



## Tesis Doctoral

Temas en la modelización de propagación de epidemias

Topics in epidemic spreading modeling

Autor

Alfonso de Miguel Arribas

Director

Yamir Moreno Vega

FACULTAD DE CIENCIAS/IUI BIFI  
2023



## Agradecimientos

Gracias en primer lugar a mis padres por tenerme, criarme, aguantarme, y por haberme apoyado psicológica y económica durante toda mi formación académica, la cual me ha acabado permitiendo llegar hasta donde he llegado.

Gracias a mi hermano, por los buenos y necesarios ratos de ocio pasados durante el doctorado y, muy importante, haber estado siempre disponible para ayudarme con cuestiones tediosas sobre informática y programación. Gracias también por haberme dado a conocer Rust y haberme facilitado los primeros pasos.

Gracias a Clara, mi pareja, por haber estado siempre ahí, incondicionalmente, a pesar de absolutamente todo. Gracias por tirar del carro en los momentos más difíciles y sostenerlo todo, por soportar mis dudas, miedos, mi errática carrera y mi desordenada rutina diaria. Gracias por acompañarme y por haber creado y compartido tantas buenas experiencias juntos. Por muchos más, por siempre.

Gracias a mis amigos de siempre, por seguir siéndolo. A pesar de mi tendencia al aislamiento y a la -buscada- soledad, gracias por seguir contando conmigo.

Gracias a Yamir, mi director de tesis, por haberme dado este billete cuando pensaba que ese tren había pasado ya para siempre. Gracias en general por todas las oportunidades de trabajo, docencia, estancias, y congresos, y por la flexibilidad y libertad de investigación que me facilitó. Gracias también a Alberto, que aunque no acabó siendo mi co-director, estuvo en primera línea de apoyo en prácticamente todos los trabajos emprendidos (incluso en los que no hubo tiempo de sacar adelante), y contribuyó enormemente en los procesos de análisis, revisión y a la hora de aportar ideas para encarrilarlos. Gracias a Mario Floría por la colaboración y supervisión durante mi primer año. Y gracias a Esteban, por la colaboración surgida en el último año, por acogerme en Boston con sus estudiantes y colaboradores, y darme a conocer la comunidad de *NetMob*.

Gracias también al resto de compañeros de trabajo, que considero amigos, y con los que he compartido muchos buenos momentos de distensión y diversión. De nuevo a Alberto, Carlos (sus cafés y sus diatribas, y su apoyo generalizado), Felipe, Javier, Guilherme, Henrique, Dan, Marco y especialmente a Tiago (*bem, calma!*), Ariadna y Mario, conviviendo intensamente durante las estancias veraniegas en Turín (y videojugando hasta la extenuación). De nuevo también, Mario, gracias además por todo el trabajo realizado en la asignatura de Física Computacional y por estar ahí de asistencia y apoyo moral constante, ya que empezamos y terminamos a la vez este viaje, de principio a fin.



## Abstract

This thesis, deeply influenced by the advent of the COVID-19 pandemic, explores the modeling of epidemics in human populations and their consequential impacts. Two underlying threads guide the presented works: i) the use of real-world data to enhance model realism, and ii) the application of computational mechanistic models to describe the spread of epidemics. The insights provided in this text aim to humbly contribute to the cumulative knowledge of the field's literature and bring insights to enrich public debate and health policy-making.

Chapters 1 and 2 comprise *Part I: Overture*. Respectively, they offer, through the biased lens of a physicist background, an extensive overview of the field and the theoretical basis to frame and understand the subsequent collection of works presented.

*Part II: Metapopulation models* contains chapters 3 and 4, focusing on the spatial spread of infectious diseases and the role of mobility, primarily in urban environments.

In Chapter 3, we examine the effectiveness of perimeter lockdowns at the urban scale, implemented in Madrid during the COVID-19 pandemic. Using this real case as inspiration, we study in an idealized setting how effective this strategy can be in containing an epidemic. Employing a metapopulation model informed by real mobility data, we discovered that in highly interconnected urban systems, diseases spread rapidly, and localized lockdowns prove ineffective. The assumptions made, including perfect surveillance and full compliance, further emphasize the limitations of such a strategy. Our research adds to the evidence suggesting that mobility restrictions are largely ineffective unless they are sufficiently stringent. Hence, the main focus should be on understanding and altering the contact patterns of the population, which drive the transmission of infectious diseases.

In Chapter 4, we delve into the role of exploration and preferential return mobility in the spatial spread of epidemics, aiming to advance standard metapopulation models that often overlook realistic mobility features. Utilizing the d-EPR model, informed by high-resolution mobility data from the Greater Boston Area, different settings and scenarios were analyzed. The results reveal that only the d-EPR model with a heterogeneous mobility distribution shows a distinct impact across mobility groups, indicating that explorers significantly contribute to the spread of the disease.

*Part III: Single-population behavioral structured models* introduces models considering age (chapter 5) and network structure (6). Here, the role of human behavior in epidemics, particularly concerning vaccine hesitancy, is assessed.

Chapter 5 shifts focus to evaluate the impact of vaccine hesitancy, exacerbated by concerns and misinformation during the COVID-19 pandemic. Leveraging U.S. surveys on vaccine acceptance and detailed age-structured contact matrices, we aimed to determine how vaccine

## IV

reluctance could potentially lead to secondary COVID-19 outbreaks, especially considering the rapid spread of the Delta variant. The results show a direct correlation between the scale of secondary outbreaks and vaccine hesitancy shares. While younger groups significantly influence attack rates, the elderly notably affect mortality rates. Crucially, the study's model aligns well with real-world vaccination and mortality data from the Delta wave.

In Chapter 6, we analyze the interplay between infectious disease dynamics, vaccination, and opinion dynamics on vaccine uptake with a threshold model on homogeneous and heterogeneous networks. Generally, as vaccination rates increase, a cascading pro-vaccination behavior emerges, aiding in controlling the spread and leading to a growing disease-free phase for a significant portion of the parameter space. However, for heterogeneous networks, vaccination rates must be exceptionally high to prevent large outbreaks. The impact of anti-vaccine attitudes on vaccination rates is minimal at lower levels but becomes significant as these rates rise. The introduction of anti-vaccine sentiment effectively raises the system's threshold, potentially leading to a higher prevalence of the disease. The study culminates by characterizing the impact on more realistic scenarios, leveraging heterogeneous survey data across all U.S. states.

Finally, *Part IV: Closure* and its corresponding chapter 7 summarize the work and offer concluding thoughts.

# Index

<b>I Overture</b>		<b>IX</b>
<b>1 Introduction</b>		<b>1</b>
1.1 No need to motivate this . . . . .		1
1.2 The physics of... Epidemic spreading? . . . . .		4
1.3 An overview of epidemic spreading modeling . . . . .		8
1.4 Focus, structure, and goals . . . . .		37
1.5 List of works . . . . .		39
<b>2 Framework</b>		<b>41</b>
2.1 Elements of epidemic spreading modeling . . . . .		41
2.1.1 Compartmental models . . . . .		41
2.1.2 The basic reproduction number . . . . .		48
2.1.3 The force of infection . . . . .		53
2.1.4 Herd immunity . . . . .		55
2.2 Models with structure: Heterogeneity enters . . . . .		57
2.2.1 Age . . . . .		57
2.2.2 The network paradigm . . . . .		61
2.3 Metapopulations: Introducing spatial structure and mobility . . . . .		68
2.3.1 The epidemic reactions . . . . .		70
2.3.2 The diffusion process . . . . .		71
2.3.3 The global invasion threshold . . . . .		75
2.3.4 Simulations . . . . .		77
2.4 Microscopic human mobility models . . . . .		81
2.4.1 Explorers and returners: The EPR family of models . . . . .		81
2.4.2 Scales in human mobility: The container model . . . . .		90

<b>II Metapopulation models</b>	<b>93</b>
<b>3 Assessing the effectiveness of perimeter lockdowns at the urban scale: the case of Madrid</b>	<b>95</b>
3.1 Introduction . . . . .	95
3.2 Material and methods . . . . .	97
3.3 Results and discussion . . . . .	101
3.4 Conclusions . . . . .	109
<b>4 Epidemic spreading in an urban environment under the d-EPR model of microscopic human mobility</b>	<b>111</b>
4.1 Introduction . . . . .	111
4.2 Material and methods . . . . .	113
4.3 Results and discussion . . . . .	117
4.4 Conclusions . . . . .	130
<b>III Single-population behavioral structured models</b>	<b>133</b>
<b>5 Exploring the impact of vaccine hesitancy in secondary outbreaks of COVID-19 in the US</b>	<b>135</b>
5.1 Introduction . . . . .	135
5.2 Material and methods . . . . .	137
5.3 Results and discussion . . . . .	141
5.4 Conclusions . . . . .	150
<b>6 Epidemic spreading in contact networks coupled to a threshold-based opinion dynamics on vaccine uptake</b>	<b>153</b>
6.1 Introduction . . . . .	153
6.2 Material and methods . . . . .	155
6.3 Results and discussion . . . . .	160
6.4 Conclusions . . . . .	177
<b>IV Closure</b>	<b>181</b>
<b>7 Closure</b>	<b>183</b>
<b>8 Bibliography</b>	<b>191</b>
<b>Figure List</b>	<b>223</b>

<i>INDEX</i>	VII
<b>Table List</b>	<b>225</b>
<b>Appendix</b>	<b>226</b>
<b>A Codes</b>	<b>229</b>



# **Part I**

## **Overture**



# Chapter 1

## Introduction

*You've met with a terrible fate, haven't you?*

— Happy Mask Salesman

### 1.1 No need to motivate this

As it is natural with every topic and conversation one wants to engage the reader in, texts on epidemiology are no less in remarking the importance of their study subject. Indeed, communicable diseases are not only an inseparable part of human societies but a by-product of their very essence: mixing and crowding.

Given my background, I was scarcely familiar with epidemiology when I started my Ph.D. in the Fall of 2019. I remember delving into the technical literature on epidemic spreading at that time. These texts often began by citing various significant epidemic outbreaks in human history and their dire consequences for both society and the individual. The usual suspects included the multiple waves of the Black Death during the fourteenth century, the Great Plague of London in 1665-1666, and the **Spanish** flu of 1918-1919 [1]. These events were truly devastating, particularly the first in Europe. The texts also mentioned more recent incidents, such as the first pandemic of the 21st century: the SARS outbreak of 2002-2004 [2], and the 2009 H1N1 swine flu, which was the first influenza pandemic of the same century [3]. Like earthquakes, different parts of the world are constantly experiencing outbreaks of varying size, geographical extent, and duration. And, similar to earthquakes, we only become aware of these events when they strike close to home or with severe impact. In less technologically advanced countries, diseases such as ebola [4], malaria [5], typhus [6], cholera [7], and tuberculosis [8] are common realities, and indeed, some are endemic in many regions [9]. However, quite egocentrically, all these stories seemed distant to me, both in space and time. Important, certainly, but not immediately relevant or dangerous in the daily life of a citizen in the so-called *developed countries*. Then, six months after I began my

Ph.D., the COVID-19 pandemic shattered our sense of normalcy [10]. The scale and duration of this event changed everything. Reading those “old” texts now, I cannot help but think: If only you knew what was coming...

The COVID-19 pandemic, which began in late 2019 in the city of Wuhan [11, 12, 13], China, unleashed an unprecedented global health crisis that swiftly gripped nearly every corner of the world [10, 14]. As the virus rapidly spread across continents thanks to our highly interconnected systems [15, 16], it exposed the vulnerabilities of healthcare systems [17, 18] and posed daunting challenges to governments and societies [19]. Hospitals were pushed to their limits as they struggled to handle the influx of patients, leading to shortages of critical medical supplies and personnel. The pandemic laid bare the need for robust and agile healthcare infrastructures capable of handling large-scale outbreaks, and even more importantly, the need for adequately assessing risks and not being prisoners of our normalcy bias.

Beyond its immediate health impact, the pandemic significantly disrupted economies and social structures worldwide [20, 21, 22]. Countries enforced strict lockdowns and travel restrictions [23, 24], leading to a halt in economic activities and widespread unemployment [25, 26]. The pandemic’s ripple effects were felt across industries, affecting sectors like tourism, hospitality, and retail. Governments faced the daunting task of balancing public health concerns with economic realities, making decisions with far-reaching consequences for their citizens’ well-being. Supply chains were tightly forced, and some sectors had to adapt to a *war* economy mode to weather the storm. Moreover, the pandemic revealed disparities in access to healthcare and resources, disproportionately affecting vulnerable populations [27]. Marginalized communities faced more significant risks due to inadequate access to medical care, limited information, and crowded living conditions [28]. It underscored the importance of addressing social determinants of health and promoting equity in healthcare access to build more resilient communities.

Admittedly, the COVID-19 pandemic has not been as lethal and devastating as some historical pandemics in terms of morbidity, yet it has caused significant disruption. Personally, I found the human response to the pandemic more shocking than the reach of the disease itself. In this regard, I would highlight two main points: (i) The profound alteration of daily life and our perception of what was normal and taken for granted. In Spain, control and mitigation measures were so stringent that they profoundly impacted our daily lives. Social relationships, jobs, regular leisure activities, even those involving no interaction, were all put on hold for what seemed like an endless succession of days and weeks. Our usual *normal* was lost, and with the relentless spread of the virus, there seemed to be no end in sight to this new reality. While some

actions were justifiable given the exceptional nature of the situation, the second-order effects of these measures on the socioeconomic system made it natural to question whether everything was justified or necessary to flatten the curve. At that time, the dilemma was between prioritizing health or the economy. For some, focusing on health was seen as a way to save the economy. For others, the matter was not so simple; overemphasis on containing the virus at any cost could have negative repercussions on employment, savings, and finances, and could potentially lead to increased poverty and social unrest. (ii) The resilience and self-organizing ability of human systems. The state of nations and the world are highly debatable and beyond the scope of this monograph. However, it is apparent that despite the size and number of shocks of different kinds that societies received during the pandemic, we somehow managed to get through without falling apart initially. As we emerge from the pandemic, many challenges exacerbated during this period remain, and others that were background issues have intensified. Yet, human systems demonstrated remarkable resilience and self-organization during these years. I would like to emphasize the majority of the population's willingness to adopt stringent measures, the professionalism of healthcare workers operating in high-risk and uncertain conditions, the coordinated efforts of global and regional organizations in maintaining international health and financial aid, and the scientific community's exceptional research efforts in analyzing the pandemic's impact and developing an effective vaccine in record time, which helped ease the reliance on strict non-pharmaceutical interventions.

Given these recent and striking events, the COVID-19 pandemic and subsequent crises should serve as a stark reminder of three things: (i) the importance of epidemiology and public health in safeguarding human well-being; (ii) the need for better-informed models that capture the interplay of human dynamics and contagion processes, including epidemic spreading, economic shocks, and opinion dynamics, to inform appropriate public policies, and (iii) the necessity to rethink our societal structures and supporting processes to withstand potentially higher risks and multi-shock events in the future.

## 1.2 The physics of... Epidemic spreading?

This dissertation is authored by a student with a B.Sc. in Physics, currently enrolled in the doctoral program in Physics at the University of Zaragoza. Laymen might wonder: What the hell is a physicist doing writing a thesis on the spread of infectious diseases in the realm of physics? Well, I am not the first to embark on this path, either here or elsewhere, and I am quite sure I will not be the last. In fact, I also hold an M.Sc. in the Physics of Complex Systems. The institute and group where I have spent the last few years focus on complexity and complex systems as one of its central themes. With this additional context, it all starts to make more sense.

Complex systems, or complexity science, is an expansive interdisciplinary field that emerged in the latter half of the 20th century [29, 30, 31]. Though its pioneers were primarily from the physical sciences [32], it quickly evolved into a diverse, interdisciplinary effort, extending beyond traditional physics problems of particles and fields. While we can find examples within the natural—forgive the redundancy—boundaries of Physics, complex systems, a term that is hard to define, encompass structures and processes as varied as those in organic systems, like the human brain [33], ecosystems [34], and socio-technical systems such as the World Wide Web [35], social dynamics [36], cities [37, 38, 39], and the financial and economic system [40, 41, 42, 43], among others. Indeed, one of the most stimulating aspects of this then-emerging field was the opportunity to tackle new challenges and cross boundaries that had historically seemed insurmountable, perhaps due to a mix of their unfamiliar nature and the absence of an appropriate framework and tools.

But after enumerating a diverse list of systems considered complex, what is *single* discipline approaching them, all the more when there are traditional areas of knowledge that studied them in the first place? The key points lie both in the object of study and in the foundations of the discipline. Without approaching the problematic definition of a complex system [44], these typically are characterized but at least some of the following items:

- Open systems. Systems that interact with their environment, exchanging information, matter, or energy.
- Interconnectedness. Elements within the system are interconnected, meaning that a change in one element can affect others.
- Feedback loops. Processes where the output of a system feeds back into itself as input, influencing future outputs.

- Nonlinearity. Small changes in input can lead to disproportionately large changes in output, making prediction challenging.
- Emergence. Behaviors and patterns emerge from simple interactions between elements of the system, which are not predictable from the properties of the individual elements.

Recognizing these characteristics in the previously mentioned examples becomes straightforward. Now that the problem is framed, there arises a need for a conceptual and operational framework to recognize, analyze, and understand these features. This is precisely what the science of complex systems has been striving to achieve since its inception.

Returning to our main topic: How does the spread of an epidemic fit into the scheme of complex systems science? [45] And what is an epidemic after all? An epidemic refers to the rapid spread of disease to a large number of hosts in a given population over a short period of time. When we focus on a specific population of hosts, we essentially have a system. This system can be characterized by observable—and therefore measurable—properties, such as the health status of the hosts at a certain scale. For instance, we can classify hosts by their health status and count the number of individuals in each category<sup>1</sup>. For simplicity, let us assume that identifying infected hosts is straightforward through their symptoms, bypassing the intricacies of characterizing infection within a host. If, as the definition of an epidemic suggests, or as we observe, there is a change involving the number of affected hosts, then the system is undergoing a dynamic process. This scenario is essentially a dynamical system. Fortunately, there is a well-established theory for representing and analyzing dynamical systems [46], which is a standard tool in both physics and complexity science.

A dynamical system can be effectively described using differential equations. Here, we consider the rates of change of the observables we are monitoring and apply some *intuition* or a well-established theory to understand how these observables behave over time and through interactions within the system. For instance, when modeling the rate of change in the infected population of hosts, we observe that this number can increase or decrease. This leads us to explore the mechanisms contributing to the rise and fall of this population. How does infection occur? Is it through interactions between healthy and infected individuals? If so, how exactly? Does it depend on host attributes like sex or age? These are the types of questions that physicists regularly investigate<sup>2</sup>.

---

<sup>1</sup>This is the essence of compartmental models, which are discussed later in 2.1.1.

<sup>2</sup>Of course, this is not exclusive to physicists, nor are they inherently more successful without a proper understanding and dedication to the specifics of the problem at hand.

Frameworks familiar to physicists, such as phase transitions and mean-field theories [47], network theory [48, 49], and the theory of stochastic processes [50], are invaluable in accurately modeling the dynamics of disease transmission and recovery. As we will discuss later, a significant contribution by the statistical physics community was the study of contagion processes on networks, challenging the traditional assumption of homogeneous mixing within a population. Networks, in their various forms and complexities, allow for the representation of intricate contact structures that more closely resemble real-world human interactions.

Furthermore, when our system involves multiple populations, the spatial propagation pattern of the disease becomes crucial. In this case, the problem can be approached as a reaction-diffusion system, another concept familiar to physicists. The “reactions” refer to epidemic-related processes (like infection and recovery), while “diffusion” represents the spatial spread of the disease across the system. This raises new questions: What are appropriate spatial representations of human settlements? How do humans move? [51] As we develop a more refined model of a real population, we might ask: Does this represent an actual human population or just a collection of wandering automata? Indeed, human behavior is a critical factor in epidemic spread [52], from individual attitudes to top-down, authority-led control strategies, especially when the disease in question can significantly alter our interaction patterns and social mixing.

Soon, what started as a relatively naive model, focused mainly on the number of infected individuals, quickly transformed into an untamed and intricate hydra of complexities. However, there is a reassuring aspect to this: not all modeling choices are necessary simultaneously to accurately characterize a specific spreading phenomenon. The approach largely depends on the specific questions we are addressing and the desired level of detail and precision. Nevertheless, one quickly realizes that closed-form analytical solutions become impractical. This is where computers come into play, providing an *in silico* platform for representing human systems under epidemic dynamics.

But the truth is, that dealing with complex models involves more than just handling their intricacies. Even though physicists may have a solid foundation in mathematics, powerful abstract frameworks, and modeling tools, they often lack detailed knowledge in areas crucial to medicine, clinical epidemiology, or microbiology. This expertise is essential to characterize the natural history of a disease, assess its impact on the human body, and understand how the disease affects people of different ages, sexes, or other attributes. Such specialized knowledge is also crucial for conceiving pharmaceutical interventions, like medicines or vaccines. Moreover, practical epidemiology can provide

methods to accurately measure key epidemiological quantities (such as transmission rate, serial interval, or basic reproduction number), which are necessary to inform mathematical models for generating precise forecasts rather than just hypothetical scenarios.

Physicists may also lack a background in human behavior and psychology, which can be invaluable in formulating models of human behavior during a disease outbreak. In heterogeneous populations, contact patterns can be highly complex, with intricate nuances among population traits. Models aiming to account for this will need a substantial amount of reliable data to inform their contact matrices, necessitating well-designed surveys by experts such as epidemiologists or demographers. When expanding the scope to multiple populations and spatial propagation, the expertise of urban scientists is required for accurate representations of human settlements, along with human mobility experts to model human movement appropriately. Furthermore, if we aim to understand the interplay between the spread of an epidemic and the economic impact on a region, we quickly realize that to develop competent, reliable models for epidemic spread—models that are relevant and useful for understanding the phenomena and for guiding public health policies—a physicist is just one member of a necessary team of experts.

