thus reaching $\mathcal{R}(t) = 1$ at larger times. On the contrary, from panel (d) we observe that the action of CT is the opposite: as f increases the decreasing trend of $\mathcal{R}(t)$ is accelerated, thus achieving $\mathcal{R}(t) = 1$ much sooner than in the case without detection. It is also remarkable that, in the case of large values of f, the long term values of the effective reproductive number remain $\mathcal{R}(t) \lesssim 1$. This explains the situation shown in panel (b) in which an almost-steady small number of new contagions are observed after the epidemic curve is bent, thus providing a large and slow discharge of new cases. This way, CT places the system in a kind of critical equilibrium that lasts as long as there is a large enough fraction of susceptible individuals to be infected.

To round off, we can use the expression for $\mathcal{R}(t)$ to monitor the impact on the reproduction number of each degree class. To this aim, we can define the effective reproduction number of the nodes of degree k, $\mathcal{R}_k(t)$, as,

$$\mathcal{R}_k(t) = \frac{\sum_{(i\,|k_i=k)} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)\mathcal{I}_i(t)}{\sum_{(i\,|k_i=k)} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)} \,. \tag{3.23}$$

Computing this expression for each degree class, we show in Fig. 3.7 the evolution of $\mathcal{R}_k(t)$ when symptomatic detection ($\delta = 1,\ f = 0$) and CT ($\delta = 0.05,\ f = 1$) are at work in panels (a) and (b) respectively. From these two plots it becomes clear the fast drop of the infective potential of super-spreaders under the action of CT compared to the case of symptomatic detection.

3.5 Adding more realism into the model: Manual CT and limited resources

To round off this chapter, we now address two features that, in real circumstances, occur in the implementation of TTI policies. On the one hand, we want to include the possibility that the healthcare system may not be able to respond in the usual time frame to resource scarcity, such as tests, due to a high demand. This situation, common during periods of high epidemic incidence, means that the probability that symptomatic agents are tested, δ, is no longer constant as we have assumed up to now. On the other hand, a second aspect to consider is the fact that, even with a low penetration of the tracing application, CT is able to operate "manually", *i.e.*, by directly asking the detected person about her usual contacts. Obviously, this version does not offer an exhaustive tracking of contacts and is limited to the most usual acquaintances with whom the detected person is aware of having interacted. The inclusion of the manual version of the CT results in a change in the probability of being detected by a contact.

3.5.1 Markovian equations

We star by generalizing the basic model of contact tracing to include both manual and digital contact tracing. To this aim, let us first define that the probability of being digitally or manually contact traced is given by the joint probability,

$$P_{ij}^{CT} = P_{ij}^{DCT} + P_{ij}^{MCT} - P_{ij}^{DCT}P_{ij}^{MCT} \tag{3.24}$$

where $P_{ij}^{DCT} = f^2$ and P_{ij}^{MCT} are the probabilities of the node i being traced in a digital and manual way respectively, provided i is a neighbor of a detected infected node j. This way, we consider that a detected j individual can directly notify those contacts she is aware of. To keep the model simple we assume that the typical number of manually traced

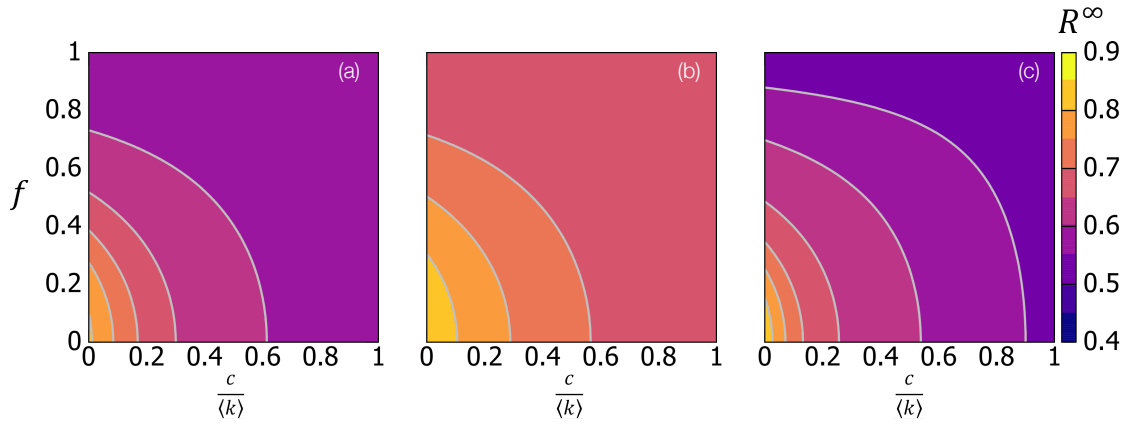


Figure 3.8: **Effects of manual and digital CT**. Attack rate R^{∞} as a function of the fraction of people that adopted the detection application, f, and the ratio between the number of contacts traced manually and the mean degree of the network, $c/\langle k \rangle$, for the close contact networks: (a) the hospital, (b) the science gallery and (c) the school.

contacts is c, thus neglecting the potential heterogeneity of real contacts in social networks. This assumption implies that the a close contact of j will be detected with probability 1 when $k_j < c$ whereas when $k_j > c$ the probability is $\frac{c}{k_j}$. With these two possibilities in mind, we can write the probabiity that j is manually traced by a detected infectious neighbor i as,

$$P_{ij}^{MCT} = 1 + \frac{1}{2k_j}((c - k_j) - |c - k_j|). \tag{3.25}$$

As discussed above, in this final part we also incorporate the constraints on detection when the health system is overwhelmed. To this end, we define a new function, the resource availability $g(t)$, which changes as a function of the extent of the outbreak. In particular, we consider that the availability of resources depends on the number of infected symptomatic individuals so that, if the number of symptomatic individuals is below a threshold, the detection capacity remains constant $g(t) = g_0$, but when a certain tolerance is exceeded the capacity of the system decreases exponentially. Thus, the resource availability $g(t)$ is defined as,

$$g(t) = g_0 \qquad if \qquad I_S(t) < N\theta, \tag{3.26}$$

$$g(t) = g_0 e^{-\lambda(I_S(t) - N\theta)} \quad if \quad I_S(t) > N\theta, \tag{3.27}$$

where $I_S(t) = \sum_{j=1}^{N} \rho_j^{I_S}(t)$ is the number of symptomatic infected people, λ can be understood as the decay rate of the system's capacity under the increase of the test demand, and θ is the tolerance threshold, i.e., the maximum demand that the system can stand before the decay of resources starts to show up.

In order to introduce the resource availability $g(t)$ in the Markovian formalism we consider $g_0 = 1$ and interpret $g(t)$ as the probability that a test demand is fulfilled. This implies that parameter δ in the original model can be substituted by $g(t)$, and the probabilities

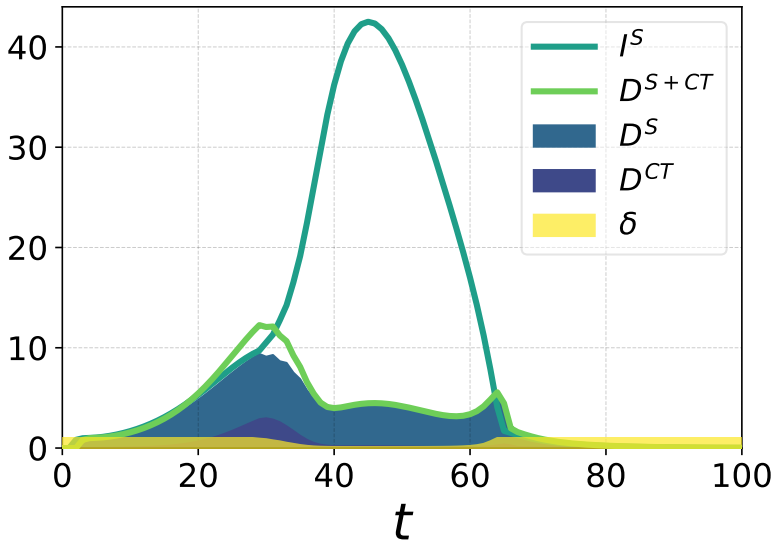


Figure 3.9: **Dynamics of detection collapse**. Time evolution of detection reached by symptomatic detection D^S and contact tracing D^{CT}. We also represent the time evolution of the number of symptomatic individuals I^S and the detection parameter δ shaped by $g(t)$. The model parameters to generate these trajectories are: $\theta = 0.01$, $\lambda = 0.2$, $f = 0.2$, $c = 0$, $\gamma = 1$, $p = 0.65$, $g_c = 0.1$ and $\beta_A = \beta_S = 0.0635$.

(3.11-3.12) of being detected transform into,

$$\Pi_i^{P\to D}(t) = \Pi_i^{I_A\to D}(t) = 1 - \prod_{j=1}^{N} \left(1 - A_{ij} g(t) P_{ij}^{CT} \rho_j^D(t)\right), \tag{3.28}$$

$$\Pi_i^{I_S\to D}(t) = 1 - (1 - g(t)) \prod_{j=1}^{N} \left(1 - A_{ij} g(t) P_{ij}^{CT} \rho_j^D(t)\right), \tag{3.29}$$

where we have already incorporated the probabilities of being identified as contact of an infectious neighbor (3.24) in the former expressions.

Before we show the effects of the limited resources in the generalize model, in Fig. 3.8, we map the effects of manual and digital CT into epidemic diagrams $R^\infty(c/\langle k\rangle, f)$ exhibiting the attack rate as a function of the number of people an infected individual can recall, $c/\langle k\rangle$, and the adoption of the CT application, f, for the proximity networks we have previously analyzed. In these diagrams we consider the resources to test and detect are unlimited so $g(t) = g_0 = 1$. Both policies have almost symmetrical effects on the attack rate, suggesting that manual CT has similar macro and micro effects as digital CT (see Sections 3.3 and 3.4) acting as a competitor in the dynamics.

To show the impact of the existence of limited resources, in Fig. 3.9 we plot the epidemic evolution in a scenario where the incidence exceeds the tolerance threshold and leads to the loss of the detection capacity. In particular, we can observe how the evolution for the number of detected (both those detected by symptoms and those detected by tracing),

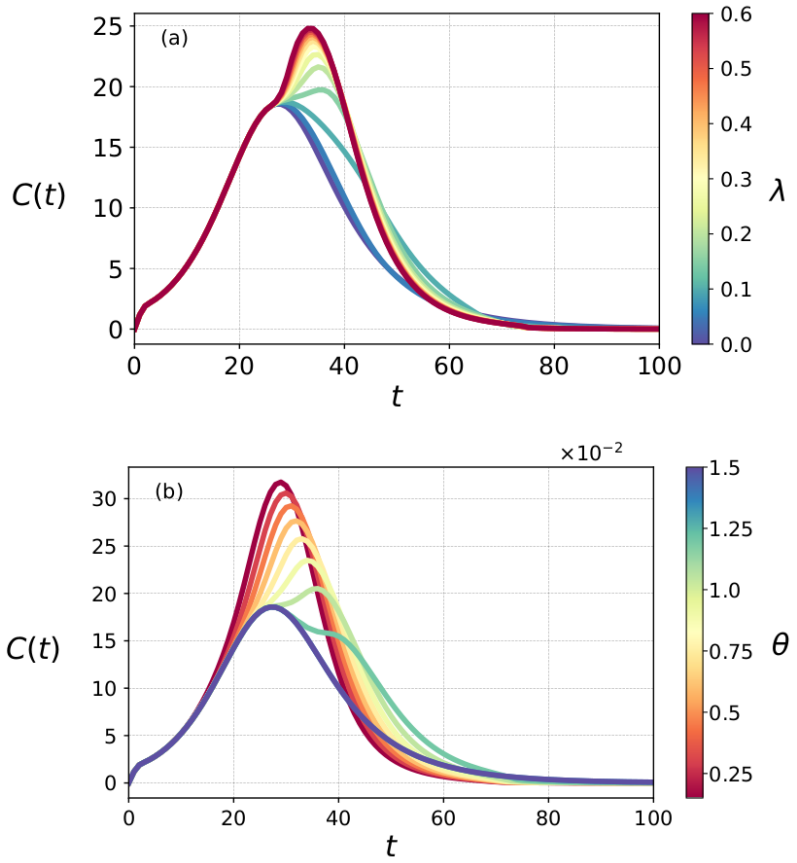


Figure 3.10: **Epidemic curve and scarce resources**. (a) Time evolution of the incidence for different values of the resources loss strigency λ with $\theta = 0.01$. (b) Epidemic curve for different values of the resources availability threshold θ with $\lambda = 0.02$. In both panels, the rest of the parameters are set to: $f = 0.2$, $c = 0$, $p = 0.65$, $\gamma = 1$, $g_r = 0$, $\beta_A = \beta_S = 0.0635$.

D^{S+CT}, shows an oscillating trend with three detection maxima. The first rise of D^{S+CT} corresponds to epidemic control that slows the spread of the pathogen until it reaches its maximum (see yellow curve for $g(t)$). Once the tolerance threshold is reached, the system fails and CT detection decays (See curve for D^{CT}). This leads to a considerable increase in incidence resulting in an epidemic wave with the only control of the detection of few symptomatic cases (second maximum of D^{S+CT}). Finally, a new small detection peak occurs at the end, indicating that CT detection system has been recovered (not the increase in $g(t)$), yielding to the third detection peak. However this recovery occurs too late, since the propagation dynamics are in the natural decline due to the depletion of the susceptible pool.

3.5.2 Collapse transition

To shed more light on the effects of the resource availability we now study how epidemic curves are shaped when different tolerance threshold θ and different decay rates for re-

sources, λ, are used in the definition of $g(t)$. In Fig. 3.10.a, we analyze the consequences of modifying λ by inspecting the evolution of the number of contagions when $\theta = 0.01$. Note that the smaller the value of λ the slower the decay of $g(t)$ and, thus, the more resilient the system is to the loss of resources. Obviously, when $\lambda = 0$ (blue line) we recover the usual scenario in which the detection capacity is not altered. As λ increases, we observe that the epidemic curve separates from the original one when large incidence values appear. For small λ, we observe that the curve becomes broader, while, for large values of λ, there is a second peak of contagion signaling that the epidemic has gone out of control.

In Fig. 3.10.b we fix $\lambda = 0.02$ and vary the value of the tolerance threshold θ. Obviously, if θ is very large (blue line) the system is always under control and CT detection operates as usual. When θ decreases we observe two different trends. For moderate values, *e.g.* $\theta = 0.0125$, the epidemic wave is broader, signalizing that CT has been lost for a short window of time (close to the maximum of contagions) while for smaller values CT is lost during the rise of the epidemic incidence and giving rise to a new epidemic wave whose maximum is much larger than the original (controlled) one.

From the former results, seems clear that, under limited testing resources, the system display two well different regimes, one in which the epidemic is under control, even if CT is weakly lost during the peak of the epidemic wave, and another one in which control is lost and the epidemic outcome appears to be similar to the one observed without any tracing policy. To test this hypothesis in the two panels of Fig. 3.11 we plot the attack rate R^∞ as a function of the infectivity ($\beta_A = \beta_S$) for different values of θ and λ. For the sake of comparison, we plot in bright red the curve of free propagation, *i.e.* that without any control measure, while the light blue line represents the epidemic curve when resources are unlimited. These two cases correspond to $\theta = 0$ and $\theta = 1$ respectively.

From the different curves in panel 3.11.a, it is clear that for large enough values of the infectivity the epidemic curve cease to be under control and an abrupt transition from the fully controlled scenario to the uncontrolled one takes place. Obviously this transition happens for larger infectivities β_A when the tolerance threshold increases. The sharpness of this transition is controlled by the decay rate of the resource availability λ as shown in panel 3.11.b. In this case, since θ is fixed, the transition occurs at the same value of β_A while the convergence towards the uncontrolled epidemic curve becomes smoother as λ decreases.

3.6 Conclusions

In this chapter, we have analyzed the effectiveness of CT by formulating its functioning as a secondary contagion dynamics that is triggered by the identification of symptomatic individuals and propagates as a detection wave. This way, CT competes with the spread

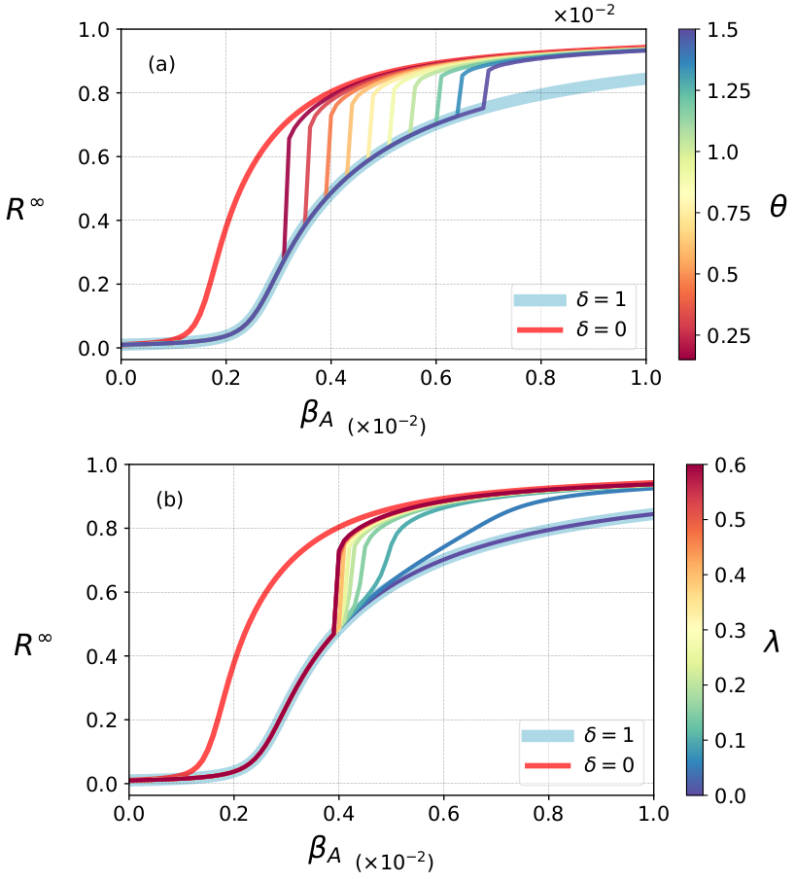


Figure 3.11: **Attack rate, infectivity and detection collapse**. (a) Characterization of the abrupt collapse phase transition in terms of θ for $\lambda = 0.6$, $f = 0.2$, $c = 0$, $p = 0.65$, $\gamma = 1$, $g_r = 0$ and $\beta_A = \beta_S$. (b) Characterization of the abrupt collapse phase transition in terms of λ for $\theta = 0.0045$, $f = 0.2$, $c = 0$, $p = 0.65$, $\gamma = 1$, $g_r = 0$ and $\beta_A = \beta_S$.

of the pathogen eliminating potential transmission chains. The proposed compartmental model has been analyzed under a Markovian framework that nicely agrees with mechanistic simulations and allows a systematic study of particular network architectures such as the proximity networks and the analysis of microscopic dynamical patterns. By using this approach, we were able to derive the evolution of the effective case reproduction number, which quantifies the average number of secondary cases that a case infects over the course of her infectious period and, therefore, can be used to monitor the spread of diseases as well as the effectiveness of contention strategies.

The results of our study suggest the importance of implementing CT in addition to symptomatic detection. An important indicator of CT's effectiveness is its ability to abruptly reverse the trend of the original epidemic curve, causing it to bend. The implementation of symptomatic detection, on the other hand, produces a softer modification of the epidemic trajectory, known as flattening, which results in a delayed epidemic peak.

The qualitative differences in the performance of CT and symptomatic detection are

microscopically rooted. We have shown that CT allows an early detection of large degree infectious nodes, well before than symptomatic detection does. This early identification of super-spreaders during their contagious cycle is fundamental to pre-emptying the spread of the virus, cut the potential transmission channels, and cause the bending of the epidemic trajectory rather than its flattening. In more general grounds, the advantage of CT in advancing an outbreak lies in the use of the same strategy as the pathogen itself: rapid spread through our social fabric by taking advantage of its heterogeneous nature.

Furthermore, in the last section of this chapter, we presented the generalize model that includes both the combination of manual and digital CT and the limitation of testing capacity. We have shown that resource scarcity could lead to an abrupt failure of the implementation of any control strategy and that the collapse of an overwhelmed system could lead to the same result as if no control strategy were implemented.

Apart from these findings, the formalism presented here allows understanding and quantifying the impact of CT strategies in particular proximity networks that are critical to protect. Examples of these networks include companies, hospitals and schools to name a few. It also allows the possibility of designing modifications of these social structures in order to both decrease the impact of potential virus transmission and enhance the efficiency of CT strategies.

The presented Markovian model's simplicity makes it possible to analyze the significance of CT and the qualitative changes that occur differently from symptom detection. However, the compartmental dynamics and the social structure can be extended in order to address the performance of CT in more realistic scenarios. For instance, here we assume that CT is implemented by means of an application whose penetration is characterized by f and the number of contacts a person can remember c. However, we have captured this duality using the same close contact network. Instead we could better represent this duality with the use of multiplex social structures in which usual contacts are distinguished from an specific environment. Other more realistic features could be added such as, the presence of delays in digital or manual CT. For instance, in many practical situations tests cannot be done immediately and thus the waiting time cooperates with the infectious period of infectious acquaintances, since they are not identified until the symptomatic case is confirmed. In addition, once confirmed, all the contacts of the detected agent are, in principle, suspects of being infectious. In those situations all the individuals are quarantined, so that, also Susceptible and Exposed neighbors are also removed from the population together with those Presymptomatic, Asymptomatic and Symptomatic. Another limitation is the absence of directionality of the model since here traced contacts can correspond to either secondary (forward CT) or prior (backward CT) infections of those detected cases, provided the traced cases are infectious. These and other features should be considered as extensions of the model presented here including temporal activity patterns.

Chapter 4

Control of vector-borne diseases through the analysis of human mobility patterns

There is nothing to fear in life, you just have to understand it. Now is the time to understand more, so we can fear less.

Marie Curie.

So far, we have examined epidemic diffusion dynamics in relation to direct interactions between individuals within a specific population or social environment. In this context, the population was represented as a network, where each node corresponded to an individual, and the links represented the interactions between individuals. In this representation, we have captured some emergent properties of close contact networks which may influence the epidemic outcomes. If we aim to understand epidemics at different scales, such as cities, urban areas, or countries, a different approach is required since the interest lies in how human mobility affects the spread of infectious diseases across regions. A useful approach to addressing this problem is to create a metapopulation network with nodes representing the geographic areas (also known as patches or subpopulations) into which the system's population is divided. In its turn, the links between patches or subpopulations represent the fluxes of individuals moving among them. So-called metapopulations are used as a natural venue to represent the spatial diffusion of a contagion process due to the mixing events of individuals from different subpopulations, such as cities, urban areas, or different geographical regions.

In the following two chapters, we will study the role that recurrent human mobility patterns and the specific demographic distribution into subpopulations of the individuals have on the epidemic outcome, with a focus on understanding how control policies should

be applied in order to suppress or mitigate epidemic outbreaks. Specifically, we examine vector-borne diseases (VBD) with the objective of deriving risk indicators that quantify the vulnerability of geographic areas to outbreaks and, additionally, identifying the areas where control resources should be deployed.

This chapter is organized as follows. First, in Section 4.1, we introduce the motivation to study vector-borne diseases, and the corresponding compartmental model is explained in Section 4.2. Then, in Section 4.3, we propose a metapopulation framework that particularizes the Markovian approach introduced in Chapter 2 to integrate into a single formalism human mobility patterns, and human and vector demographics. Once introduced and validated, we make use of the Markovian metapopulation framework to calculate the epidemic threshold and to explore the effect of mobility patterns and asymptomaticity on disease spread. In Section 4.4, we apply our model to one of the most rapidly spreading mosquito-borne viral diseases in the world, dengue, by generating a data-driven epidemic vulnerability indicator. The theoretical epidemic risk indicator is applied to find the most dengue-endangered areas in the city of Santiago de Cali, Colombia, where dengue is an endemic disease. To round off, in Section 4.5 we use a bio-control method introducing a competition dynamics between immune mosquitoes and mosquitoes that can acquire and transmit dengue to generate a targeted control strategy to decrease the impact of dengue in Santiago de Cali.