Thus, the term *physics of* in the title should not be misleading. For me, the *physics of* is about delving deep into the conception and analysis of models that explore the underlying mechanisms of observed phenomena. When it comes to complex problems typical in fields like biology, sociology, and economics, these challenges cannot be tackled by single-domain experts alone. Individuals with relevant backgrounds and skills are needed to contribute to the discourse. The physics of life, society, and the economy, therefore, are not newly conquered territories for Physics to dominate; rather, they are fertile grounds for diverse collaboration and exciting discoveries. The study of epidemics thus epitomizes the science of complex systems, where we confront complex phenomena that require multi- and interdisciplinary expertise at various stages of the research process, along with a diverse set of tools and frameworks from fields such as physics, mathematics, statistics, and computer science.

## 1.3 An overview of epidemic spreading modeling

Mathematical epidemiology has evolved into a vast and fertile field with robust branches since its early days, dating back to Daniel Bernoulli's contribution in 1760 [53, 54]. This phenomenon of rapid growth is not unique to this field; many areas of science today seem to be expanding almost uncontrollably, reminiscent of unchecked exponential growth at the onset of an outbreak. This is particularly true in the realm of mathematical epidemiology, where the COVID-19 pandemic has significantly accelerated scientific production. The interdisciplinarity that characterizes this field can sometimes yield groundbreaking new insights, but at other times, it may lead to frustrating miscommunications. Indeed, crafting a comprehensive review of the entire field is a formidable task.

Therefore, let this serve as an excuse for providing a somewhat biased overview of the field, with a focus specifically on the main developments in *mathematical* and *computational* epidemiology. This overview will particularly highlight the literature underpinning the work in this dissertation, emphasizing the very topics that form its core.

### The pioneers

The book *Natural and Political Observations Made upon the Bills of Mortality* by John Graunt is possibly one of the first pieces of knowledge, if not the first, concerned with methods of public health statistics. These *Bills of mortality* consisted of weekly records of numbers and causes of death in London parishes, dating from 1592 and kept continuously for decades later (Figure 1.1). Graunt analyzed the various causes of death and offered a method to estimate comparative risks of dying from various diseases, giving the first approach to a theory of competing risks.

From 17th-century London, we travel to late 18th-century Basel, where a prolific polymath, Daniel Bernoulli, published in 1776 a mathematical model to analyze mortality from smallpox. Bernoulli's main objective was to calculate the adjusted life table if smallpox were to be eliminated as a cause of death. Bernoulli defined two current standard epidemiological parameters: the force of infection (the rate of acquiring an infection during a certain time window) and the case fatality (the proportion of infections resulting in death) and worked out a formula for estimating herd immunity with better estimates than some standard schemes still used today. In the words of epidemiologist Klaus Dietz, Bernoulli *was far ahead of modern epidemiology*.<sup>3</sup>

---

<sup>3</sup>A modern revision of Bernoulli's work is covered by Dietz et al. [55].

1606.

**A TABLE of the  
CHRISTENINGS and MORTALITY  
For the Year 1605 and 1606.\***

	Weeks. Month.	Days of the Month.	Christ.	Bur.	Pla.	Par. infecc.		Weeks. Month.	Days of the Month.	Christ.	Bur.	Pla.	Par. infecc.	
	1 Dec.	26.	100	116	5	5		28 July	3.	109	110	25	12	
	2 January	2.	117	151	6	5		29	10.	111	134	33	18	
	3	9.	130	138	4	4		30	17.	115	146	50	22	
	4	16.	124	138	3	2		31	24.	96	140	46	26	
	5	23.	143	121	6	4		32	31.	132	178	66	29	
	6	30.	124	101	3	2		33 August	7.	131	181	67	29	
	7 Febr.	6.	122	105	5	5		34	14.	141	197	75	33	
	8	13.	131	118	7	6		35	21.	133	189	85	28	
	9	20.	126	109	12	6		36	28.	125	207	85	29	
	10	27.	102	117	9	8		37 Septem.	4.	123	241	116	32	
	11 March	6.	110	98	7	4		38	11.	134	216	105	28	
	12	13.	126	137	9	7		39	18.	121	214	92	36	
	13	20.	123	133	14	11		40	25.	132	204	87	35	
	14	27.	134	123	17	8		41 October	2.	121	256	141	40	
	15 April	3.	123	114	13	9		42	9.	134	218	106	38	
	16	10.	132	145	27	11		43	16.	142	227	117	37	
	17	17.	139	129	12	8		44	23.	131	224	109	38	
	18	24.	118	110	11	7		45	30.	124	226	101	34	
	19 May	1.	92	136	17	10		46 Novem.	6.	136	183	68	27	
	20	8.	116	103	13	11		47	13.	125	162	41	20	
	21	15.	128	94	13	8		48	20.	121	145	28	11	
	22	22.	113	132	14	9		49	27.	143	123	22	13	
	23	29.	94	98	9	7		50 Decem.	4.	155	160	45	17	
	24 June	5.	129	112	16	8		51	11.	135	137	38	20	
	25	12.	127	112	19	14		52	18.	136	132	28	15	
	26	19.	121	119	15	10		53	25.	134	135	38	19	
	27	26.	132	126	24	16								
	The Totals													
	Christened — — — 6614													
	Buried — — — 7920													
	Whereof of the Plague 2124													

\* BELL'S London's Remembrancer.

**A TABLE**

Digitized by Google

Original from  
THE OHIO STATE UNIVERSITY

Figure 1.1: Bill of Mortality for 1606. Retrieved from The Collection of the Bills of Mortality from the archive from the The Ohio State University. Image under public domain retrieved from [https://en.wikipedia.org/wiki/John\\_Graunt](https://en.wikipedia.org/wiki/John_Graunt).

Returning to Britain in the mid-19th century, we encounter William Farr, who, in 1840, underscored the importance of empirical observation in uncovering the latent causes of epidemics. He observed, during a smallpox epidemic, that the plot of the number of deaths followed a roughly bell-shaped or “normal” curve (Figure 1.2). Farr

also noted that recent epidemics of other diseases had exhibited a similar pattern<sup>4</sup>. A few years later, John Snow identified the source of the 1854 cholera outbreak as a public water pump. His empirical analysis of the disease's pattern led local authorities to disable the well pump, which subsequently helped curtail the outbreak. Snow's thorough investigation and documentation of the outbreak later became a foundational work in the history of public health and geography.

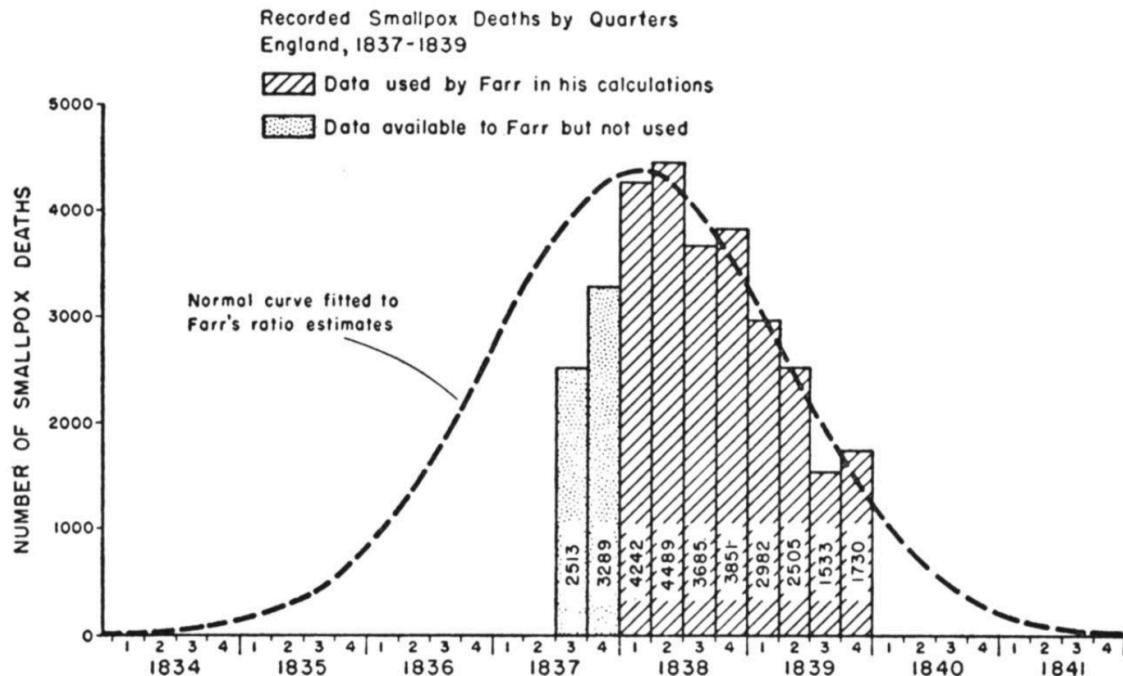


Figure 1.2: **Farr's curve depiction.** Farr's empirical description of the decline of the smallpox epidemic in England, 1837-1839. From Robert E. Serfling's *Historical review of epidemic theory* [56] (<https://www.jstor.org/stable/41449058>).

Although Bernoulli employed mathematical expressions to tackle an epidemiological problem, his model admittedly cannot be considered a conventional epidemiological model for informing the spread of infectious diseases, as it lacked dynamical evolution, for instance. Instead, the credit for pioneering the modern approach belongs to the Russian physician Pyotr Dimitrievich En'ko. In the late 1880s, his work on probabilistic modeling - specifically the chain binomial model - and data analysis foreshadowed the more renowned contributions of Lowell Reed and Wade Hampton Frost in the 1920s. Originating outside the *Western world*, En'ko's contributions are less well-known, primarily highlighted and elucidated by Dietz in [57]. Dietz made

<sup>4</sup>See [56] for a detailed account of Farr's work, as well as some of the earliest mathematical developments by inquirers such as John Brownlee and Ronald Ross, at the dawn of mathematical epidemiology. The reference also covers further advances and refinements throughout the first half of the 20th century.

a compelling argument for En'ko's paper being the first to discuss the elements of a genuine epidemic model [58].

### Classical epidemiology

At the turn of the nineteenth to the twentieth century, Sir Ronald Ross, often credited as the father of modern mathematical epidemiology, stands as the first Nobel Prize (1902) awarded to work related to this area of knowledge. In particular, he was credited for his pioneering work on malaria and his discovery of the mosquito-human transmission dynamics of the disease. In 1911, he went on to show that malaria can persist only if the number of mosquitoes is above a certain threshold. Therefore it is not necessary to kill all mosquitoes to eradicate malaria but just a certain fraction. This was the first account of the concept of epidemic threshold [59] and a rudimentary version of the basic reproduction number (nowadays denoted as  $R_0$ ), arguably the most relevant parameter in modern epidemiology. Field trials supported Ross' conclusion and led to sometimes brilliant successes in malaria control. His seminal trilogy (in collaboration with Hilda Hudson) [60, 61, 62] introduced compartmental models expressed in terms of a set of ordinary differential equations as we know them today. In fact, these models were so general that Ross coined the term "A Theory of Happenings" for them, pointing out that his results could have applications in economics, sociology, and other fields [56]. Unfortunately, during Ross's time, the mathematical modeling of infectious diseases was not widely accepted, hindering rapid progress in the field.

Another seminal contribution at the time, 1906, came from epidemiologist William Heaton Hamer [63], who postulated that the probability of an infection in the next period of time (in a discrete-time model) should be proportional to the number of susceptible individuals and the number of infectious individuals, introducing thus non-linearity in the modeling of epidemics. Clearly, what he suggested was a mass action law for the rate of new infections, and this idea would become ubiquitous in compartmental models for basically a century. The mass-action principle was at the time a well-known concept in chemistry, which was conceived as early as the 1670s by Robert Boyle [64]<sup>5</sup>. Actually, the idea was originally formulated in a discrete-time model, but in 1908 Ross translated the problem into a continuous time framework, later published in his aforementioned trilogy [66].

The turn of the century introduced cornerstone concepts and methods to the field of epidemiology and the mathematical modeling of infectious diseases. While all of these developments required further refinement and research, they laid the groundwork. Then, in 1927, came what is considered the most seminal contribution to the field.

---

<sup>5</sup>See [65] for a historical perspective of the mass-action law in epidemiology.

This development firmly established mathematical epidemiology and gave it strong momentum for the rest of the century.

William Ogilvy Kermack and Anderson Gray McKendrick, in their joint work *A contribution to the mathematical theory of epidemics* [67], following the tracks of Ross, Hudson, and Hammer, published an epidemic compartmental model that expanded on the ideas of the previous names and included susceptible, and infected and removed individuals for the first time. This model was based on more general assumptions than were used by Ross. This health-status-based distinction is what nowadays constitutes the renowned SIR model but, actually, the original model was more intricate than the current standard epidemic model. The Kermack-McKendrick model was based on an age-of-infection scheme and as such involved an integro-differential set of equations. Their model did not include vital dynamics (i.e. natural birth and death rates) in their first work, but it would do it in later follow-ups. To capture epidemic modeling of diseases that can become established in a population and persist, Kermack and McKendrick followed their original paper with Part II [68] in 1932, and Part III [69] in 1933<sup>6</sup>. In their last works, Kermack and McKendrick would give up their age-of-infection model for a simplified Markovian version, and those types of model would not be reborn until their usage for modeling the HIV epidemics in the eighties [72].

The problem addressed by Kermack and McKendrick in their first paper was a topic of great interest but also the source of much controversy at the time. This controversy was centered on the factors that determined both the magnitude of the epidemic and the causes of termination in a population. They arrived at the conclusion that the course of an epidemic is not necessarily terminated by the exhaustion of healthy individuals but that there exists a threshold density of the population below which the introduction of one or more infected individuals does not give rise to an epidemic, whereas if the population is only slightly more dense a small epidemic occurs. These results clearly established the non-linearity of the spreading dynamics of infectious diseases, by unveiling that the success of spreading is linked to a threshold behavior that depended on the disease parameters and the population density. Here again, reappears a threshold parameter, but is still not identified as  $R_0$ , nor given a specific symbol. This was quite unexpected at the time since medical circles considered two explanations for the die-out of an epidemic: (i) the supply of healthy people had been exhausted and (ii) during the course of the epidemic the virulence of the infectious

---

<sup>6</sup>Some years later, the trilogy would be expanded by Part IV [70] and Part V [71] even though with less commercial success. Because of their seminal importance to mathematical epidemiology, the Kermack–McKendrick fundamental trilogy of papers was reprinted in 1991.

agent had decreased [66]. Their work did not only lay robust theoretical foundations but also put a great deal of effort into parameter estimation from data and compared their models' predictions with observations [66].

Interestingly, McKendrick also published a contribution in 1926 to stochastic epidemic modeling [73]. Around this time, the year 1928, mathematician Lowell Reed and physician Wade Hampton Frost developed a binomial chain model for disease propagation, used in their biostatistics and epidemiology classes at the Johns Hopkins University [74, 75]<sup>7</sup>. The Reed-Frost formulation is an alternative to the mass action principle based on a “probability of effective contact” [76]. Both the mass-action and Reed-Frost approaches to infection transmission have been common, and preferred depending on the specific inclinations of the modeler (mathematically or statistically inclined). This last one, though, would be more appropriate for small populations due to the relevance of random effects at that scale.

It would take around a couple of decades more for the modern theory of stochastic modeling to be formally established. Subsequently, stochastic epidemic modeling took off. We can highlight here Maurice Stevenson Bartlett and Norman T.J. Bailey as the first greatest contributors in this area. The first one studied a continuous time stochastic SIR model [77], deriving into a frenzy on stochastic modeling formalism and methods for the next decades [78, 79, 80, 81, 82]. The stochastic modeling of epidemics leaves with important results such as the finite probability of extinction of an outbreak even when conditions are ripe for it (the epidemic threshold is above unity), and that the final size of an epidemic distributes normally around a mean value. Stochastic effects are of great relevance when the spreading occurs in small enough populations.

When Bailey published in 1975 the second edition of the classic review, *The Mathematical Theory of Infectious Diseases*, the bibliography already contained 539 papers, being the 62% published between 1964 and 1973. One can consult this information in a later review by Lisa Sattenspiel [76], where the author appears overwhelmed by the exponential pace that research in mathematical epidemiology had achieved at the time<sup>8</sup>. As a promising field with central relevance for human lives and a lot of questions pending answers and improvements, the following years and decades saw an impressive amount of work, both devoted to the theoretical modeling aspects, but also, a growing body of knowledge dedicated to the practical application of the principles learned and the models devised to real epidemics.

The pervasive assumption of homogeneous mixing expressed through the

---

<sup>7</sup>As we exposed previously, En'ko had already worked out these very same ideas in the previous century. Interestingly, Reed and Frost never published their work. It was Helen Abbey [74] who gave a detailed account of it in 1952.

<sup>8</sup>I can totally relate as I write these lines.

mass-action transmission term in the models was developed in the early times out of mathematical convenience. This assumption allowed easy modeling, analytical tractability, and immediate intuitions. However, indeed, it was a recognized fact that populations do not mix randomly. Consequently, we start to see an increasing interest in incorporating *structure* into the models and consider features like nonrandom mixing and heterogeneous transmission, based on traits like age [83, 84, 85, 86, 87, 88] or sex [89, 86, 90, 91, 92, 93]. We also see other departures from linearity, through the conception of nonlinear incidence rates [94, 95, 96], or the introduction of seasonality [97, 98]. In particular, in studies about mixing, crude assumptions had to be made on the structure of the mixing matrices, since empirical data on mixing patterns stratified by age were very limited. This is one of the things that would greatly improve in the new century. On the more practical side, the mathematical modeling of infectious diseases gained importance in the 1980s with the emergence of the HIV epidemic [99, 100, 101, 102]. We also start to see the use of structured models to attempt to describe the spread of measles [103, 104, 105, 106, 107], influenza [108, 109, 110], smallpox [111], and hepatitis A [112, 113].

As a culmination of an era, *Infectious Diseases of Humans: Dynamics and Control* by Roy M. Anderson and Robert M. May [114] represents a monumental compilation of the advances and applied knowledge of the time<sup>9</sup>.

During the nineties, the focus on heterogeneity and structure doubled down [116]. Special attention deserves the work of O. Diekmann and Hans Heesterbeek on extending the notion of the reproduction basic number to heterogeneous populations [117, 118, 119]. They not only extended the notion, actually, but developed an operational definition to compute this quantity, through the so-called next-generation operator. By the way, even though anticipated by Ross and Kermack-McKendrick as we told, it was not until 1957, with G. McDonald, also in a work related to malaria, that the pervasive threshold quantity would be baptized by the name “basic reproduction rate” and denoted by  $z_0$ . Later, the name and notation would change to what we know today as  $R_0$ [119]<sup>10</sup>.

Other significant developments in the nineties, related to the introduction of structure in models, built upon research lines opened in the previous decade, such as household models [120, 121, 122, 123, 124, 125]. These studies on households were motivated by the need to distinguish household sources of infection from community sources in observational studies. Particularly influential on the transmission

---

<sup>9</sup>For readers interested in the advances in epidemic modeling up until the late eighties, we also recommend exploring the works of Dietz and Schenzle [115], and Sattenspiel [76].

<sup>10</sup>We will devote a full subsection to  $R_0$  in the next chapter.

within households, a debate emerged on the appropriate modeling of the force of infection term, ultimately related to the impact of contact rate in the mixing process [126, 127, 128, 129].

Additionally, during this decade, there was a growing interest in modeling spatial epidemics [130, 131, 132, 133, 134], incorporating the metapopulation framework from ecological sciences. In subsequent years, with the availability of extensive human mobility data and the advancement of complex network frameworks, metapopulations became a standard and powerful tool for analysis and prediction in an increasingly interconnected world.

\* \* \*

At this point, the *epidemic* of research on epidemics has gone deeply into an exponential phase of production. As a biased overview of the field, then, in the next sections entering the 21st century, I will present some of the modeling frameworks, developments, and results on selected subareas or domains within the vast literature of mathematical epidemiology and epidemic spreading modeling. I will concentrate on (i) network epidemiology, (ii) epidemics and mobility, (iii) epidemics and human behavior, (iv) the application of models to real epidemics, and (v) lastly, the modeling efforts and aftermath of COVID-19<sup>11</sup>.

### The advent of the network revolution

The concept of a network - specifically, contact networks - and their relevance in the mixing patterns of a population was recognized even before the emergence of network epidemiology in the early 2000s. References to contact networks and some pioneering models exploring this idea can be found a decade earlier in works like [135, 136, 66]. Indeed, rooted in graph theory and dating back to Leonhard Euler's famous Königsberg bridge problem, significant developments occurred during the 1960s and 1970s. However, the field truly exploded in the late 1990s and early 2000s, driven by the advent of the internet and the availability of large-scale data for analysis. Seminal papers by Albert-László Barabási et al. [137, 48], demonstrating properties of networks with high heterogeneity of connections, Duan J. Watts and Steven H. Strogatz [138], characterizing the small-world effect, and eventually, M.E.J. Newman's exploration of contagion dynamics on networks [139, 140], laid the groundwork for a surge in research on complex networks and dynamical systems. It is worth noting that the application

---

<sup>11</sup>In the next chapter, sections will be devoted to a more pausing treatment of some technical aspects of complex networks, spreading on networks, and metapopulation models.

of networks in epidemic spreading was recognized not only by physicists (like Barabási or Newman) but also by epidemiologists, as evidenced by the work of Keeling et al. [141], who were on track just before the field gained widespread attention.