4.1 Vector-borne diseases

Vector-borne diseases (VBD) are infections spread by cross-contagion between an arthropod species and humans. Vectors are usually bloodsucking insects such as mosquitoes, ticks, triatomine bugs, sandflies, and blackflies that pick up viruses, bacteria, or parasites by biting infected people. When they bite again, they can transmit the pathogen to the next person for the rest of their life during each subsequent bite/blood meal. The WHO estimates more than 700,000 annual deaths for all vector-borne diseases worldwide. Malaria, dengue, chikungunya, yellow fever, Zika, and other 22 illnesses are listed as VBD representing 17% of all infection diseases [91–93].

Even though the burden of these diseases is highest in tropical and subtropical areas affecting disproportionately the poorest populations [94, 95], within the past two decades, VBD have spread to new parts of the world. The increasing threat of VBD is primarily associated with population displacement, rapid unplanned urbanization [96], globalization of travel and trade [97], and climate change[98, 99]. All of this helps to vectors carrying diseases spread to take up residence in new areas widening the threat to global public health, both in terms of the number of people affected and their geographical spread.

The WHO estimates that more than half the world's population is in danger of con-

tracting a VBD [92]. In particular, dengue is emerging as a serious public health concern. Dengue has turned into the most widespread VBD in the world, during the past 50 years dengue cases have 30-fold [92] with staggering human and economic costs. More than 3.9 billion people living in 129 countries are at risk of contracting dengue fever [100, 101] and currently, 96 million cases are reported every year, a quantity that is estimated to represent around 25% of all real cases [102, 103].

Regarding prevention and control of VBD, current techniques lay emphasis on entomological surveillance and source reduction of mosquito breeding due to the lack of effective therapeutics or vaccines. Therefore, the suppression of mosquito (or vector) populations by removal of urban breeding habitats and insecticide/larvicide treatments have been the most popular control response. However, most species of vectors, including *Aedes aegypti*, the primary vector carrying dengue, have acquired insecticide resistance[104].

In 2009, it was discovered that an infection caused by *Wolbachia* bacteria in *Aedes aegypti* mosquitoes prevents them from transmitting viruses that cause dengue and other diseases such as Zika and Chikungunya [105]. This high-level control method has been proven to be the most self-sustain in the long run. However, introducing *Wolbachia* in all places where is needed is not an easy task and requires strategic planning to efficiently distribute the limited resources to protect the most people as possible. The aim of this chapter is to create an effective plan to distribute resources that help to control or mitigate VBD. First, by understanding the epidemic evolution dynamics, then calculating the most threatened areas, and finally, making the most of the limited resources by efficient distribution policies.

4.2 Ross-Macdonald model

The Ross-Macdonald (RM) model captures mosquito-borne disease transmission. It is based on the quantitative foundations of epidemiology written in the seminal work of Ronald Ross first proposed in 1911 [27, 28] and then revised by George Macdonald between 1950 and 1969 [106, 107]. The RM compartmental model assumes that both vectors and potential hosts (humans) can be either susceptible or infectious as shown in Fig 4.1. Susceptible vectors contract the disease after biting an infected person with a transmission rate $\widetilde{\lambda}^{HM}$. The vectors feed on blood with a frequency β. Likewise, susceptible people get infected with a transmission rate $\widetilde{\lambda}^{MH}$ by being bitten by an infected vector. Infected people become susceptible again after recovery with a rate $\widetilde{\mu}^{H}$, while vectors die and are replaced by susceptible newborns with rate $\widetilde{\mu}^{M}$.

To construct the mean-field approximation we consider a system with N_H potential hosts (humans) and N_M vectors. Thus, the temporal evolution of the density of infected vectors ($\rho^M = I^M/N_M$) and humans ($\rho^H = I^H/N_H$) can be written as ordinary differential

equations of gain and loss,

$$\dot{\rho}^H = -\rho^H \widetilde{\mu}^H + (1-\rho^H)\gamma\beta\widetilde{\lambda}^{MH}\rho^M, \tag{4.1}$$
$$\dot{\rho}^M = -\rho^M \widetilde{\mu}^M + (1-\rho^M)\beta\widetilde{\lambda}^{HM}\rho^H, \tag{4.2}$$

the first term in the right-hand side of both equations represents the loss of infectious cases due to human hosts' recovery and the renewal of vectors respectively. The second term in the right-hand side of the equations represents the gain of infectees due to the interaction between humans and vectors. Observe, that the mixing events that give rise to a human infection are limited by γ, defined the ratio between vectors abundance and human population (*i.e.* $\gamma = N_M/N_H$).

Now that we have obtained the dynamics of the system governed by the coupled former ODEs, we aim to calculate the average number of new infections caused by an infected individual over the susceptible population following the next generation matrix (NGM) method [108]. Let us first define a vector $\overrightarrow{\rho} = (\rho^H, \rho^M)'$, where the prime denotes transpose. The former ordinary differential equations can be separated in the sum of two arrays; the matrix corresponding to transmissions $\mathbf{T}$ (gain) and the matrix corresponding to transitions $\mathbf{\Sigma}$ (loss), then,

$$\dot{\overrightarrow{\rho}} = (\mathbf{T}+\mathbf{\Sigma})\overrightarrow{\rho}, \tag{4.3}$$
$$\tag{4.4}$$

where,

$$\mathbf{T} = \begin{pmatrix} 0 & \gamma\beta\widetilde{\lambda}^{MH} \\ \beta\widetilde{\lambda}^{HM} & 0 \end{pmatrix}, \tag{4.5}$$

$$\mathbf{\Sigma} = \begin{pmatrix} -\widetilde{\mu}^H & 0 \\ 0 & -\widetilde{\mu}^M \end{pmatrix}, \tag{4.6}$$

finally, the effective reproductive number can be obtain with the largest eigenvalue of the NGM that is defined $\mathbf{K} = -\mathbf{T}\mathbf{\Sigma}^{-1}$ [108] then,

$$\mathbf{K} = \begin{pmatrix} 0 & \gamma\beta\widetilde{\lambda}^{MH} \\ \beta\widetilde{\lambda}^{HM} & 0 \end{pmatrix} \begin{pmatrix} 1/\widetilde{\mu}^H & 0 \\ 0 & 1/\widetilde{\mu}^M \end{pmatrix} = \begin{pmatrix} 0 & \gamma\beta\widetilde{\lambda}^{MH}/\widetilde{\mu}^M \\ \beta\widetilde{\lambda}^{HM}/\widetilde{\mu}^H & 0 \end{pmatrix}, \tag{4.7}$$

now we can obtain the effective reproductive number with the maximum eigenvalue of the NGM,

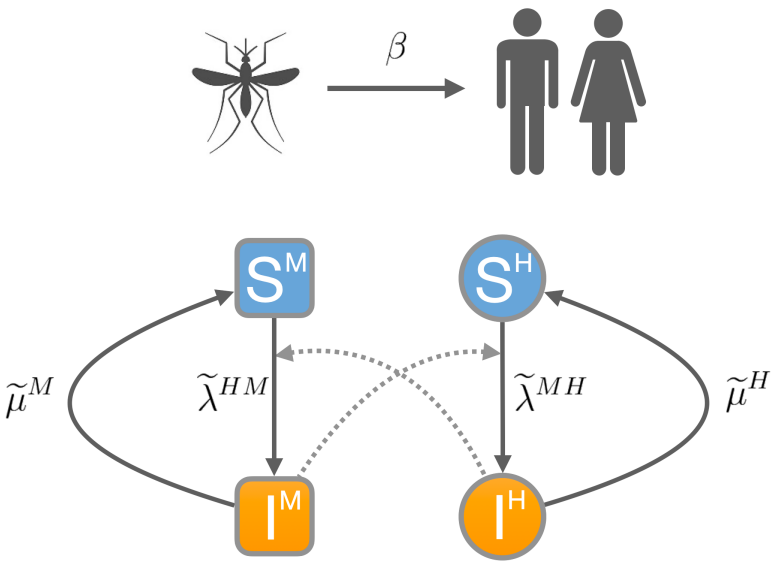


Figure 4.1: **Ross-Macdonald cross-contagion model**. Illustration of the RM model processes which identifies five key quantities: *(i)* the rate at which an infected vector transmits the disease to a susceptible individual after biting him/her to feed, $\widetilde{\lambda}^{MH}$, *(ii)* the rate at which an infected human transmits the disease to a susceptible vector after the vector has feed on him/her, $\widetilde{\lambda}^{HM}$, *(iii)* the feeding rate of vectors β, *(iv)* the probability that an infected human recovers, $\widetilde{\mu}^{H}$, and *(v)* the mortality rate of vectors, $\widetilde{\mu}^{M}$. The diagram shows the cross contagion between susceptible (**S**) and infected (**I**) species.

$$\mathcal{R}_0 = \Lambda_{max}\left(K\right) = \sqrt{\frac{\beta^2 \gamma \widetilde{\lambda}^{MH} \widetilde{\lambda}^{HM}}{\widetilde{\mu}^{H} \widetilde{\mu}^{M}}}, \tag{4.8}$$

let us recall that the transition between disease-free or an endemic equilibrium is given by the critical point given by $\mathcal{R}_0 = 1$.

4.3 Metapopulation model for VBD

In the previous section, we explained the RM model with a mean-field approach which assumes that both humans and vectors are homogeneously distributed. It, therefore, ignores any spatial correlations between their interactions. Such a description cannot explain the heterogeneous impact of VBD in urban areas, cities or countries. Likewise, this approach cannot be used to get spatial information to guide the optimal allocation of the limited resources available to control VBD. To solve those problems, we theoretically incorporate the heterogeneous spatial distribution of humans and vectors and the human mobility patterns.

We use a metapopulation framework based on the MIR model presented in Chapter 2 to theoretically analyze the epidemic impact of human mobility flows and demographics. Let us define a population of N_H humans spatially distributed in N_p geographical areas

(neighborhoods, cities, countries, etc.). We assumed that the number of residents in each patch i is denoted by n_i, therefore,

$$N_H = \sum_{i=1}^{N_p} n_i \ . \tag{4.9}$$

Those individuals residents in i may daily move and interact in other patches to do their activities and then come back to their residence. The human recurrent mobility patterns are the interactions between patches, that are represented in a weighted and directed network encoded in a row-stochastic matrix $\mathbf{R}$ (*the sum across columns is equal to 1 for every row and the entries are non-negative*), whose entries R_{ij} account for the probability to travel from patch i to patch j. We obtain this matrix $\mathbf{R}$ from real mobility data-sets capturing the observed number of trips between each pair of nodes i and j, W_{ij} in a daily basis. With this information at hand one can calculate the elements of $\mathbf{R}$ as,

$$R_{ij} = \frac{W_{ij}}{\sum_{l=1}^{N_p} W_{il}} \ . \tag{4.10}$$

While humans move as described above, the demographic distribution of the N_M vectors remain constant at each patch, being m_i the population of vectors at patch i. The assumption of the vectors do not move across the patches is valid for many VBD such as dengue, Zika, or Chikungunya since their carriers, *Aedes* mosquitoes, typically fly an average of 400 meters [109, 110].

Now that we have constructed the metapopulation network, we integrate the human mobility dynamics with the cross-contagion processes dictated by the RM model. Given the daily nature of human mobility flows, we adopted a time-discrete version of RM dynamics that allows to couple both dynamical processes in a simple way. First, we consider that at each time step, t, susceptible individuals ($\mathbf{S}$) decide to move from their residence i to the neighboring patch j encoded by R_{ij} with probability p or remaining in their residence with probability $(1-p)$. Moreover, as symptoms associated with some VBD can range from severe to mild and even asymptomatic, we also include the possibility that infected individuals move with probability αp, being $\alpha \in [0,1]$ a parameter that accounts for the fraction of the asymptomatic infected individuals. This way, human mobility induces the heterogeneous mixing between distant geographical areas.

Once the movement stage has been performed the interaction one comes into play. In this stage, we assume that all the vectors and humans placed in the same patch interact with the rules dictated by the RM model. Namely, susceptible humans become infected with probability λ^{MH} after being bitten by an infected vector whereas healthy vectors become infectious with probability λ^{MH} when interacting with an infected human. In addition, we assume that each vector makes a number of β contacts with (healthy or

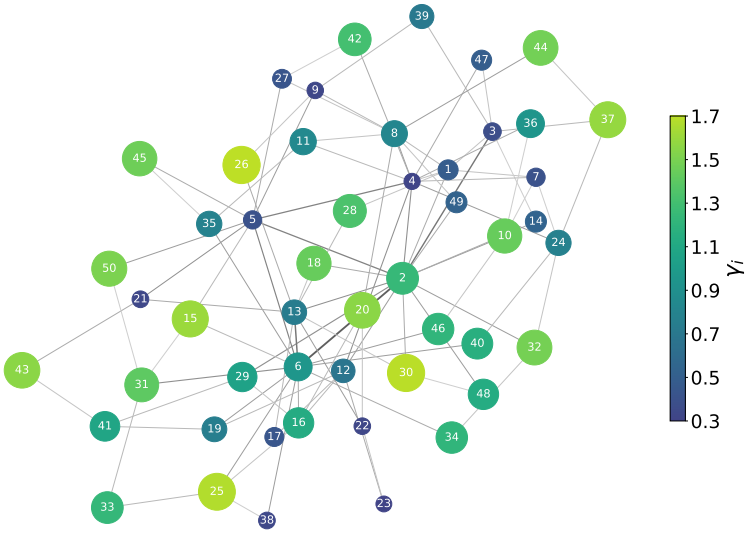


Figure 4.2: **Synthetic metapopulation**. Barabási-Albert network with $N = 50$ patches and average degree $\langle k \rangle = 4$. The human population is homogeneously distributed with 1000 people in each patch. The ratio between mosquitoes and humans given by $\gamma_i = m_i/n_i$, is randomly taken from a uniform distribution within the range $\gamma_i \in [0.3, 1.7]$. The colors and sizes of the nodes are proportional to its ratio.

infected) humans. Finally, also during this interaction stage, infected humans become susceptible with probability μ^H, while (healthy or infected) vectors die with probability μ^M, being replaced by newborn healthy ones. Let us note, that the parameters in this stage that replace the original rates in the original RM are due to the time discrete nature of the metapopulation framework.

To round of the microscopic processes that occur at each time step, we set a return stage in which each individual moves back to their residence. Then, the process starts again at time $t + 1$.

4.3.1 Model Equations

The previous description is generalization of the discretization of the Movement-Interaction-Return model for VBD and can be tackled by mechanistic simulations, as well as with a mathematical framework in the form of equations of the fraction of infected humans and vectors associated to each patch.

To this aim, for each patch i ($i = 1, ..., N_p$) we write the time discrete evolution of the fraction of hosts and vectors in the infectious state, denoted hereinafter by ρ^H and ρ^M respectively. Taking into account the microscopic rules defined above, both quantities read,

$$\rho_i^H(t+1) = \rho_i^H(t)(1-\mu^H) + (1-\rho_i^H(t))I_i^H(t), \quad (4.11)$$

$$\rho_i^M(t+1) = \rho_i^M(t)(1-\mu^M) + (1-\rho_i^M(t))I_i^M(t), \quad (4.12)$$

where μ^H and μ^M are the probabilities of human recovery and vectors' renewal respectively, and $I_i^H(t)$ and $I_i^M(t)$ are the probabilities that a susceptible human with residence in subpopulation i and a susceptible vector associated to i are infected at time t respectively. The former infection probability for humans reads,

$$I_i^H(t) = (1-p)P_i^H(t) + p\sum_{j=1}^{N_p} R_{ij}P_j^H(t)\ , \quad (4.13)$$

where $P_i^H(t)$ is the probability that an individual placed in population i at time t is infected. This probability can be written as,

$$P_i^H(t) = 1 - \left(1 - \lambda^{MH}\rho_i^M \frac{1}{n_i^{\text{eff}}(\rho^H(t),\alpha,p)}\right)^{\beta m_i}\ , \quad (4.14)$$

where λ^{MH} is the probability that a human contract the disease after being bitten by an infected mosquito and $betam_i$ represents the possible interactions, where β is the number of bites per unit of time. Those interactions depend on the ratio between mosquitoes and the number of humans placed in (but not necessarily residing in) population i, *i.e.* $n_i^{\text{eff}}(\rho^H(t),\alpha,p)$. This latter effective population can be expressed as,

$$\begin{aligned} n_i^{\text{eff}}(\rho_H(t),\alpha,p) &= \left[1 - p\left(1-(1-\alpha)\rho_i^H(t)\right)\right] n_i \\ &+ p\sum_{j=1}^{N_p} R_{ji}\left(1-(1-\alpha)\rho_j^H(t)\right) n_j\ . \end{aligned} \quad (4.15)$$

In the same fashion, the probability of mosquito infection summarized in the expression $I_i^M(t)$ in Eq. (4.12) reads,

$$I_i^M(t) = 1 - \left(1 - \lambda^{HM}\frac{i_i^{\text{eff}}(t)}{n_i^{\text{eff}}}\right)^{\beta}\ , \quad (4.16)$$

where λ^{HM} is the probability of a susceptible mosquito contract the pathogen after interact with an infected human by biting him/her β times at each time step. In this way, $i_i^{\text{eff}}(t)$ is the number of infected humans placed in population i at time t necessary to obtain the ratio of infected humans, and can be written as,

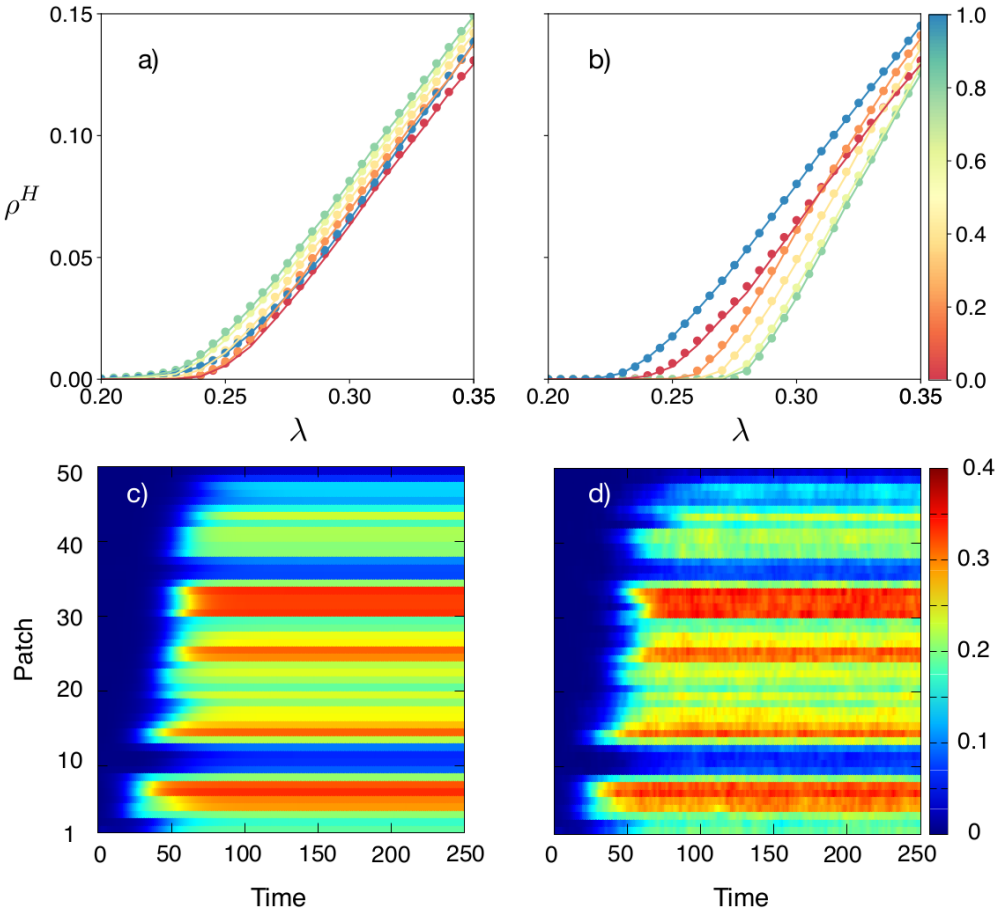


Figure 4.3: **Validation of the Markovian formalism of the VBD spreading**. Top: Epidemic size ρ^H as a function of the contagion probabilities between vectors and humans $\lambda^{HM} = \lambda^{HM} = \lambda$ and the human mobility p (color code). Dots correspond to results of averaging 50 stochastic realizations whereas solid lines represent the theoretical predictions obtained by iterating the equations of the formalism. Model parameters have been fixed to $(\beta, \mu^H, \mu^M) = (1, 0.3, 0.3)$. The values assumed for the restriction of the mobility of infected agents are (a) $\alpha = 0$ and (b) $\alpha = 1$. Bottom: Temporal evolution of the fraction of infected agents (color code) inside each patch obtained by iterating Eqs. (4.11-4.12) in Panel (c), and by tracking the individual state of each agent while performing a single Monte Carlo simulation in Panel (d). Infectivity is fixed to $\lambda = 0.3$ whereas mobility parameter p is set to $p = 0.3$. The rest of the parameters are the same as in Panel (a).

$$i_i^{\text{eff}}(t) = (1 - \alpha p) n_i \rho_i^H(t) + \alpha p \sum_{j=1}^{N_p} R_{ji} n_j \rho_j^H(t) \ . \tag{4.17}$$

The above equations describe the time evolution for the VBD incidence, $\vec{\rho}^H(t) = \{\rho_i^H(t)\}$ and $\vec{\rho}^M(t) = \{\rho_i^M(t)\}$, in a collection of connected patches with arbitrary demographic, $\vec{n}$, and vector, $\vec{m}$, distributions. Thus, by iterating Eqs. (4.11) and (4.12) starting from a given initial condition $\vec{\rho}^H(0)$ and $\vec{\rho}^M(0)$, we can monitor the spatio-temporal propagation of VBD and evaluate the steady epidemic prevalence on each geographical area.

4.3.2 Validation of the model equations with mechanistic simulations

Once the effective equations for the evolution of the epidemic incidence have been introduced, we now checked their validity by comparing their results with those obtained with mechanistic simulations. First, we study the steady epidemic prevalence across patches, $\vec{\rho}^H$, as a function of the contagion probabilities λ^{HM}, λ^{HM} and the degree of human mobility p in the population. To reduce the number of parameters and without loss of generality, let us define $\lambda^{HM} = \lambda^{HM} = \lambda$. We analyze the case in which human mobility is governed by a Barabási-Albert network (BA) of $N = 50$ patches and average degree $\langle k \rangle = 4$ (see Fig. 4.2). We consider the metapopulation matrix entries are $R_{ij} = 1/k_i$, where the population is homogeneously distributed with $n_i = 1000$ individuals. Concerning vectors' distribution, let us consider that the ratio between vectors and human populations inside a patch i, denoted in the following as $\gamma_i = m_i/n_i$, is randomly drawn from a uniform distribution within the range $\gamma_i \in [0.3, 1.7]$.

In Fig. 4.3, we corroborate the proposed Markovian formulation agrees with agent-based mechanistic Monte Carlo simulations in which the dynamics of each single human and vector is simulated according to the RM dynamics and the mobility flows are dictated by the matrix $\mathbf{R}$. The top figures, (a) and (b), show the steady-state of the metapopulations's density of infected hosts (ρ^H) as a function of the infectivity λ, this is,

$$\rho^H = \frac{\sum_{i=1}^{N_p} n_i \rho_i^H(\infty)}{\sum_{i=1}^{N_p} n_i} \, . \tag{4.18}$$

The color scale represents different values of p while lines accounts for the solution of the Markvovian equations, points are the result of averaging 50 mechanistic Monte Carlo simulations. While in Fig. 4.3 (a) we consider $\alpha = 0$. In Fig. 4.3 (b) we fully re-scale the mobility of asymptomatic people with $\alpha = 1$. In both panels the agreement between the effective Markovian and Monte Carlo simulations is clear.

To verify that the formalism is able to capture the spatio-temporal dynamics, in Figs. 4.3 (c-d) we plot the temporal evolution of the incidence of each patch, starting from the same initial condition according (c) the Markovian formulation is represented and in (d) the Monte Carlo simulations. Note that despite the noise induced by the stochastic nature of the mechanistic simulations shown in (d), the distribution of infected hosts agrees with the results from the theoretical framework in (c). Let us remark that the iteration of the Markovian equations save computational cost by iterating $2 \times N$ equations instead of performing lengthy agent-based simulations in which each microscopic process related with contagion dynamics and the mobility of the host are simulated.

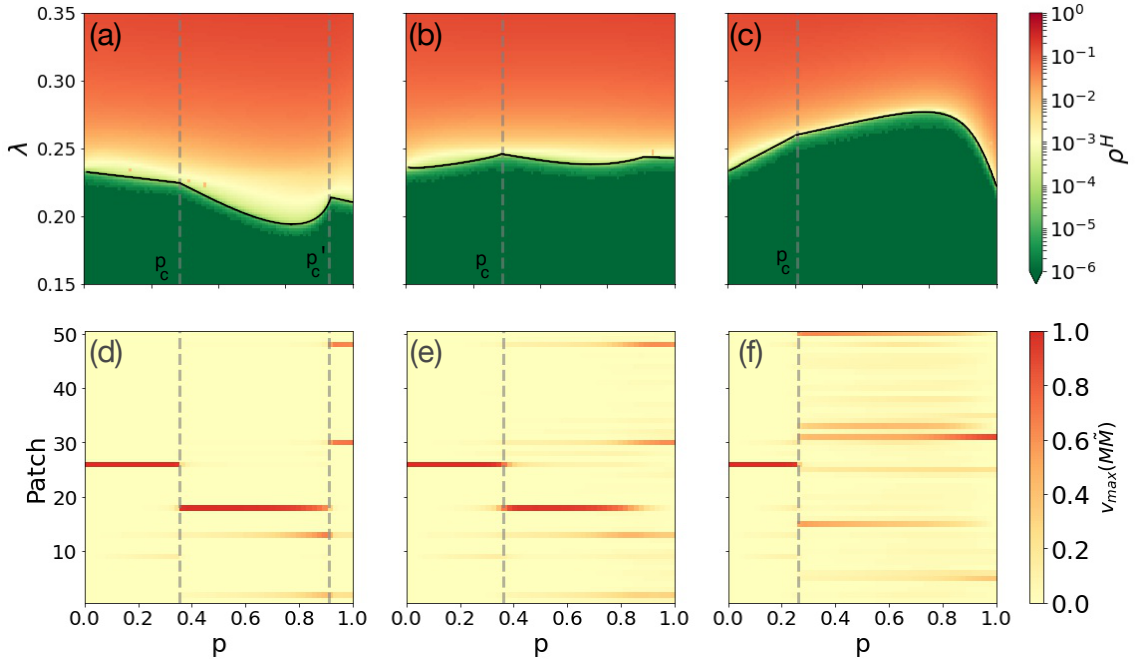


Figure 4.4: **Epidemic threshold and evolution of the leading eigenvector of matrix $\widetilde{\mathbf{MM}}$.** Panels (a)-(c) in the top show three epidemic diagrams $\rho^H(p, \lambda)$ in a synthetic metapopulation of $N = 50$ patches. Each panel corresponds to a different re-scaling value (α) for the mobility of infected humans, namely: (a) $\alpha = 0$, (b) $\alpha = 0.5$ and (c) $\alpha = 1$. In addition, we have set $\lambda^{HM} = \lambda^{MH} = \lambda$ while the rest of the RM parameters are: $\mu^H = 0.3$, $\mu^M = 0.3$, and $\beta = 1.0$. The color code show the incidence ρ^H as obtained from agent-based simulations while the solid curve represent the prediction for the epidemic threshold, λ_c, calculated from Eq. (4.30). The bottom panels (d)-(f) show the evolution, as a function of p, of the N components of the eigenvector of matrix $\widetilde{\mathbf{MM}}$ corresponding to maximum eigenvalue $\Lambda_{max}(\widetilde{\mathbf{MM}})$.

4.3.3 Epidemic threshold

The validity of the effective Markovian equations allow us to analytically study the epidemic threshold that separates the absorbing phase, in which the epidemic activity vanishes, and the active phase, in which an endemic epidemic state prevails (*i.e.* a fraction of the population is always infected). These two phases can be observed in Figs. 4.3 (a-b) as a function of the spreading infection rate λ.

Let us assume that the disease has reached the stationary state so $\rho_i^H(t+1) = \rho_i^H(t) = \rho_i^H$, and $\rho_i^M(t+1) = \rho_i^M(t) = \rho_i^M \ \forall i$. In this case, Eqs. (4.11)-(4.12) turn into,

$$\mu^H \rho_i^H = (1 - \rho_i^H) I_i^H, \tag{4.19}$$

$$\mu^M \rho_i^M = (1 - \rho_i^M) I_i^M, \tag{4.20}$$

As we are interested in computing the epidemic threshold, we now assume that the impact of the disease is finite but negligible for all geographical areas, which implies that

$\rho_i^x = \epsilon_i^x \ll 1 \ \forall i$, with $x = \{H, M\}$ standing for humans and vectors respectively. So, we can linearize the equations Eqs. (4.14)-(4.16),

$$P_i^H(t) \simeq \lambda^{MH}\beta\frac{m_i}{n_i^0}\epsilon_i^M \ , \tag{4.21}$$

$$I_i^M(t) \simeq \frac{\lambda^{HM}\beta}{n_i^0}\sum_{j=1}^{N_p}[\delta_{ij}(1-\alpha p)n_i + \alpha p R_{ji} n_j]\epsilon_j^H(t) \ , \tag{4.22}$$

similarly we can rewrite the Eq. (4.13) as,

$$I_i^H(t) \simeq \lambda^{MH}\beta\sum_{j=1}^{N_p}\left(pR_{ij}\frac{m_j}{\tilde{n}_j^0} + (1-p)\delta_{ij}\frac{m_i}{\tilde{n}_i^0}\right)\epsilon_j^M \ , \tag{4.23}$$

where $\tilde{n}_i^0 = n_i^{\text{eff}}(\rho^H(t) = 0, \alpha, p)$ and δ_{ij} is the Konecker delta. Substituting in Eqs. (4.19-4.20) while just keeping the linear terms we obtain,

$$\epsilon_i^H = \sum_{j=1}^{N}\frac{\lambda^{MH}\beta}{\mu_H}\underbrace{\left(pR_{ij}\frac{m_j}{\tilde{n}_j^{\text{eff}}} + (1-p)\delta_{ij}\frac{m_i}{\tilde{n}_i^{\text{eff}}}\right)}_{M_{ij}}\epsilon_j^M, \tag{4.24}$$

$$\epsilon_i^M = \sum_{j=1}^{N}\frac{\lambda^{HM}\beta}{\mu_M}\underbrace{\left(\alpha p R_{ji}\frac{n_j}{\tilde{n}_i^{\text{eff}}} + (1-\alpha p)\delta_{ij}\frac{n_i}{\tilde{n}_i^{\text{eff}}}\right)}_{\widetilde{M}_{ij}}\epsilon_j^H. \tag{4.25}$$

Note that in the second term on the right hand of the last two equations we can pack the metapopulation information in a new matrix, whose entries M_{ij} and $\widetilde{M}_{ij}$ respectively, depends on both the mobility properties (p, $\mathbf{R}$) and the demographic distribution of both humans and vectors ($\vec{n}$, $\vec{m}$). Observe that both matrices $\mathbf{M}$ and $\widetilde{\mathbf{M}}$ do not depend on the dynamics. For the sake of clarity, let us write the former system of equations in a more compact way,

$$\begin{pmatrix}\vec{\epsilon}^H \\ \\ \vec{\epsilon}^M\end{pmatrix} = \begin{pmatrix}0 & \frac{\beta\lambda^{MH}}{\mu^H}M \\ \\ \frac{\beta\lambda^{HM}}{\mu^M}\widetilde{M} & 0\end{pmatrix}\begin{pmatrix}\vec{\epsilon}^H \\ \\ \vec{\epsilon}^M\end{pmatrix} \ . \tag{4.26}$$

Equation (4.26) makes evident the bipartite nature of the processes involved in the spread of VBD with matrices M and $\widetilde{M}$ capturing vector-to-human and human-to-vector infections, respectively. Thus, in order to quantify indirect infections between humans mediated by vectors and vice versa we should iterate Eq.(4.26) obtaining,

$$\begin{pmatrix}\vec{\epsilon}^H \\ \\ \vec{\epsilon}^M\end{pmatrix} = \frac{\beta^2\lambda^{MH}\lambda^{HM}}{\mu^M\mu^H}\begin{pmatrix}M\widetilde{M} & 0 \\ \\ 0 & \widetilde{M}M\end{pmatrix}\begin{pmatrix}\vec{\epsilon}^H \\ \\ \vec{\epsilon}^M\end{pmatrix} \ . \tag{4.27}$$

From Eq. (4.27) it becomes clear that nontrivial solutions for $\vec{\epsilon}^H$ correspond to the eigenvectors of matrix $\mathbf{M}\widetilde{\mathbf{M}}$. Specifically, given a metapopulation defined by $\vec{n}$, $\vec{m}$, $\mathbf{R}$ and

p, the stationary solutions with infinitesimal incidence correspond to eigenvectors of $\mathbf{M}\widetilde{\mathbf{M}}$ whose eigenvalues can be written as,

$$\Lambda_i = \frac{\mu^M \mu^H}{\beta^2 \lambda^{MH} \lambda^{HM}} \ . \tag{4.28}$$

Under these conditions, the maximum eigenvalue $\Lambda_{max}(\mathbf{M}\widetilde{\mathbf{M}})$ encodes the combination of the RM parameters that corresponds to the epidemic threshold, namely:

$$\frac{\beta^2 \lambda^{MH} \lambda^{HM}}{\mu^M \mu^H} \Lambda_{max} = 1 \ . \tag{4.29}$$

The former equation reveals the minimum infectivities, either λ^{HM} or λ^{MH}, that trigger the epidemic outbreak. To derive a simple critical infectivity one can set $\lambda^{MH} = \delta \lambda^{HM}$, so that,

$$\lambda_c^{MH} = \sqrt{\frac{\mu_H \mu_M}{\delta \beta^2 \Lambda_{max}(\mathbf{M}\widetilde{\mathbf{M}})}} \tag{4.30}$$

To test the validity of Eq. (4.30) we carried out numerical simulations in the synthetic metapopulation used in the previous section (a BA network with 50 nodes). In Figs. 4.4(a-c) we plot the epidemic diagrams computing the fraction of infected hosts, ρ^H (color code), as a function of λ and p. For each value of p; the active population moving to other subpopulations; there exists a critical value λ_c, such that for all $\lambda > \lambda_c$ an epidemic phase prevails (*i.e.* endemic prevalence). The solid black curves given by the Eq. (4.30) show that our theoretical approximation to the epidemic threshold accurately captures the dependence of λ_c with p.

In these figures, the solid black curves are given by the solution of the Eq. (4.30). It is clear that our theoretical approach epidemic threshold accurately captures the dependence of λ_c with p.

4.3.4 Epidemic detriment & abrupt transitions of leading pacthes

So far we have introduced a theoretical framework that adds human mobility to the diffusion processes of VBD. We have proven that our theoretical analysis represent the spatio-temporal dynamics and the epidemic threshold dependence with p. In this section, we discuss the implications of adding human mobility patterns and re-scaling it according to epidemiological characteristics, such as asymptomaticity α . In Figs. 4.4(a-c) we show the epidemic diagrams $\rho^H(\lambda, p)$ of the synthetic metapopulation of 50 nodes. From these figures, the interdependence between mobility and the epidemic threshold for three different values of α is clear. In the absence of asymptomatic hosts, *i.e.* $\alpha = 0$, Fig. 4.4(a) shows that the more people move (as p increases) the better the disease diffusion is (as λ_c decreases). This effect appears due to the inability of infected individuals to move from high endangered zones or hotspots reported in [111], thus enhancing the probability of infection

in a higher volume of the population. However, the decreasing trend λ_c is broken for large mobility values, giving rise to the so-called epidemic detriment driven by mobility.

The epidemic detriment effect is amplified as the fraction of asymptomatic individuals increases, *i.e.* $\alpha \to 1$, as shown in Fig. 4.4(b-c). The detrimental effect on the epidemic threshold was found for SIR and SIS diseases in networked metapopulations [51, 112] as a result of the redistribution of the effective populations across patches due to mobility. In the case of VBD, this process corresponds to a homogenization of the effective ratios between vectors and humans, so that a high-risk patch with large $\gamma_i = m_i/n_i$ tends to decrease its effective value due to the increase of n_i^{eff} caused by human mobility. Furthermore, the mobility of asymptomatic infected individuals from the high risk areas to the low ones increases the former relaxation of the contagion process.

Apart from the different evolution of the curves $\lambda_c(p)$ as the fraction of asymptomatic infectious hosts changes, the plots in Figs. 4.4(a-c) show another interesting result. In particular, note that the derivative $\dfrac{d\lambda_c}{dp}$ is not continuous. The sudden changes of slope of the epidemic threshold curves pinpoint critical values of the mobility p in which the microscopic distribution of the epidemic incidence across patches changes dramatically.

In particular, these abrupt changes are the product of crossovers between the two maximum eigenvalues of matrix $\mathbf{M}\widetilde{\mathbf{M}}$ as p varies. The two maximum eigenvalues interchange their order at some critical mobility value p_c [113]. These crossovers do not have a strong impact on the epidemic threshold since the function $\lambda_c(p)$ is continuous. However, they are the fingerprint of a sudden change in the form of the eigenvector corresponding to the maximum eigenvalue, $\vec{V}_{max}$, of the matrix $\mathbf{M}\widetilde{\mathbf{M}}$.

The evolution of the components of $\vec{V}_{max}$ as a function of p for a $\lambda > \lambda_c$ are shown in the bottom panels, Figs. 4.4(d-f). The critical values, p_c, are signalized by grey dotted lines in the top and bottom panels of the Figs. 4.4. As anticipated above Figs. 4.4(d-f) shows that abrupt changes on the slope $\lambda_c(p)$ that correspond to abrupt changes of $\vec{V}_{max}$.

The former abrupt transitions encode a change in the distribution of those patches that are more relevant in an epidemic outbreak. In particular, in the three cases reported in Figs. 4.4 patch number 26 is the one causing the epidemic onset for $p = 0$ and $p \ll 1$. This is obvious since patch 26 is the one with largest ratio $\gamma_i = m_i/n_i$ in our synthetic metapopulation (see Fig. 4.2). However, as p increases, the leading patch changes, being replaced by patch 18 in the case of $\alpha = 0$ (d) and $\alpha = 0.5$ (e) while for $\alpha = 1$ (f) the leading patch is replaced by a collection of them. Remarkably, the case $\alpha = 0$ shows a second abrupt transition at $p_c' \simeq 0.92$. These abrupt changes are extremely important for the design of targeted control strategies, since the geographical areas that are more important to protect/control change at critical values p_c.

4.4 Epidemic risk: a data-driven vulnerability indicator

As the spread of the pathogen can be mitigated by reducing Λ_{max}, tackling those patches with the highest contribution in values of $\vec{V}_{max}$ might help to create effective and efficient control policies. In other words, focusing the immunization resources in those patches responsible for the epidemic onset should thoroughly help to suppress the outbreak at whole population level as we will show in Section 4.6. However, the contribution of the patches to the epidemic onset can be different from their vulnerability.

A different issue of the contribution of each patch to the onset of an epidemic outbreak concerns the vulnerability of each area once the epidemic outbreak has unfolded. In this regard, we can generate a prevalence indicator by recalling the meaning of matrix $\mathbf{M}\widetilde{\mathbf{M}}$, where the element $\left(\mathbf{M}\widetilde{\mathbf{M}}\right)_{ij}$ is the virtual interactions that a person from patch i receives from individuals in patch j. These virtual interactions act as contagion pathways as they are mediated by the vector populations in all the patches. Thus, we refer to $\mathbf{M}\widetilde{\mathbf{M}}$ as the *mixing matrix* of our metapopulation. Each term of the mixing matrix is the result of the multiplication of matrices $\mathbf{M}$ and $\widetilde{\mathbf{M}}$ and explicitly yields,

$$\begin{aligned}\left(\mathbf{M}\widetilde{\mathbf{M}}\right)_{ij} &= \delta_{ij}(1-\alpha p)(1-p)\frac{m_i n_i}{(\tilde{n}_i^{\text{eff}})^2} \\ &+ \alpha p(1-p)R_{ji}\frac{m_i n_j}{(\tilde{n}_i^{\text{eff}})^2} \\ &+ (1-\alpha p)pR_{ij}\frac{m_j n_j}{(\tilde{n}_j^{\text{eff}})^2} \\ &+ \alpha p^2\sum_k R_{ik}R_{jk}\frac{m_k n_j}{(\tilde{n}_k^{\text{eff}})^2}\,. \end{aligned} \tag{4.31}$$

The explicit expression of $\left(\mathbf{M}\widetilde{\mathbf{M}}\right)_{ij}$ shows that there are 4 possible mechanisms of contagion that an individual of patch i can receive from the individuals from patch j, each one encoded by the terms of Eq. (4.31). Namely,

(i) when an infected host from patch i remains at their residence, transmits the disease to a vector in i which then transmits the disease to another resident in i;

(ii) when an infected host from patch j visits patch i and infects a vector that will later pass the disease to a resident of patch i that remain in patch i;

(iii) when a resident of patch j infects a vector there and a susceptible human traveling to from i to j gets infected by the former vector in j;

(iv) when both the infected individual from j and the susceptible individual from i travel to a contiguous third patch k where the infection takes place mediated by a vector from k.

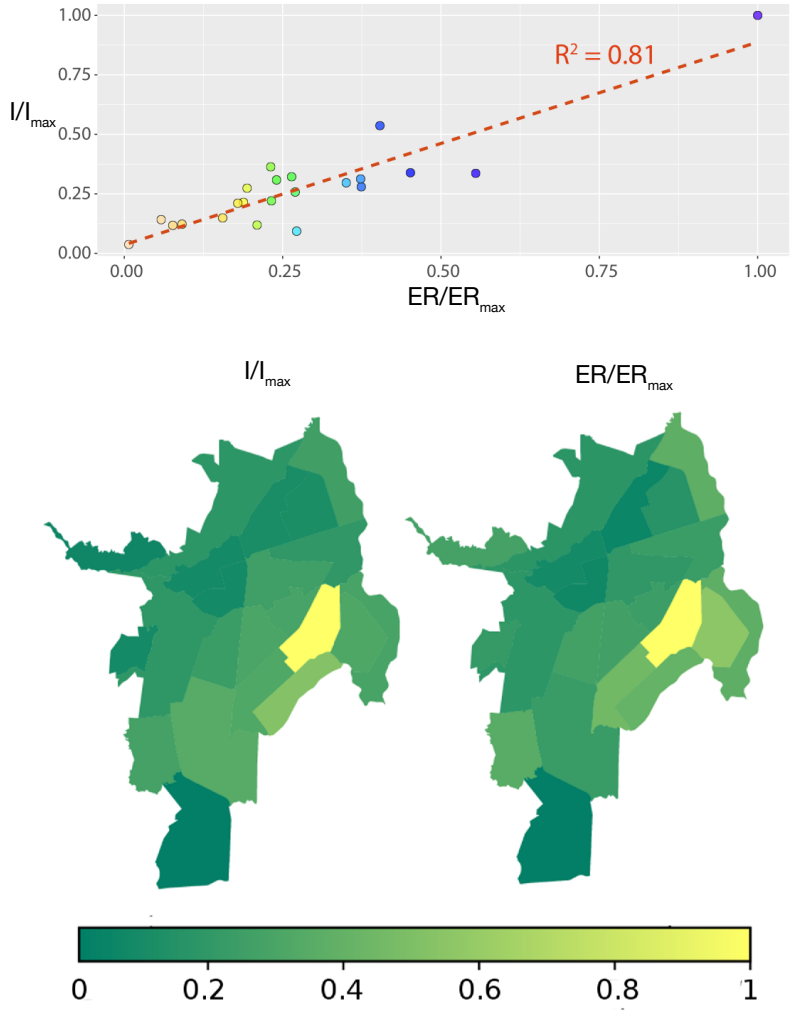


Figure 4.5: **Real dengue incidence vs estimated epidemic risk in the city of Cali, Colombia**. Top: Normalized epidemic risk (ER/ER_{max}) *versus* normalized dengue incidence (I/I_{max}) for each of the 22 districts of Cali. Color encodes the Epidemic Risk, from the lowest (yellow) to the highest (blue).The correlation between the two variables yields a coefficient of determination of $R^2 = 0.81$. Bottom: Spatial distributions of the normalized dengue incidence in the city of Cali (left) and the normalized epidemic risk (right) according to Eq. (4.32). The parameters concerning agents' mobility have been set to $(p, \alpha) = (0.36, 0.75)$.

Thus, to make predictions about the impact of the disease on a geographical area, i, we must account for all the possible infections from each patch of the metapopulation and weight the resulting number by the population of i. Then, the epidemic risk indicator for each patch i, in the following denoted as ER_i, can be defined as,

$$ER_i = n_i \sum_{j=1}^{Np} \left(\mathbf{M}\widetilde{\mathbf{M}}\right)_{ij} \quad . \tag{4.32}$$

Importantly, to compute the epidemic risk ER_i, there is no need for either numerical simulations or the integration of the Markovian equations. In fact, as the Eq. (4.32)

suggests, the only information needed is the demography, the vector distribution, and mobility patterns.

We validate the Epidemic Risk indicator by adding the available data about the dengue incidence of the city of Santiago de Cali (Colombia). Its geographical position and climate make Cali a dengue hiper-endemic city whose historical records of incidence are easy to find. In 2016, Cali reported about 20,000 cases of dengue, which accounted for about 20% of the total cases reported throughout Colombia.

To compare the observed dengue incidence across the Cali's 22 communes during the years 2015-2016 [114, 115], we collected the demographic information of the 2 million residents and mobility data sets. Mobility flows connecting communes are extracted from urban commuting surveys [116]. As a result, more than 10^5 trajectories were recorded, which suppose a representative sample of Cali's commuting flows. Once all the data have been gathered, we constructed an origin-destination matrix, encoded in our formalism by the matrix $\mathbf{R}$, a weighted directed network encoding the probability that an agent visits other neighborhoods different from its residence (see 4.10).

Apart from the mobility network, the distribution of vectors across the city also plays a crucial role in the outcome of the disease. The number of vectors inside a geographical region is strongly linked to environmental features such as altitude, temperature and humidity but also to human-dependent factors like health and economic conditions. To model vectors' distribution across communes, we use as a proxy the so-called *recipient index*. This quantity encodes the probability of finding vector pupae in different recipients which have been previously distributed across the city. A high value of the index means a higher probability of finding vectors. For this reason, we assume that the ratio between the number of vectors and humans inside each patch in our model is directly proportional to its recipient index, which is extracted from entomological data of the year 2015 [117].

Furthermore, it is known that the fraction of asymptomatic dengue cases is around $\alpha = 0.75$ [102, 103, 118–120] and we leave p as a free parameter to be calibrated. The calibration maximizes the correlation between both the empirical and the theoretical risk indicators, yielding p = 0.36.

In Fig. 4.5 we compare the real *vs* theoretical spatial dissemination of dengue in Cali. We normalized the values of both epidemic risk and dengue incidence by their maximum observed value (in both cases that of district 13). The found agreement ($R^2 = 0.81$) leads us to conclude that using the data-driven epidemic risk indicator, using high-resolution mobility patterns and demographic data, is enough to make an accurate prediction of the vulnerability under the spread of VBD of each patch in a metapopulation.

4.5 Immunization ranking from spectral analysis of the mixing matrix

In the introduction of this chapter, we have mentioned that the most long-lasting and efficient way to prevent and control dengue and other VBD is through the introduction of the *Wolbachia*-infected mosquitoes into endangered areas. It has been shown that the transmission of the dengue virus into endemic areas is reduced when mosquitoes carry *Wolbachia* bacteria. Unfortunately, *Aedes aegypti*, the primary vector carrying the dengue virus, do not acquire this bacterium in their natural environment and, consequently, Wolbachia has to be introduced in vitro into the mosquito eggs in a laboratory. Once a massive production of 'safe' mosquitoes, *i.e.* mosquitoes infected with *Wolbachia*, is performed they are deliberately released into their natural habitat taking over the local mosquito population.

The success of this bio-control method relies on the bias of the natural competition that favors the proliferation of *Wolbachia*-carrying mosquitoes over wild-type vectors [121]. This evolutionary advantage of *Wolbachia*-carrying vectors underlies the cytoplasmic incompatibility [122, 123]. The fertilized eggs resulting from the mating of an uninfected female mosquito and an infected male mosquito do not hatch, whereas the offspring of the mating of an infected female mosquito and any male mosquito (regardless of its wilde type or wolbachia infection status) will carry *Wolbachia*.

Despite that *Wolbachia* is safe and self-sustaining as control method, introducing it in all places where is needed is not an easy task and requires strategic planning to efficiently distribute the limited resources to protect the most people as possible. Thus, our aim in this final section of the chapter is to consider the information at hand about human mobility and the demographic distribution of vectors and humans to produce an immunization raking that can be applied to inform health authorities when deciding the best targets to implement the release of Wolbachia-carrying mosquitoes. In the previous section, we have found a data-driven indicator of the epidemic risk of a geographical area. One could think that this indicator is a good proxy to chose which areas are more convenient to immunize first, or where to focus the limited resources. However, in this section we will show that the epidemic risk indicator differs from an immunization ranking.