But what exactly is a network? A network, or more classically termed a graph, is an object consisting of vertices (nodes) and edges (links)<sup>12</sup>. The most common mathematical representation of networks is the adjacency matrix  $\mathbf{A}$ , where, in a simple network, the element  $A_{ij}$  of the matrix indicates the existence of a connection between node  $i$  and node  $j$  if it is 1; otherwise, it is 0. A fundamental property of networks is their degree distribution, with the degree being the number of connections each node has. Networks can be classified based on the heterogeneity of their degree distribution. Networks, where nodes have a characteristic degree, are called homogeneous, whereas networks without a characteristic scale in their degree distribution are referred to as scale-free and are mathematically described by power-laws [142]<sup>13</sup>.

Real-world networks [146, 147] markedly differ from the regular lattices that many physicists were accustomed to at the inception of the field. These networks exhibit dynamic self-organization and statistical heterogeneity, often displaying a hierarchical structure with a few nodes serving as *hubs*. These hubs are located at the far end of the degree distribution, rare but characterized by a significantly high number of connections. As a result, degree distributions in real-world networks tend to be heavy-tailed, skewed and span several orders of magnitude. Both social and infrastructure networks typically organize into communities with tightly interconnected nodes. Although randomness is invariably present, real-world networks, in contrast to random graphs [148], are defined by organizing principles and correlations in connectivity patterns [143]. These characteristics profoundly influence their dynamical evolution and the behavior of dynamical systems operating within these structures.

In epidemiology, networks are utilized to represent peer-to-peer contact networks [149, 150], where nodes symbolize hosts and links denote the social and ultimately physical connections between individuals. The introduction of complex network theory in epidemiology was revolutionary, providing a framework to move beyond the traditional well-mixing (mass-action) assumption and to represent individual interactions in a detailed manner [151].

One of the earliest and most notable findings in the study of contagion processes on networks was the discovery that the epidemic threshold vanishes for scale-free networks (with exponent  $2 < \gamma < 3$ ) in the thermodynamic limit ( $N \rightarrow \infty$ )

---

<sup>12</sup>In the following discussion, I will use the terminology of networks, nodes, and links, following the physics tradition rather than the mathematical one.

<sup>13</sup>Despite being a relatively new field, the literature on networks is extensive. For both good introductions and comprehensive coverage, see [143, 144, 145].

[152, 153, 154, 155, 156]. This implied a finite probability of a sizable outbreak at any transmission rate. Although real networked systems are not infinite, critical socio-technical systems, like the airport network [157], do exhibit scale-free properties. Even far from thermodynamic conditions, this topology still facilitates the spread of epidemics. We have learned that in these systems, spreading quickly reaches the highest degree nodes, the hubs, from which it easily propagates throughout the rest of the system [158, 159]. This situation is a double-edged sword, but it can be advantageous if we prioritize the isolation/immunization of these superspreaders [160, 161, 162].

The dynamics of spreading on networks are formulated in terms of systems of ordinary differential equations representing a mean-field theory, with a primary aim to compute the epidemic threshold or critical point of the system [163]. This evolved from the simplest homogeneous mean-field theory, equivalent to classical deterministic epidemic models, to theories acknowledging increasingly higher levels of heterogeneity [164]. The opposite extreme of the homogeneous mean field would involve dynamic equations for every node, representing the probability of being in a certain state. However, solving these systems is unfeasible for even simple scenarios, not to mention real-life cases, due to high dimensionality and recursive relationships. This necessitates enforcing closures [165, 166] on the interaction terms of the equations, leading to various levels of approximation and mean-field theories [167], such as individual-based or quenched mean-field [168, 164], pair-quenched mean-field [169], or heterogeneous (degree-based) mean field [152, 154], among others [164]. As these theories represent different degrees of approximation, their epidemic threshold predictions differ [170, 163]. Simulations, particularly stochastic micro-simulations [171], either discrete-time [172] or continuous-time based on the Gillespie algorithm [173, 174], are invaluable for simulating complex systems and advancing theoretical models or when insights from their analysis are insufficient.

While the spatial substrate of early networked epidemic models mainly consisted of static and simple networks, sometimes directed or weighted, significant extensions of the original formalism were developed about a decade later<sup>14</sup>. One such extension is the temporal network formalism [175], and another is multilayer/multiplex networks [176, 177, 178]. Temporal networks allow for dynamic connections between nodes, reflecting a natural aspect of real-world networks where connections can be created, destroyed, or rewired with diverse variability. In epidemiology, the consideration of inter-event times, contact duration distributions, and temporal correlations is particularly important [179]. The static network formalism assumed homogeneity in these aspects and lacked

---

<sup>14</sup>For an in-depth technical exploration of the field of epidemic spreading or contagion processes in complex networks, we recommend the reviews [167, 163] and the book [164].

consideration of the third, but empirical data show that distributions of inter-event times and contact duration can be long-tailed [180]. The temporal behavior of social interactions is often characterized by heavy-tailed and skewed statistical distributions, as seen in studies on mobile phone communications [181] and face-to-face interactions [182]. Temporal networks also encompass adaptive networks, essential for modeling behavioral changes and non-pharmaceutical interventions. A characteristic of spreading processes on temporal networks is their slower progression compared to static networks [167].

In the realm of multiplex/multilayer networks [183], each layer of this collection of networks may represent a specific subsystem within a broader system. For example, fully representing mobility in a country would require considering various types of mobility networks (e.g., road, train, and air), necessitating a multilayer approach to capture this complexity. In multiplex networks, although several layers are present, the nodes remain the same across layers, typically changing the topology, i.e., the connections between nodes in each layer. These systems can represent an individual's contact networks in different settings (household, workplace, community) or a combination of physical connections and social media interactions. This framework facilitates the exploration of competing contagion dynamics, such as an epidemic process in one layer and information spreading in another.

Recently, the broad field of dynamical processes on complex topologies, far from stagnating after two decades of explosive growth, is undergoing novel and exciting advancements. The current trend focuses on higher-order interactions, moving beyond the dyadic nature of contacts inherent in traditional network models. Exploring these interactions, new tools like simplicial complexes and hypergraphs have emerged and are garnering significant attention [184, 185, 186, 187, 188, 189]. Although still in its infancy, this field is already yielding novel insights. Higher-order interactions often introduce new sources of non-linearity not present in standard network approaches [190]. While currently theoretical, this approach holds promise for the development of spreading dynamics models that enhance our understanding of real-life phenomena.

### Epidemic spreading in a highly interconnected world

The Black Death devastated Europe in the 14th century. At that time, travel options were limited, and journeys typically covered short distances within the scale of a day. The spatial propagation of infectious diseases before the 20th century can be primarily considered as a linear spatial diffusion phenomenon. Rough estimates from historical data indicate that the spread of the Black Death through Europe progressed from south to north, with an invasion front moving at an approximate velocity of 200-400

miles per year [191]. Mathematically, this process can be represented by a SIR model with a diffusion term. When the traveling front velocity is calculated from this model, it aligns with the estimated range from historical data [192]. This simple calculation suggests that spatial diffusion was likely the main driver of spreading dynamics in that era.

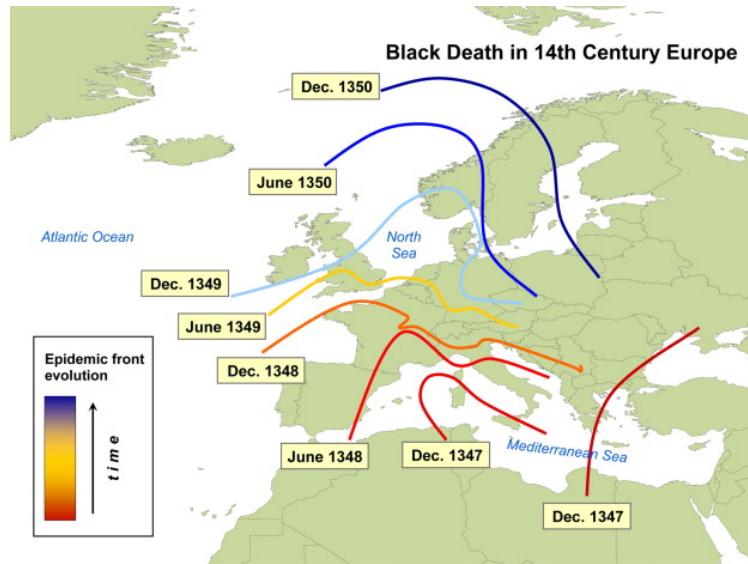
In contrast, modern travel and transportation modes, their volume and speed, and the connections they establish between different areas have dramatically transformed due to the utilization of powerful energy sources and subsequent technological advancements. This modern phase enables diseases to spread to virtually any point in the world within a few days. COVID-19 serves as a paradigmatic example of this phenomenon<sup>15</sup>. Fortunately, COVID-19, despite its rapid spread, was not as lethal as it could have been, and societies were generally well-equipped to mitigate high mortality rates. However, this experience should not lead to complacency or overconfidence in facing future pandemics. It serves as a reminder of the need for vigilance and preparedness for potentially more dangerous diseases that may emerge in our highly interconnected world.

The modeling of infectious disease spread traditionally focused on propagation within a single, isolated population. While a standard diffusion model may have sufficed in a world devoid of long-range spatial correlations, the realities of the 20th century introduced new complexities that demanded innovative approaches [192]. A critical factor that significantly altered the global spread of infectious diseases was the growth and development of the commercial aviation network, necessitating its inclusion in disease spread models [15]. The application of complex network theory, already successful in modeling the heterogeneity of real-life contact networks, could be extended to accommodate the heterogeneous spatial distribution and connectivity of human settlements. This need led to the adoption of the metapopulation framework, a model that accounts for multiple interacting subpopulations connected through networks of movement or interaction, such as those facilitated by modern transportation systems.

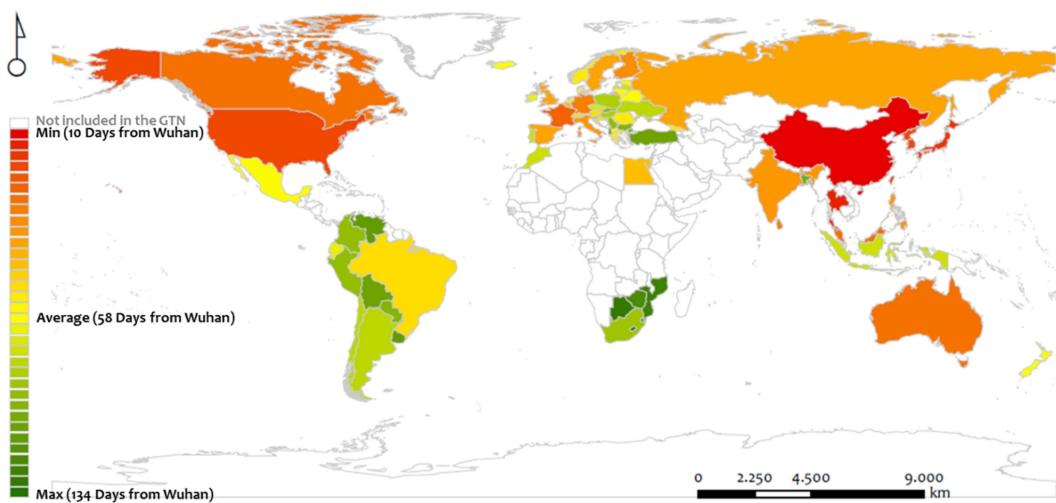
Actually, the concept of metapopulation originated in the ecology literature through the work of Richard Levins [194]. A *metapopulation* is understood as a group of populations of the same kind that live in spatially isolated areas but there is some level of interaction among them through migration processes. These isolated areas occupied by each population were initially referred to as *patches*. This paradigm turned out to be extremely useful in the case of infectious diseases. Here, the spatial structure, the patches, can be areas that, depending on the particular focus, may range from countries or regions to cities, neighborhoods, households, or very specific

---

<sup>15</sup>Refer to Figure 1.3 for a comparison of Black Death and COVID-19.



(a) Black Death spread from Colizza et al. [192] (<https://doi.org/10.1016/j.crvi.2007.02.014>).



(b) COVID-19 spread from Tsiotas et al. [193] (<https://doi.org/10.1038/s41598-021-04717-3>).

Figure 1.3: **Black Death and COVID-19 spatial spread.** Panel (a) illustrates the diffusion of the Black Death across 14th-century Europe, while Panel (b) presents an estimation of the spatial distribution of COVID-19's temporal spread. This estimation measures the number of days from the emergence of the first case in Wuhan. Unlike the Black Plague, which diffused from Southeast to North Europe over approximately three years, COVID-19 spread globally in just a few months. This stark contrast highlights the dramatic acceleration in the global spread of infectious diseases, reflecting the profound changes in transportation and connectivity over the centuries.

locations where the contact and contagion processes are relevant. This hints at the versatility of the framework. However, it only really makes sense to consider a set of patches as entities of the same metapopulation if these are connected, although loosely, through human mobility (or any disease-transmitting vector for vector-borne diseases). Therefore, appropriate human mobility models are an important modeling feature when working with a metapopulation. Metapopulation models comprise several fundamental ingredients<sup>16</sup>: (i) the spatial structure of the system, that is, the network topology, and the mobility model connecting it, this some specific accounting or characterization of human mobile flows, (ii) the population structure, which may include attributes such as the age and the mixing patterns, and (iii) the epidemic model, representing the epidemiological processes, given by the specific disease under study, at every node. The progressive availability of massive quantities of data related to these diverse ingredients has allowed for major theoretical and practical advances in the field.

Within the epidemiology literature, the pioneering attempt to model the spatial (global) spread of an infectious disease was carried out by L. Rvachev and I. Longini [109], in which 52 major cities worldwide were connected through an intercity aviation transportation network. The authors applied the model to simulate the global spread of the 1968-1969 Hong Kong (H3N2) flu, capitalizing on previous studies of the Russian airline network [108]. In successive years, similar modeling approaches, although limited by very partial knowledge of the worldwide transportation network, were developed to reproduce specific outbreaks such as pandemic influenza [110, 195, 196], HIV [197], or SARS [198, 199, 200].

The beginnings of the metapopulation approaches were mainly led by the public health and mathematical epidemiology communities. We consider the work of Lisa Sattenspiel as the first truly introducing metapopulations in the context of epidemic modeling in 1987 [112]. In that landmark paper, Sattenspiel considered two different types of interactions between individuals, local ones occurring within a given patch, and social ones connecting individuals originating from different locations on the system. This idea was later expanded by Sattenspiel and Dietz to include the effects of mobility [201] thus laying the foundations for the development of epidemic models at the global scale. Along the development of the field, some of these models were mechanistic, explicitly taking into account the movement of individuals [108, 109, 110, 197, 201, 202, 195, 203], whereas others were based on effective coupling approaches where the diffusion process expressed as a force of infection coupling different subpopulations [131, 133, 204, 205, 206, 207]. Before the advent of the complex

---

<sup>16</sup>In the next chapter, section 2.3 we will give a more detailed account of the technical and mathematical aspects of the metapopulation approach.

network framework, some works compared the effect of non-local human anomalous diffusion with that of ordinary diffusion behavior. For instance, Brockmann et al. unraveled that long-range human mobility and interactions generate novel irregular spreading patterns without an apparent wavefront [208]. The availability of the complete worldwide airport network dataset (WAN) and the research carried out on its topology and properties [157], eventually allowed a full-scale computational study of global epidemics. Throughout the first decade of the 20th century, the combination of increased data availability, computational power, and the theoretical knowledge of the developing area of complex networks, really made the metapopulation approach and the analysis of the spatial propagation of infectious diseases explode. This is very well embodied in the work of Vittoria Colizza, Alessandro Vespignani, and collaborators. On the more practical side, we highlight the development of large-scale computational models, like the GLEaM model, which defined the state of the art in metapopulation modeling and has been widely applied to large-scale pandemic forecasting [209, 210, 211, 212, 213, 214]<sup>17</sup>. In some early works [200], Colizza et al. found that topological heterogeneity reduces the predictability of a pandemic forecast, whereas high-level heterogeneity of traffic flows improves it.

On more theoretical grounds, the authors also developed an analytical framework for networked metapopulations and unveiled critical properties in spatial epidemics [45, 192, 215, 216, 217]. Conceptually, and from a physicist's point of view, metapopulation models can be expressed in the language of reaction-diffusion models [50], where agents in the system can be regarded as chemical reactants, that indeed experience reactions represented by the different epidemic processes of the given disease at every node, and, obviously, the diffusion of particles across the system is represented by human mobility [215]. Similarly to the first mathematical spatial models, we can build a series of mean-field equations for locations where the contributions to changes in populations may come through epidemiological reactions or the diffusion of particles. Now, the crucial contribution from the network community is the introduction of the specific network arrangement for the mobility of agents, instead of the classical spatial diffusion term. This allows for the consideration of non-trivial topologies connecting locations. Additionally, the use of mathematical approximations, like the degree-based mean-field approach, easily allowed us to obtain relevant analytical insights into the systems' behavior [216].

One of the most relevant theoretical results obtained was the realization of the existence of a global epidemic threshold [216, 217]. Similarly, as the threshold behavior observed in single populations and characterized by the basic reproduction number

---

<sup>17</sup>Later in this section, we will give a more detailed account of the GLEaM model.

$R_0$ , we can characterize global outbreaks in a metapopulation by the so-called global invasion threshold,  $R_*$ . When  $R_* < 1$ , a global invasion (what we could term a pandemic if simulating a global spreading) cannot propagate, otherwise, conditions are prone for it. An important effort when formulating these metapopulation models is to come up with some analytical expression for the global invasion threshold, which depending on the model can get quite convoluted. Overall, though, what defines this threshold can be separated into the following elements: the disease parameters, the population characteristics, the topology, and the mobility parameter. What is generally found is that for real-life standard situations, and assuming a disease with  $R_0 > 1$ , the value of the mobility parameter that characterizes overall flows in the whole system, should be reduced to unrealistically low levels, in order to avoid generalized spatial spreading.

Successive theoretical works on metapopulations have focused on posing some variation on the basic metapopulation recipe (network/mobility + population structure + epidemiological model), and then developing mean-field equations plus an analytical computation of the global invasion threshold. Regarding mobility, metapopulation models typically assume indistinguishable agents and simple Markovian random walks to represent human mobility. Here we can highlight some works, though, that implemented recurrent mobility [218, 219, 220, 221, 222, 223, 224], to simulate commuting, which is a typical behavior of regular human mobility. Naturally, commuting modifies the global invasion threshold in the metapopulation system. Balcan and Vespignani find in [219] that with a high return rate, the sojourn time (i.e., duration of stays) of infected commuters might be too short to transmit the infection to susceptibles in adjacent unaffected locations. In a more sophisticated model with recurrent mobility [220], Balcan et al. embedded commuting networks in several countries into the worldwide air transportation network. The introduction of short-range commuting mobility enhances the synchronization of the spread of the epidemic for locations geographically close. Interestingly, in [221], it is found a mobility regime in which higher mobility is detrimental to the spread of disease. Recurrent mobility is also explored through the MIR model (mobility-interaction-return) in a multiplex setting [225], where the layers represent agent classes with different mobility patterns, akin to the socioeconomic class of the agents. Social mixing between classes and mobility crucially determines critical areas that may trigger the epidemic outbreak at the critical point.

Regarding population structure, the pervasive assumption in metapopulation models at every subpopulation is again the homogeneous or well-mixing approach. Remarkably, departures from this assumption have been considered in very few works

until now [226, 227, 228, 229, 224], when it comes to toy metapopulation models and not large-scale sophisticated models such as the GLEAM. We highlight here the work of Apolloni et al. [228], where the framework developed by Colizza and Vespignani is extended to heterogeneous networked metapopulations with two classes of agents that interact at every location through a contact matrix. The agents are taken as children and adults, and the epidemic threshold and the fraction of invaded subpopulations are explored under different mixing scenarios and approximations. The framework, though, is general enough to consider other types of heterogeneity classes and to be extended to a higher number of groups. More recently, the MIR model has been extended to introduce contact heterogeneity in [224] by distinguishing agents through a contact number class  $k$ ,  $k$  being drawn from a heterogeneous distribution. The model is explored on small star graphs but shows some rich non-trivial behavior in the vein of [221], where it is found the emergence of epidemic detriment when enhancing mobility.

Another possibility where heterogeneity may kick in is through inter-location (as opposed to intra-location) differences. Location-specific factors can be the potential drivers resulting in substantial variations of disease incidence between populations [230, 231]. Based on this, Wang et al. [232, 233] introduced two categories of location-specific human contact patterns into a phenomenological reaction-commuting metapopulation model. One of these categories is the destination-driven scenario, where individual contact features are determined by the visited locations, and the other is an origin-driven scenario, where the contacts of individuals are relevant to their subpopulation of residence. Due to the mixing of individuals with heterogeneous contact capacities in each subpopulation, the location-specific contact patterns reduce the epidemic threshold significantly and thus favor disease outbreaks in contrast to the traditional homogeneous mixing scenario. Inter-location contact heterogeneity has also been explored in an urban-rural model [234], where the epidemic threshold is found to be a non-trivial function of local parameters (reaction rates), global parameters (network's topology), and the cross-over population sizes separating urban and rural settings. In Tizzoni et al. [235], the authors introduce contact rates that scale with subpopulation sizes. The authors found that the heterogeneity of contact rates promotes the spreading of an epidemic and such effect is enhanced when the distribution of the mobility flows between subpopulations is heterogeneous.

Additionally, some other forms of heterogeneity have been explored, such as heterogeneous sojourn times [236], heterogeneous infection rates [237], or heterogeneous vaccination rates [238]. Finally, behavioral changes or coupled dynamical processes to the spreading of the disease have also been considered in some other works [239, 240, 241]. For a more detailed account of works related to the networked

metapopulation framework, we recommend the reader to follow a review [242].