4.5.1 Competition dynamics among vector populations

We start by analyzing the competition dynamics between wild and *Wolbachia*-infected mosquitoes in the same habitat, which is equivalent to a single patch or subpopulation in the metapopulation formalism. To this aim, we used the iterative logistic growth of the invasion dynamics of *Wolbachia*-infected mosquitoes presented in [124]. The ecological growth model includes the shortening of the life cycle of *Wolbachia*-infected mosquitoes. The probability

of birth of the *Wolbachia*-infected mosquitoes (r_w) is lower than the probability of birth of the wild species (r_m). The *Wolbachia* infection also increase the death probability (α) compared to the wild species, *i.e.* $\alpha_w > \alpha_m$. The time-discrete dynamic evolution of the wild-type mosquitoes $m_i(t)$ and *Wolbachia*-infected mosquitoes $w_i(t)$ populations read,

$$m_i(t+1) = m_i(t)\left[1 + r_m \frac{m_i(t)}{m_i(t)+w_i(t)} - \alpha_m - \beta_m \left(m_i(t)+w_i(t)\right)\right], \quad (4.33)$$

$$w_i(t+1) = w_i(t)\Big[1 + r_w - \alpha_w - \beta_w \left(m_i(t)+w_i(t)\right)\Big]. \quad (4.34)$$

The second terms on the right hand side of these equations represent the mosquitoes' renewal and they contain the key ingredient for the long-term prevalence of Wolbachia-infected mosquitoes over the wild-type ones. In particular, while female *Wolbachia*-carrying mosquitoes successfully reproduce by any mating interaction with rate r_w, the reproduction rate of wild mosquitoes is affected by the interaction of *Wolbachia*-infected male mosquitoes since their offspring eggs will not hatch, so the term $\frac{m_i(t)}{m_i(t)+w_i(t)}$ accounts for the probability of interacting with wild-type males. The third and fourth terms on the right hand side of the former equations represent the removal due to mortality and the competition between the two species respectively. We assume that the competition parameters, β_m for the wild-type, and β_w *Wolbachia*-infected mosquitoes, are given by the ability of each species to survive in the absence of the other. Under this assumption, we set the competition parameters such that each population in the absence of the other species yields to $m_i^* = w_i^* = \gamma_i n_i$ in the steady state.

When *Wolbachia*-infected mosquitoes are introduced in a region the species coexist till the competition reaches an attracting fixed point. In Fig. 4.6(a) we plot different trajectories in the phase space of the system that illustrate the ecological invasion of *Wolbachia*-infected mosquitoes over the wild-type population. Note that, regardless of the initial distribution of *Wolbachia*-infected mosquitoes, the wild-type population will eventually vanish, so that ($m^* = 0,\ m_w^* = \gamma_i n_i$) is the only stable fixed point. In Fig. 4.6(b) we show the temporal evolution of an initial release of *Wolbachia*-carrying mosquitoes representing just 10% of the wild-type population. Note that before *Wolbachia*-infected mosquitoes take over the habitat the population of wild-type vectors is already extinct.

4.5.2 Geographical contagion dynamics & immunization

In Section 4.3, we have considered a metapopulation model for VBD such that the population of mosquitoes remain constant in time, since (susceptible and infected) mosquitoes were considered to be replaced by new (susceptible) with probability μ^M (see Eq. (4.12)).

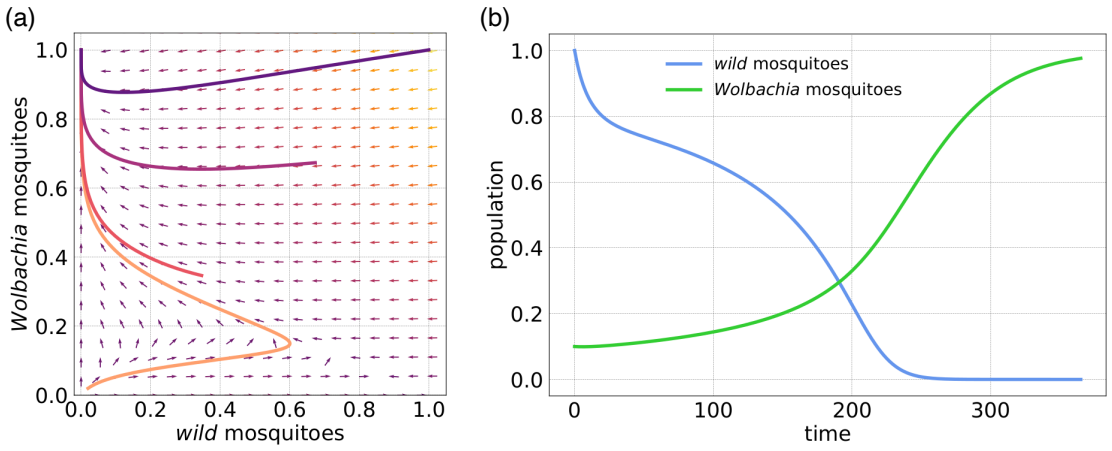


Figure 4.6: **Competition dynamics among vector populations**. (a) Phase plane of the competition dynamics between wild-type and *Wolbachia*-infected mosquitoes. Note that populations appear normalized to their maximum possible value. All the trajectories start with a similar population of the two types of vectors. The parameters used in Eqs. (4.33)-(4.34) are chosen according to current estimates: $r_m = 0.14$ [125, 126], $r_w = r_m/2$ [127], $\alpha_m = 0.2r_m$, $\alpha_w = 1.5\alpha_m$ [128], $\beta_m = (r_m - \alpha_m)/n_i\gamma_i$, $\beta_w = (r_w - \alpha_w)/n_i\gamma_i$, and $n_i\gamma_i = 5 \cdot 10^4$. (b) Time evolution of the populations of wild-type and *Wolbachia*-infected mosquitoes when introducing a small initial population of the latter.

However, in order to incorporate the competition dynamics between the two populations of vectors and the immunizing effects of *Wolbachia*-infected mosquitoes, we need to reformulate our metapopulation formalism.

First, to ease the formulation of the metapopulation equations under competition dynamics let us change to absolute quantities. This way, the composition of the vector population in a patch i is defined as a set of three variables $\{m_i^S(t),\ m_i^I(t),\ w_i(t)\}$, that correspond to the total number of susceptible wild-type mosquitoes, the number of wild-type infectious mosquitoes, and the number of *Wolbachia*-infected mosquitoes respectively. Note that the sum of the first two variables correspond to the total number of wild-type mosquitoes in patch i, $m_i^S(t) + m_i^I(t) = m_i(t)$, whose evolution together with that of $w_i(t)$ is given by Eqs. (4.33)-(4.34). Then, the time evolution of the population of (wild) vectors able to transmit the VBD in each patch depends on the competition dynamics with Wolbachia-carrying mosquitoes, so that the cross-contagion events between humans and vectors become less frequent as the reduction of the wild-type mosquitoes takes place in a patch with the presence of Wolbachia-carrying competitors.

By writing the evolution equations for each of the three variables, $m_i^S(t)$, $m_i^I(t)$ and $w_i(t)$, describing the vector composition of a patch i we note that the population of mosquitoes that have acquired *Wolbachia*, $w_i(t)$, remain the same as in Eq. (4.34). However, the evolution of both $m_i^S(t)$ and $m_i^I(t)$ are obviously affected by the contagion dynamics dictated by the RM model. We can write the evolution equations for the number

of wild-type mosquitoes in each epidemiological compartment as,

$$\begin{aligned} m_i^S(t+1) &= (1-\alpha_m)\, m_i^S(t) + r_m \frac{[m_i(t)]^2}{m_i(t)+w_i(t)} \\ &- \beta_m m_i^S(t)\,(m_i(t)+w_i(t)) - \Pi_i^M(t) m_i^S(t), \qquad (4.35) \\ m_i^I(t+1) &= (1-\alpha_m) m_i^I(t) - \beta_m m_i^I(t)\,(m_i(t)+w_i(t)) \\ &+ \Pi_i^M(t) m_i^S(t), \qquad (4.36) \end{aligned}$$

where $\Pi_i^M(t)$ accounts for the probability that a susceptible wild-type vector in patch i is infected at time t,

$$\Pi_i^M(t) = 1 - \left(1 - \lambda^{HM} \frac{i_i^{\text{eff}}(t)}{n_i^{\text{eff}}(t)}\right)^{\beta} , \qquad (4.37)$$

where n_i^{eff} represents the the effective population placed in patch i (after the movement stage has taken place) at time t, while $i_i^{\text{eff}}(t)$ represents the number of infected humans placed in the same patch (below we write the explicit expressions for these quantities).

It is worth emphasizing that we assume that contracting dengue does not alter the mosquito's life dynamics. Likewise, as it is typically assumed, the form of Eqs. (4.35)-(4.36) implies that newborn wild-type mosquitoes are not carriers of dengue virus even in the case their parents were infected by the virus. However, there is no total consensus on the absence of vertical transmission of the virus, with different empirical evidence for and against this hypothesis. We refer the reader to [129] for a comprehensive review on the topic.

Now, to capture the contagion dynamics in the human population, we assign to each patch i of the metapopulation one variable, $n_i^I(t)$, that accounts for the absolute number of infectious residents of patch i at time t. This single variable completely characterizes the epidemiological state of residents in patch i since the number of susceptible residents at time t is $n_i - n_i^I(t)$. The time-discrete evolution of the number of infectious residents of patch i, $n_i^I(t)$, obeys the following set of equations,

$$n_i^I(t+1) = n_i^I(t)(1-\mu^H) + (n_i - n_i^I(t))\Pi_i^H(t)\ , \qquad (4.38)$$

where the first term accounts for being infected at the residence patch i while the second term captures the probability that the infection occurs in any of the possible commuting destinations reached from patch i. Thus, in the former expression $\Pi_i^H(t)$ accounts for the probability that a susceptible human with residence in patch i is infected at time t, and can be written as,

$$\Pi_i^H(t) = (1-p)P_i^H(t) + p\sum_{j=1}^{N_p} R_{ij} P_j^H(t)\ , \qquad (4.39)$$

where $P_i^H(t)$ is the probability that an agent placed in population i at time t is infected. This probability reads,

$$P_i^H(t) = 1 - \left(1 - \lambda^{MH} \frac{1}{n_i^{\text{eff}}(t)} \frac{m_i^I(t)}{m_i(t)}\right)^{\beta m_i(t)} , \qquad (4.40)$$

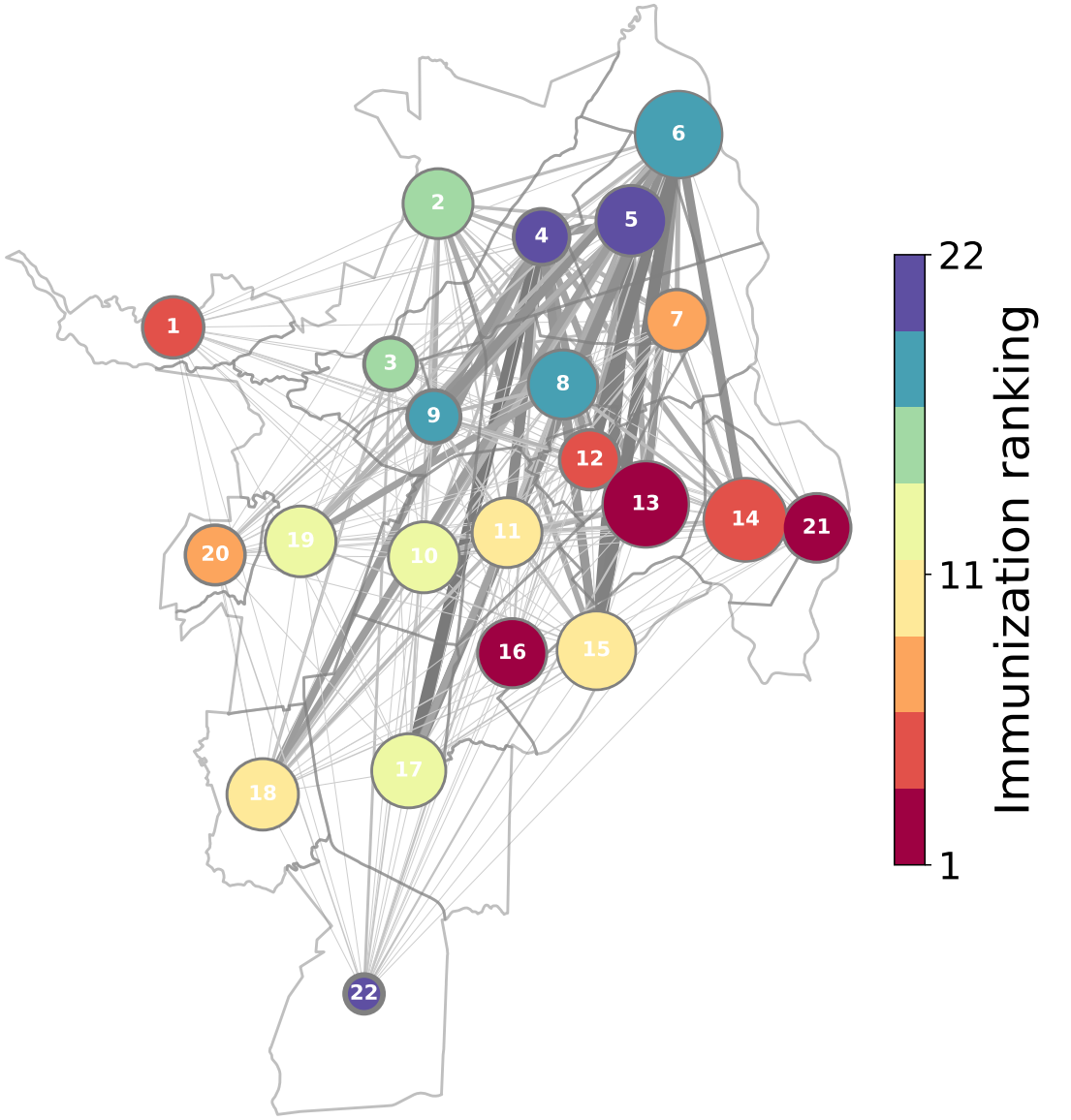


Figure 4.7: **Santiago de Cali immunization ranking**. The map represents the metapopulation of Santiago de Cali as a network in which the size of the nodes is proportional to the human population of each district whereas the colors are set according to their importance in the immunization ranking. Links account for the structure of the Origin-Destination matrix **R**.

where, as stated above, $n_i^{eff}(t)$ is the effective population of a patch i, i.e. the number of residents that remain in patch i plus the visitors from other patches. The former equation accounts for the probability that, once in a patch i at time t with $m_i(t)$ wild-type mosquitoes, a susceptible agent is infected by at least one bite from an infectious mosquito, being $\beta m_i(t)$ the total number of bites distributed among $n_i^{eff}(t)$ humans. Equation (4.40) contains three time dependent variables that directly depend on: (i) the epidemiological state of humans ($n_i^{eff}(t)$) in the whole population, (ii) the competition dynamics between *Wolbachia*-infected and wild-type vectors ($m_i(t)$), and (iii) the infection of wild-type mosquitoes by humans ($m_i^I(t)$).

For what concerns the first variable, $n_i^{eff}(t)$, we consider that a fraction α of infected individuals are able to move, since they are asymptomatic [120] while still being contagious [130]. Taking this into consideration the effective population of a patch i can be calculated as,

$$\begin{aligned} n_i^{\text{eff}}(n_i^I(t),\alpha,p) &= (1-p)n_i + p(1-\alpha)n_i^I(t) \\ &+ p\sum_{j=1}^{N_p} R_{ji}\left(n_j - (1-\alpha)n_j^I(t)\right) , \end{aligned} \tag{4.41}$$

In the same fashion, we can obtain the effective number of infected humans placed in population i at time t as,

$$i_i^{\text{eff}}(t) = (1-\alpha p)n_i^I(t) + \alpha p\sum_{j=1}^{N_p} R_{ji}n_j^I(t) . \tag{4.42}$$

4.6 Targeted release of *Wolbachia*-infected mosquitoes

The process of immunizing one patch was already shown in Fig. 4.6. *Wolbachia*-infected mosquitoes will take over the wild mosquitoes natural habitat after some time of being released, thus preventing contagion in the immunized patch. Obviously, if we have a set of N_p patches the easiest and fastest way to control dengue in a metapopulation would consist in immunizing all the patches at once. However, in reality, resources to produce and distribute *Wolbachia*-infected mosquitoes are quite limited, so we need to devise strategies to make the most of those scarce resources to reach and protect as many people as possible. Thus, equipped with the new metapopulation framework that includes the *Wolbachia* immunization dynamics, here we aim to find the best strategy to distribute the limited resources among the set of patches that define a metapopulation. In Section 4.4, we obtained the epidemic risk associated to a metapopulation. We could naively expect that the immunization of the most vulnerable patches would provide the optimal immunization strategy. However, the vulnerability indicator accounts for the expected prevalence once an outbreak has occurred and the disease has spread over the network while here, instead,

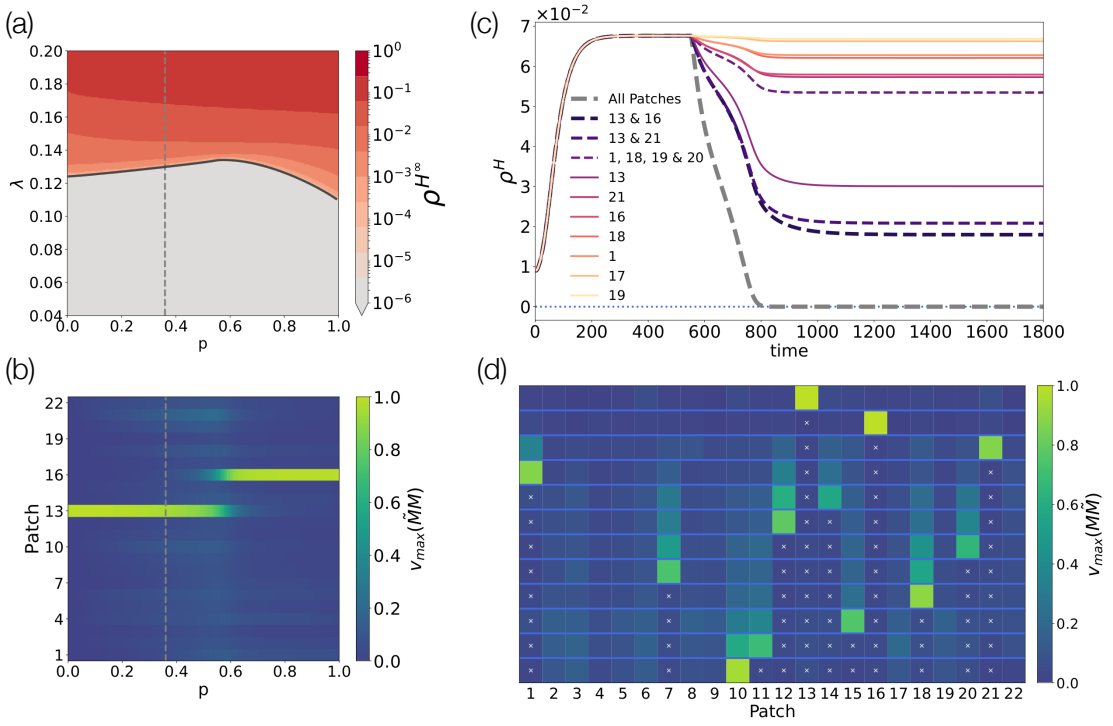


Figure 4.8: **Immunization strategies**. (a) Prevalence ρ^H as a function of p and $\lambda = \lambda^{HM} = \lambda^{MH}$ for the city of Cali. The vector distribution is obtained from real data [117] and the rest of the parameters of the model are: $\beta = 1$, $\mu = 0.3$ [131] and $\alpha = 0.75$ [120]. (b) Evolution of the components of the leading eigenvector $\vec{V}_{max}$ of matrix $\mathbf{M}\widetilde{\mathbf{M}}$ as a function of p. (c) Mitigation effect over the steady prevalence ρ^H when *Wolbachia*-infected mosquitoes are released in different patches (in this panel $p = 0.36$ [132]). (d) Illustration of the iterative method for ranking patches according to their relevance for mitigation. The first row shows the contribution of each component of $\vec{V}_{max}$. Then, we remove the patch corresponding to the largest component of $\vec{V}_{max}$ (patch 21) in matrix $\mathbf{M}\widetilde{\mathbf{M}}$ and compute the new leading eigenvector, shown in the second row (note that deleted patch 21 appears marked with a cross), and find its largest component (patch 16). This way, each row, say x, shows the composition of the leading eigenvector after removing the $x-1$ most relevant patches in $\mathbf{M}\widetilde{\mathbf{M}}$.

we are interested in pinpointing those areas whose simultaneous immunization provides the larger decrease in the prevalence of the whole system.

To identify the best immunization targets, let us first analyze the dengue prevalence in the absence of *Wolbachia* ($w_i = 0 \ \forall \ i$). In Fig. 4.8(a) we plot the the steady-state of the density of infected hosts as a function of the active population p and the contagion probability λ (for the sake of simplicity we have set $\lambda^{HM} = \lambda^{MH} = \lambda$). As we have previously defined $\rho^H = \lim_{t\to\infty} \sum_{i=1}^{N_p} n_i^I(t)/N_H$ is the fraction of dengue-infected humans in the steady-state. In this figure we plot the curve $\lambda_c(p)$ that separates the disease-free and the endemic regions for each value of p. It is precisely at λ_c where we can extract analytically the information about those patches that play a key role in the unfolding of an

epidemic outbreak [112]. As we have studied, in Section 4.3, this is achieved by inspecting the spectral properties of the mixing matrix $\mathbf{M}\widetilde{\mathbf{M}}$. Let us recall that this $N_p \times N_p$ matrix rules the evolution of the system for $\lambda \gtrsim \lambda_c$. In particular, while the maximum eigenvalue of the mixing matrix yields λ_c, the components of the associated eigenvector $\vec{V}_{max}(\mathbf{M}\widetilde{\mathbf{M}})$ quantify the contribution of each patch to an epidemic outbreak.

In Fig. 4.8(b) we show the evolution of the components of $\vec{V}_{max}$ as we change the fraction of the people moving from its patch p. In Section 4.4, we have found that the fraction of the active population in the city of Cali is close to 0.36[132]. At this value of p (signaled by a grey line in Fig. 4.8(a-b) the three most influential patches according to $\vec{V}_{max}$ are 13, 21 and 16, being 13 the most important compared to the following two. We validate how relevant are these patches by immunizing one at the time. To this aim, we simulate the release of a small population of *Wolbachia*-infected mosquitoes in a single patch and monitor the effect on the global prevalence ρ^H. We choose the two most influential patches according to our critical matrix (13 and 21) and 4 additional areas, specifically patches 1, 18, 19, 20, which correspond to the districts selected by the initiative World Mosquito Program (WMP) to implement *Wolbachia* immunization in Cali. Note that the eigenvector centrality $\vec{V}_{max}$ predicts little to no relevance of these areas. We select the two most influential patches (13 and 21) and 4 additional areas whose influence is very low as dictated by $\vec{V}_{max}$, specifically patches 1, 18, 19 and 20, since these districts were selected by the initiative World Mosquito Program [133] (WMP) to implement *Wolbachia* immunization in Cali. The results (solid curves in Fig. 4.8(c) confirm that the release in those poorly influential areas has little impact in decreasing the overall prevalence ρ^H while immunizing influential patches has a great mitigation effect, especially when the release is implemented in district 13.

When a set of patches is to be simultaneously immunized, as it is the case in usual campaigns [134], the mitigation effect increases. As an example, in Fig. 4.8(c) we show the decrease of the prevalence when the four areas (1, 18, 19, and 20) chosen by the WMP are immunized at a time. In the context of simultaneous immunization, one can apply an immunization strategy based on the components of $\vec{V}_{max}$ to immunize in the most efficient way. Then, given that the resources for x patches are available, one should implement the release in those patches corresponding to the x-largest components of $\vec{V}_{max}$. This strategy, considering the first two patches (13 and 21), is shown in Fig. 4.8(c). However, in the same plot, we also present the simultaneous immunization of the first and the third patches (13 and 16) according to their relevance in $\vec{V}_{max}$. From this plot, it is clear that this latter choice outperforms the former one causing a larger mitigation of the endemic level ρ^H.