### The behavioral component

Typically, epidemic models are portrayed as operating in an already *dead* society, that is, without a glimpse of a reaction to the spreading due to risk-aversion, disease incapacity, or control and prevention strategies. Indeed, models introducing mobility include human behavior, and also models with vaccination or other kinds of interventions, represent a reaction to the spreading. But here we are referring to a more genuine, detailed, micro-based accounting of human behavior and attitudes, both from an individual and a social point of view. Standard models without any or little consideration of the human behavioral component can be successful enough for mild epidemics, like the yearly influenza outbreaks in some countries. However, for disruptive situations like the one induced by COVID-19, it was clear enough that the spreading dynamics dramatically impacted our lives, and in turn, the behavioral response importantly shaped the contagion dynamics in multiple ways. In fact, the spreading of an epidemic can trigger changes in a population in varied non-exclusive ways. Either through fear or risk-aversion mechanisms, the individuals may consider taking prophylactic measures like wearing face masks, for instance, or avoiding certain risk situations, like crowded places. By looking at how our environment reacts, a cascading social behavior can be spurred through self-organization. All these behavioral changes will affect the fate of the spreading, helping to *bend the curve*. In turn, the evolution of the outbreak will affect our perception and habits, thus setting a feedback loop between the epidemic spreading and human behavior.

In the absence of a well-established, unified framework, we briefly review a selection of approaches and works that include modeling human behavior. We will focus on self-initiated (bottom-up) human behavior with a social component, excluding top-down approaches.

Firstly, there are several key features to consider when modeling human behavior coupled with an epidemic process. This list may include [243]: (i) Direction of societal response. This can be top-bottom, from authorities to the individuals through obligation or recommendation. Or bottom-up, emerging from the individual in itself and propagating throughout society. Of course, it can be both, which is what normally happens. (ii) Source of information. Individuals can be exposed to information on the disease and the disease-spreading evolution that will lead to changes in their behavior. We can speak of globally available information when the information is accessible to anyone through mass media or social media. We speak too of locally available information when this spreads by word of mouth, acquaintances, or in a

local community. Thus, depending on the exposition of the individuals, they will base their behavioral choice on global, local information, or some combination of both sources. (iii) Type of information. Here it is considered whether the information is based on objective rational grounds, or it is rather something unrelated or comes from some subjective assessment. We could ponder objective information as coming from the disease incidence or prevalence in the population, or the risks to health that may impinge. On the contrary, personal beliefs would belong to the subjective information group. (iv) Specific response. This includes all kinds of responses that the individual may initiate, ranging from the wearing of face masks, reduction of contacts, and reduction of mobility, to vaccination, to name a few. Depending on all these factors, the modeling approach can be hugely affected. From having a relatively simple model where only specific parameters like the contact rate or the transmission rate have to be tweaked, to more structured models that demand the introduction of more compartments, or the explicit modeling of information spreading or opinion dynamics together with the disease dynamics.

In addition to these specific considerations regarding the behavioral component, a now standard modeling option is whether to simulate the epidemic in well-mixed populations or within networked systems. This choice is significant, as human behavior is strongly influenced by the attitudes of our neighbors.

In well-mixed populations, a notable framework was established by Perra et al. in 2011 [244]. They conceptualized the behavioral and epidemic dynamics as two competing contagion processes: the infectious disease itself and the 'fear of the disease.' The fear of disease induces behavioral changes in the population, creating a distinct susceptible compartment. Individuals in this compartment respond to the disease spread, which translates into a reduction in the disease's transmission rate. This model, accounting for various information mechanisms like belief-based and prevalence-based, and contrasting global versus local information, exhibits rich phenomenology. It demonstrates multiple waves, transition points, discontinuous transitions, and memory effects of behavioral changes induced by the epidemic, highlighting the complexity of the interplay between disease and behavior.

During the same period, other research focused on the spread of awareness (i.e., fear of infection), extending the analysis to complex networks. A pivotal study by Funk et al. [245] formulated a model for the spread of awareness in both well-mixed and networked populations. In well-mixed settings, coupling the dynamics of the epidemic and awareness can reduce the outbreak size, but does not affect the epidemic threshold. However, in populations on a triangular lattice, behavioral responses can halt disease spread if the infection rate is below a certain threshold. Specifically,

the authors showed that the impact of locally spreading awareness is amplified if the social network of potential infection events and the communication network over which individuals communicate overlap, especially if the networks have a high level of clustering. Similarly, Wu et al. [246] explored several types of awareness: contact-based, increasing with the number of contacts, local-based, increasing with the fraction of infected contacts in the neighborhood, and global-based, increasing with the overall disease prevalence. The authors found that global awareness cannot decrease the likelihood of an epidemic outbreak while both local awareness and contact awareness do it. Generally, individual awareness of an epidemic contributes toward the epidemic spreading inhibition. The fact individual behavioral responses suppress epidemic spreading has been also found in other studies like Zhang et al. [247], in which the authors focused on an epidemic response model where individuals respond to the epidemic according to the number of infected neighbors in the local neighborhood. By studying both SIS and SIR epidemiological models with the behavioral response rule in scale-free networks, they found that individual behavioral response can in general suppress epidemic spreading, due to the crucial role played by the hub nodes who are more likely to adopt protective response to block the disease spreading path.

The multilayer network framework offers a solid foundation for exploring the interaction of competing dynamical processes such as disease spreading and behavior or information contagion. For instance, Granell et al. [248] proposed a multilayer model where one layer represents the physical contact network for disease spread, and the other is a virtual layer (e.g., a social network) where disease awareness is disseminated. They discovered that an increase in the transmission rate could lower long-term disease incidence while raising the epidemic's outbreak threshold. When awareness leads to total immunization, they found a metacritical point at which the epidemic's onset depends on the completion of the awareness process. In a more nuanced version of this model [249], the authors relaxed some assumptions and introduced mass media awareness broadcasting, which led to the disappearance of the metacritical point. The concept of social dynamics was further expanded to an awareness cascade model in [250], where agents exhibit herd behavior based on the actions of others. Intriguingly, an approximate local awareness ratio of 0.5 can cause a two-stage effect on the epidemic threshold and impact epidemic sizes, regardless of network structure. These findings offer a new perspective on realistic contagion prevention. Additional research [251] has explored self-awareness induced by infected neighbors, showing that coupling this process with disease spreading can reduce infection density without increasing the epidemic threshold, regardless of the information source.

Temporal networks, particularly adaptive networks, are another effective framework

for modeling human behavior and responses. Studies on contact switching as a protection strategy [252, 253, 254] demonstrate its efficacy in controlling outbreaks and reveal complex behaviors like rich nonlinear behavior such as bistability, hysteresis, first-order transitions, and epidemic reemergence. Implementing vaccination, immunization, and quarantine in adaptive networks has shown that vaccination is more effective in adaptive than static networks, irrespective of disease dynamics [255]. However, immunization and quarantine strategies have yielded counterintuitive results [256], indicating that timing is crucial and not necessarily “the earlier, the better”. The optimal effect for these is obtained when a strong community structure exists.

While these approaches offer versatile frameworks, they often overlook the costs associated with prophylactic responses. There are some other approaches, not without their own limitations, whose focus precisely relies on assessing these costs. Economic epidemiology models, focusing on utility maximization, address this aspect [257, 258, 259, 260, 261]. Individuals in these models aim to maximize utility through dynamic optimization problems. These models show us that when considering health outcomes, one must be acutely aware of the welfare costs associated with self-protective behavior or implementing disease mitigation policies. Measures like encouraging infectious individuals to self-quarantine may be counterproductive, actually causing a rise in disease prevalence due to susceptible individuals feeling less threatened by infection and subsequently abandoning their own self-protective behavior [260]. Also, a population who is given a pessimistic outlook of an epidemic may in fact cause the disease to spread more rapidly [259], as in the self-fulfilled prophecy effect.

Another widely adopted series of approaches incorporating individual-based decisions through cost assessment are the so-called *vaccination games*, which are grounded in the concepts and tools of game theory. Vaccination games aim to establish a specific framework for modeling the vaccination dilemma [262, 263, 264, 265, 266]. In brief, while society would generally agree on the goodness of eradicating or mitigating disease, at the individual level, one could expect that others take action. If a critical mass of people relies on others, the vaccination coverage may fail to reach the population’s herd immunity level, thus provoking the disease to propagate or re-emerge. Overall, vaccination game models agree that (i) it is impossible to eradicate the spread of disease under voluntary vaccination unless specific conditions are met, and (ii) the vaccination level that is best for self-interest is always well below the optimal level needed by the community to achieve “herd immunity” unless vaccination cost is sufficiently low. This occurs because non-vaccinated individuals can expect to escape the infection if the overall vaccination coverage is sufficiently high to curb the epidemic by reaching the herd immunity threshold. Such protection may require

extremely high coverage of vaccine uptake and may be difficult to sustain because individuals tend to exploit the temporary herd immunity (i.e. free-riding) and choose not to vaccinate. If each individual acts purely in a selfish fashion, herd immunity is unattainable without externalities (e.g. subsidies, regulation, etc.) [262, 263, 267]. The possibility of vaccination is also dependent on the transmissibility of the disease and is also proportional to the number of contacts the individuals have if they make decisions based on memories [268]. Generally, self-interest behaviors are not aligned with social optimum [269] unless the vaccine is sufficiently inexpensive [270], in which case both self-interest and social optimum converge to herd immunity. A typical phenomenon is wave-like vaccination behavior. When vaccination coverage is high, the disease prevalence is depressed; subsequently, people have less incentive to vaccinate, and the prevalence level quickly recovers when the vaccination coverage drops [267]. Similarly, when social distancing is adopted, if the number of susceptible individuals is not adequate to sustain the epidemic, the epidemic peak decreases below the threshold value; while if the reproduction number for people not adopting social distancing is still greater than 1, the spreading recovers, creating wave-like oscillations [271]. When vaccination games have been explored on complex networks, it has been found that vaccination coverage can be dramatically impacted by the network topology. The affectation of the network’s topology on the dynamics is a well-known result of network theory from its conception and vaccination games on networks are no exception. Vaccination coverage and herd immunity are affected by the topology but this effect is sensitive to the increase in the vaccination cost. A cost threshold exists beyond which small topological changes cause a significant reduction in vaccination coverage and a rise in infected cases, compared to homogeneous populations. This threshold-related effect is observed in both random and scale-free networks. For a more recent, general, and also comprehensive expositions on the topic we recommend the reader to follow [272, 273].

A less explored avenue is the coupling of disease spread with human behavior through explicit opinion dynamics models [36], such as the voter model, majority-rule model, or other threshold models. In these models, the “opinion” refers to the adoption of vaccination or protective behaviors. Some studies in this area include [274, 275, 276]. These models can exhibit rich behaviors, including sudden transitions, bistability, or network segregation.

Incorporating the behavioral component in epidemic spreading models is crucial, especially for epidemics that, even moderately, disrupt regular human behavior, thereby altering epidemic dynamics. Although the literature covers a wide range of models and scenarios, most approaches share some common challenges. The foremost among

these is the lack of real-world, quantitative data on behavioral changes in populations during epidemic outbreaks. This scarcity of data is, without doubt, the major hurdle in effectively integrating behavior-disease models. Accurate and comprehensive behavioral data is essential for these models to reflect real-world scenarios and provide valuable insights into epidemic control and prevention strategies.

### The rise of computational and data-driven epidemiology

For much of the 20th century, the advancements made by mathematical epidemiology were largely overlooked by disease control and prevention authorities. These developments were more often used retrospectively to validate policies rather than assist in their creation [277, 115, 76]. Generally, the mathematical approach was rarely adopted in planning infectious disease control. It was only towards the end of the century that mathematical modeling began to be more widely utilized in public health policymaking [72]. For example, during the first two decades of the AIDS pandemic, modeling approaches were increasingly employed to predict its evolution and identify effective prevention strategies. With the turn of the century, the influence of mathematical modeling on public health became more pronounced, particularly in evaluating intervention strategies for newly emerging and reemerging pathogens. Initially, the threat of bioterrorism using the smallpox virus spurred the use of mathematical modeling for contingency planning [278]. Later, the SARS outbreak, caused by a newly emerging pathogen, initiated the use of mathematical modeling for real-time analysis of infectious disease outbreak data to assess the effectiveness of intervention measures [279].

It was also in the mid-2000s that sophisticated large-scale models, like the aforementioned GLEaM model, started being developed. GLEaM integrates global population estimates with comprehensive airline transportation and commuting databases. The model divides the globe into 3362 transportation basins (patches or nodes), each centered around an airport and representing major metropolitan areas. While most human mobility does not occur through flights, the airport network is crucial for long-range travel and thus facilitates the rapid global spread of diseases. Additionally, the commuting network within each basin complements this mobility, enabling connections between neighboring basins. This combination allows GLEaM to accurately depict both international and regional disease spread. Within each basin, GLEaM employs the homogeneous mixing approximation but can incorporate a detailed compartmental structure, offering more nuance than a simplistic standard mixing approach. Furthermore, GLEaM can include age structure [280], further refining population dynamics, and can model interventions and mobility reductions,

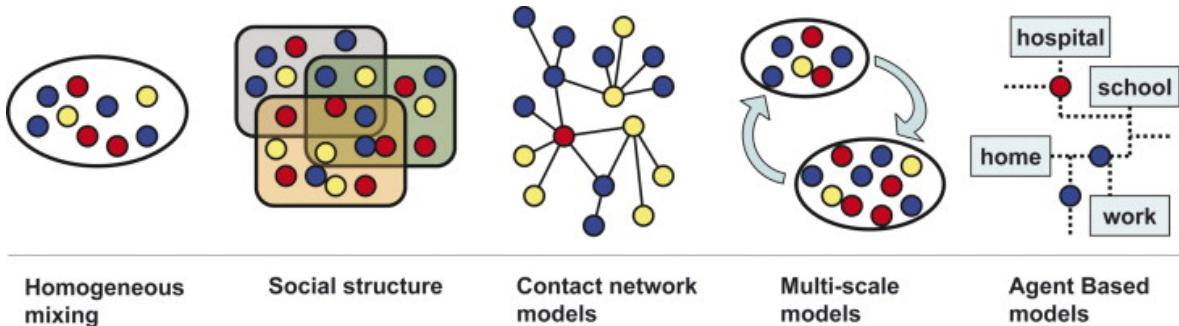
making it a versatile framework for exploring policies and counterfactual scenarios. GLEaM was first tested during the 2009 H1N1 pandemic, successfully predicting that the pandemic’s peak would occur in October and November 2009, with these predictions published in early September 2009 [212]. A later comparison between these forecasts and official data from health authorities [281] confirmed GLEaM’s efficacy in real-time epidemic forecasting. The model accurately predicted the peak week for 87% of the countries in the northern hemisphere with accessible data. In other cases, the maximum deviation was only two weeks. GLEaM has also been used during the COVID-19 pandemic, informing the effectiveness of travel restrictions [16] and for real-time forecasting [282, 283].

As indicated by the name, GLEaM specializes in large-scale epidemic forecasting but does not offer fine-grained resolution, particularly at the individual level. Models that do differentiate individual agents within a system are known as agent-based models (ABMs)<sup>18</sup>. While the concept of ABMs dates back to the 1940s, their widespread use only began in the 1990s due to their intensive computational demands. ABMs have applications across various fields, including biology, economics, and social dynamics. When modeling the spread of an epidemic at this detailed scale, attributes often overlooked in broader models, such as age, gender, workplace, residence, family structure, etc., become significant. This approach leads to the creation of synthetic populations that undergo specific dynamics.

To ensure reliability, synthetic populations must be statistically representative of the actual population under study, a task that is far from trivial. Much effort is devoted to data collection and developing methods to accurately construct these populations. The process typically starts with creating households based on size and age distribution, scaling from neighborhoods to the country level. This is followed by assigning occupations and workplaces (or schools for children) and incorporating other relevant attributes. In the realm of ABMs based on synthetic populations, one of the pioneering models was TRANSIMS [284], which generated a synthetic population for Portland, Oregon (USA). It included a route planner to determine each individual’s daily travel, modeling the impact on Portland’s transportation infrastructure and the effects of disruptions or modifications on the population’s daily life. This work led to the development of EpiSims [285], an extension of TRANSIMS applied to epidemiology. EpiSims utilized the TRANSIMS infrastructure to generate a contact network among individuals in Portland, facilitating the study of disease spread within this synthetic population.

---

<sup>18</sup>See Figure 1.4 for a look at the different modeling frameworks reviewed throughout this literature overview.



**Figure 1.4: Different scales structure used in epidemic modeling.** Circles represent individuals and each color corresponds to a specific stage of the disease. From left to right: homogeneous mixing, in which individuals are assumed to homogeneously interact with each other at random; social structure, where people are classified according to demographic information (age, gender, etc.); contact network models, in which the detailed network of social interactions between individuals provide the possible virus propagation paths; multi-scale (or metapopulation) models, which consider sub-population coupled by movements of individuals, while homogeneous mixing is assumed on the lower scale; agent-based models, which recreate the movements and interactions of any single individual on a very detailed scale (a schematic representation of a city is shown). Figure from Colizza et al. [192] (<https://doi.org/10.1016/j.crvi.2007.02.014>).

In recent years, modelers have successfully expanded ABMs beyond the scope of single cities. One of the early models, developed by Ferguson et al. [286, 287], was used to devise containment strategies for pandemic influenza. This model heavily relies on detailed data for constructing the synthetic population and for modeling disease evolution. It features stochastic infection dynamics and considers various transmission settings (household, workplace, and community). However, unlike GLEaM, it does not explicitly model human mobility dynamics. Subsequently, FluTE emerged as one of the most realistic large-scale agent-based epidemic models for the continental United States. It simulates international seeding of infections based on the traffic at the 15 largest international airports in the US, infecting individuals each day proportionally to the traffic at these locations. FluTE, a refinement of a previous model [288], further enhances the modeling of infectious processes by varying the infectiousness of individuals over time in their considered SIR model. Due to the model's complexity, special attention is required for calibrating disease parameters [289]. FluTE set a precedent in transparency by releasing its source code, enhancing verifiability and reproducibility in computational modeling. It has been successfully employed to study influenza spread and assess intervention impacts in Los Angeles County [290] and at the national level in the United States [289].

While ABMs provide a more detailed representation of the system compared to standard metapopulation models, they also incur a high computational cost in terms

of performance and memory usage. Another significant drawback of ABMs, which can be also an issue for less sophisticated models, is their intensive data requirements. As models become more complex and layered, parameters cannot be arbitrarily generated; they must be accurately derived if we aim to accurately describe real-life phenomena. This process involves not only generating a synthetic population but also constructing mobility networks and utilizing epidemiological surveillance data. Achieving this level of detail demands high-quality methods and infrastructure for data collection. However, even in developed countries, surveillance data is often imprecise, incomplete, subject to large backlogs, reporting delays, and based on small sample sizes. Moreover, data coverage within the same country can be uneven and lack standardization. These factors severely limit the predictive capabilities of epidemic models calibrated with surveillance data. Another major issue is the spatial granularity of data, which is generally aggregated at the country or regional level. The absence of more granular ground truth data hinders the selection and training of realistic epidemic models. To advance the state of the art in epidemiological modeling for forecasting and informing public health policies with maximum accuracy, a greater emphasis must be placed on the importance of data in modeling. This necessitates concerted efforts to collect high-quality, standardized data from surveillance systems and public health authorities.

Complementary to these advances and needs, the exploitation of the massive amount of information shared nowadays through the Internet lies at the core of the new science of digital epidemiology [291, 292, 293, 294]. Thanks to the mass adoption of smartphones in the last decades and their virtual interactions in social networks, an incredible amount of information is being created that conveys epidemiologically relevant indicators. This can be reflected in users complaining about catching a cold before the weekend on Facebook or Twitter, searching for symptoms of particular diseases on search engines, or Wikipedia, and canceling their dinner reservations on online platforms are just some examples. An intense research activity, across different disciplines, is clearly showing the potential, as well as the challenges and risks, of such digital traces for epidemiology [295]. This new approach allows for the early detection of disease outbreaks [296], the real-time monitoring of the evolution of a disease with an incredible geographical granularity [297, 298, 299], the access to health-related behaviors, practices, and sentiments at large scales [300, 301], inform data-driven epidemic models [302, 303], and development of statistical-based models with prediction power [304, 305, 306, 301, 307, 308, 309, 310, 311, 312]. The search for epidemiological indicators in digital traces can be based on active and passive methodologies. In the former, users are asked to share their health status using apps and web-based platforms [313]. Examples are Influzenanet which is available in different European

countries [298], and Outbreaks Near Me (ex Flu Near You) in the USA [299] that engage tens of thousands of users that together provide the information necessary for the creation of interactive maps of ILI in almost real-time. In passive data collection, instead, information about individuals' health status is mined from other available sources that do not require the active participation of users. News articles [297], queries on search engines [304], posts on online social networks [305, 306, 301, 307, 308], page view counts on Wikipedia [309, 310] or other online/offline behaviors [311, 312] are typical examples. The original growth of digital epidemiology was largely fueled by the rapidly increasing amounts of data generated on the internet, particularly on social media. Google Flu Trends (GFT) was among the earliest well-known examples of digital epidemiology, leveraging symptomatic search queries for the purpose of tracking influenza-like illnesses [304]. The specific problems with GFT have been well described and discussed [314, 295], but a larger problem was the private ownership of the underlying data, which meant that the algorithm could not be reproduced and investigated in an independent way. Twitter and Wikipedia have also been exploited for surveillance purposes [305, 306, 301, 307, 308]<sup>19</sup>.