The roots behind the former counter-intuitive result lie in the double mitigation effect caused by the immunization of a single patch. First, at the local level, immunization considerably decreases the contagion of residents, since only imported cases can occur but

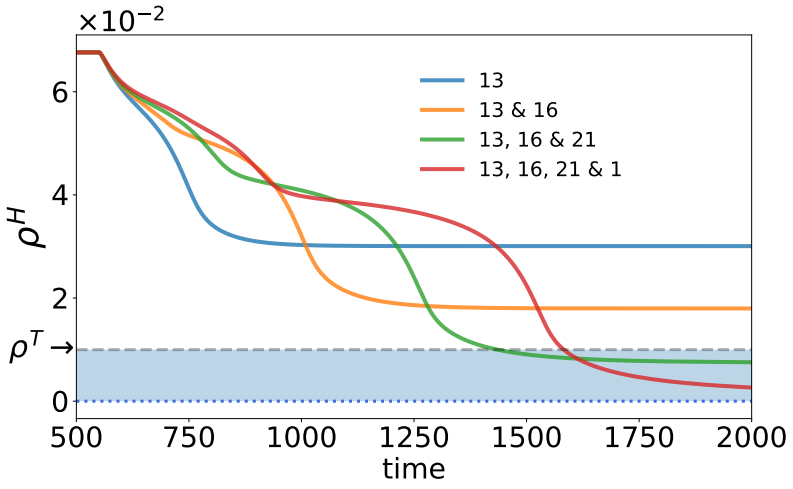


Figure 4.9: **Resources distribution**. Time evolution of the global prevalence ρ after $5 \cdot 10^4$ *Wolbachia*-infected mosquitoes are released and evenly distributed among different patches (color code). The target prevalence is $\rho^T = 10^{-2}$.

they will not cause secondary infections within the patch. In addition, at the global level, it also prevents the importation of new cases to those areas connected with the immunized patch. Thus, removing the first patch (21) alters the influence of the remaining patches, changing the rank obtained with $\vec{V}_{max}$. A solution to this problem is to choose the x^{th} patch to immunize by removing the contribution of the previously chosen $x^{th} - 1$ patches. This way, instead of choosing the x^{th} largest component of $\vec{V}_{max}$ as the immunization target, one finds the largest component of the leading eigenvector of a $(N_p - x + 1) \times (N_p - x + 1)$ matrix resulting from the removal of the $x - 1$ previously immunized patches in $\widetilde{\mathbf{MM}}$. The iterative method used to find out the x most influential patches is illustrated in Fig. 4.8(d) and the complete rank provided by this method is shown in Fig. 4.6(c). Applying the iterative method we validate that the two most important patches to immunize at a time are 13 and 16, rather than the choice 13 and 21 resulting from the inspection of $\vec{V}_{max}$.

Our previous analyses have focused on identifying the most important patches to concentrate control policies in a few areas. Now, we apply the classification to a more realistic scenario where resources are limited, *i.e.*, we only have a fixed amount of *Wolbachia*-infected vectors to release. Obviously, the more patches we distribute this number of mosquitoes, the greater the final mitigation effect. However, distributing these resources among a large number of patches implies releasing small numbers of *Wolbachia*-infected vectors and, therefore, a very long lead time to obtain the desired decrease in prevalence. Therefore, in these circumstances, the question is how many and which patches we should consider achieving a mitigation target ρ^T in the shortest possible time. We can use the ranking obtained and consider different scenarios. Namely, as shown in Fig. 4.9, we can: (i) concentrate all resources on patch 13, (ii) distribute them equally between the two most

important 13 and 16, (iii) spread them between 13, 16, and 21, and so on. For a given $\rho^T = 10^{-2}$ the correct choice corresponds to distribute resources among the three most relevant patches.

4.7 Conclusions

In this chapter, devoted to the control of infectious diseases in metapopulations, we have explored the interplay between human mobility, epidemiological features, such as asymptomaticity, demographics, and the spreading behavior of infectious diseases. By focusing on VBD with their characteristic cross-contagion contagion phenomena between humans and vectors, we added the vector distribution on the geographical places. With all these inseparable ingredients playing a role in the onset of the epidemic outbreak, we aimed to understand their interplay to finally find the best strategy to control VBD.

First, our main goal has been to highlight the need of incorporating real human mobility patterns into the design of containment policies targeting specific geographical areas, for efficient policies can turn useless due to a small variation of human mobility habits. To this aim, we have derived an analytic expression of the epidemic threshold that captures the critical conditions that lead to the active phase of epidemics. The study has examined the detrimental effect of mobility on the spread of VBD, as well as the abrupt changes in how epidemics unfold due to geographic context and relevance. In this respect, we have shown that patches leading to epidemic onset may suddenly change as human mobility varies, highlighting the need for a careful examination of the so-called mixing matrix before considering control policies designed to reduce mobility such as confinement.

Once derived the interplay between human mobility and the demography of humans and vectors through the expression of the mixing matrix, we have estimated the epidemic risk associated with different connected regions. Interestingly, to obtain an epidemic risk indicator it is not necessary to perform any simulation or integration of the equations, for it can be readily calculated from the data-driven expression of the mixing matrix. We have shown that the epidemic risk when applied to Santiago de Cali, where dengue is an endemic disease, correlated very well with the observed dengue prevalence across the different patches that compose this city.

Finally, in the last section, we have incorporated the competition dynamics between *Wolbachia*-infected and wild-type mosquitoes into our metapopulation model. This reformulation of the model has allowed us to study the effects of this novel immunization strategy when applied in a selective manner over a metapopulation. In particular, we have focused on pinpointing those geographical areas that provide the best epidemiological outcome after the release of a limited amount of Wolbachia-infected vectors. To do so we have focused on the spectral properties of the mixing matrix and found that the best way to

find out the set of patches to control is to perform the analysis of the leading eigenvector of the former matrix in an iterative way, i.e. by analyzing successive versions of the matrix in which the patches with the largest contribution to the leading eigenvector have been removed. The roots behind this approach are that immunization of a single patch also provides "partial" immunity to the other patches, with this second-order protective effect being less or more depending on the relationship that each patch bears to the one it has been chosen to protect.

In conclusion, the framework in this chapter paves the way for planning targeted interventions in urban areas, especially those dealing with scarce and difficult means of control, such as Wolbachia-infected mosquitoes, for which a careful study of a population as a set of interacting units (here patches) is much needed to achieve a systemic goal such as the mitigation or suppression of an endemic disease.

Chapter 5

Reactive epidemic control through mobility & social reaction

Scientific process has two motives: one is to understand the natural world and the other is to control it.

C.P. Snow

In the previous chapters, we have developed several theoretical frameworks aimed at studying different epidemiological scenarios. These frameworks take advantage of incorporating into their formulation real data of social activity, ranging from face-to-face interactions to daily human mobility patterns, an ingredient of utmost importance when dealing with communicable diseases and their control. In this chapter, we develop an epidemic model that combines the insights obtained from previous chapters about the influence that both human activity and the social adherence to restrictions have on the success of epidemic control policies. In particular, the new theoretical framework explores the interplay between the spread of infectious diseases and the modification of human interactions by mandatory mobility restrictions, as well as, the social response to those closure restrictions.

The metapopulation model we rely on in this chapter has an important difference with the model developed in Chapter 4: the inclusion of two different scales for human interactions. In particular, we focus on the interaction due to human mobility between different geographic areas and, also, the local (community) interactions given by the economic activities in each area. It is on the limitation of these two types of interaction that we will base the containment policies. It is also important to note that, unlike the model developed in Chapter 4, in which restrictions were imposed in an exogenous way, here the activation of contention measures is endogenous and reactive. In this way, control measures are strengthened or relaxed depending on the observed disease incidence, much like the evolution of contention measures during the COVID-19 pandemic in European countries.

This chapter is organized as follows. In Section 5.1 we motivate the importance of our adaptive response model in the light of the global socioeconomic challenges that the COVID-19 pandemic posed to most of the countries in the world. We argue the need for adaptive control responses as a function of the social needs of the most affected sectors, the health and the economic systems. Then, in Section 5.2, we introduce the reactive metapopulation model, describe its basic ingredients, most of them motivated by the SARS-CoV-2 transmission patterns, and construct the Markovian equations that characterize the epidemiological evolution. In Section 5.4, we analyze different epidemic scenarios with the aim of illustrating the effects that the reactive control policy and the social response to restrictions have on both the epidemiological evolution and the disruption of the socioeconomic fabric. Finally, in Section 5.5 we study the interplay between the proposed control dynamics and the particular structure of human mobility flows and the demographic distribution of a population. By focusing on the autonomous communities of Spain we show that some territories may afford less restrictive measures than others as a consequence of the length of the shortest paths between the most crowded city and the rest of the populations. Thus, this chapter highlights the need to use systemic approaches to achieve an optimal trade-off between the strength of the control policies for the benefit of the health system and their associated socioeconomic cost.

5.1 The aim of the joint control policies

In early 2020, the world was disrupted by a new contagious virus, the SARS-CoV-2. Little was known about the epidemiological features of the new highly contagious pathogen, no drugs or vaccines were available. Furthermore, when the epidemic outbreak had had already an elevated impact on the population, the world authorities resorted to non-pharmaceutical interventions (NPIs) to keep the epidemic outbreak under control [135]. Of course, the specific intervention was chosen to cope with the pandemic and the degree of accomplishment of the different measures notoriously varied from one country to another, depending on different social, economic, political, or geographic features [136, 137]. While some countries located in South-East Asia or Oceania were capable of maintaining low levels of community transmission via the timely combination of strict border closure policies, robust contact tracing systems, and punctual local lockdowns to eradicate localized outbreaks, most of the countries worldwide reacted with population-wide lockdowns as an urgent solution to the exponentially growing number of local contagions [138].

However, widespread lockdowns to contain the COVID-19 pandemic disrupted the world economy [139]. School and business closures, restrictions on international travel, and trade led to massive job losses and the worst GDP contraction in decades in most of these countries [140–142]. Even though vaccines gave some relief to both health and

economic systems by relaxing restrictions in some countries, the unequal distribution and adoption of vaccines worldwide[143, 144], along with their reduced efficiency to the new variants of the virus [145, 146], have prevented countries from completely abandoning NPIs as efficient tools against the ongoing pandemic.

Moreover, fearing the extending human toll with renewed waves caused by new variants of the virus, policymakers have drawn on lockdowns, stay-at-home and self-quarantine programs over and over again [147]. The multi-layered crisis provoked by the pandemic that primarily encompasses the health and economic systems requires joint effective policies to prevent worse outcomes. An optimal policy response has to face the challenge of reducing the daily incidence of the disease allowing health care systems to cope while reshaping their economies by adjusting the mobility restrictions.

Several models have addressed targeted lockdowns to optimize the economic losses and the epidemic control [148–150]. Others have attempted to promote sustainable policies by alternating [151] or intermittently implementing [152] control measures that allow the economic system to cope while giving relief to the health systems. The different policy scenarios were simulated to analyze both health and economic effects [153]. However, these models do not account for the importation of new cases from other places that may spark a new outbreak or epidemic wave, most of them do not include human mobility, and the important effect of social reactions to the policies implemented [154–156].

With all that in mind, in the next section, we introduce a theoretical framework to explore the interplay between the spread of infectious diseases and closure interventions promoted to keep the outbreaks under control and their economic consequences.

5.2 The reactive control model

As we have previously seen in Chapters 2 and 4 metapopulations capture the influence of the heterogeneous spatial structures on the epidemic dynamics, where the mixing events due to human mobility between geographical areas allow significant perturbations in the community transmission of the virus at each patch. Here, we heuristically create an extension of the Movement-Interaction-Return (MIR) model [51] in which we add interactions taking place at the household level, as presented in [157], and a compartmental dynamics that, as it was the case in Chapter 3, is inspired in the transmission of SARS-CoV-2 [79]. Additionally, we add reactive control policies that are activated or deactivated as a tolerance threshold is reached.

5.2.1 Basic building blocks of the reactive compartmental model

Before explaining in detail the characteristics and parameters governing the evolution of our reactive metapopulation model, let us briefly describe the three main features that characterize its functioning.

The contagion dynamics. As usual, we draw the course of the disease by means of a compartmental model that contains the main epidemiological stages. In particular, the compartments are: Susceptible (S), Exposed (E), Presymptomatic infectious (P), Asymptomatic infectious (A), symptomatic Infectious (I), Detected (D), and Recovered (R). The transition probabilities between each of the former compartments are represented Fig. 5.1-A, and explained in detail below.

Human mobility. Being a metapopulation framework, our model assumes that the whole population of N individuals is distributed into N_P subpopulations, each one associated with a specific geographic location (patch) i and of size n_i ($N = \sum_{i=1}^{N_P} n_i$). The population n_i of each patch i corresponds to the number of residents associated with the corresponding geographic area as dictated by real census information. As in the previous Chapter 4, we assume that the population mobility is dominated by recurrent (back-and-forth) patterns that are captured in Origin-Destination (OD) matrices that, in turn, allow us to construct the row stochastic matrix $\mathbf{R}$ whose elements, say R_{ij}, account for the probability that a resident in patch i moves to patch j as shown in Fig. 5.1-B.

The control policy. The main novelty of the metapopulation model resides in its reactive nature. This way, we incorporate autonomous control policies which are either strengthened or relaxed as a function of the dynamical advance of the outbreak. Specifically, at each patch i, we define a parameter $p_i(t)$ that determines the share of the population that carries out its usual (unrestricted) social and labor activity. As shown in Fig. 5.1-C this parameter depends on the fraction of detected residents, hereinafter denoted by $\rho_i^D(t)$. Since, some of detected individuals are in need of the health system, when this fraction is exceeds a certain tolerance threshold, ϵ, different interventions are activated to reach the desirable control of the disease, being the extent of the contention measures captured by the value of $p_i < 1$. Obviously, when $\rho_i^D(t) < \epsilon$ subpopulation i follows a normal socioeconomic activity level ($p_i = 1$). Mimicking different real policies implemented during the course of the COVID-19 pandemic, we implement in our model the multiscale interventions as follows:

1. At the metapopulation level, usual mobility patterns, encoded in matrix $\mathbf{R}$, are altered to minimize the interactions with the population from those areas with high incidence numbers ($\rho_j^D > \epsilon$). This way the probability that a patch j with high incidence at time t (and hence low $p_j(t)$) receives commuters from patch i decreases as $p_j(t)R_{ij}$.

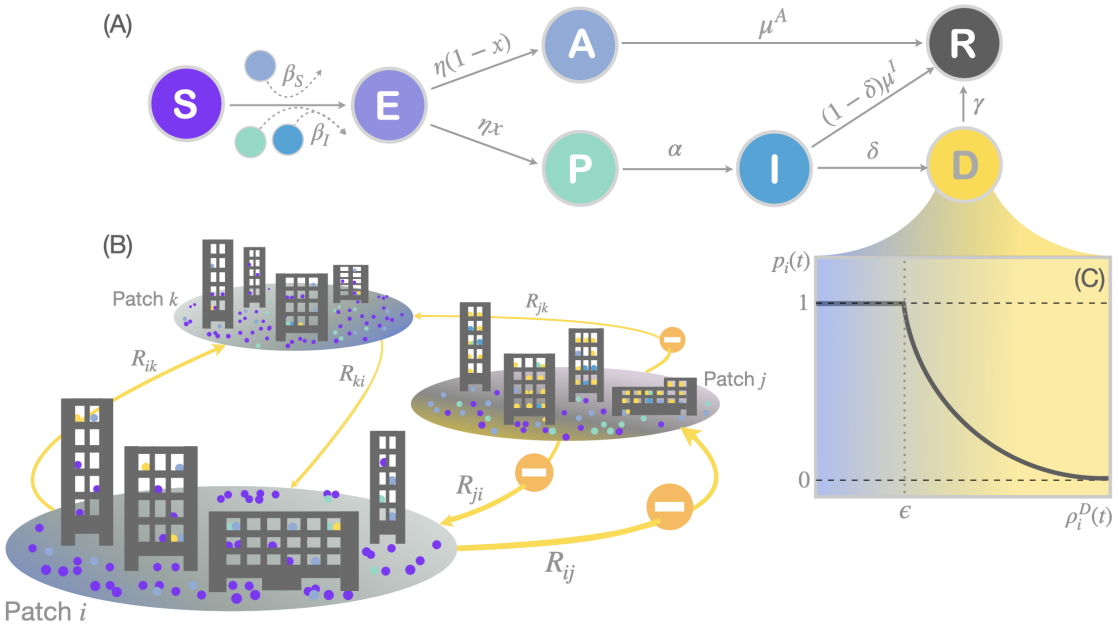


Figure 5.1: **Scheme of the multiscale reactive control model**. A. The compartmental model comprises 7 compartments. Susceptible individuals (S) can be infected by infectious agents in asymptomatic (A), presymptomatic (P) and symptomatic (I) compartments. Once infected, agents pass to an exposed (E) compartment before passing to either asymptomatic (A) or symptomatic. In the latter case agents transit first a presymptomatic (P) stage before developing symptoms (I). A symptomatic agent can be detected by the need of the health system (D) before being removed (R) or, otherwise, directly pass to R, which is the final compartment for those asymptomatic. B. In the metapopulation model, individuals interact both inside and outside households. The activity outside households takes place in the residential patch or in a different one through the movement of individuals. Mobility flows between patches are ruled by matrix $\mathbf{R}$ whose elements R_{ij} account for the probability that a resident in patch i visits j. C. When the fraction of D in a patch i exceeds some threshold ϵ the activity level of the patch p_i decreases, affecting both the mobility and the social contacts of agents.

2. Apart from receiving less visitors, the economic activity of a patch j with high incidence is modulated according to the extent of the disease in its population. This way, a fraction $p_j(t)$ of residents in patch j follows a normal socioeconomic activity while the remaining $(1-p_j(t))$ fraction are set to inactive and reduce their interactions to those taking place at the household level.

3. Finally, the reduction of the economic activity of the population has been usually accompanied with measures aimed at creating social bubbles preventing the transmission of the disease between members from different households. Note that this mechanism demands the creation of a new compartment CH to accommodate those susceptible individuals without infectious individuals in their social bubbles and, therefore, protected from the disease. Likewise, to model the efficiency of this mech-

anism, we introduce a new parameter, the social permeability ϕ, denoting the fraction of households whose members are not efficiently isolated [158], despite being inactive from the socioeconomic point of view.

5.3 Coupling contagion and commuting dynamics

In the following we give a more accurate description of the afore presented ingredients by formulating the Markovian framework. As it has been shown along the previous chapters of this thesis, the Markovian formalism allow us to make a mathematical representation of the interplay between contagion dynamics, the mobility patterns, and, in this particular model, the reactive containment policies.

5.3.1 Coupling spreading dynamics and human mobility

As shown in Fig. 5.1-A our compartmental model includes 7 epidemiological compartments in an attempt to capture those epidemiological states that are relevant for addressing SARS-CoV-2 transmission and the associated non-pharmaceutical interventions. The most relevant aspects of the different transitions between compartments are as follows:

(i) First, susceptible (S) agents can be infected by contact with presymptomatic (P), asymptomatic (A) and infected (I) agents with probability β^c, $c \in \{P, A, I\}$.

(ii) If infected, susceptible agents turn exposed (E), for an average period of η^{-1} days. It is then assumed that a fraction x of the population becomes symptomatic while the remaining fraction $(1 - x)$ becomes asymptomatic (A).

(iii) The presymptomatic but infectious stage last for an average period of α^{-1} days before becoming symptomatic and infectious (I). Symptomatic and asymptomatic infectious individuals are assumed to overcome the disease with probabilities μ^I and μ^A respectively.

(iv) In the case of symptomatic individuals, we assume that they can be detected (D) with probability δ and, once identified, detected individuals are perfectly isolated and assumed not to transmit the disease accordingly.

(v) Finally, to control the time one detected individual is identified as an active case, we assume that they pass to the R compartment with probability γ.

Let us note that the presented model share many issues with the one used in Chapter 3 for implementing the competition between the spread of SARS-CoV-2 and contact tracing policies. However, apart of not incorporating the tracing mechanism for individuals in

compartments P and A, it is worth to note that the present model does not incorporate the presynmptomatic stage for asymptomatic individuals, at variance with the compartmental dynamics presented in Chapter 3.

The interplay between contagion processes contained in the former epidemiological dynamics and human mobility is modeled with a Microscopic Markov Chain Approach [43]. To this aim, it is assumed that each time step corresponds to a day and each patch i is characterized by a set of variables $\rho_i^m(\mathrm{t})$ that govern the fraction of the residents in compartment m at time t. Following the rules of the compartmental dynamics and the recurrent nature of human mobility patterns, we can write the following equations,

$$\begin{aligned}
\rho_i^S(t+1) &= \left[\rho_i^S(t)+\rho_i^{CH}(t)\right]\left[p_i(t)(1-\Pi_i^{act}(t)) + (1-p_i(t))(1-sh_i)(1-\Pi_i^{in}(t))\right] , && (5.1)\\
\rho_i^{CH}(t+1) &= \left[\rho_i^S(t)+\rho_i^{CH}(t)\right](1-p_i(t))sh_i , && (5.2)\\
\rho_i^E(t+1) &= \left[\rho_i^S(t)+\rho_i^{CH}(t)\right]\left[p_i(t)\Pi_i^{act}(t)+(1-p_i(t))(1-sh_i)\Pi_i^{in}(t)\right]+(1-\eta)\rho_i^E(t) , && (5.3)\\
\rho_i^P(t+1) &= (1-\alpha)\rho_i^P(t)+\eta x\rho_i^E(t) , && (5.4)\\
\rho_i^A(t+1) &= (1-\mu^A)\rho_i^A(t)+\eta(1-x)\rho_i^E(t) , && (5.5)\\
\rho_i^I(t+1) &= (1-\delta)(1-\mu^I)\rho_i^I(t)+\alpha\rho_i^P(t) , && (5.6)\\
\rho_i^D(t+1) &= (1-\gamma)\rho_i^D(t)+\delta\rho_i^I(t) , && (5.7)\\
\rho_i^R(t+1) &= \rho_i^R(t)+(1-\delta)\mu^I\rho_i^I(t)+\mu^A\rho_i^A(t)+\gamma\rho_i^D(t) , && (5.8)
\end{aligned}$$

where sh_i represents the probability that a confined susceptible resident of patch i is isolated in a completely susceptible household. This happens when all the members of the household are not infectious and not mixed with the rest of households. Therefore,

$$sh_i(t) = \left(1-\rho_i^E(t)-\rho_i^P(t)-\rho_i^A(t)-\rho_i^I(t)\right)^{\sigma_i-1}(1-\phi) , \tag{5.9}$$

where σ_i is the average size for the households located inside the population i.

The former set of equations contains the terms Π_i^{act} and Π_i^{in} that account for the probability of active and inactive individuals with residence in patch i contracting the disease respectively. Importantly, these two terms capture the mixing between contagion processes and recurrent mobility between patches. In the following, we describe their functional form.