### The moment of truth for epidemic modeling: The COVID-19 pandemic

We have already outlined the significant disruption caused worldwide by the COVID-19 pandemic. What began as a minor curiosity in a remote location on the early morning news quickly escalated. It turned into a situation where entire cities, home to millions, were put on standby. Then, within just a couple of months, the crisis was at our doorstep, infiltrating our daily lives. The epidemic, now a pandemic, was not only spread across the globe but also struck particularly hard in developed countries that typically avoid such calamities.

Consequently, news and daily conversations became singularly focused on the COVID-19 situation, both nationally and internationally. Prime-time media and mainstream journals, usually dominated by politics and current affairs, were overtaken by pandemic-related discussions. Standard commentators were replaced by a variety of experts [317], and even technical concepts such as compartmental models [318, 319], “flattening the curve” [320], the basic reproductive number [321, 319], and superspreaders [322], became common knowledge to the general public.

As evident, epidemic modeling has increasingly become significant in public health policy formulation over the past decades. With the onset of the COVID-19 crisis, the importance of this field saw a considerable rise [323, 324, 325, 326]. These models have been instrumental in projecting the pandemic's trajectory, evaluating the impact

---

<sup>19</sup>We recommend the following references for a more exhaustive revision of the field [292, 315, 316].

of health interventions like lockdowns, border closures, and mobility restrictions, and anticipating potential side effects, among other uses. Moreover, given the scarcity of alternative decision-making tools, such as robust experimental evidence, particularly in the pandemic’s early stages, these models emerged as the primary guides for political decision-making [327].

In January 2020, models focusing on travelers from Wuhan highlighted significant underreporting of early COVID-19 case rates both within China [328] and internationally [329], leading to swift border closure measures. China imposed a travel lockdown in Wuhan on January 23rd, followed by partial restrictions in various countries throughout February, escalating to more comprehensive measures in March. Many governments relied on travel risk models to project case counts with and without border closures (as seen in Shearer et al. [330]). Advanced interconnected meta-population travel models provided deeper insights. Retrospective analyses revealed that while the Wuhan lockdown had minimal impact on delaying the outbreak within China, it significantly affected the spread to other countries [331]. These models also predicted the shift of outbreak epicenters from Asia to Europe, then from the USA to South America and Africa, reflecting regional interconnectedness [332], aiding in enhanced monitoring in vulnerable nations. Initial models typically estimated a high reproduction number, averaging two to three in China before interventions [333], highlighting the pandemic’s severity. These assessments consistently indicated that an uncontrolled epidemic would heavily burden healthcare systems and result in significant loss of life [334, 328, 335]. Notably, in the United Kingdom, the government’s response was significantly influenced by epidemiological models from London’s Imperial College. A key model revision regarding ICU bed occupancy led to a substantial increase in the projected fatalities, directly contributing to the decision to implement the initial lockdown on March 23rd [336]. Similarly, in France, the scientific committee guiding public decisions during the pandemic relied extensively on epidemiological models [337], with their mid-March recommendations leading to measures aimed at reducing social interactions.

If the area of epidemic propagation modeling had long since entered a trend of exponential production, the COVID-19 pandemic broke its limits and has produced a veritable explosion of scientific content around various aspects of it. Going into details or an in-depth review of the work of mathematical modeling during the COVID-19 pandemic exclusively is beyond the scope of this thesis.

The pandemic has thrust mathematical modeling of infectious diseases, along with related medical and biosciences fields, into the spotlight of both scientific and general news. The interdisciplinary, real-time work undertaken to understand its effects on

human health, address the health crisis, unravel the complex spread of SARS-CoV-2, and develop effective vaccines rapidly, is commendable. Nevertheless, there are aspects that warrant critique and improvement. From our field’s perspective, the urgent need for quality, standardized, high-granularity data to feed increasingly sophisticated models is clear. Such models are essential for exploring specific interventions to maximize the health of individuals and, in turn, minimize the socioeconomic impact. This data collection should primarily be a task for public health surveillance services, but it requires a dialogue with scientists and modelers to better align each other’s needs and capabilities. Another significant challenge is achieving advances in modeling and incorporating human responses during an epidemic similar to the one we have experienced.

## 1.4 Focus, structure, and goals

As the reader may have appreciated from the field overview, the modeling of epidemics incorporates various elements, some of which originate from very disparate worlds and knowledge domains. Clearly, without pathogens, there is no spreading, unless one broadens the definition of contagion to include information, rumors, knowledge, or cultural memes, which is not the focus here. We concentrate on disease toy models for influenza-like illnesses, exemplified by the paradigmatic SIR compartmental model. Despite its simplicity, this model can provide important insights and reasonably accurate predictions for more complex diseases like COVID-19. Another crucial aspect of spreading phenomena is the host population experiencing the epidemic outbreak. Pertinent questions and modeling efforts here involve the interplay between the spread of the disease and the structure and heterogeneity of the susceptible population. In this thesis, we explore a diverse range of scenarios, from models involving homogeneous populations to those introducing varying degrees of heterogeneity, such as age, behavioral response, or mobility patterns. Our research aims to address questions relevant to public health policymaking and to gain deeper insights into specific epidemic phenomena. Our primary methods will be *in silico* experiments conducted through extensive stochastic mechanistic simulations, meaning our approach will be predominantly computational rather than analytical, informed by real data whenever possible.

This thesis is structured in several parts. *Part I: Overture* contains the introduction and a detailed presentation of the frameworks underpinning our work. *Part II: Metapopulation models* includes a selection of works within the metapopulation framework, emphasizing systems where spatial structure and mobility are key. In Chapter 3, we assess the effectiveness of perimeter lockdowns in containing urban-scale epidemics, using Madrid during the COVID-19 pandemic as a case study. In Chapter 4, we couple SIR spreading dynamics with the d-EPR model of human mobility to study the role of heterogeneous mobility patterns in urban epidemic outbreaks. *Part III: Single-population behavioral structured models* presents two works focusing on heterogeneous single-population models with human behavior features. In Chapter 5, we explore the impact of secondary COVID-19 outbreaks across the heterogeneous landscape of age and hesitancy in the US using an age-structured SIR model. In Chapter 6, we continue exploring vaccine hesitancy by investigating a model that couples an epidemic with a vaccination campaign and opinion dynamics on vaccine uptake in a network-structured population. *Part IV: Closure* is dedicated to conclusions, where Chapter 7 summarizes our main findings and offers perspectives for

future research, extensions, and improvements.

Ultimately, this thesis has a dual purpose. First and foremost, it aims to make novel, albeit modest, contributions to the extensive body of knowledge in epidemiology and epidemic spreading modeling, enriching the public discourse on the topic and aiding in public health policy decision-making. Secondly, on a more personal note, this thesis has also been an exciting and challenging journey of learning for myself.

## 1.5 List of works

The content of chapters 5 and 3 correspond entirely and chronologically to the following publications:

- de Miguel-Arribas, A., Aleta, A., & Moreno, Y. (2022). Impact of vaccine hesitancy on secondary COVID-19 outbreaks in the US: an age-structured SIR model. *BMC infectious diseases*, 22(1), 1-12 [338].
- de Miguel Arribas, A., Aleta, A., & Moreno, Y. (2023). Assessing the effectiveness of perimeter lockdowns as a response to epidemics at the urban scale. *Scientific Reports*, 13(1), 4474 [339].

The content of chapters 4 and 6 is not yet published and we are currently working in framing and writing a draft for future publication.

Additionally, included in this period but not as a part of the main body of work:

- de Miguel-Arribas, A., Morón-Vidal, J., Floría, L. M., Gracia-Lázaro, C., Hernández, L., & Moreno, Y. (2023). Contests in two fronts. arXiv preprint arXiv:2303.18109 [340].



# Chapter 2

## Framework

*I simply wish that, in a matter which so closely concerns the well-being of mankind, no decision shall be made without all the knowledge which a little analysis and calculation can provide.*

— Daniel Bernoulli

*As a matter of fact, all epidemiology, concerned as it is with the variation of disease from time to time or from place to place, must be considered mathematically, however many variables as implicated, if it is to be considered scientifically at all.*

— Sir Ronald Ross

*España no va a tener, como mucho, más allá de algún caso diagnosticado de COVID-19.*

— Fernando Simón

### 2.1 Elements of epidemic spreading modeling

#### 2.1.1 Compartmental models

Compartmental models are fundamental in modeling the spread of infectious diseases. Their basic principle involves identifying the relevant states of a population undergoing a dynamical process and categorizing it into corresponding compartments. While this approach is broadly applicable beyond epidemiology, our specific interest lies in the health statuses of the population under study.

To illustrate, let us consider a hypothetical influenza-like illness spreading within a human population. Individuals can be classified into one of the following three health statuses or microstates:

- Susceptible ( $S$ ): Individuals that had no previous contact with the disease.

- Infected/Infectious ( $\mathcal{I}$ ): Individuals that are hosting the pathogen and with the ability to transmit the disease to others.
- Removed ( $\mathcal{R}$ ): Individuals that suffered the disease but are no longer infected or infectious.

At any given time  $t$ , we can categorize individuals by their microstate, leading to three macrostates or state variables representing the total number of individuals in each compartment. The SIR model, a standard model for diseases that confer permanent (or long-lasting) immunity, is based on this three-state compartmentalization<sup>1</sup>.

Having defined the compartments, the next step is to determine how individuals transition between them. During an outbreak, susceptibles may become infected, while infected individuals may recover or die, exiting the infectious stage. These transitions can be schematically represented using the language of chemical reactions [50]. For the SIR model, the transitions are as follows:



The first reaction, expressed as (2.1), represents the infection process where a susceptible individual becomes infected after contact with an infected individual. This occurs at a constant rate  $\beta$ , commonly known as the infection or transmission rate. The second reaction, (2.2), models the recovery or removal of infected individuals. This is a spontaneous process happening at a constant rate  $\mu$ , which is the recovery or removal rate. The inverse of  $\mu$  gives us the average infectious period  $T_I$  characteristic of the disease.

In this model, the microstates are dynamic variables, as are the macrostates. Developing a model that explains and predicts the changes in these quantities is the essence of the physics of epidemic spreading. In seeking a basic deterministic approach to this problem, differential equations become our tool of choice, serving as the natural language for expressing and analyzing compartmental models. These models start by abstracting away individual differences, focusing instead on state-based population counts and the flows between these states.

To build a deeper understanding of the modeling process, let us derive the ordinary differential equations representing the SIR model from first principles. Denote the

---

<sup>1</sup>The specific compartmentalization depends on the disease being modeled and the aspects we aim to analyze. For example, COVID-19's complexity is not fully captured by an SIR model, making it inadequate for precise real-time forecasting. However, it might suffice for understanding general effects or rough impact estimates. Another standard model in epidemiology is the SIS model, where infected individuals become susceptible again after recovery, making it suitable for diseases like the common cold or certain sexually transmitted infections.

populations in each compartment at time  $t$  as  $S(t)$ ,  $I(t)$ , and  $R(t)$ . We will analyze the net change in the number of individuals in each compartment over a small time interval  $[t, t + \Delta t]$ . This net change, represented by  $\Delta X(t)$  with  $X = S, I, R$ , is simply the difference between the number of individuals entering and leaving the compartment during this interval. By dividing these changes by the length of the interval and taking the limit as  $\Delta t \rightarrow 0$ , we obtain the rate of change of each compartment:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\text{new infections rate}, \\ \frac{dI(t)}{dt} &= +\text{new infections rate} - \text{removal rate}, \\ \frac{dr(t)}{dt} &= +\text{removal rate},\end{aligned}$$

Now, importantly, how the right-hand terms in the above expressions behave as functions of  $S(t)$ ,  $I(t)$ , and  $R(t)$  is something dependent on the hypotheses on the particular characteristics of the population and the spreading process undergoing. As we adopted a deterministic continuous-time approach, it is normally more convenient to express the dynamical equations in terms of the population *densities* (as normalized by the total population) than as integer numbers, and thus we make the following simple transformations:  $s(t) \equiv S(t)/N$ ,  $i(t) \equiv I(t)/N$ , and  $r(t) \equiv R(t)/N$ , where  $N$  is the total population of the system, which will stay constant throughout the whole dynamics, satisfying  $N = S(t) + I(t) + R(t)$  or equivalently  $1 = s(t) + i(t) + r(t)$ . The standard SIR model then reads:

$$\frac{ds(t)}{dt} = -\beta \langle k \rangle s(t)i(t), \quad (2.3)$$

$$\frac{di(t)}{dt} = \beta \langle k \rangle s(t)i(t) - \mu i(t), \quad (2.4)$$

$$\frac{dr(t)}{dt} = \mu i(t). \quad (2.5)$$

Before going into details about its landscape of solutions, it is worth taking a dive into some crucial assumptions that this model carries:

- Population homogeneity. The population is assumed composed of homogeneous or identical individuals. In fact, there is no such thing as the *individual* in compartmental models. This allows for a characteristic transmission rate in the infection process  $\beta$  and a characteristic removal rate  $\mu$  for the whole population. This is a first strike at realistic and potentially relevant individual traits of a population such as sex, age, or socioeconomic status.

- Well-mixed interactions. The term  $\beta i(t)$  (or  $\beta I(t)/N$ ) is called the per-capita *force of infection* and has the form of the so-called *mass action law* which gives rise to a mean-field of infection experienced by the average susceptible individual, serving the final blow to individuality in this approach. Well-mixed interactions assume a random mixing among the population and thus the contact between any two individuals in the system is totally random.
- Markovian approximation. The state of the system is independent of the dynamical history of the system but the immediately previous state. This property is also called memoryless. This is expressed in the assumption of exponentially distributed times for the infectious period.

These are the most relevant assumptions underlying these questions. However, this model carries some other simplifications worth mentioning.

- Simplified disease's natural history. The SIR model omits certain crucial aspects. For instance, it assumes no latent period, no fatalities due to the disease, and permanent immunity post-infection. Some of these faults are corrected in other toy models like SEIR (latency) or the SIRS (immunity loss), and surely more compartments are added when modeling real diseases with accurate forecasts in mind.
- Single population (no spatial structure). The model does not consider any spatial structure, ignoring spatial propagation, mobility patterns, and the connectivity of human settlements. It also assumes a constant population size without accounting for births and deaths.
- Absence of human behavior. Unless symptoms go unnoticed or are very mild, typically humans react, either through self-organization or following some hierarchical top-down control measures, to avoid disease contagion or mitigate its consequences. None of this is considered in the standard models.

These limitations have been recognized since the earliest epidemiological compartment models by Kermack and McKendrick and have been progressively addressed in the field, as discussed in our field overview (Section 1.3). For instance, the well-mixed assumption, though still common, has been refined through the introduction of structure and contact matrices and superseded by network epidemiology. Spatial structure and human mobility have been successfully incorporated via the metapopulation framework, and human behavior has increasingly been factored into models to explore intervention effects and epidemic-behavior feedback loops.

Returning to the SIR model's dynamics, two key questions naturally arise:

- Under which conditions does an epidemic outbreak occur?
- Given such an outbreak happens, what is the expected final size?

The first question relates to the early time of the epidemic unfolding whereas the second one obviously refers to the late time dynamical evolution. To answer the first question we can explore under which conditions  $dI(t)/dt|_{t=0} > 0$ . It is assumed, additionally, that when a new pathogen starts propagating host-to-host in a population, this will be fully susceptible to it, and also initially, only a very small fraction of individuals will be infected. All in all, we have:

$$\left. \frac{dI}{dt} \right|_{t=0} > 0 \quad \longrightarrow \quad \beta S(0) \frac{I(0)}{N} - \mu I(0) > 0 \quad \longrightarrow \quad \beta - \mu > 0 \longrightarrow ; \quad \frac{\beta}{\mu} > 1. \quad (2.6)$$

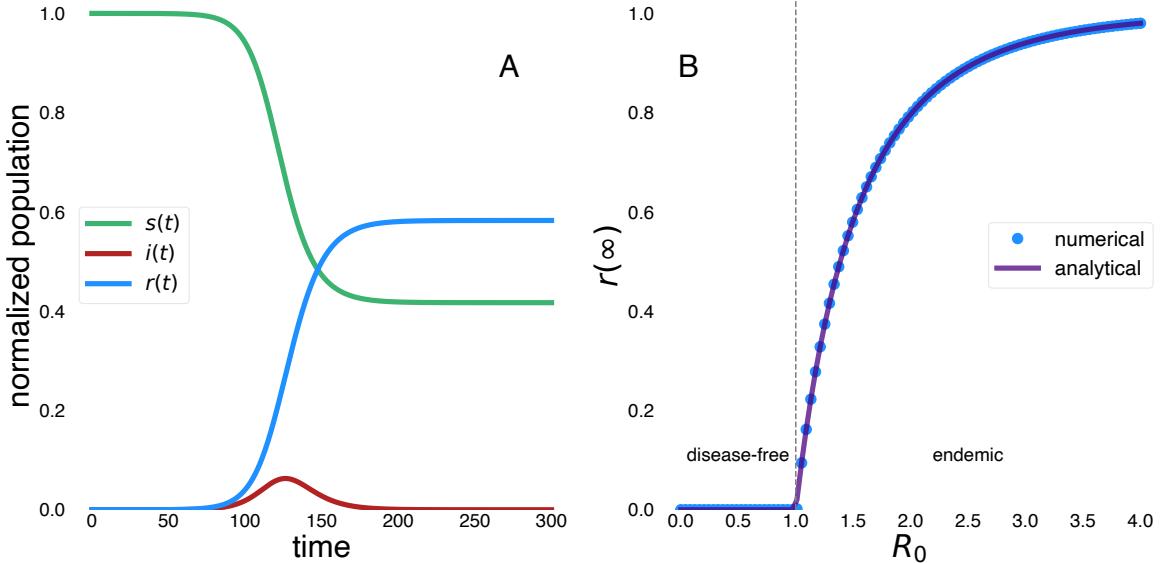
The two last relationships tell us in different ways about the possibilities of the growth of the epidemic. The first one,  $r \equiv \beta - \mu$  is the exponential growth rate, and the second one,  $R_0 \equiv \frac{\beta}{\mu}$  is the so-called basic reproduction number. This quantity, which we will treat in more detail in the coming Section 2.1.2, is defined as the average number of cases generated by the typical infectious host in an otherwise fully susceptible population. Then, if  $R_0 > 1$ , translates into  $dI/dt > 0$ , because on average the number of cases generated is greater than one, whereas if  $R_0 < 1$ , a sizeable outbreak cannot be sustained since contagion chains are unable to generate more than one case. We see then clearly a threshold behavior in this model at  $R_0^c = 1$ , which receives the name of the epidemic threshold, and marks a bifurcation point in the solution landscape.

When  $R_0 < 1$ , the only stable solution in the system is  $r(\infty) = 0$ , which we refer to as the *disease-free* solution, whereas if  $R_0 > 1$ , conditions are ripe for an outbreak. The SIR dynamics presents an absorbing state (an infinity of them, actually) when the population in the infected compartment reaches zero, something that is typically represented by  $i(\infty) = 0$ . This means that once the system exhausts the infected population, the dynamics grinds to a halt and the epidemic dies out. Answering the basic second question posed, it can be shown after basic algebra that the final epidemic size or prevalence in the population at the absorbing state, can be obtained from the following self-consistent equation:

$$r(\infty) = 1 - e^{-R_0 r(\infty)}, \quad (2.7)$$

where it has been assumed that  $r(0) = 0$ , even though this is not a necessary condition to arrive at a solution. Above the epidemic threshold, the disease-free solution turns globally unstable and a new family of solutions emerges referred to as the *endemic* solution. These solutions are infinite in number since they depend on  $r(\infty)$ , and are globally stable. It is worth mentioning that even under a highly *aggressive* contagion

as given by  $R_0$ , the disease will not infect the whole system's population. Figure 2.1 depicts the essential behavior of the SIR model, both dynamically (Panel A) and at the absorbing state (Panel B). For more details about the stability analysis of the SIR model and other standard compartmental models, the reader is invited to follow [341].



**Figure 2.1: SIR model.** Panel A: Numerically-solved time evolution of the normalized populations by health status: susceptible  $s(t)$ , infected  $i(t)$ , and removed/recovered  $r(t)$ , with  $R_0 = 1.5$ . Panel B: Phase diagram for the normalized disease prevalence or removed/recovered fraction at the absorbing state  $r(\infty)$ . At  $R_0 = 1$ , a bifurcation occurs, moving from the disease-free solution to the endemic solution. The numerical integration (dots) is compared with the solution of equation 2.7 (line).

This exposition has provided a basic introduction to the SIR compartmental model, outlining what can be expected from its application. We have introduced the dynamical equations, examined the threshold behavior, and discussed the formula for computing the epidemic size of an outbreak (also known as the attack rate). Additionally, we have highlighted the fundamental assumptions underlying the model and, consequently, its limitations. For a more exhaustive introduction to conventional epidemiological compartmental models, we recommend consulting [342].

As we conclude this subsection, it is important to acknowledge another implicit assumption that has been present since the outset, albeit somewhat concealed. This assumption is the use of ordinary differential equations, representing a continuous-time deterministic approach to modeling the spreading dynamics. This choice implies a specific perspective on how the disease spreads and interacts with the population, and it shapes the conclusions and insights that can be drawn from the model.

Actually, a more general and realistic approach is to conceive the dynamical processes involved here as stochastic processes and formulate the problem accounting

for the genuine uncertainty in the knowledge of the system's state. The reactions previously presented are stochastic in nature, and thus infection and recovery transitions are modeled as jump (Markovian) stochastic processes, shifting the quest from finding  $(S(t), I(t), R(t))$  to rather  $P(S, I, R; t)$ , where this  $P$  is the joint probability for the system of being at state  $(S, I, R)$  at time  $t$ . The equation that accounts for the evolution of  $P(S, I, R; t)$  is the so-called master equation. Solving this equation, even for the steady or absorbing state, can be far from a trivial task, but one can get good enough insights from the system behavior through a series of approximations or looking for the moment distribution evolution equations. In fact, and here we connect the dots between formalisms, it is not a difficult (maybe a bit tedious) task to arrive at the result that the evolution equation for the first moments of the  $P(S, I, R; t)$  distribution, corresponds to the dynamical equations of the deterministic model formulation [171]. In other words, the standard deterministic dynamical equations of a given model represent an average value of the underlying stochastic model. Stochastic formulations are favored in the context of simulations and also when formulating mean-field theories in network models [164, 163]. When dealing with very small systems, stochasticity due to finite size effects is crucial, and thus adopting a stochastic formulation becomes a more adequate option. Importantly, in the specific context of epidemic models, even when  $R_0 > 1$ , there is a finite non-negligible probability of ending up without a sizeable outbreak, and thus in the corresponding free-disease state. In fact, this probability is approximately given by  $(1/R_0)^{I(0)}$  in the standard SIR model, where  $I(0)$  is the initial number of infected individuals [343]. Finally, a complementary stochastic approach early adopted in epidemic modeling is that of branching processes [343] and the generating function, which has been used in various situations, to deal with early-time epidemics [140] or compute the basic reproduction number in single population systems [344], networks [164], and metapopulations [217, 228].

Beyond stochastic and deterministic schemes, one also has the freedom to adopt a continuous-time or discrete-time approach. An example of a stochastic and discrete-time approach is the Microscopic Markov Chain, a formalism suitable for dynamical processes on networks [345], or any computational Monte Carlo implementation of a dynamical system. As for deterministic discrete-time approaches, these are not as popular as their continuous-time counterparts, but still, several works have dedicated some attention to them [346, 347, 348, 349, 350]. Notably, Allen [346] found that even though the regular SI and SIR models can exhibit chaotic behavior, a logical restriction to remain physically plausible (i.e., have positive solutions) smashed that option. However, in the case of discrete SIS or SIR models with births and

deaths, period-doubling and chaos were possible thanks to the positive feedback to the susceptible class. More than two decades later, Fennell et al. [348], in the context of epidemic models on networks, criticized some conclusions in Allen [346], pointing out that the appearance of complex behavior, like period doubling and chaos, is no more than an artifact of the time discretization. The authors' main goal in [348] is to frame the limitations of conceiving continuous-time contagion dynamics as a discrete-time process. It is then found that the accuracy of such methods is rather poor if state transition probabilities are too large, leading to noticeable deviations from the underlying continuous-time process. Some other authors have reflected on the problem of the equivalence between continuous-time vs. discrete-time approaches and its limitations in different contexts [347, 349, 351]. It remains an open problem to fully characterize how these approaches compare and should be used.

### 2.1.2 The basic reproduction number

The origins of the basic reproduction number can be traced back to demography [352, 353, 354, 355] as earlier as prior to the 20th century. In epidemiology, it was independently studied for vector-borne diseases such as malaria [59, 356] and directly transmitted human infections [67, 357, 358].

#### Definition and computation

Previously, we calculated  $R_0$  for the standard SIR model in a straightforward manner. However, that derivation was somewhat informal and lacked generality in its definition. There are several methods for formally defining or deriving the basic reproduction number  $R_0$  more broadly and for more complex models. Among these, the *survival function method* is considered the “gold standard” for determining this key epidemiological quantity.

Consider a large population and let  $F(a)$  be the probability that a newly infected individual remains infectious for at least a time since infection  $a$ . This is what is referred to, intuitively, as the survival probability. Also, let  $b(a)$  denote the average number of newly infected that an infectious individual will produce per unit of time when infected for the total time since infection  $a$ .

Then,  $R_0$  can be obtained as:

$$R_0 \equiv \int_0^\infty b(a)F(a)da. \quad (2.8)$$

The survival function method is adept at handling scenarios where infectivity varies with the time since infection or when transmission probabilities between states change

over time. Consequently, the derivation of  $R_0$  using this method is not confined to systems described solely by ordinary differential equations. This method can also be naturally extended to describe models in which a series of states are involved in the “reproduction” of an infected individual, such as vector-borne disease and the epidemic modeling of malaria [359].

It is worth remarking that  $R_0$  is a measure of the extent of an epidemic, not the speed at which the infection grows (given by the growth rate  $r$ ).

In heterogeneous systems that encompass multiple classes of infected individuals, the survival function method’s natural extension is the next-generation operator or matrix. This approach was first introduced by Diekmann et al. [117]. It is particularly applicable in scenarios involving more than one class of infectives, where the population is segmented into discrete, non-overlapping classes. This method has been employed in models with underlying age structure [360] or spatial structure [228], among other applications. In typical implementations, continuous variables within the population are represented by several discrete classes. This approach presumes that transmission probabilities between states follow Markovian processes, or equivalently, that residence times in each state are exponentially distributed.

The next generation method defines  $R_0$  as the spectral radius of the next generation operator, that is, the largest eigenvalue. Then, if  $\mathbf{R}$  is this next generation operator and we denote by  $\rho(\cdot)$  the spectral radius of an operator,  $R_0$  is given by  $R_0 = \rho(\mathbf{R}')$ . In the case of a SIR model in a heterogeneous population, when there is a characteristic transmission rate  $\beta$  and removal rate  $\mu$  in the system, through a simple transformation, one can refactor the next-generation operator and express  $R_0$  as

$$R_0 = \frac{\beta}{\mu} \rho(\mathbf{R}). \quad (2.9)$$

This way, we can directly appreciate that for the heterogeneous population,  $R_0$  is the product of the homogeneous or well-mixed SIR case times the next generation spectral radius accounting for the now genuine heterogeneity in the system. The reader is invited to consult [344] for a complete exposition of the method and explicit construction of the next-generation operator, and [361] for an illustrative application of this method to a series of standard compartmental epidemic models.

There exist other quantities that play a similar threshold role as  $R_0$  but that do not have a biological meaning or cannot be defined precisely as the typical number of secondary infections. These can be reviewed in [359] and they are the Jacobian and its stability conditions, the conditions for the existence of the endemic equilibrium (transcritical bifurcation), and the constant term of the characteristic equation.

### Empirical estimation and the effective reproduction number $R_t$

Some of the theoretical methods proposed to compute  $R_0$  involve a previous knowledge of parameters that are hard to estimate from direct measurements, such as the contact or transmission rate. For that reason,  $R_0$  is rarely computed from the survival method or the next-generation method. There exist, though, empirical methods relying on epidemiological data to directly estimate  $R_0$ . However, these methods typically involve simplifying assumptions to reduce the number of unknown parameters. Some of these methods are (i) the susceptibles at endemic equilibrium, (ii) the average age of infection, (iii) the final size equation, and (iv) the intrinsic growth rate. For a detailed account of these methodologies, including their flaws or limitations, the interested reader may follow [359] and references therein.

An important limitation of  $R_0$  is that, as its definition clearly shows, only applies in the very early time of the emergence of an epidemic outbreak, since relies on the vague concept and hard assumption of a fully susceptible population. A related quantity to compute and assess beyond that limited initial time window is the effective reproduction number,  $R_t$ <sup>2</sup>. Indeed, this is the proper monitoring quantity to look at when conceiving and assessing the effectiveness of the intervention measures deployed to mitigate or control an epidemic [362]. Similarly, as with  $R_0$ ,  $R_t$  is defined as the actual average number of secondary cases per primary case at calendar time  $t$  (for  $t > 0$ ) [363, 364, 365, 366, 367, 279].  $R_t$  shows time-dependent variation due to the decline in susceptible individuals (intrinsic factors) and the implementation of control measures (extrinsic factors). If  $R_t < 1$ , it suggests that the epidemic is in decline and may be regarded as being under control at time  $t$  (vice versa, if  $R_t > 1$ ). For the standard SIR model, the effective reproduction number is simply computed as:

$$R_t = \frac{\beta}{\mu} s(t), \quad (2.10)$$

where  $\beta$  is the disease's transmission rate,  $\mu$  is the removal rate, and  $s(t) = S(t)/N$  is the susceptible density at time  $t$ . Indeed, either naturally through the decrease of  $s(t)$  as the epidemic marches on, or through interventions that modify the transmissibility  $\beta$ ,  $R_t$  changes and will eventually cross the unit threshold, going  $R_t < 1$ , and signaling that the epidemic is under control. For a more detailed account of this quantity, we refer the reader to [368].

---

<sup>2</sup>This quantity is often denoted by  $R(t)$ , but this can bring confusion with the standard quantity denoting the number of recovered/removed individuals in time within a population.

### Relevance and general properties

In the context of the mathematical modeling of infectious diseases, one of the main goals when devising a new epidemic model is the computation of the corresponding  $R_0$ , in order to characterize the threshold behavior and its consequences for the model solutions and control strategies.

Throughout this revision, we have already introduced the threshold property of  $R_0$ , which is the most important and useful aspect of this concept. For  $R_0 < 1$ , clearly, each infected individual produces, on average, less than one new infected individual, making the spreading die out. If  $R_0 > 1$ , the pathogen will be able to successfully invade the susceptible population. Apart from the threshold behavior, several other properties have been attributed to  $R_0$ . It is not only the threshold  $R_0 = 1$  separating the endemic and the disease-free solutions that is important but its magnitude. The separation of  $R_0$  from 1 conveys information about the efforts required to contain an emerging infectious disease. Moreover, knowing  $R_0$  can inform us about the final epidemic size in a population, the required level of herd immunity, or the probability of extinction of an outbreak [369]. Admittedly, though, these estimations rely on highly simplifying assumptions, and care must be taken when used. In fact, this warning connects with the dark side of the concept  $R_0$  which will be exposed in the next lines.

Undoubtedly,  $R_0$  has long become a key concept not only in mathematical epidemiology but also in public health epidemiology. It has been described as “one of the foremost and most valuable ideas that mathematical thinking has brought to epidemic theory” [119]. Surely, it has helped the understanding of epidemic-related phenomena both theoretically and empirically and has served to inform control and mitigation strategies in real-life epidemics and pandemics. However, life is not so rosy for  $R_0$ , and the concept [370], estimations, and applications [371, 372, 373] come with several flaws recognized in the literature.

### Flaws

On the practical side of things, one of the first things to notice is that the definition of  $R_0$  speaks of a *totally* susceptible population. On the one hand, hardly ever an epidemic of newly emerging diseases can be observed from its inception. On the other hand, populations can show a non-negligible level of immunity for recurrent epidemics. Regarding the estimation of  $R_0$ , one of the common mistakes committed is to assign general  $R_0$  values for specific diseases, like if carved in stone. This assumes that estimations are unequivocally attached to diseases/pathogens and allows for extrapolation. The truth is that estimations of  $R_0$  for a particular infectious

disease have limited practical value outside the population from which the disease data originated [371]. Moreover, it has long been established that the factors shaping  $R_0$  are not disease-specific but rather contextual, involving the structure of the population in different aspects: age structure, social mixing, mobility modes, or behavioral attitudes [374, 362].

As an average, either  $R_0$  or  $R_t$  miss crucial heterogeneities as the spatial one or those inherent to the population. For instance, averages can miss regional clusters of infection. Conversely, high incidences of infection among a spatially distinct smaller subsection of a population can sway a larger region's  $R$  value [372]. Local  $R_t$  numbers become less accurate as they are applied to smaller populations, especially when absolute infections are low. Regarding population, at least for certain diseases, the existence of superspreaders can totally change how we face control measures. It was found during the COVID-19 spreading that as few as 10-20% of infected people seem to contribute to 80% of new case generation [375]. Actually, acknowledging these heterogeneities can greatly help us to focus efforts and conceive more efficient mitigation strategies, like banning events where superspreading events have a higher chance of happening.

Some fundamental properties commonly attributed to  $R_0$  are: (i) an endemic infection can persist only if  $R_0 > 1$ , (ii)  $R_0$  provides a direct measure of the control effort required to eliminate the infection. The first claim can fail due to the presence of backward bifurcations, even though, admittedly, this can be a very rare phenomenon [374]. More naturally, due to stochastic effects, an epidemic can die when  $R_0 > 1$ . The second can fail when control is applied unevenly across different host types (the problem of  $R_0$  being an average value) [374]. Conceptually, it has been argued [376] that in order to associate an  $R_0$  to a model of ODEs, an individual-level model (ILM) compatible with the ODE model must be developed, and only in those situations can the ILM  $R_0$  be unambiguously calculated. The author went on to show that two individual-level models having exactly the same expectations of the corresponding population-level variables may yield different  $R_0$  values. Obtaining  $R_0$  from empirical contact-tracing data collected and using this  $R_0$  as a threshold parameter for a population level could produce misleading estimates of the infectiousness of the pathogen, the severity of an outbreak, and the strength of the medical and/or behavioral interventions necessary for control.

## New perspectives

Recently, efforts have been addressed to offer new solutions, either alternatives or complements to  $R_0$ , to improve the theoretical underpinnings of models and the practice

of outbreak analysis and forecasting. In [375], the authors try to go beyond  $R_0$  to forecast the size of an epidemic by leveraging on the distribution of secondary infections, not only on its average ( $R_0$ ) but on the underlying heterogeneity. They show that epidemics with lower  $R_0$  can be larger if they spread more homogeneously, recognizing that  $R_0$  is insufficient in fully determining the final size of an outbreak. Both  $R_0$  and the full secondary case distribution are not properties of the disease itself, but are instead set by properties of the pathogen, the host population, and the context of the outbreak. These factors, hard to ponder and properly model, crucially shape the outcomes of spreading beyond a single quantity.

### 2.1.3 The force of infection

The force of infection (FoI) is the per-capita rate at which susceptible individuals contract a disease during an epidemic. As a dynamic quantity, it varies throughout the course of an epidemic. Historically, two main conceptualizations of FoI have emerged: frequency-dependent (FD) and density-dependent (DD) transmission. Some studies have also proposed models that stand in an intermediate regime of both these approaches.

Historically, the FoI or the transmission interaction in ODE models had been represented as  $\beta SI$ , where  $\beta$  is the transmission rate, and  $S$  and  $I$  represented spatial densities of the hosts. This is what was known as mass-action transmission, borrowing the term from chemistry, by metaphorically treating host populations as chemical reagents [377]. This had been largely unquestioned until 1995 when de Jong et al. [126] published a paper that was widely interpreted as claiming that the term  $\beta SI$  was indeed a model of “pseudo mass action”, and that transmission should follow “true mass action”, suitably represented by a term of the form  $\beta SI/N$  instead. Since then, models were divided into those using either form of transmission which also sparked some confusion in the terminology. Indeed,  $\beta SI$  is sometimes described as “mass-action”, sometimes as “density-dependent” transmission; and  $\beta SI/N$  is sometimes referred to as “mass action”, and sometimes as “frequency-dependent”. To read more details about the controversy, we redirect the reader to [126, 377, 129]. In this subsection, we will delve into the origins of these differing expressions for the force of infection.

Let us consider a spreading dynamics in a single homogeneously mixed population of size  $N$  within in a physical area  $A$ . Focusing on an SI dynamic (no recovery process), but without loss of generality, we concentrate on the infectious population’s time evolution, represented only by the gain term, which is the force of infection itself.

The transmission term can be expressed as:

$$\frac{dI}{dt} = c \times p \times v \times S. \quad (2.11)$$

Now, what does this product of factors mean? We have expressed the FoI as the product of three contributions:

1. The contact rate  $c$ ,
2. The probability,  $p$ , that a susceptible-infectious host contact occurs, and
3. The probability,  $v$ , for that contact to successfully lead to disease transmission.

The last term,  $v$ , is usually assumed constant for any given host-pathogen combination. The second term,  $p$ , is usually assumed to be  $p = \frac{I}{N}$ . This clearly depends on the assumption that what applies globally, as a proportion, to the whole population also applies “locally”, as a probability, to the given susceptible individuals in a neighborhood. Then, the root of the differences between the FD and the DD FoI lies in the interpretation of the contact process in itself.

The first, and most frequently assumed possibility is that  $c = \kappa N/A$ ; that is, the contact rate increases directly with the population density  $N/A$ , scaled by a constant  $\kappa$ , which may vary with the host-pathogen combination. In fact, in standard ODE models, the transmission rate  $\beta$  actually encodes the product  $\kappa \times v$  (sometimes  $\kappa$  is expressed as  $\langle k \rangle$ , mainly in network epidemiology and related areas). All this leads to:

$$\text{DD : } \frac{dI}{dt} = S\kappa \frac{N}{A} \frac{I}{N} v = \beta \frac{SI}{A}. \quad (2.12)$$

Thus, not only does the contact rate increase with the overall density of the host population  $N/A$ ; but the per capita FoI also increases with the density of the infected population  $I/A$ . This expression represents what is often said to describe a density-dependent transmission.

A commonly assumed alternative is to consider that the contact rate is constant,  $c = \eta$ , irrespective of the population density. To avoid confusion, mainly because units differ, we can then refer to the transmission rate under this scheme as  $\beta' = v\eta$ , leading us to

$$\text{FD : } \frac{dI}{dt} = S\eta \frac{I}{N} v = \beta' \frac{SI}{N}. \quad (2.13)$$

Here, the per capita force of infection increases with the prevalence of infection,  $I/N$ , which might also be called the “frequency” of the infected population. Then, this expression is what is often said to describe a frequency-dependent force of infection.

The question then naturally emerges: which one is correct (if at all)? And what are their ranges of application? In [378] it is underscored that the selection of a

mechanism holds no sway when examining a singular population of unchanging size due to an appropriate recalibration of the parameters. However, it is noted that disease propagation patterns will differ when modeling a large heterogeneous spatial area. The reproduction number  $R_0$  for frequency-dependent transmission is fixed regardless of population density, and therefore, does not have a population invasion or persistence threshold for the given parameters because  $R_0 > 1$ . For density-dependent transmission, the actual reproduction number depends on the population size and an invasion threshold. The mechanism, however, assumes indefinite linear contact rate growth with the population size, and it is questionable whether an individual has more contacts in a population of  $10^6$  than  $10^5$  [379, 380, 381]. Several laboratory studies have found that the DD model is inadequate for describing pathogen transmission, but also, for specific cases, alternatives other than the FD perform better [377]. The FD model remains the default modeling option in a vast majority of works within the scientific literature and for the most diverse situations, but some other mathematical forms have been explored [377, 129, 382]. A particular context where there is clearly an open problem regarding the adequate transmission form is that of household models. Given the specific features of households and interactions within, one might think that a DD FoI would be reasonably appropriate, and whereas some models have considered this type of transmission [383, 384], others maintain an FD approach [385], or have sought for alternative mechanisms [386, 387]. In any case, when looking at household attack rates for a series of diseases we see mixed results. Some studies have found that, counterintuitively, there is a peak in the attack rate for middle-sized households, and households beyond a certain size show a decreasing prevalence [388].

#### 2.1.4 Herd immunity

Herd immunity is another key concept in epidemiology and public health, and is intimately related to vaccination programs [389, 390]. It is stated that herd immunity occurs when a sufficient fraction of a population has become immune to a disease, whether through virus exposure in previous infections or through vaccination, thereby reducing the likelihood of further outbreaks and infections for individuals who lack immunity.

Once herd immunity has been reached during the course of an outbreak, the disease gradually disappears from the affected population. This might lead to eradication or at least a permanent reduction of infections to zero if achieved worldwide. It has been demonstrated that artificial induction of herd immunity via vaccination has contributed to the reduction of several diseases [391]. Conversely, losses of herd immunity levels due to vaccine hesitancy, have led to the resurgence of diseases in places that were once

suppressed [392].

We can apply some of the fundamental concepts reviewed in this section to illustrate the theoretical basis of the herd immunity concept. The critical value, or threshold, in a given population, is the point where the disease reaches an endemic steady state, meaning that the infection level is neither growing nor declining exponentially. For a well-mixed population, this threshold can be calculated from the effective reproduction number  $R_t$  by setting it at the critical value, that is, equal to 1. Indeed,

$$R_t = R_0 \times s(t) = 1, \quad (2.14)$$

which means that the epidemic growth is halted. Being  $s(t)$  the susceptible population density, we can rewrite it as  $1 - p$ , where  $p$  is the proportion of the population that acquires immunity, either *naturally* or *artificially*. Then, just by rearranging terms:

$$p_c = 1 - \frac{1}{R_0}. \quad (2.15)$$

Here we have renamed  $p$  as  $p_c$ , and represents the critical fraction of the population required to be immune in order to halt disease transmission. This is what is known as the herd immunity threshold (HIT) [393]. Thus, we can see also the relevance of  $R_0$  in the herd immunity effect and its role in setting the HIT: mildly expanding diseases (low  $R_0$ ) require a lower HIT, while more highly reproducible diseases demand a higher HIT to be reached. Moreover, if conditions change for  $R_0$ , pre-established levels of HIT will shift, either facilitating the outbreak suppression or, on the contrary, potentially triggering secondary outbreaks [393].

When an epidemic outbreak runs unmitigated, peak incidence marks the inflection point where  $dI/dt = 0$  and the dynamics shifts from a growing infected population to a decreasing one. This situation corresponds indeed with  $R_t = 1$ , meaning that herd immunity is a phenomenon that is naturally induced through the spreading dynamics. In the end, the final epidemic size  $r(\infty)$  is guaranteed to be  $r(\infty) > p_c$ . The reason is that, even when the HIT is reached, there is an overshoot of infected individuals that contribute to the generation of newly infected cases in the population and thus increase the prevalence at the time from  $p_c$  to  $r(\infty)$ . To be clear, then, when the HIT is reached, the number of additional infections does not suddenly drop to zero, and the spreading stops, but rather the point at which each infected individual infects fewer than one additional individual on average.