The term Π_i^{act} in Eqs. (5.1) and (5.3) depends on interactions made both inside the household and in the usual workplace of each individual. Taking into account the modified mobility patterns as a result of the outermost layer of intervention, hereinafter denoted by $\tilde{\mathbf{R}}$ (and defined below in Section 5.3.2), this probability reads as,

$$\Pi_i^{act} = 1 - P_{H,i}(t)\sum_{j=1}^{N_P}\tilde{R}_{ij}(t)P_{O,j}(t) , \tag{5.10}$$

where $P_{H,i}(t)$ and $P_{O,i}(t)$ represent the probability of not contracting the disease at time t inside a household or a workplace located at patch i respectively. The former probability

reads as,

$$P_{H,i}(t) = (1-\beta^I)^{z^h \sigma_i (\rho_i^P(t)+\rho_i^I(t))}(1-\beta^A)^{z^h \sigma_i(\rho_i^A(t))} \ . \tag{5.11}$$

Note that here we assume the number of contacts inside each household to be proportional to its size σ_i and we introduce a scaling factor z_h to ensure that the average number of contacts across the metapopulation matches the estimations obtained from the literature, denoted by $\langle k_h \rangle$. Accordingly,

$$z^h = \frac{N\langle k_h \rangle}{\sum_{j=1}^{N_P} n_j \sigma_j} \ . \tag{5.12}$$

In its turn, the probability that a susceptible active individual does not contract the disease inside the patch i at time t, $P_{O,i}(t)$, is given by,

$$P_{O,i}(t) = \prod_{j=1}^{N_P} (1-\beta^I)^{z^O f\left(\frac{n_i^{eff}}{a_i}\right)\frac{n_{j\to i}^P + n_{j\to i}^I}{n_i^{eff}}} (1-\beta^A)^{z^O f\left(\frac{n_i^{eff}}{a_i}\right)\frac{n_{j\to i}^A}{n_i^{eff}}} \ , \tag{5.13}$$

where $n_{j\to i}^P$, $n_{j\to i}^A$ and $n_{j\to i}^I$ are the number of presymptomatic, asymptomatic and infectious agents going from j to i respectively. The effective population of patch i after population movements is encoded as n_i^{eff}. Taking into account the mobility patterns, these quantities can be expressed as,

$$n_{j\to i}^m = n_j \rho_i^m(t) p_j(t) \tilde{R}_{ji}(t) \ , \tag{5.14}$$

$$n_i^{eff} = \sum_{j=1}^{N_P} n_j p_j(t) \tilde{R}_{ji}(t) \ . \tag{5.15}$$

Note that we assume that the contacts of the active population inside each patch i are governed by a monotonically increasing function f_i of the effective density of each patch i, $d_i^{eff} = n_i^{eff}/a_i$, being a_i its area. In particular, the form chosen for this function is,

$$f_i(d_i^{eff}) = 2 - e^{-d_i^{eff}/\langle d \rangle} \ , \tag{5.16}$$

where $\langle d \rangle$ is the population density across the entire metapopulation. Here the scaling factor z_O is introduced to set the average number of contacts of the active population outside the household to $\langle k_O \rangle$. Therefore,

$$z^O = \frac{\langle k_O \rangle \sum_{j=1}^{N_P} n_j^{eff}}{\sum_{j=1}^{N_P} n_j^{eff} f\left(\frac{n_j^{eff}}{a_j}\right)} \ . \tag{5.17}$$

Finally, the probability of contracting the disease for the inactive population must only account for the potential contagions occurring when interacting with the household members. Consequently,

$$\Pi_i^{in}(t) = 1 - P_{H,i}(t) \ . \tag{5.18}$$

Parameter	Value	Description	Reference
β_I	0.07	$S \xrightarrow{P,I} E$	[158]
β_A	$\beta_I/2$	$S \xrightarrow{A} E$	[158]
η	2.6^{-1}	$E \to A, P$	[159]
x	0.35	Fraction of Symptomatic	[160]
α	2.6^{-1}	$P \to I$	[158, 159]
δ	3^{-1}	$I \to D$	Assumed
μ_A	6.8^{-1}	$A \to R$	[158, 159]
μ_I	4.2^{-1}	$I \to R$	[158]
γ	14^{-1}	$D \to R$	Assumed
$\langle k_H \rangle$	3.2	Average number of contacts at the household	[49]
$\langle k_O \rangle$	8.6	Average number of contacts outside the household	[49]
σ	2.5	Average family members per house	[161]

Table 5.1: **Epidemiological parameters and interaction constants**. Note: The incubation period is 5.2 days [159] but half of this time infected individuals are not contagious (E stage). The infectious window is the same for both asymptomatic and symptomatic infections ∼6.8 days [158]. In the case of the symptomatic individual, this infectious window is divided into P and I stages.

The former equations contain a number of epidemiological parameters governing the flows between the different compartments of the epidemiological model and those social features that play a role in contagion dynamics. The exact values of these parameters are extracted from the literature and are reported in Table 5.1.

5.3.2 Adding the control policies

Mathematically, we add the manageable activity levels with the reactive function $p_i(t)$ that intrinsically depends on detected residents inside the patch ρ_i^D. For simplicity, we use a piece-wise function, so that,

$$p_i(t) = \begin{cases} 1 & if\ \rho_i^D < \epsilon \\ e^{-\chi((\rho_i^D(t)/\epsilon)-1)} & if\ \rho_i^D \geq \epsilon \end{cases} \tag{5.19}$$

where χ quantifies the strength of the reaction. Unless specified, we fix $\chi = 10$ ensuring a sharp decrease in the economic activity once the policy is activated. Remember that ϵ is the threshold alarm of the health system to avoid be overwhelmed. At the metapopulation level, usual mobility patterns, encoded in matrix $\mathbf{R}$, are altered to minimize the interactions with the population from those areas with high incidence numbers ($\rho_j^D > \epsilon$). From these

assumptions, the elements of the modified mobility matrix $\tilde{\mathbf{R}}$ are given by,

$$\tilde{R}_{ij}(t) = \begin{cases} p_j(t)R_{ij} & if\ i \neq j \\ 1 - \sum_{j \neq i} p_j(t)R_{ij} & if\ i = j \end{cases} \tag{5.20}$$

5.4 Epidemic and socioeconomic activity evolution under different interventions

Once we have introduced our metapopulation reactive model and its Markovian formulation we now present its use. The needed ingredients to implement the Markovian equations in a general scenario are: information on the spatial demographic distribution of the population, their recurrent mobility patterns, and the structure of their contacts in different contexts. By introducing this information into the reactive model we can quantify the reduction of sociability as a result of confinement policies and, thus, evaluate the impact of these non-pharmaceutical interventions.

From now on, we focus on addressing the spread of diseases across Spain. We divide the geographical territory by Spanish municipalities with at least 500 residents. We extracted demographic information from official reports published annually by the National Institute of Statistics of Spain (INE) whereas mobility patterns are obtained by INE from surveys conducted in 2011 in which the population indicated their usual places of work [161]. The information about the average number of contacts a person has per day was obtained from a study reporting the social mixing patterns in 152 countries [49]. With this information at hand, we can construct a metapopulation with $N_P = 4293$ patches and $L = 59722$ weighted and directed links.

To illustrate a controlled epidemic outbreak, we analyze the spatiotemporal dynamics shaped by the autonomous multiscale control policies here implemented. Unless specified, we use $\epsilon = 7 \cdot 10^{-4}$, ensuring a rapid response of the system when the number of detected symptomatic cases exceeds 70 individuals per 10^5 inhabitants (in the Spanish national health system, the average number of hospital beds per 10^5 people is around 240 [162] aiming to prevent the saturation of the health system capacity).

In Fig. 5.2 we represent, for different values of the permeability ϕ (the social reaction to the policy, seeing as the local stringency), the time evolution of the detected active cases in the 200 most populous municipalities in Spain (grey lines) and the average in the whole country (black line), for an outbreak whose initial seed is placed in the most populous city, Madrid. The emergent self-regulating system leads to different epidemic waves resulting from the activation and deactivation of the control policies to respond to the advance of the outbreak in each municipality. The length and frequency of the epidemic waves are strongly influenced by the social permeability allowed ϕ. High permeability values represent the

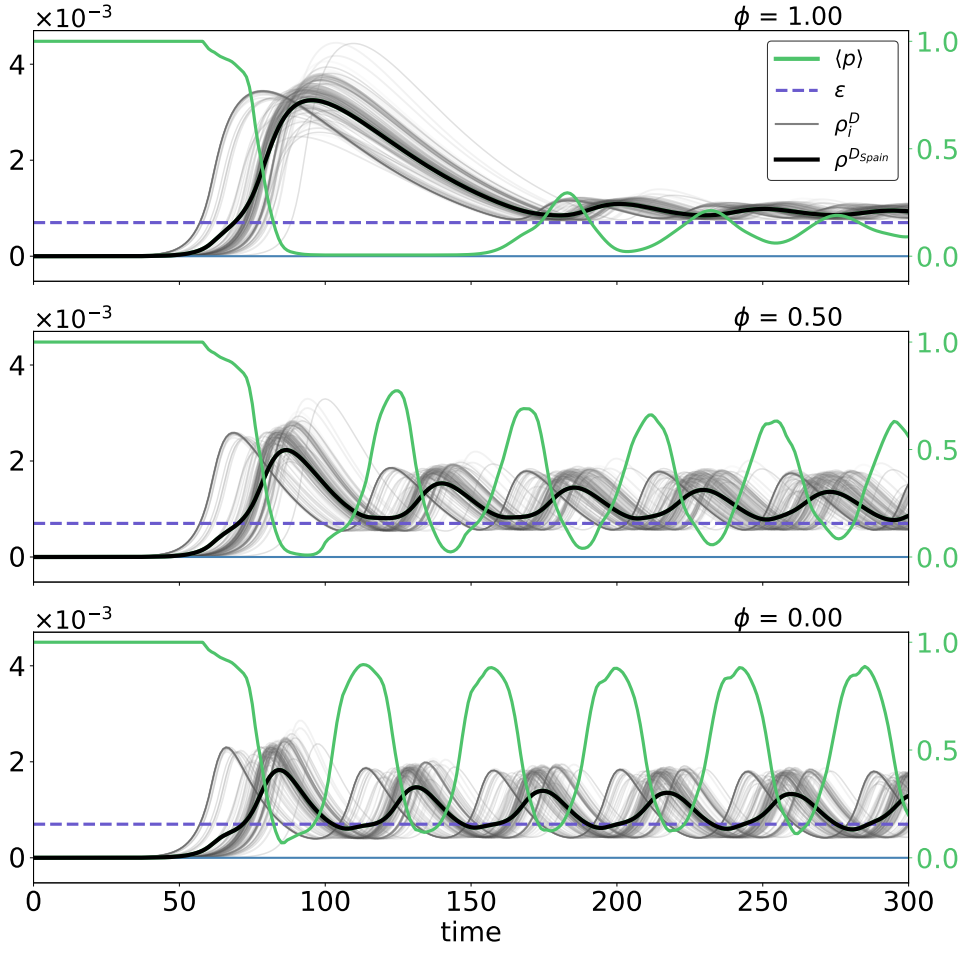


Figure 5.2: **Epidemic evolution & socioeconomic activities**. Evolution of symptomatic detected individuals (black and grey lines) and the mean value of the socioeconomic activity level $\langle p \rangle$ (green line) for three different permeability values. Grey lines represent the evolution of the 200 most populous Spanish cities, whereas the black line shows the average number of detected individuals at the national level. The dashed line represents the value of the tolerance threshold ϵ. The green line represents the average mobility allowed by the reactive policy.

failure of household isolation, giving rise to a large and long first epidemic wave followed by smaller secondary outbreaks. Instead, when permeability decreases, the creation of social bubbles becomes more successful and we observe a less severe first epidemic wave followed by more frequent and shorter subsequent epidemic outbreaks.

Apart from the epidemiological evolution (here represented by the evolution of number of symptomatic detected individuals), in the three scenarios explored we represent (green lines) the evolution of the average socioeconomic activity level allowed, $\langle p \rangle(t)$, computed across the different municipalities of the metapopulation. Besides the observed anti-phase oscillation with respect to the epidemic waves, comparing the patterns of $\langle p \rangle(t)$ for different values of the permeability ϕ yields a counter-intuitive result: the more stringent local

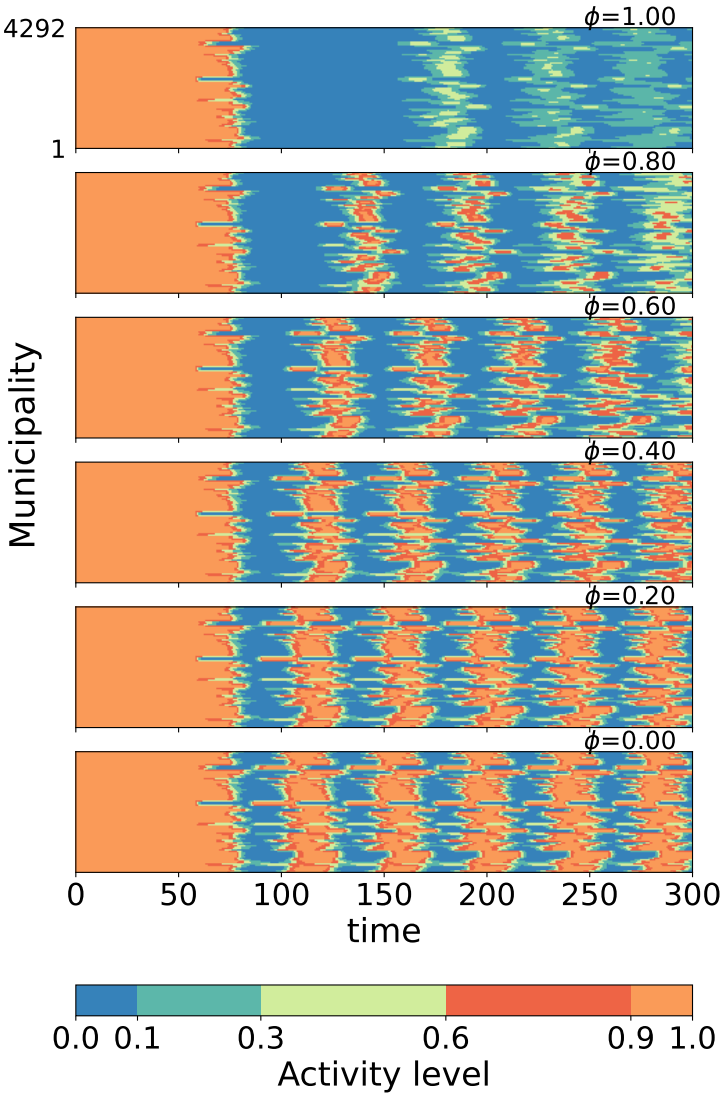


Figure 5.3: **Patch policy stringency**. Spatiotemporal activity level allowed over 300 days for each patch of the metapopulation of Spain, for different permeability values. The values of activity or mobility, $p_i(t)$ range from 0 to 1, *i.e.* from complete lockdown to the usual human activity flows respectively.

confinements are (the smaller ϕ) the larger socioeconomic activity levels reached. The same results can be observed in Fig. 5.3 in which we show the time evolution of the activity level for all the 4292 Spanish municipalities when 5 different values of the permeability ϕ are used.

In the scenarios shown in Fig. 5.2 the number of active detected cases falls above the desirable incidence of the disease as a result of the existence of different infectious compartments inducing a lag between the deployment of an intervention and its observed outcome. Regarding the impact of the local policies, the creation of social bubbles entails an instantaneous depletion of the pool of susceptible individuals, making those living in

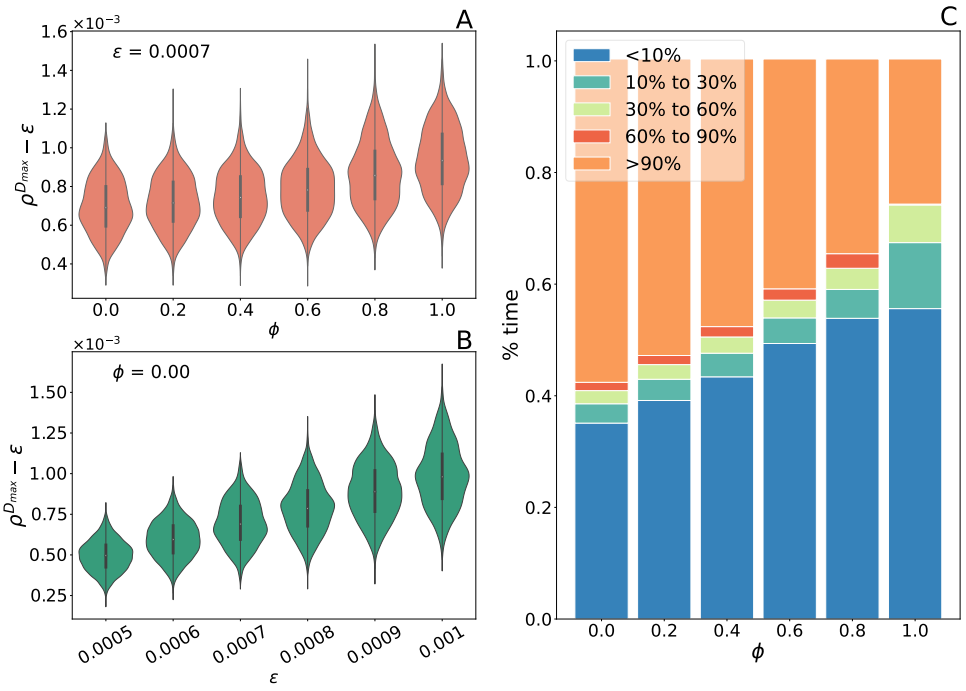


Figure 5.4: **Epidemic peak, activation threshold, permeability & socioeconomic activities**. A. Distribution of the maximum value of detected individuals in each municipality for different values of ϕ. B. Effect of the activation threshold ϵ on the distribution of the maximum value of detected. C. Time spent in each activity level allowed over 300 days (normalized to 1) for different permeability values.

healthy households unreachable to other infectious peers. This is translated into an abrupt decrease in the effective reproduction number $\mathcal{R}$, allowing the *suppression* of the outbreak. In the absence of social bubbles, all the individuals are exposed to the disease, but the outbreak advances at a slower speed due to the reduced sociability, corresponding to a *mitigation* strategy.

In epidemiological terms, Fig. 5.4-A shows that the distribution of the value of the first epidemic peaks of all Spanish municipalities is reduced as the permeability ϕ decreases, *i.e.* as the stringency of the local control policy increases. Obviously, larger values of the first peak results in a higher contribution to the fraction of recovered individuals, see Fig. 5.5, which explains the small amplitude of the subsequent epidemic waves for large values of ϕ as a result of the reduced susceptible population that remains. In addition, the effect of reducing the activation threshold ϵ is clear from Fig. 5.4-B: the lower the activation threshold, the milder the first epidemic wave.

The economic outcome from the mobility allowed by the different control policies is shown in Fig. 5.4-C. Each bar in the histogram represents the share of time during which the population of each municipality maintains a given level of mobility/economic activity corresponding to different ranges of p values. This plot highlights the effect that was

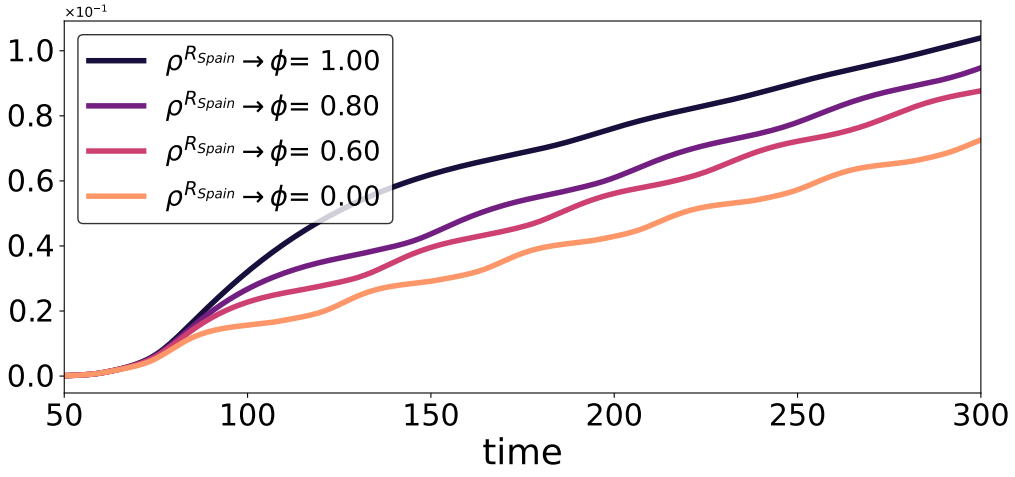


Figure 5.5: **The overall effect of the reactive policy**. Time evolution of the fraction of recovered individuals in the whole contry ρ^{Spain} for different permeability values.

observed in the evolution of $\langle p \rangle(t)$ in the different panels of Figs. 5.2 (green lines): short and strict control policies are more beneficial in economic terms than long and lax interventions. This result shows that it is possible to strike a trade-off between controlling an outbreak while minimizing the economic cost of the intervention by implementing stringent measures at the local level.

5.5 The influence of social structures on the stringency of interventions

Once we have understood that the stringency of local control policies plays a decisive role in the success of multiscale NPIs implemented, we now explore the interplay between a particular mobility network of a certain population and the functionality of the aforementioned local measures. In other words, we aim to answer how the interconnectivity between the different urban areas affects the outcome of the reactive policies for different permeability values. To do so, we divide and separate the different highest administrative divisions of Spain, called autonomous communities. It is relevant to mention that, by dividing the Spanish territory, we do not lose much information, since only the 2% of the total volume of the daily flows involve movements between different autonomous communities. From a political point of view, this administrative separation is also relevant for the design and planning of containment policies in Spain, given that health competencies, and therefore the deployment of control policies and the distribution of health resources, are in the hands of these territories.

To get some insights into the role of the mobility network, we reproduce the analysis

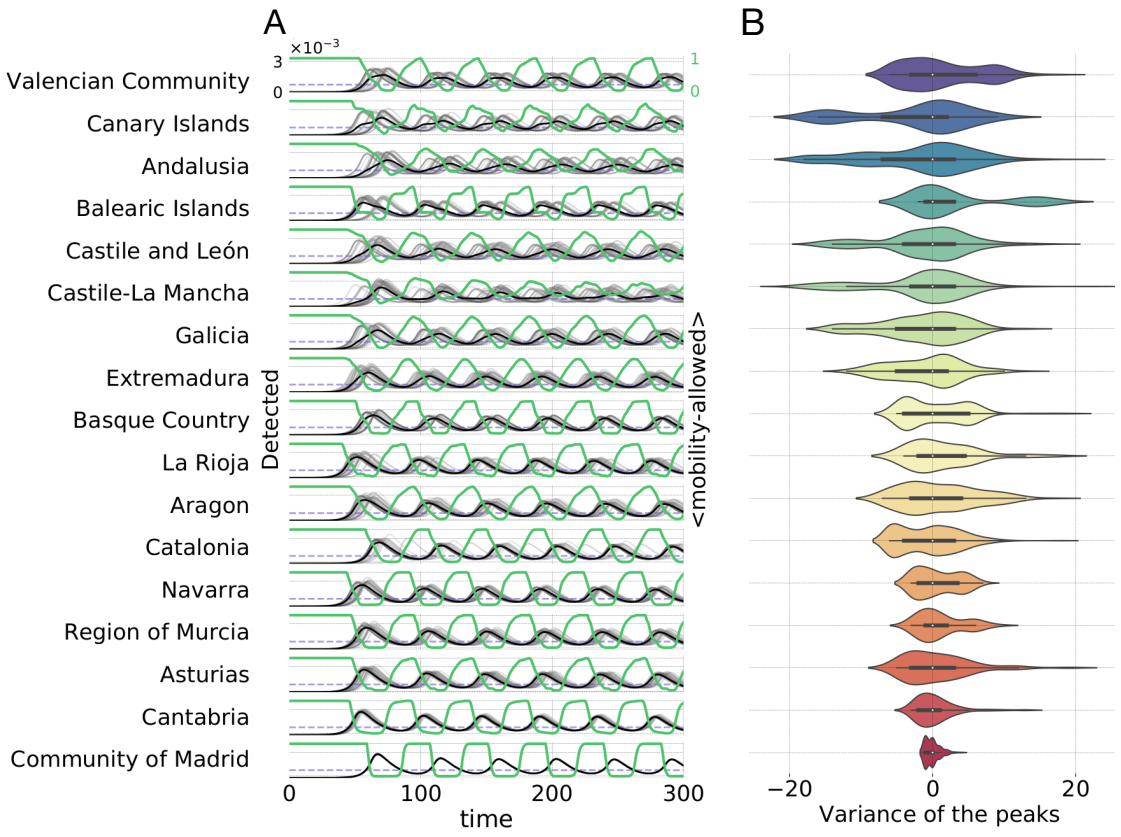


Figure 5.6: **Epidemic wave synchronization for different metapopulation structures**. A. Time evolution of the fraction of infected and detected agents for $\phi = 0$, the grey lines correspond to the 50 most populated municipalities of each autonomous community, while the black curve accounts for the average in each community. The green lines represent the average mobility in each autonomous community. B. The violin plots represent the distribution of the times at which each municipality reaches its peak in each autonomous community. Note that the origin of the curve has been displaced to match the average of these distributions.

shown in Figs. 5.2 by now focusing on each individual autonomous community. To this aim, we set $\phi = 0$, i.e. perfect household isolation, and place an infectious seed in the most populous city. In Fig. 5.6-A we represent the spatio-temporal distribution of detected cases in each territory along with the average socioeconomic activity level. Interestingly, we observe important heterogeneities concerning the shape of the different epidemic waves in each territory: while autonomous communities such as Cantabria and the Community of Madrid are characterized by very synchronized epidemic curves, triggering a sharp coherent collective evolution, the outbreaks in municipalities located in other areas such as Canary Island display important time lags leading to less coherent epidemic waves. These heterogeneities become more evident in Fig. 5.6-B, where we represent the time distribution of the position of the first epidemic peak inside each autonomous community, finding

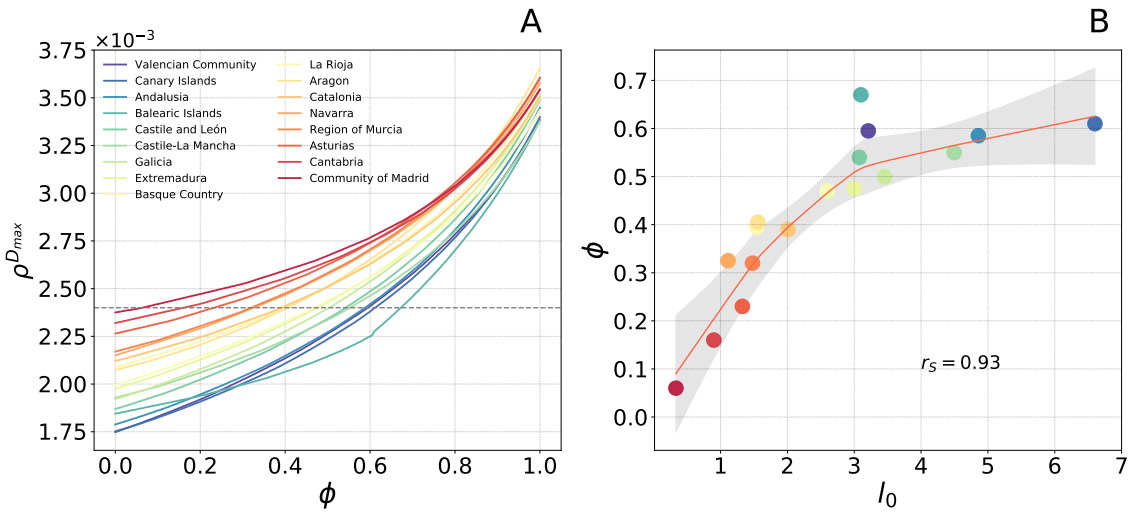


Figure 5.7: **Permeability, saturation of the health system & structural correlations**. A. The maximum number of detected individuals during the first epidemic wave $\rho^{D^{max}}$ as a function of the social permeability ϕ allowed for each autonomous region. The dashed line indicates the threshold used, 2.4×10^{-3}, to calculate the permeability that can be afforded by each territory. B. Correlation between the affordable permeability of each community ϕ^{aff} and the average shortest contagion path connecting a randomly picked individual with the infection source in each territory l_0. Solid red line shows the LOESS regression of the data whereas the shadowed region showed the 95% confidence interval obtained by bootstrapping 1000 samples.

important differences in terms of their variance.

From a systemic perspective, the synchronization of the individual epidemic curves has deep implications for the temporal evolution of the overall number of cases in the entire metapopulation. To check this hypothesis, we represent in Fig. 5.7-A how the magnitude of the peak of individuals detected in the first epidemic wave evolves as a function of the social permeability allowed for each autonomous community. Interestingly, we observe how this indicator, somehow related to the highest expected burden to the health system over the course of the disease, strongly varies between the different territories of Spain. Specifically, the affordable permeability values to keep the cumulative incidence over 14 days below 240 cases per 10^5 inhabitants range from $\phi^{aff} = 0.06$ for the Community of Madrid to $\phi^{aff} = 0.67$, corresponding to the Balearic Islands.

The heterogeneity between different territories observed in the former analysis reveals that the different scales governing the reaction-diffusion processes that cause the spatial diffusion of an epidemic outbreak are highly intertwined. As a result, the outcome of the local policies, such as the modulation of the economic activity regulated by $p(t)$ or the social bubbles mechanism governed by ϕ, are highly influenced by global aspects such as the demographic distribution of the population or how they move across a given territory. Qualitatively, we observe a coherent epidemic wave inside those territories requiring

more strict policies (smaller ϕ^{aff} values). In contrast, the municipalities located in those autonomous communities for which loose policies are affordable exhibit an asynchronous evolution. As a consequence, the internal policies are activated at different times across the territory, allowing a progressive control of the outbreak, which results in the flattening of the aggregated epidemic curve and explains the lower magnitude of the peak.

Finally, we aim at finding a heuristic argument to connect the features of the underlying metapopulation with the affordable permeability values for each territory. Providing an initial infectious individual is seeded in a single patch, the degree of coherence observed for the time evolution of the ensemble of municipalities in each territory has to be related to the communication of the source of infection with the rest of the network through the population movements. To find such relation, we first define a new matrix $\mathbf{M}$ whose elements M_{ij} determines the expected number of interactions of one individual placed in patch i with the population living in patch j. Specifically, under the model assumptions, these elements read [111],

$$M_{ij} = \sum_{l=1}^{N_P} R_{il} f_l \frac{n_j R_{jl}}{\tilde{n}_l} \ . \tag{5.21}$$

In dynamical terms, it seems logical to assume that a strong interaction between two patches, as dictated by the matrix $\mathbf{M}$, boosts the diffusion of infectious individuals across them. Therefore, we can define a social proximity measure d_{ij} between two patches i and j as,

$$d_{ij} = M_{ij}^{-1} \ . \tag{5.22}$$

Note that the elements d_{ij} somehow reflect the time taken by an epidemic outbreak, seeded in patch i, to invade patch j. Our hypothesis is that the degree of synchronization of the epidemic curves from the different autonomous communities, and consequently their affordable permeability, can be inferred from the structural features of the matrix $\mathbf{d}$. Namely, shorter effective distances from every single node of the metapopulation to the source of the outbreak should give rise to a higher degree of synchronization. Following this rationale, we define our structural metric l_0 as the average shortest path length connecting a randomly picked individual with the source of an outbreak over the proximity matrix $\mathbf{d}$. Accordingly,

$$l_0 = \frac{\sum_{i=1}^{N_P} n_i l_{i0}}{\sum_{i=1}^{N_P} n_i} \ , \tag{5.23}$$

where l_{i0} constitutes the shortest path length connecting the location i with the source of the outbreak in the proximity matrix $\mathbf{d}$.

Considering the former definitions in Fig. 5.7-B we represent the affordable permeability ϕ^{aff} as a function of the value of l_0 for each autonomous community. In this plot we

observe a strong positive correlation between both indicators. Moreover, the LOESS (Locally Estimated Scatterplot Smoothing [163]) regression of the data yields a monotonically increasing function relating ϕ^{aff} and l_0. Taken together, these results indicate that regions with dispersed mobility networks may not require local control policies as strict as densely communicated territories.

5.6 Conclusions

In this chapter, we have introduced a metapopulation model in which restrictions on social activity depend on the epidemiological evolution. In this sense, the model allows the (local) social fabric to be restored to normal levels once the epidemiological situation manages to fall below a tolerance threshold, in attempt of achieving the difficult trade-off between the social functioning and the relief of overwhelmed health-care systems. The contention measures act at multiple scales: at the global level, mobility decreases to and from high-incidence areas, at each patch the economic activity decreases and, at the household level, social bubbles are created, shielding the susceptible living with non-infectious individuals.

The effectiveness of the former measures depends on both their severity and the social adhesion to them. This way, our analyses have focused on how the stringency of measures at the local level influences the ability to control epidemic waves. In this regard, our work yields three main results. First, as expected, a higher stringency of local measures facilitates the control provided by the policies implemented at larger scales, resulting in a train of epidemic waves manageable from a health point of view rather than a huge first epidemic outbreak. Secondly, we note how, in addition to serving better outbreak management, the implementation of stringent measures counter-intuitively provides less disruption to the socioeconomic fabric as societies can afford more periods of normal activity than in the case of lax implementation of local measures.

Finally, the third result concerns the influence that different social architectures (*i.e.* the demographic spatial distribution and the human mobility network) has on the severity of local rules needed to achieve a given control of the epidemic. By focusing on the different autonomous communities of Spain, we have shown that their different architectures are determining (decisive) for the outcome of the spreading behavior and its control. In particular, while some autonomous communities should be more aggressive with local restrictions, forcing household isolation to more than 95% of the population, others may allow a certain degree of local activity as they only need that around 35% of individuals are inside safe social bubbles. To explain the heterogeneity in the stringency of local restrictions needed, and taking a similar avenue to that followed in Chapter 4, we have relied on a new data-driven metric that combines the different aspects of social mixing that interplay in a metapopulation. In particular, by computing the average contagion distance

between the infected seed and any potential susceptible individual, we have shown that those autonomous communities for which this distance is small are the ones that need more stringent measures at the local level. Naturally, this result is reflected in the synchronous epidemic dynamics observed for territories with a small contagion distance: the more synchronized epidemic waves of individual patches are, the more restrictive local measures should be. Nonetheless, the economic activities in the synchronized autonomous communities get benefited. Thus, there is a trade-off that may be beneficial in any situation.

Chapter 6

Conclusions

Communicable diseases are taking advantage of our increasingly interconnected world to rapidly spread on a global scale with unprecedented ease. Population growth, rapid urbanization, climate change, and our increased mobility at all scales are creating a complex new battleground in the fight against new and old diseases. For this reason, in order to be able to face, and even anticipate, epidemic outbreaks, we must be able to develop tools that, given the various factors involved in the spread of a pathogen, allow us to identify which containment strategies are the most appropriate given the epidemic risk, or how and where to implement them so that we make effective use of the available means.

As we have seen throughout this thesis, epidemic modeling allows us to study the properties of the spread of an infectious disease by drawing statistical scenarios in different circumstances. Furthermore, epidemic modeling not only takes advantage of our understanding of the mechanisms underlying the spread of contagion of a specific infectious disease, but also allows us to introduce the patterns of human behavior involved in the transmission of infectious diseases. In fact, access to large databases of human interactions and mobility patterns has, in recent years, brought about a paradigm shift in the way we model epidemics. Consequently, we have been able to develop forecasting frameworks with greater spatio-temporal reliability and, above all, opening the door to the evaluation of non-pharmacological control strategies, i.e., based on intervention on human behavioral patterns.

Despite great efforts to understand propagation phenomena, many fundamental and technical challenges and, of course, limitations remain to be overcome. In this thesis, we have designed data-driven models that aim at projecting scenarios under different control intervention strategies and, above all, have allowed us to shed some light on how different aspects of our human behavior intervene in the dissemination of pathogens. Taking advantage of this latter aspect allows us to use control strategies that directly influence human behavior more effectively.

Following this goal, we have developed data-driven models that enable to analyze the mitigating effects of control interventions at different scales, from the creation of social bubbles at the household level to mobility at a regional scale. In the following, we review the main results derived from the frameworks developed in this thesis.

In Chapter 3, we have studied the effects of contact tracing (CT) in social networks. The theoretical framework was proposed as a competition between CT chains and the spread of the pathogen, triggered by the identification of symptomatic individuals. Once the CT system is activated, a wave of detection is generated that eliminates the transmission chains of the pathogen. We have shown, that, while the sole symptomatic detection flattens the curve causing a delay in the epidemic curve, the macroscopic effect of CT is reversing the trend of the epidemic incidence by bending it. In addition, we have shown the microscopic differences between symptomatic detection and CT. As hubs are more likely to be infected, they are also more likely to be detected by CT dynamics. Thus, the effectiveness of CT relies on the early detection of super-spreaders and silent infections. We have also found that mitigation strategies might completely fail in the scarcity of resources, showing that an overwhelmed system could lead to the same result where no control strategy is implemented.

In Chapter 4, we present a theoretical framework that encompasses the interplay between human mobility, epidemiological features of vector-borne diseases, and demographics. Combining all these ingredients into a metapopulation model has allowed us to portray the vulnerability of geographical areas to vector-borne diseases. This means that we can estimate the effect of an epidemic outbreak, the so-called epidemic risk, in endemic areas before they occur. We have also shown that epidemic outcomes are highly dependent on recurrent mobility patterns and disease features such as asymptomaticity. Our findings indicate that, although mobility might be beneficial to disease propagation, the interplay of asymptomaticity and mobility have a detrimental effect on the epidemic outcome. Furthermore, an analysis of the spectral properties of the mixing matrix has allowed us to develop an immunization ranking. Importantly, this analysis should be recurrent since the immunization of one area confers partial immunity to other patches that are interconnected through human mobility, thus modifying the spectrum of the mixing matrix.

In Chapter 5, we have proposed a framework that takes into account human mobility patterns, demographics, and epidemiological features. In this model, a reactive control policy is integrated, being activated or deactivated as the extent of the epidemic outbreak reaches a certain threshold. This threshold represents the health system capacity and the control strategy directly acts on mobility restrictions imposed to avoid close contacts as much as possible. The model relies on three control scales: selective global restrictions on travel from and to the affected place, selective mesoscopic restrictions that reduce internal mobility of high incidence areas, and the microscopic social reaction to restrictions. We

have shown that the implementation of reactive lockdowns would help the health system to cope while restructuring the socio-economic fabric in a fast way. Furthermore, we have found that highly restrictive social bubbles not just 'maximize' the effectiveness of the control strategy in the final outcome of the epidemic, but allow to ease the control measures for longer periods of time, being more beneficial to the economic system than less restrictive policies. Another interesting result found in this chapter has come from the comparison of different autonomous regions in Spain. We have shown that the severity of local control strategies depends on the precise network structure of human mobility flows within each specific region.

In summary, we have developed three different frameworks integrating control strategies that, despite working at different scales, aim at reshaping human interactions to control epidemic outbreaks. Apart from the derived results, the frameworks developed are examples of how complex systems thinking can be used to guide our future understanding of how infectious diseases spread in our modern interconnected world, in which health policies cannot ignore the impact on human behavior and the economy of nations.

Conclusiones

Las enfermedades transmisibles se aprovechan de nuestro mundo cada vez más interconectado para propagarse rápidamente a escala mundial con una facilidad sin precedentes. El crecimiento de la población, la rápida urbanización, el cambio climático y nuestra creciente movilidad a todas las escalas, están creando un nuevo y complejo campo de batalla en la lucha contra enfermedades contagiosas, nuevas y antiguas. Es por ello que, para poder hacer frente a los brotes epidémicos, e incluso anticiparnos a ellos, necesitamos desarrollar herramientas que, dados los diversos factores que intervienen en su propagación, nos ayuden a determinar qué estrategias de contención son más eficaces en función de sobre el riesgo de epidemia, o cómo y dónde implementarlos para que podamos maximizar los recursos disponibles.

Como hemos visto a lo largo de esta tesis, el modelado de epidemias nos permite estudiar las propiedades de propagación de una enfermedad infecciosa, dibujado escenarios estadísticos bajo diferentes circunstancias. Los modelos epidemiológicos no solo aprovechan la comprensión que se tiene de los mecanismos de contagio, si no que también es posible introducir los patrones de comportamiento humano involucrados en la transmisión. De hecho, el acceso a grandes bases de datos de interacciones humanas y patrones de movilidad ha provocado, en los últimos años, un cambio de paradigma en la forma en que modelamos las epidemias. En consecuencia, hemos podido desarrollar marcos de previsión con mayor fiabilidad espacio-temporal y, sobre todo, abriendo la puerta a la evaluación de estrategias de control no farmacológicas, es decir, basadas en la intervención sobre patrones de comportamiento humano.

A pesar de los grandes esfuerzos para comprender los fenómenos de propagación, quedan por superar muchos desafíos técnicos y fundamentales y, por supuesto, limitaciones. En esta tesis hemos diseñado modelos basados en datos que pretenden proyectar escenarios en diferentes estrategias de intervención o control y, sobre todo, nos han permitido esclarecer cómo intervienen distintos aspectos de nuestro comportamiento humano en la diseminación de patógenos. Aprovechar este último aspecto nos permite utilizar estrategias de control que influyen directamente en el comportamiento humano de manera más efectiva.

Siguiendo este objetivo, hemos desarrollado modelos basados en datos que permiten analizar los efectos mitigadores de intervenciones de control a diferentes escalas, desde la creación de burbujas sociales a nivel del hogar hasta la movilidad a escala regional. A continuación, revisamos los principales resultados derivados de los marcos desarrollados en esta tesis.

En el Capítulo 3, hemos estudiado los efectos del rastreo de contactos (RC) en redes sociales. El marco teórico se planteó como una competencia entre cadenas del RC y la propagación del patógeno, desencadenada por la identificación de individuos sintomáticos. Una vez activado el sistema de RC, se genera una onda de detección que elimina las cadenas de transmisión del patógeno. Hemos demostrado que, mientras que la sola detección sintomática aplana la curva provocando un retraso en máximo disminuido, el efecto macroscópico del RC invierte 'rápidamente' la tendencia de la incidencia epidémica al doblarla. Además, hemos mostrado las diferencias microscópicas entre la detección sintomática y el RC. Como es más probable que los hubs se infecten, también es más probable que la dinámica del RC los detecte. Por lo tanto, la efectividad del RC se basa en la detección temprana de super propagadores e infecciones silenciosas. También hemos encontrado que las estrategias de mitigación pueden fallar por completo ante la escasez de recursos, lo que demuestra que un sistema abrumado podría conducir al mismo resultado donde no se implementa una estrategia de control.

En el Capítulo 4, presentamos un marco teórico que abarca la interacción entre la movilidad humana, las características epidemiológicas de las enfermedades transmitidas por vectores y la demografía. La combinación de todos estos ingredientes en un modelo de metapoblación nos ha permitido retratar la vulnerabilidad de las áreas geográficas a las enfermedades transmitidas por vectores. Esto significa que podemos estimar el efecto de un brote epidémico, el llamado riesgo epidémico, en áreas endémicas antes de que ocurra. También hemos demostrado que los resultados de la epidemia dependen en gran medida de los patrones de movilidad recurrentes y de las características de la enfermedad, como la asintomaticidad. Nuestros hallazgos indican que, aunque la movilidad podría ser beneficiosa para la propagación de la enfermedad, la interrelación de la asintomaticidad y la movilidad tiene un efecto benéfico en el resultado de la epidemia. Además, un análisis de las propiedades espectrales de la matriz de mezcla nos ha permitido desarrollar un ranking de inmunización. Es importante destacar que este análisis debe ser recurrente ya que la inmunización de un área confiere inmunidad parcial a otros parches que se encuentran interconectados a través de la movilidad humana, modificando así el espectro de la matriz de mezcla.

En el Capítulo 5, hemos propuesto un marco que tiene en cuenta los patrones de movilidad humana, la demografía y las características epidemiológicas. En este modelo se integra una política de control reactivo, activándose o desactivándose a medida que la extensión

del brote epidémico alcanza un umbral determinado. Este umbral representa la capacidad del sistema de salud. La estrategia de control actúa directamente sobre las restricciones de movilidad impuestas para evitar en lo posible los contactos estrechos. El modelo se basa en tres escalas de control: restricciones globales selectivas en los viajes desde y hacia el lugar afectado, restricciones mesoscópicas selectivas que reducen la movilidad interna de las áreas de alta incidencia y la reacción social microscópica a las restricciones. Hemos demostrado que la implementación de bloqueos reactivos ayudaría al sistema de salud a hacer frente mientras se reestructura el tejido socioeconómico de manera rápida. Además, hemos encontrado que las burbujas sociales altamente restrictivas no solo 'maximizan' la efectividad de la estrategia de control en el resultado final de la epidemia, sino que permiten suavizar las medidas de control por períodos de tiempo más largos, siendo más beneficiosas para el sistema económico que políticas menos restrictivas. Otro resultado interesante encontrado en este capítulo proviene de la comparación de diferentes comunidades autónomas en España. Hemos demostrado que la severidad de las estrategias de control local depende de la estructura de red precisa de los flujos de movilidad humana dentro de cada región específica.

En resumen, hemos desarrollado tres marcos diferentes que integran estrategias de control que, a pesar de trabajar a diferentes escalas, tienen como objetivo remodelar las interacciones humanas para controlar los brotes epidémicos. Además de los resultados derivados, los marcos desarrollados son ejemplos de cómo se puede utilizar el pensamiento sistémico complejo para guiar nuestra comprensión futura de cómo se propagan las enfermedades infecciosas en nuestro mundo interconectado moderno, en el que las políticas de salud no pueden ignorar el impacto en el comportamiento humano y la economía de naciones.

Appendix

A

In this appendix, we complete the derivation presented in Section 3.4.1, for the calculation of the effective reproduction number $\mathcal{R}^c(t)$ (Eqs.3.20 and 3.22) from the Markovian equations. To this aim, we show the calculation of the probability that an individual i infected at time t spend times τ_E, τ_P, and τ_S or τ_A in compartments E, P, and I_A or I_A respectively. Considering the Markovian equations the probability for those transiting the symptomatic phase is:

$$\begin{aligned}
\mathcal{P}_i(\tau_E,\tau_P,\tau_S|t) &= (1-\eta)^{\tau_E-1}\eta \\
&\times \left\{\delta_{\tau_S,0}\cdot\Pi_i^{CT}(t+\tau_E+\tau_P)+(1-\delta_{\tau_S,0})\cdot\alpha\left(1-\Pi_i^{CT}(t+\tau_E+\tau_P)\right)\right\} \\
&\times \prod_{s=t+\tau_E+1}^{t+\tau_E+\tau_P-1}(1-\alpha)\left[1-\Pi_i^{CT}(s)\right] \\
&\times \left\{\delta_{\tau_S,0}+(1-\delta_{\tau_S,0})\prod_{s=t+\tau_E+\tau_P+1}^{t+\tau_E+\tau_P+\tau_S-1}(1-\mu)(1-\delta)\left[1-\Pi_i^{CT}(s)\right]\right. \\
&\times \left.\left\{\delta+(1-\delta)\left[\Pi_i^{CT}(t+\tau_T)+\left(1-\Pi_i^{CT}(t+\tau_T)\right)\mu\right]\right\}\right\}, \qquad \text{(A.1)}
\end{aligned}$$

while for those asymptomatic reads:

$$\begin{aligned}
\mathcal{P}_i(\tau_E,\tau_P,\tau_A|t) &= (1-\eta)^{\tau_E-1}\eta \\
&\times \left\{\delta_{\tau_A,0}\cdot\Pi_i^{CT}(t+\tau_E+\tau_P)+(1-\delta_{\tau_A,0})\cdot\alpha\left(1-\Pi_i^{CT}(t+\tau_E+\tau_P)\right)\right\} \\
&\times \prod_{s=t+\tau_E+1}^{t+\tau_E+\tau_P-1}(1-\alpha)\left[1-\Pi_i^{CT}(s)\right] \\
&\times \left\{\delta_{\tau_A,0}+(1-\delta_{\tau_A,0})\left[\Pi_i^{CT}(t+\tau_T)+\left(1-\Pi_i^{CT}(t+\tau_T)\right)\mu\right]\right\} \\
&\times \prod_{s=t+\tau_E+\tau_P+1}^{t+\tau_E+\tau_P+\tau_A-1}(1-\mu)\left[1-\Pi_i^{CT}(s)\right], \qquad \text{(A.2)}
\end{aligned}$$

where τ_T is defined as the total duration of the infectious period of the infected agent, *i.e.* $\tau_T=\tau_E+\tau_P+\tau_S$ for symptomatic individuals and $\tau_T=\tau_E+\tau_P+\tau_A$ for asymptomatic

patients. In both expressions $\Pi_i^{CT}(t)$ is the probability of being detected through CT and is equal to $\Pi_i^{P\to D}(t)$ and $\Pi_i^{I_A\to D}(t)$ as written in Eq. (10).

Computing Eqs. (A.1) and (A.2) requires to save the time evolution along the epidemic trajectory of the following quantities: $\Pi_i^{CT}(t)$ and $\rho_i^S(t)$ for all the nodes (i = 1,...,N). These two sets of quantities, together with the evolution of the infection probabilities of each node $\Pi_i^{S\to E}(t)$, allow us to obtain the effective case-reproduction number, $\mathcal{R}^c(t)$ in Eq. (3.20).

To illustrate the validity of the former expressions, lets suppose that CT is absent, so that $\Pi_i^{CT}(t) = 0\ \forall\ t$ and i. Then the conditional probabilities are identical for all the nodes:

$$\begin{aligned}\mathcal{P}(\tau_E,\tau_P,\tau_S|t) &= [\delta+(1-\delta)\mu]\,(1-\mu)^{\tau_S-1}(1-\delta)^{\tau_S-1}\\ &\times\ \alpha(1-\alpha)^{\tau_P-1}(1-\eta)^{\tau_E-1}\eta\,,\end{aligned}\tag{A.3}$$

and

$$\begin{aligned}\mathcal{P}(\tau_E,\tau_P,\tau_A|t) &= \mu(1-\mu)^{\tau_A-1}\alpha(1-\alpha)^{\tau_P-1}\\ &\times\ (1-\eta)^{\tau_E-1}\eta\,,\end{aligned}\tag{A.4}$$

and the factorization of the two probabilities and their time independence becomes clear.