Even though the well-mixed setting is a pervasive assumption that may show good results in certain conditions, populations can show a high degree of structure and heterogeneity and this has implications in our standard estimations on the required HITs to halt a disease. [394, 395].

## 2.2 Models with structure: Heterogeneity enters

The SIR model introduced at the beginning of this chapter does not account for any type of heterogeneity within the population affected by disease propagation. But what exactly do we mean by heterogeneity in this context?

In any modeling framework, be it in epidemiology, economics, or another field, the standard approach often begins with heavily simplified assumptions, such as homogeneous mixing. This is done to glean insights into basic mechanisms in idealized situations before introducing more complex and realistic features. We understand that real populations do not behave like uniformly programmed automata randomly interacting with each other. While simple assumptions can sometimes be effective, they are often inadequate for capturing more complex phenomena. Recognizing that individuals and populations exhibit heterogeneity in various relevant factors is crucial for a deeper understanding and more accurate modeling. Therefore, a quest to model heterogeneity and to understand its implications becomes essential.

Populations consist of individuals with diverse characteristics - different sexes, ages, susceptibilities to diseases, levels of social and economic activity, locations, and so on. Recognizing and integrating some of these traits into our models is vital if we aim to advance our knowledge and offer more precise predictions and realistic public health policies. This effort to model heterogeneity has been a focus of epidemiology for the past several decades.

In this section, we will concentrate on two particularly relevant aspects of heterogeneity in the context of disease spreading: age and contact patterns. Other significant traits, like spatial structure and mobility patterns, will be discussed in detail in sections 2.3 and 2.4.

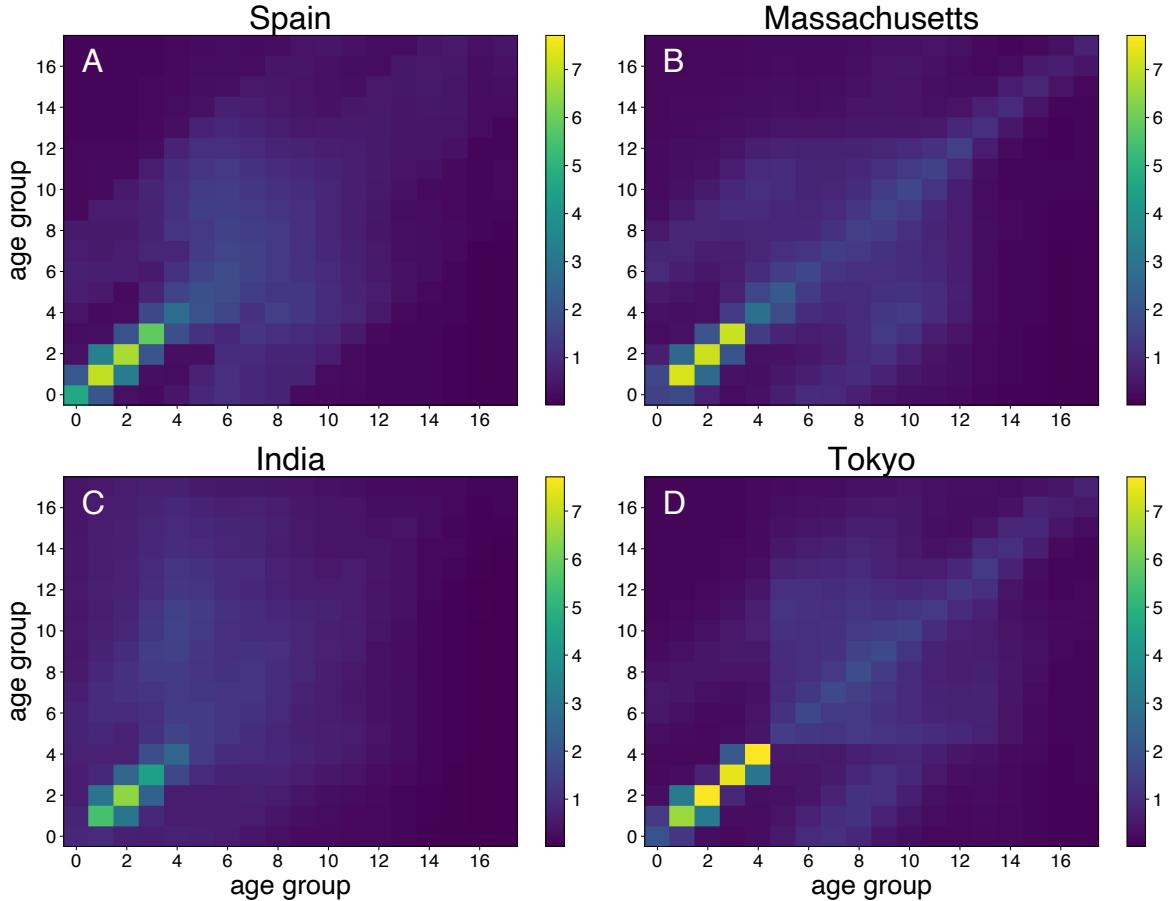
### 2.2.1 Age

Age is a crucial factor in epidemic modeling for several reasons. Firstly, it can significantly influence susceptibility to diseases and the progression of illnesses [396, 397]. Secondly, age plays a key role in shaping contact patterns within a population [398, 399, 400, 401, 402]. In recent years, numerous studies have dedicated considerable effort to gathering data on age-based mixing patterns. This data is used to construct contact matrices, which are instrumental in informing structured models for more accurate forecasting and detailed analysis of impacts across different age groups [403, 404, 405, 406, 360]<sup>3</sup>. Furthermore, models with higher resolution enable the exploration of interventions targeted at specific groups, offering advantages over

---

<sup>3</sup>Figure 2.2 provides examples of these contact matrices, as obtained from [360].

the more generalized approaches of unstructured models.



**Figure 2.2: Age-structured contact matrices.** Example of contact matrices for different countries and regions with 18 age groups resolution. Every cell represents an element  $C_{aa'}$  of the contact matrix, with an estimation of the average number of contacts between age group  $a$  and  $a'$ . Selected examples from Spain (Panel A) Massachusetts, USA (Panel B), India (Panel C), and Tokyo, Japan (Panel D). These matrices have been downloaded from <https://github.com/mobs-lab/mixing-patterns>, a repository related to the research performed by Mistry et al. [360]. The work done in Chapter 5 relies on this data for the United States, but utilizes the higher resolution matrices of 85 age groups.

Consider a single closed and constant population of  $N$  individuals. These individuals have age as an attribute, which we can assume is a discrete quantity sorted in  $A$  age groups. To answer fundamental questions about the progression of an epidemic in this population, we can still use compartmental models as a tool, but now we have to resolve every health-based compartment into further  $A$  subdivisions, each belonging to a different age group  $a$ . Then, continuing with the SIR model as our standard example, we can write populations in compartments as  $X = \sum_a^A X_a$ , for  $X = S, I, R$ , where  $S, I, R$  are the total populations in the system related to the health-based state, and  $S_a, I_a$ , and  $R_a$  are the age-based subdivisions. Now, every susceptible group  $S_a$

could potentially be in contact, at different rates, with any other infectious group  $I_a$ . In order to suitably express the force of infection exerted on every group, we have to move beyond the simple term  $SI/N$  and somehow express the heterogeneity of the population. This is accomplished through the construction of the contact matrix.

### The contact matrix and the force of infection

The contact matrix for a system with a single class of heterogeneity, age, conformed by  $A$  different groups, can be expressed as:

$$\mathbf{C} = (C_{aa'}) . \quad (2.16)$$

The element  $C_{aa'}$  represents the contact rate of individuals of group  $a$  with those of group  $a'$ . Assuming a standard definition of what a direct physical pairwise contact is, the contact matrix is symmetrical, and thus  $C_{aa'} = C_{a'a}$ .

Now, let us dissect the term  $C_{aa'}$  to get further insight into the mixing process of the population groups. Consider that in our population of  $N$  agents, every group  $a$  shares a fraction  $f_a$ , so that  $N = \sum_a^A f_a N$ , and  $N_a = f_a N$ . Every individual in the population has the property of making dyadic contact with others independently of the group they belong to. Then, let  $k_i$  be the number of contacts of an individual  $i$ . These contacts can be distributed among the different groups as  $k_i = \sum_a^A k_{i,a}$ , where  $k_{i,a}$  is the number of contacts that individual  $i$  makes with individuals in group  $a$ . If individual  $i$  belongs to group  $a$ , we add up all  $a$ -group individuals and obtain the total number of contacts performed by  $a$ -group population,  $K_{a,T}$ :

$$K_{a,T} \equiv \sum_{i \in a}^{N_a} k_i. \quad (2.17)$$

Decomposing  $k_i$  into  $k_i = \sum_a^A k_{i,a}$  we readily arrive to

$$\sum_{i \in a}^{N_a} \left( \sum_a^A k_{i,a} \right) = \sum_{a'}^A K_{aa'}. \quad (2.18)$$

Now,  $K_{aa'}$  stands for the number of contacts performed by all the individuals in group  $a$  among individuals of group  $a'$  only. As we said, we assume that contacts are symmetric, and thus by definition, it holds that  $K_{aa'} = K_{a'a}$ . From  $K_{aa'}$  and  $K_{a,T}$ , we can get the per group share of contacts:

$$p_{aa'} \equiv \frac{K_{aa'}}{K_{a,T}}, \quad a, a' = 1, \dots, A. \quad (2.19)$$

These quantities satisfy  $\sum_{a'}^A p_{aa'} = 1$  for every group  $a$ . Additionally, as a consequence of the relationship of the symmetry of contacts, it also holds that  $p_{aa'} K_{a,T} = p_{a'a} K_{a,T}$ .

For every group  $a$ , considering its population amounts to  $N_a$ , we can also state that their average number of contacts is

$$q_a = \frac{K_{a,T}}{N_a}, \quad a = 1, \dots, A. \quad (2.20)$$

Taking into account these definitions, we can express the (per capita) contact matrix as:

$$\mathbf{C} = (C_{aa'}) = \frac{1}{N} \left( \frac{p_{aa'} q_a}{f_a} \right). \quad (2.21)$$

We see that the term  $C_{aa'}$  stems from the contributions of the average social activity of group  $a$ ,  $q_a$ , their share in the total population,  $f_a$ , and their mixing share with other groups  $a'$ . Of course, we have to make clear that even though we have taken notice of the system's heterogeneity and introduced an age-based mixing structure, we are averaging the *contact* attribute over groups and thus we are still missing a more detailed picture of the contact structure.

Let us move to express the force of infection in this heterogeneous population. For the susceptible population in group  $a$ ,  $S_a$ , this is simply expressed as a linear superposition of all the contributions for all the infectious groups in the system,  $I_a \forall a$ . Now, the key point is that these groups do not interact in the same way, or at the same rate, and this is something given in effect by the contact matrix. All in all, the per capita force of infection on group  $a$ ,  $\lambda_a$  can be expressed as:

$$\lambda_a = \frac{1}{N} \sum_{a'}^A \beta_{aa'} C_{aa'} I_{a'}. \quad (2.22)$$

Here,  $\beta_{aa'}$  is the transmission rate between group  $a$  and  $a'$ . For generality, it is shown as inter-group dependent, and thus as a matrix, but depending on the particular disease and data at hand, it might not be necessary to account for such a fine-grained depiction of the transmission rate.

For completeness, here is how it would look the full system of this age-structured SIR model:

$$\frac{dS_a(t)}{dt} = -\frac{1}{N} S_a(t) \sum_{a'}^A \beta_{aa'} C_{aa'} I_{a'}, \quad (2.23)$$

$$\frac{dI_a(t)}{dt} = \frac{1}{N} S_a(t) \sum_{a'}^A \beta_{aa'} C_{aa'} I_{a'}(t) - \mu_a I_a(t), \quad (2.24)$$

$$\frac{dR_a(t)}{dt} = \mu_a I_a(t). \quad (2.25)$$

The last quantity to be defined here is  $\mu_a$ , which in this case is the age-dependent decay rate to the removed compartment or, equivalently, the inverse of the infectious period. Even though some analytical solutions exist for the steady-state prevalence

in the simplest versions of this system, the standard procedure to find a whole solution (including the dynamics) is to resort to numerical schemes as the 4th-order Runge-Kutta method.

Here, I focused on age as a relevant defining trait in the population for illustrative purposes, since an age-structured model is used in Chapter 5 of this thesis, but this model building and analysis could be applied to other potential relevant attributes in the population<sup>4</sup>.

### 2.2.2 The network paradigm

Classical epidemiological models have traditionally assumed minimal heterogeneity and rarely incorporated any specific contact structure, with the homogeneous well-mixed approach predominantly used. However, largely driven by advancements in the field of statistical physics, the emerging study of *complex networks* has opened new avenues. This development has enabled the exploration and testing of well-known epidemiological dynamics within more intricate topologies, moving beyond the confines of classical regular lattice models.

#### What are networks?

A network is defined as a graph  $G$  given by a pair  $(V, E)$ , where  $V$  is a set, the elements of which are called nodes, and  $E \subset V \times V$  is a set of pairs of nodes, called links. For a finite network with  $N$  nodes, the node set is taken simply as  $V = \{1, 2, \dots, N\}$ . The set  $E$  can be encoded in the so-called adjacency matrix, which is given by:

$$(\mathbf{A})_{ij} = \begin{cases} 1 & \text{if } i \text{ and } j \text{ are connected,} \\ 0 & \text{otherwise.} \end{cases} \quad (2.26)$$

This definition can be easily generalized to *weighted* networks, now represented by matrix  $\mathbf{W}$ , and where in case of a connection between nodes  $i$  and  $j$ , is represented more generally as  $(W)_{ij} = w_{ij} \in \mathbb{R}$ . Thus, this extension allows us to describe networked systems where it is not only relevant whether two parts of it are connected, but to what extent<sup>5</sup>. Another generalization of the most elemental generic network model is conferring a direction to connections, then we talk of *directed* networks. Whereas

---

<sup>4</sup>In this exposition we introduced a model of discrete compartmentalization, but age is also considered as a continuous variable in the literature and modeled through differential partial equations schemes [88, 407].

<sup>5</sup>An example of this could be the mobility flows between different districts of an urban system. It is not just relevant to know whether there are citizens moving between two districts in a city, something rather expected considering the spatial scales involved, but what particular mobility volume are we witnessing.

undirected networks have a symmetric adjacency matrix,  $\mathbf{A} = \mathbf{A}^T$ , the same does not necessarily apply to directed networks<sup>6</sup>.

Nodes are characterized by their number of connections to other nodes, and their degree  $k$ . Subsequently, one can expect diverse degree distributions  $P(k)$  depending on how many connections the constituent nodes have in a network. This by itself is the most elementary detailed attribute of a network and can have tremendous consequences when dynamical processes occur in networked systems [408]. One can broadly distinguish between two main types of networks according to the degree distribution, homogeneous and heterogeneous networks [145]. The former type has a characteristic average degree from which small deviations are rather rare. Examples of homogeneous networks are the trivial cases of complete networks, regular networks, or random graphs, also famously referred to as Erdős-Rényi networks. On the contrary, heterogeneous networks, where the most prominent examples are the so-called scale-free networks, are mathematically represented by power-laws  $P(k) \propto k^{-\gamma}$ , with  $2 < \gamma < 3$  [142], and show no characteristic degree. Whereas the majority of nodes accumulate in range with a very small degree, there is a finite sizeable probability of finding nodes with a very high degree  $k$ , and thus these systems are also referred to as *heavy-tailed*.

### Generative models

As Chapter 6 explores an epidemic dynamics on networks, we provide here descriptions of the network models employed to furnish the necessary background.

The Erdős-Rényi (ER) model is one of the foundational models for random graphs, introduced by Paul Erdős and Alfréd Rényi in the 1960s. In this model, a network is built by starting with  $N$  isolated nodes and adding edges between pairs of nodes with a fixed probability  $p$ . The result is a graph where the degree distribution,  $P(k)$ , follows a binomial distribution (see Figure 2.3 A1 and A2):

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}. \quad (2.27)$$

For large  $N$ , this distribution approximates a Poisson distribution:

$$P(k) \approx \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}, \quad (2.28)$$

where  $\langle k \rangle$  is the average degree of the network,  $\langle k \rangle = (N-1)p$ . One striking property of the ER model is the emergence of a giant connected component as  $p$  surpasses a

---

<sup>6</sup>Continuing with the mobility example, indeed we are typically interested in the origin and destination of trips, and thus in the sense of direction. On the contrary, physical contact networks could be expected to be symmetrical for the most part, since pairwise contact relationships are reciprocal.

particular threshold. However, the ER model does not reproduce many properties of real-world networks, like the common presence of hubs or clusters of connections.

The Barabási-Albert (BA) model, proposed by Albert-László Barabási and Réka Albert in 1999, aims to explain the emergence of scale-free networks. Unlike the ER model, many real-world networks display a heterogeneous connectivity distribution where a few nodes (hubs) possess a disproportionately large number of connections. The BA model begins with a few nodes and grows the network by preferentially attaching new nodes to those with higher degrees. The probability of the new node  $i$  to connect with an existing node  $j$  in the network is proportional to the number of connections (degree) of  $j$ ,

$$P(i \leftrightarrow j) = \frac{k_j}{\sum_u k_u}. \quad (2.29)$$

This “preferential attachment” leads to a power-law degree distribution (see Figure 2.3 B1 and B2):

$$P(k) \sim k^{-\gamma}, \quad (2.30)$$

where  $\gamma$  is typically between 2 and 3 for many real-world networks. By construction, the BA model follows a degree distribution with  $\gamma = 3$  in the thermodynamic limit ( $N \rightarrow \infty$ ) [32]. This model captures the essence of many empirical networks such as the World Wide Web, certain social networks, and biological systems, emphasizing the role of growth and preferential attachment in network evolution.

Figure 2.3 emphasizes the differences between homogeneous (ER) and heterogeneous (BA) networks. While in the former, there is a well-defined degree scale and the probability of finding nodes with  $k$  rapidly drops off for  $k \gg \langle k \rangle$ , in BA networks, we find a non-null probability of finding nodes with a disproportionately high degree (the hubs).

The configuration model [409], represents a gold standard for generating heterogeneous networks, including uncorrelated scale-free networks with degree exponents between 2 and 3. It is particularly versatile due to its ability to create a wide array of network types, given that the input is a specified degree sequence. This sequence can be of any form, as long as the sum of all degrees, or “stubs”, is even. For additional details on the Erdős-Rényi and Barabási-Albert models, the configuration model, and other canonical models and algorithms for constructing networks, readers are encouraged to consult the references provided in this section.

## Metrics

Naturally, different topological configurations in a network, that is, how nodes are connected among them, yield different properties to it. From the degree distribution of

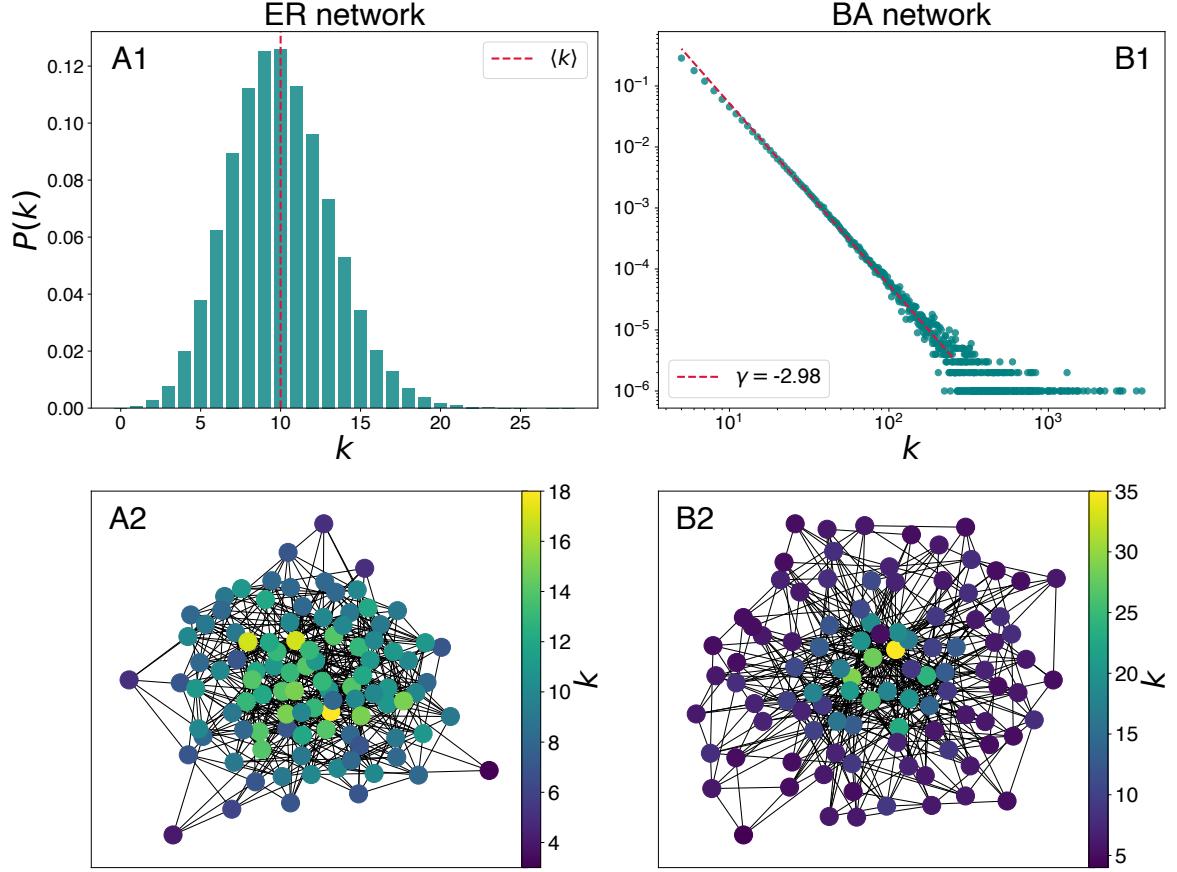


Figure 2.3: **ER and BA degree distributions and networks.** Degree distribution for an Erdős-Rényi network of size  $N = 10^5$  (Panel A1), and for a Barabási-Albert network of size  $N = 10^6$  (Panel B1), both with average degree  $\langle k \rangle = 10$ . Visual representation of an ER network in Panel A2 and of a BA network in Panel B2 (reduced system sizes to  $N \sim \mathcal{O}(10^2)$  for visualization purposes).

a network, one can obtain global measures as the average degree  $\langle k \rangle = \sum_k kP(k)$ , the variance  $\text{Var}(k) = \langle k^2 \rangle - \langle k \rangle^2$ , or any moment of order  $m$  associated with the degree distribution  $P(k)$ ,  $\langle k^m \rangle = \sum_k k^m P(k)$ .