Bibliography

[1] P. W. Anderson, Science **177**, 393 (1972).

[2] E. N. Lorenz, Journal of Atmospheric Sciences **20**, 130 (1963).

[3] F. Dyson, Science **200**, 677 (1978).

[4] P. Bak, C. Tang, and K. Wiesenfeld, Physical Review Letters **59**, 381 (1987).

[5] P. Bak, C. Tang, and K. Wiesenfeld, Physical Review A **38**, 364 (1988).

[6] N. H. Packard, J. P. Crutchfield, J. D. Farmer, and R. S. Shaw, Physical Review Letters **45**, 712 (1980).

[7] R. M. May, in *The Theory of Chaotic Attractors* (Springer, 2004), pp. 85–93.

[8] Z. Olami, H. J. S. Feder, and K. Christensen, Physical Review Letters **68**, 1244 (1992).

[9] B. Drossel and F. Schwabl, Physical Review Letters **69**, 1629 (1992).

[10] R. M. Anderson and R. M. May, *Infectious diseases of humans: dynamics and control* (Oxford University Press, 1992).

[11] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[12] J. Piret and G. Boivin, Frontiers in microbiology **11**, 631736 (2021).

[13] S. Correia, S. Luck, and E. Verner, Public Health Interventions do not: Evidence from the (1918).

[14] N. Madhav, B. Oppenheim, M. Gallivan, P. Mulembakani, E. Rubin, and N. Wolfe (2018).

[15] F. Brauer, Infectious Disease Modelling **2**, 113 (2017), ISSN 2468-0427.

[16] D. J. Gubler, Archives of Medical Research **33**, 330 (2002).

[17] A. T. Price-Smith, *Plagues and politics: Infectious Disease and International Policy* (Springer, 2001).

[18] N. LePan, *Visualizing the history of pandemics*, `https://www.visualcapitalist.com/history-of-pandemics-deadliest/`, accessed: 2022-06-21.

[19] V. Colizza, M. Barthélemy, A. Barrat, and A. Vespignani, Comptes Rendus Biologies **330**, 364 (2007).

[20] G. P. Wormser and B. Pourbohloul, *Modeling infectious diseases in humans and animals by matthew james keeling and pejman rohani princeton, nj: Princeton university press, 2008. 408 pp.* (2008).

[21] R. Pastor-Satorras, C. Castellano, P. Van Mieghem, and A. Vespignani, Reviews of Modern Physics **87**, 925 (2015).

[22] W. Cota, D. Soriano-Paños, A. Arenas, S. C. Ferreira, and J. Gómez-Gardeñes, New Journal of Physics **23**, 073019 (2021).

[23] D. Soriano-Paños, W. Cota, S. C. Ferreira, G. Ghoshal, A. Arenas, and J. Gómez-Gardeñes, Annalen der Physik p. 2100482 (2022).

[24] J. Snow, John Churchill, New Burlington Street, London, England (1855).

[25] D. Bernoulli, Histoire de l'Acad., Roy. Sci.(Paris) avec Mem pp. 1–45 (1760).

[26] W. H. Hamer, *The milroy lectures on epidemic diseases in england: The evidence of variability and of persistency of type; delivered before the royal college of physicians of london, march 1st, 6th, and 8th, 1906* (Bedford Press, 1906).

[27] R. Ross, Nature **87**, 466 (1911).

[28] R. Ross, *The prevention of malaria* (John Murray, 1911).

[29] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **115**, 700 (1927).

[30] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **138**, 55 (1932).

[31] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **141**, 94 (1933).

[32] R. Pastor-Satorras and A. Vespignani, Physical Review Letters **86**, 3200 (2001).

[33] P. Erdős, A. Rényi, et al., Publ. Math. Inst. Hung. Acad. Sci **5**, 17 (1960).

[34] A. D. Broido and A. Clauset, Nature Communications **10**, 1 (2019).

[35] R. Albert and A.-L. Barabási, Reviews of Modern Physics **74**, 47 (2002).

[36] T. Opsahl, F. Agneessens, and J. Skvoretz, Social Networks **32**, 245 (2010).

[37] W. Wang, M. Tang, H. E. Stanley, and L. A. Braunstein, Reports on Progress in Physics **80**, 036603 (2017).

[38] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical processes on complex networks* (Cambridge University Press, 2008).

[39] J. T. Matamalas, A. Arenas, and S. Gómez, Science Advances **4**, eaau4212 (2018).

[40] B. Karrer and M. E. Newman, Physical Review E **82**, 016101 (2010).

[41] A. S. Mata and S. C. Ferreira, EPL (Europhysics Letters) **103**, 48003 (2013).

[42] L. A. Braunstein, Z. Wu, Y. Chen, S. V. Buldyrev, T. Kalisky, S. Sreenivasan, R. Cohen, E. López, S. Havlin, and H. E. Stanley, International Journal of Bifurcation and Chaos **17**, 2215 (2007).

[43] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, EPL (Europhysics Letters) **89**, 38009 (2010).

[44] D. Brockmann and D. Helbing, Science **342**, 1337 (2013).

[45] H. Barbosa, M. Barthelemy, G. Ghoshal, C. R. James, M. Lenormand, T. Louail, R. Menezes, J. J. Ramasco, F. Simini, and M. Tomasini, Physics Reports **734**, 1 (2018).

[46] G. K. Zipf, American Sociological Review **11**, 677 (1946).

[47] S. A. Stouffer, American Sociological Review **5**, 845 (1940), ISSN 00031224.

[48] F. Simini, M. C. González, A. Maritan, and A.-L. Barabási, Nature **484**, 96 (2012).

[49] K. Prem, A. R. Cook, and M. Jit, PLoS computational biology **13**, e1005697 (2017).

[50] S. office of the European Union, *Passenger mobility statistics* (2021), URL `https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Passenger_mobility_statistics`.

[51] J. Gómez-Gardeñes, D. Soriano-Paños, and A. Arenas, Nature Physics **14**, 391 (2018).

[52] M. J. Keeling and P. Rohani, in *Modeling infectious diseases in humans and animals* (Princeton university press, 2011).

[53] G. Chowell, J. M. Hyman, L. M. Bettencourt, C. Castillo-Chavez, and H. Nishiura, *Mathematical and statistical estimation approaches in epidemiology* (Springer, 2009).

[54] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, and A. B. Gumel, Infectious Disease Modelling **5**, 293 (2020), ISSN 2468-0427.

[55] R. O. J. H. Stutt, R. Retkute, M. Bradley, C. A. Gilligan, and J. Colvin, Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences **476**, 20200376 (2020).

[56] K. Prem, Y. Liu, T. W. Russell, A. J. Kucharski, R. M. Eggo, N. Davies, S. Flasche, S. Clifford, C. A. Pearson, J. D. Munday, et al., The Lancet Public Health **5**, e261 (2020).

[57] P. Block, M. Hoffman, I. J. Raabe, J. B. Dowd, C. Rahal, R. Kashyap, and M. C. Mills, Nature Human Behaviour **4**, 588 (2020).

[58] C. M. Peak, L. M. Childs, Y. H. Grad, and C. O. Buckee, Proceedings of the National Academy of Sciences **114**, 4023 (2017).

[59] J. O. Lloyd-Smith, A. P. Galvani, and W. M. Getz, Proceedings of the Royal Society of London. Series B: Biological Sciences **270**, 1979 (2003).

[60] M. Serafino, H. S. Monteiro, S. Luo, S. D. S. Reis, C. Igual, A. S. L. Neto, M. Travizano, J. S. Andrade, and H. A. Makse, PLOS Computational Biology **18**, e1009865 (2022).

[61] S. Kojaku, L. Hébert-Dufresne, E. Mones, S. Lehmann, and Y.-Y. Ahn, Nature Physics **17**, 652 (2021).

[62] G. Bianconi, H. Sun, G. Rapisardi, and A. Arenas, Physical Review Research **3**, L012014 (2021).

[63] A. Barrat, C. Cattuto, M. Kivelä, S. Lehmann, and J. Saramäki, Journal of The Royal Society Interface **18**, 20201000 (2021).

[64] M. J. Keeling, T. D. Hollingsworth, and J. M. Read, Journal of Epidemiology & Community Health **74**, 861 (2020), ISSN 0143-005X.

[65] J. A. Firth, J. Hellewell, P. Klepac, S. Kissler, M. Jit, K. E. Atkins, S. Clifford, C. J. Villabona-Arenas, S. R. Meakin, C. Diamond, et al., Nature Medicine **26**, 1616 (2020).

[66] J. Hellewell, S. Abbott, A. Gimma, N. I. Bosse, C. I. Jarvis, T. W. Russell, J. D. Munday, A. J. Kucharski, W. J. Edmunds, F. Sun, et al., The Lancet Global Health **8**, e488 (2020).

[67] M. E. Kretzschmar, G. Rozhnova, M. C. J. Bootsma, M. van Boven, J. H. H. M. van de Wijgert, and M. J. M. Bonten, The Lancet Public Health **5**, e452 (2020).

[68] C. M. Peak, R. Kahn, Y. H. Grad, L. M. Childs, R. Li, M. Lipsitch, and C. O. Buckee, The Lancet Infectious Diseases **20**, 1025 (2020).

[69] A. J. Kucharski, P. Klepac, A. J. K. Conlan, S. M. Kissler, M. L. Tang, H. Fry, J. R. Gog, W. J. Edmunds, J. C. Emery, G. Medley, et al., The Lancet Infectious Diseases **20**, 1151 (2020).

[70] A. Aleta, D. Martín-Corral, A. Pastore y Piontti, M. Ajelli, M. Litvinova, M. Chinazzi, N. E. Dean, M. E. Halloran, I. M. Longini Jr, S. Merler, et al., Nature Human Behaviour **4**, 964 (2020).

[71] B. Karrer and M. E. Newman, Physical Review E **84**, 036106 (2011).

[72] C. Granell, S. Gómez, and A. Arenas, Physical Review Letters **111**, 128701 (2013).

[73] J. Sanz, C.-Y. Xia, S. Meloni, and Y. Moreno, Physical Review X **4**, 041005 (2014).

[74] F. Darabi Sahneh and C. Scoglio, Physical Review E **89**, 062817 (2014).

[75] O. Kogan, M. Khasin, B. Meerson, D. Schneider, and C. R. Myers, Physical Review E **90**, 042149 (2014).

[76] C. Poletto, S. Meloni, A. Van Metre, V. Colizza, Y. Moreno, and A. Vespignani, Scientific Reports **5**, 7895 (2015).

[77] L. Hébert-Dufresne and B. M. Althouse, Proceedings of the National Academy of Sciences **112**, 10551 (2015).

[78] D. Soriano-Paños, F. Ghanbarnejad, S. Meloni, and J. Gómez-Gardeñes, Physical Review E **100**, 062308 (2019).

[79] E. Estrada, Physics Reports **869**, 1 (2020).

[80] A. Vespignani, H. Tian, C. Dye, J. O. Lloyd-Smith, R. M. Eggo, M. Shrestha, S. V. Scarpino, B. Gutierrez, M. U. G. Kraemer, J. Wu, et al., Nature Reviews Physics **2**, 279 (2020).

[81] E. H. Y. Lau and G. M. Leung, Nature Medicine **26**, 1534 (2020).

[82] *Covid-19 pandemic planning scenarios*, `https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html`, accessed: July 1, 2020.

[83] K. K.-W. To, O. T.-Y. Tsang, W.-S. Leung, A. R. Tam, T.-C. Wu, D. C. Lung, C. C.-Y. Yip, J.-P. Cai, J. M.-C. Chan, T. S.-H. Chik, et al., The Lancet Infectious Diseases **20**, 565 (2020).

[84] S. Gómez, A. Arenas, J. Borge-Holthoefer, S. Meloni, and Y. Moreno, EPL (Europhysics Letters) **89**, 38009 (2010).

[85] B. Guerra and J. Gómez-Gardeñes, Physical Review E **82**, 035101 (2010).

[86] S. Gómez, J. Gómez-Gardeñes, Y. Moreno, and A. Arenas, Physical Review E **84**, 036105 (2011).

[87] P. Vanhems, A. Barrat, C. Cattuto, J.-F. Pinton, N. Khanafer, C. Régis, B.-a. Kim, B. Comte, and N. Voirin, PloS One **8**, e73970 (2013).

[88] L. Isella, J. Stehlé, A. Barrat, C. Cattuto, J.-F. Pinton, and W. Van den Broeck, Journal of Theoretical Biology **271**, 166 (2011), ISSN 0022-5193.

[89] M. Salathé, M. Kazandjieva, J. W. Lee, P. Levis, M. W. Feldman, and J. H. Jones, Proceedings of the National Academy of Sciences **107**, 22020 (2010).

[90] A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, American Journal of Epidemiology **178**, 1505 (2013), ISSN 0002-9262.

[91] *World mosquito program*, `https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases`, accessed: 2022-25-2.

[92] W. H. Organization et al., Tech. Rep., World Health Organization (2014).

[93] W. H. Organization et al., Tech. Rep., WHO Regional Office for South-East Asia (2014).

[94] V. R. Preedy and R. R. Watson, *Handbook of disease burdens and quality of life measures*, vol. 4 (Springer, 2010).

[95] S. Lozano-Fuentes, D. Elizondo-Quiroga, J. A. Farfan-Ale, M. A. Loroño-Pino, J. Garcia-Rejon, S. Gomez-Carro, V. Lira-Zumbardo, R. Najera-Vazquez, I. Fernandez-Salas, J. Calderon-Martinez, et al., Bulletin of the World Health Organization **86**, 718 (2008).

[96] A. B. Knudsen and R. Slooff, Bulletin of the World Health Organization **70**, 1 (1992).

[97] E. A. Casman and H. Dowlatabadi, *The contextual determinants of malaria* (Routledge, 2010).

[98] W. Tabachnick, Journal of Experimental Biology **213**, 946 (2010).

[99] C. Caminade, K. M. McIntyre, and A. E. Jones, Annals of the New York Academy of Sciences **1436**, 157 (2019).

[100] *Vector-borne diseases*, `https://www.who.int/news-room/fact-sheets/detail/vector-borne-diseases`, accessed: 2021-25-11.

[101] J. D. Stanaway, D. S. Shepard, E. A. Undurraga, Y. A. Halasa, L. E. Coffeng, O. J. Brady, S. I. Hay, N. Bedi, I. M. Bensenor, C. A. Castañeda-Orjuela, et al., The Lancet Infectious Diseases **16**, 712 (2016).

[102] S. Bhatt, P. W. Gething, O. J. Brady, J. P. Messina, A. W. Farlow, C. L. Moyes, J. M. Drake, J. S. Brownstein, A. G. Hoen, O. Sankoh, et al., Nature **496**, 504 (2013).

[103] S. Swaminathan and N. Khanna, International Journal of Infectious Diseases (2019).

[104] I. R. Montella, A. J. Martins, P. F. Viana-Medeiros, J. B. P. Lima, I. A. Braga, and D. Valle, The American Journal of Tropical Medicine and Hygiene **77**, 467 (2007).

[105] L. A. Moreira, I. Iturbe-Ormaetxe, J. A. Jeffery, G. Lu, A. T. Pyke, L. M. Hedges, B. C. Rocha, S. Hall-Mendelin, A. Day, M. Riegler, et al., Cell **139**, 1268 (2009).

[106] G. Macdonald et al., The Epidemiology and Control of Malaria. (1957).

[107] D. L. Smith, K. E. Battle, S. I. Hay, C. M. Barker, T. W. Scott, and F. E. McKenzie, PLoS Pathogens **8**, e1002588 (2012).

[108] O. Diekmann, J. A. P. Heesterbeek, and M. G. Roberts, Journal of the Royal Society, Interface **7**, 873 (2010).

[109] L. E. Muir and B. H. Kay, The American Journal of Tropical Medicine and Hygiene **58**, 277 (1998).

[110] P. F. Verdonschot and A. A. Besse-Lototskaya, Limnologica **45**, 69 (2014), ISSN 0075-9511.

[111] S. Hazarie, D. Soriano-Paños, A. Arenas, J. Gómez-Gardeñes, and G. Ghoshal, Communications Physics **4**, 1 (2021).

[112] D. Soriano-Paños, L. Lotero, A. Arenas, and J. Gómez-Gardeñes, Physical Review X **8**, 031039 (2018).

[113] S. Gomez, A. Diaz-Guilera, J. Gomez-Gardenes, C. J. Perez-Vicente, Y. Moreno, and A. Arenas, Physical Review Letters **110**, 028701 (2013).

[114] *Boletín Epidemiológico. Semana epidemiológica (Municipio Santiago de Cali)* (Secretaría de Salud Pública Municipal de Cali, 2015).

[115] *Boletín Epidemiológico. Semana epidemiológica (Municipio Santiago de Cali)* (Secretaría de Salud Pública Municipal de Cali, 2016).

[116] G. Escobar-Morales, *Cali En Cifras 2013* (Departamento Administrativo de Planeación, 2013).

[117] Alcaldía Municipal de Santiago de Cali, *Indices entomológicos en sumideros del programa prevención y control enfermedades transmitidas por vectores - etv*, `http://datos.cali.gov.co/dataset/indices-entomologicos-en-sumideros` (Data from 2013), accessed: 2020-06-20.

[118] *Boletín Epidemiológico. Semana epidemiológica (Valle del Cauca)* (Secretaría de Salud del Valle del Cauca, 2015).

[119] *Boletín Epidemiológico. Semana epidemiológica (Valle del Cauca)* (Secretaría de Salud del Valle del Cauca, 2016).

[120] C. Chastel, Frontiers in Physiology **3**, 70 (2012).

[121] T. Walker, P. Johnson, L. Moreira, I. Iturbe-Ormaetxe, F. Frentiu, C. McMeniman, Y. S. Leong, Y. Dong, J. Axford, P. Kriesner, et al., Nature **476**, 450 (2011).

[122] A. A. Hoffmann, S. L. O'Neill, and J. H. Werren, *Influential passengers: inherited microorganisms and arthropod reproduction* (Oxford University Press, 1997).

[123] M. Turelli and A. A. Hoffmann, Nature **353**, 440 (1991).

[124] D. Vicencio, O. Vasilieva, and P. Gajardo, arXiv preprint arXiv:2011.10501 (2020).

[125] B. W. Alto and S. A. Juliano, Journal of Medical Entomology **38**, 548 (2001).

[126] M. M. Sowilem, H. A. Kamal, E. I. Khater, et al., Trop Biomed **30**, 301 (2013).

[127] T. Ruang-Areerate and P. Kittayapong, Proceedings of the National Academy of Sciences **103**, 12534 (2006).

[128] M. S. Blagrove, C. Arias-Goeta, A.-B. Failloux, and S. P. Sinkins, Proceedings of the National Academy of Sciences **109**, 255 (2012).

[129] V. H. Ferreira-de Lima and T. N. Lima-Camara, Parasites & Vectors **11**, 1 (2018).

[130] V. Duong, S. Ly, P. Lorn Try, A. Tuiskunen, S. Ong, N. Chroeung, A. Lundkvist, I. Leparc-Goffart, V. Deubel, S. Vong, et al., PLoS Neglected Tropical Diseases **5**, e1244 (2011).

[131] *Factsheet about dengue*, `https://www.ecdc.europa.eu/en/dengue-fever/facts`, accessed: 2021-11-25.

[132] D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni, and J. Gómez-Gardeñes, Physical Review Research **2**, 013312 (2020).

[133] *World mosquito program*, `https://www.worldmosquitoprogram.org/es/avances-nivel-mundial/colombia/cali`, accessed: 2021-25-1.

[134] A. Utarini, C. Indriani, R. A. Ahmad, W. Tantowijoyo, E. Arguni, M. R. Ansari, E. Supriyati, D. S. Wardana, Y. Meitika, I. Ernesia, et al., New England Journal of Medicine **384**, 2177 (2021), pMID: 34107180.

[135] Y. Bo, C. Guo, C. Lin, Y. Zeng, H. B. Li, Y. Zhang, M. S. Hossain, J. W. Chan, D. W. Yeung, K. O. Kwok, et al., International Journal of Infectious Diseases **102**, 247 (2021).

[136] T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, et al., Nature Human Behaviour (2021).

[137] N. Perra, Physics Reports (2021).

[138] *COVID-19 Stringency Index*, `https://ourworldindata.org/grapher/covid-stringency-index`.

[139] T. Laing, The Extractive Industries and Society **7**, 580 (2020), ISSN 2214-790X.

[140] R. Ong, *COVID-19: Remittance Flows to Shrink 14% by 2021* (The World Bank, 2020).

[141] P. Verma, A. Dumka, A. Bhardwaj, A. Ashok, M. C. Kestwal, and P. Kumar, *A Statistical Analysis of Impact of COVID19 on the Global Economy and Stock Index Returns*, vol. 2 (SN Computer Science, 2021).

[142] G. Gopinath, *The great lockdown: Worst economic downturn since the great depression*, vol. 14 (International Monetary Fund, 2020).

[143] E. Mathieu, H. Ritchie, E. Ortiz-Ospina, M. Roser, J. Hasell, C. Appel, C. Giattino, and L. Rodés-Guirao, Nature Human Behaviour **5**, 947 (2021).

[144] N. Cumming-Bruce, *Unequal vaccine access is widening the global economic gap, a u.n. agency says.*, `https://www.nytimes.com/2021/10/27/world/united-nations-coronavirus.html`.

[145] J. Lopez Bernal, N. Andrews, C. Gower, E. Gallagher, R. Simmons, S. Thelwall, J. Stowe, E. Tessier, N. Groves, G. Dabrera, et al., N Engl J Med pp. 585–594 (2021).

[146] Pfizer and BioNTech, *Pfizer and biontech provide update on omicron variant*, `https://www.pfizer.com/news/press-release/press-release-detail/pfizer-and-biontech-provide-update-omicron-variant`.

[147] N. Haug, L. Geyrhofer, A. Londei, E. Dervic, A. Desvars-Larrive, V. Loreto, B. Pinior, S. Thurner, and P. Klimek, Nature Human Behaviour (2020).

[148] D. Acemoglu, V. Chernozhukov, I. Werning, M. D. Whinston, et al., *A multi-risk SIR model with optimally targeted lockdown*, vol. 2020 (MA: National Bureau of Economic Research, 2020).

[149] D. Acemoglu, V. Chernozhukov, I. Werning, and M. D. Whinston, American Economic Review: Insights **3**, 487 (2021).

[150] L. Brotherhood, P. Kircher, C. Santos, and M. Tertilt (2020).

[151] D. Meidan, N. Schulmann, R. Cohen, S. Haber, E. Yaniv, R. Sarid, and B. Barzel, Nature Communications **12**, 220 (2021).

[152] F. Della Rossa, D. Salzano, A. Di Meglio, F. De Lellis, M. Coraggio, C. Calabrese, A. Guarino, R. Cardona-Rivera, P. De Lellis, D. Liuzza, et al., Nature Communications **11**, 5106 (2020).

[153] P. C. Silva, P. V. Batista, H. S. Lima, M. A. Alves, F. G. Guimarães, and R. C. Silva, Chaos, Solitons & Fractals **139**, 110088 (2020), ISSN 0960-0779.

[154] B. Faucher, R. Assab, J. Roux, D. Levy-Bruhl, C. T. Kiem, S. Cauchemez, L. Zanetti, V. Colizza, P.-Y. Boëlle, and C. Poletto, medRxiv (2021).

[155] Y. Yan, A. A. Malik, J. Bayham, E. P. Fenichel, C. Couzens, and S. B. Omer, Proceedings of the National Academy of Sciences **118** (2021), ISSN 0027-8424.

[156] G. De Meijere, V. Colizza, E. Valdano, and C. Castellano, Physical Review E **104**, 044316 (2021).

[157] C. Granell and P. J. Mucha, Physical Review E **97**, 052302 (2018).

[158] A. Arenas, W. Cota, J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños, and B. Steinegger, Physical Review X **10**, 041055 (2020).

[159] S. A. Lauer, K. H. Grantz, Q. Bi, F. K. Jones, Q. Zheng, H. R. Meredith, A. S. Azman, N. G. Reich, and J. Lessler, Annals of Internal Medicine **172**, 577 (2020).

[160] *COVID-19 Pandemic Planning Scenarios*, `//www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html`.

[161] Instituto Nacional de Estadística, *Demographics and mobility patterns*, `https://www.ine.es` (2021).

[162] *National Helath System capacity of Spain*, `https://www.mscbs.gob.es/estadEstudios/sanidadDatos/tablas/tabla22.htm`.

[163] W. S. Cleveland and S. J. Devlin, Journal of the American Statistical Association **83**, 596 (1988).