When analyzing the connections between nodes with varying degrees in a network, one key measure is the Pearson correlation coefficient calculated for the degrees of nodes at either end of each link. This coefficient helps determine the nature of node interconnectivity in the network. If the Pearson correlation coefficient is above zero, the network is described as assortative. In assortative networks, nodes with similar degrees tend to connect with each other. Conversely, if the coefficient is below zero, the network is considered disassortative, meaning nodes with higher degrees are more likely to connect with nodes of lower degrees. A coefficient close to zero suggests no significant degree correlation in the network. It is important to note that the Pearson coefficient measures only linear correlations.

Another critical metric in network analysis is the clustering coefficient or the number

of triangles within the network. This metric is particularly relevant in the context of spreading phenomena, as high clustering can lead to a faster depletion of susceptible nodes around infected ones. The clustering coefficient provides insight into the local network structure and the potential for rapid spread or containment of contagion within closely-knit groups.

Finally, another set of metrics that can help characterize a network are the so-called centrality measures. For instance, the most elemental one is the degree of centrality, which simply reflects the number of connections a node has. Thus, the nodes with the higher number of connections - known as hubs, especially in scale-free networks- would be the most central ones. Nodes with higher degrees are more central. Betweenness centrality measures how often a node lies on the shortest paths between other nodes. It identifies nodes that bridge different parts of the network. Eigenvector centrality considers a node's importance based on its connections to other well-connected nodes. A node is more central if it is connected to other highly central nodes.

For thorough expositions on the topic of classical complex networks, metrics, models, algorithms, and beyond, we refer the reader to references [143, 49, 145]. To delve into more advanced topics, like multilayer/multiplex, the interested reader can take a look at [177, 410], and to [411] for the extension to temporal networks. As for the topic of general dynamical processes on networks, we recommend [408, 172, 412].

### Epidemic spreading on complex networks

As we explained in the introduction, the network framework was a revolution within mathematical epidemiology that spurred waves of research on the topic of how complex and nontrivial topologies could impact the dynamics and the overall impact of an infectious disease. Here we will just limit ourselves to a brief exposition of an example of a networked SIR model. We choose the degree-based or heterogeneous mean-field theory to build the dynamical equations for the SIR process running on a generic network.

At odds with our treatment of the standard well-mixed SIR model in Section 2.2, here there is a distinctive property related to each individual, and this is their degree  $k$ , that is, the number of neighbors with which interact. Importantly, under the well-mixing approximation, any individual could virtually interact with any other with the same probability, but now, interactions extend only to first neighbors. In the spirit of the heterogeneous mean-field theory, what we do then is to collect the individuals in the population by health status, but also by their degree. This way, we focus on the *degree block* quantities  $s_k = S_k/N_k$ ,  $i_k = I_k/N_k$ ,  $r_k = R_k/N_k$ , where  $S_k$ ,  $I_k$ , and  $R_k$  are the number of susceptible, infected, and removed individuals in the degree class  $k$ ,

respectively.  $N_k$  represents the total number of nodes in the degree class  $k$ . The global averages- that is, the total populations by health status- are given by  $s = \sum_k P(k)s_k$ ,  $i = \sum_k P(k)i_k$ ,  $r_k = \sum_k P(k)r_k$ . After these definitions, the HMF theory dynamical equations read:

$$\frac{ds_k(t)}{dt} = -s_k(t)\lambda_k(t), \quad (2.31)$$

$$\frac{di_k(t)}{dt} = s_k(t)\lambda_k(t) - \mu i_k(t), \quad (2.32)$$

$$\frac{dr_k(t)}{dt} = \mu i_k(t). \quad (2.33)$$

As it is evident, now we are neither looking at the global variables  $s(t)$ ,  $i(t)$ ,  $r(t)$ , but neither at the dynamical evolution of the individual health microstates, but what the HMF did was to adopt a compromise in between. Here,  $\lambda_k$  is the force of infection, now dependent on the  $k$ -class, which is defined as  $\lambda_k \equiv \beta k \Theta_k$ , where  $\gamma$  is the infection rate per contact; this relates to the well-mixing SIR transmission rate  $\beta'$  through  $\beta' = \beta k$ . Then,  $\Theta_k$  describes the density of infected neighbors of nodes in the degree class  $k$ . Intuitively, this density is a function of the conditional probability that a node  $k$  is connected to any node  $k'$  and proportional to the number of infected nodes in each class  $k'$ :  $\Theta_k = \sum_{k'} P(k'|k)i_{k'}$ . For uncorrelated networks, the probability of finding a node of degree  $k'$  in the neighborhood of a node in degree class  $k$  is independent of  $k$ . Therefore,  $\Theta_k = \Theta = \sum_{k'} (k' - 1)P(k')i_{k'}/\langle k \rangle$  where the term  $k' - 1$  is due to the fact that at least one link of each infected node points to another infected vertex.

Now let us compute the basic reproduction number  $R_0$  for this mean-field theory and see how it compares with the classical well-mixed SIR, which we recall was given by  $R_0 = \beta/\mu$ . To accomplish that, let us consider the early time limit  $t \rightarrow 0$  of the spreading process and assume that we are dealing with a mostly susceptible population. This implies  $s_k \gg i_k$  and  $r_k \approx 0 \forall k$ . The equation for the infected compartment then becomes  $di_k(t)/dt = \beta k \Theta(t) - \mu i_k(t)$ . Multiplying both sides for  $P(k)$  and summing over all values of  $k$  we obtain the dynamical equation for the global quantity, the infected density  $i(t)$ ,  $di(t)/dt = \beta \langle k \rangle \Theta(t) - \mu i(t)$ . In order to understand the behavior of  $i(t)$  when  $t \rightarrow 0$ , let us consider an equation built by multiplying both sides of the last equation by  $(k - 1)P(k)/\langle k \rangle$  and summing over all the degree classes. We obtain then:

$$\frac{d\Theta(t)}{dt} = \beta \left( \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) \Theta(t) - \mu \Theta(t). \quad (2.34)$$

The fraction of infected individuals for each  $k$  class increases if and only if  $d\Theta/dt > 0$ . This sets the condition for epidemic growth and therefore [408]:

$$\frac{d\Theta}{dt} > 0 \quad \rightarrow \quad \beta \left( \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} \right) - \mu > 0 \quad \rightarrow \quad \frac{\beta}{\mu} \frac{\langle k^2 \rangle - \langle k \rangle}{\langle k \rangle} > 1. \quad (2.35)$$

From this, we obtain that:

$$R_0 = \frac{\beta}{\mu} \frac{\langle k^2 \rangle - \langle k \rangle^2}{\langle k \rangle}. \quad (2.36)$$

Compared to the well-mixed SIR model  $R_0$ , we see that a factor related to the network's topology now enters the calculation<sup>7</sup>. This factor is especially important for scale-free networks with exponent  $2 < \gamma < 3$ . Note that we can write the threshold condition as:

$$\frac{\beta}{\mu} > \frac{\langle k \rangle}{\langle k^2 \rangle - \langle k \rangle^2}. \quad (2.37)$$

This way, for fixed  $\mu$ , we have an expression for the critical transmission rate  $\beta$ . For an uncorrelated scale-free network with exponent  $2 < \gamma < 3$ , in the thermodynamic limit, the right-hand side of the inequality diverges, meaning that any value of  $\beta$  can trigger an epidemic outbreak and thus infecting a finite fraction of the system.

Of course, real networks are not infinite systems but they do show a heavy-tailed nature, and thus present fluctuations in the number of contacts (large  $\langle k^2 \rangle$ ) that are significantly larger than the average degree  $\langle k \rangle$  resulting in very small thresholds. Hubs play a significant role in the rapid spread of infections. These hubs establish direct connections among nodes that would otherwise be distant from each other. Once these pivotal hubs are infected, diseases exploit this connectivity to reach a substantial portion of nodes within the network. The vulnerability of real interaction networks to disease transmission is a concerning discovery; however, it also opens the door to effective strategies for outbreak control and mitigation. Hubs, being central and influential nodes in network connectivity [145], offer an efficient target for vaccination. By immunizing a small fraction of these hubs, it is possible to promptly halt disease propagation and safeguard the wider population. Notably, in practical scenarios, the awareness of network structures is often limited. Identifying hubs might not be straightforward, necessitating the use of indirect methods.

Interestingly, the same hub feature that contributes to disease diffusion also facilitates their identification. Due to their connectivity, high-degree nodes link to numerous lower-degree nodes. Thus, a random selection of a node, denoted as A, in the network, followed by exploring one of its connections to reach another node, B, is likely to lead to a hub. The underlying concept is analogous to the “friend paradox” [145], where, on average, your friends have more friends than you do. This approach enables the prioritization of immunization for node B, which proves to be considerably more impactful than targeting node A. This counterintuitive strategy remains highly effective even when applied to networks with rapid changes [413, 414, 415].

---

<sup>7</sup>For more general mean-field theories in networks, it can be shown that  $R_0$  relates to the largest eigenvalue of the adjacency matrix of the network where the dynamics occurs [163].

For a very exhaustive exposition on the topic of epidemics on complex networks, we recommend the book by Kiss et al. [164], and the review by de Arruda et al. [163] for an extension of the network mean field theory to multilayer and multiplex networks.

## 2.3 Metapopulations: Introducing spatial structure and mobility

Incorporating structure into our basic single-population models opens up various avenues for extension. Some approaches involve characterizing the population by adding relevant features, such as age. Others aim to more accurately represent how individuals mix and make contact, for which the network framework has proven to be a significant enhancement. Ultimately, another key method of extending our models is by introducing spatial structure. As discussed in the introduction chapter, the metapopulation framework has been particularly successful in this regard [242].

A metapopulation can be conceptualized as a network (refer to Figure 2.4). In this model, the nodes<sup>8</sup> represent geographical units, often corresponding to human settlements. These nodes could be countries, regions, provinces, cities, or even finer-grained locations like urban points of interest. The links connecting these nodes represent human mobility flows between these locations.

Dynamical systems within a metapopulation framework have been formulated as reaction-diffusion systems [215, 217], where epidemiological processes are the “reactions” and human mobility is the “diffusion”<sup>9</sup>.

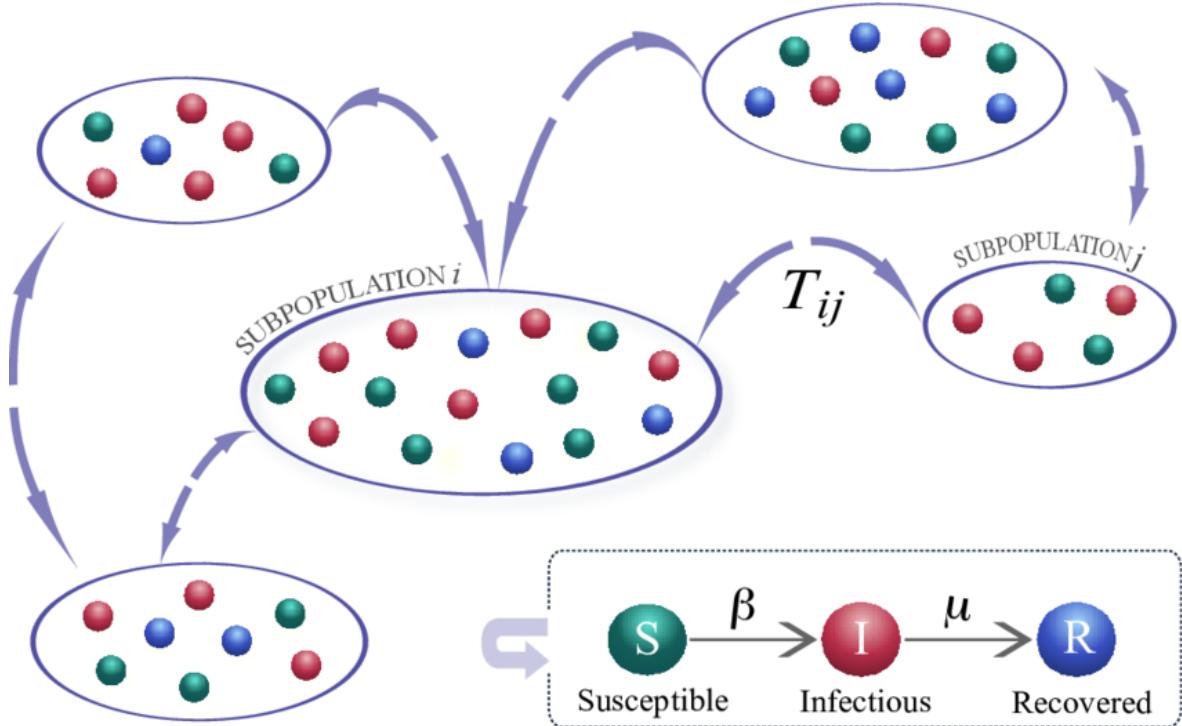
A metapopulation is characterized by several parameters. Regarding the spatial structure, we have the size of the network  $V$ , that is, the number of locations and then the specific network topology. Regarding the population inhabiting it, we have the number of individuals dwelling in every location  $N_l$ ,  $l \in V$ , which adding up gives the total population in the system  $N = \sum_l N_l$ . Individuals interact within the nodes as given by the disease under study, and move or diffuse along them, following some particular mobility rules. Dealing with a more complex system than a single population, a metapopulation model allows for multiple levels of description. For a typical SIR epidemic process, we can consider:

- Global state: This would be given by the total populations of susceptible, infected, and removed individuals added over every location:  $S, I, R$ .

---

<sup>8</sup>Nodes in metapopulation systems are typically referred to as *subpopulations*. However, in this exposition, they will be simply referred to as locations, without any loss of generality.

<sup>9</sup>Human mobility is more complex than physical diffusion. In its simplest form, Markovian mechanistic mobility processes can be analogous to classic diffusion processes [417, 50].



**Figure 2.4: Metapopulation scheme.** Nodes, referred to as subpopulations, represent spatial locations where individuals interact, whereas the links connecting the subpopulations represent mobility flows. In this example, the SIR process occurs inside each subpopulation, where homogeneous mixing is assumed. Also at each time step, individuals move between neighboring subpopulations  $i$  and  $j$  according to a mobility matrix  $T_{ij}$ . Reference: Scheme of the epidemic and mobility models in a metapopulation by Ventura et al. [416], used under CC BY 4.0. License: <https://creativecommons.org/licenses/by/4.0/>.

- Local state: Since now there are well-defined separate locations, we can also characterize the impact at the local scale and thus have:  $S_l, I_l, R_l, \forall l \in V$ .
- Individual state: The standard assumption in metapopulation models is to consider homogeneous populations of individuals at every node, but this does not prevent us from considering some kind of heterogeneity related to the individual.

In the case of large metapopulation systems with a high level of (inter-location) heterogeneity, the analytical description in terms of specific features of each single location is extremely complicated. To achieve some analytical treatment, similarly, as it was originally conceived in contact network epidemiology, the degree-block description or heterogeneous mean-field theory [152], can be adopted also here. This assumption can be largely justified based on the empirical evidence of the statistical equivalence of locations that have the same degree [217]. Under the degree-based mean-field approach,

we can write the populations at every location as,

$$N_k = \frac{1}{V_k} \sum_{i|k_i=k} N_i, \quad (2.38)$$

where  $V_k$  is the number of nodes with degree  $k$  and the sums run over all nodes  $i$  having degree  $k_i$  equal to  $k$ . The degree block variable  $N_k$  represents, therefore, the average number of individuals in all subpopulations with degree  $k$ .

To get a better grasp of how the epidemic and mobility dynamics work in metapopulations, we will now give a more detailed account of each of these main processes.

### 2.3.1 The epidemic reactions

Regular metapopulation models assume that the contagion process or any other epidemic-related process takes place within the nodes of the networks where the population mixes homogeneously. Working with the SIR scheme, this is not much more different from the single population model shown before. The dynamics can be identified with the following set of reaction equations:



Here  $\beta$  and  $\mu$  are the already familiar transmission and removal rates related to the SIR process, respectively. These reactions occur simultaneously at every node, with individuals capable of moving throughout the system. We assume that the primary distinguishing feature of each location is its degree, which is crucial for the in-and-out mobility flows. Consequently, in terms of the contagion process, all locations behave in a uniformly homogeneous manner. However, empirical studies have explored variations in the transmission process, such as how it might scale with a location's population size [235]. Additionally, some works have incorporated heterogeneity within populations, whether by age [228] or by the degree of the nodes [224].

Returning to our metapopulation model, the per-capita force of infection acting on each location can be conceptualized similarly to that in a well-mixed, single-population SIR model:

$$\lambda_l(t) = \beta \frac{I_l(t)}{N_l(t)} \quad \forall l \in V, \quad (2.41)$$

where now  $I_l(t)$  is the population of infected individuals, and  $N_l(t)$  is the total population at location  $l$  and time  $t$ , respectively. Within the context of

reaction-diffusion processes [215], the contagion interaction is often presented in a generic way as “reaction kernel” denoted by  $\Gamma$  and defined by:

$$\Gamma_l(t) = \frac{I_l(t)S_l(t)}{N_l(t)}, \quad (2.42)$$

which is no other thing than the per-capita force of infection divided by  $\beta$  and multiplied by the total population of susceptibles at the corresponding location. Adopting the degree-block representation, this is simply:

$$\Gamma_k(t) = \frac{I_k(t)S_k(t)}{N_k(t)}. \quad (2.43)$$

Independently of the representation adopted, one can easily identify that we are dealing with the so-called frequency-dependent force of infection. Other types of interaction can be introduced directly just by using a different reaction kernel [215, 217, 235].

As in previous schemes, once we have set the relevant states of the system, the dynamical transition that can occur, and how they occur, the next step would lead to building the dynamical equations governing those system states at every location (or for every  $k$ -class, in this case). But now we have to keep in mind that apart from the epidemic process, there is a diffusion or mobility process going on simultaneously, and this will have to be properly accounted for in those equations. Thus, prior to that, let us delve into the mobility aspect of the metapopulation system<sup>10</sup>.

### 2.3.2 The diffusion process

Now, independently of the spreading process, consider that individuals move from a location with degree  $k$  to another with degree  $k'$  with a diffusion rate  $d_{kk'}$ . The rate at which individuals leave a location with degree  $k$  is then given by  $p_k = k \sum_{k'} P(k'|k)d_{kk'}$ . One can represent then the population dynamics by a mean-field dynamical equation of the form:

$$\frac{\partial N_k(t)}{\partial t} = -p_k N_k(t) + k \sum_{k'} P(k'|k)d_{k'k}N_{k'}(t). \quad (2.44)$$

The first right-hand side term is negative and thus represents a loss term. It tells us about the loss of local population driven by the mobility rate  $p_k$ . The second term is a positive sum of contributions, each one representing an incoming flow of individuals from neighboring locations  $k'$ . These contributions are the result of several factors:

---

<sup>10</sup>The details of the mean-field equations for the metapopulation SIR model under some of the specific mobility models introduced in the next section, as well as the solutions for early-time dynamics, can be found in [217]. While analytical equations bring some insights, in complex systems as contagion processes on metapopulations are, one needs to resort to *in silico* experiments, that is, simulations, both as a check for the theory and most importantly, to extend the theory and achieve a bigger picture of the system’s behavior. We explain how these simulations work in Section 2.3.4.

(i) the conditional probability  $P(k'|k)$ , which encodes all the topological correlations of the network, (ii) the diffusion rate matrix  $d_{kk'}$ , which informs about the rate at which individuals from locations with degree  $k'$  move to locations with degree  $k$ , and (iii)  $N_{k'}(t)$ , which is the population of locations with degree  $k'$  at time  $t$ . The overall term is proportional to the number of links  $k$  times the average number of individuals coming from each neighboring location.

In order to solve the above set of equations one has to bring in more details about the specific network topology and the mobility model playing out. In the following, I will briefly show the solutions for different diffusion processes that consider diffusion rates depending on the traffic of each node or on the population size of each location.

### Traffic dependent mobility rates

Here it is assumed that the rate at which an individual leaves a given location is independent of its degree  $k$ , yielding  $p_k = p \forall k$ . Assuming also homogeneous diffusion along any given connection, individuals have the same probability to move along any one of the links departing from the location at which they are located. This is a networked version of the traditional random walk problem. In this case, the diffusion rate along any given link of a node with degree  $k$  will be simply equal to

$$d_{kk'} = \frac{p}{k}. \quad (2.45)$$

Of course, this is an oversimplification that hardly ever applies in real-life systems.

A more realistic process considers the movement of individuals to be proportional to the traffic intensity along a given edge. This is an example of heterogeneous diffusion based on the degrees of the involved locations:

$$d_{kk'} = p \frac{w_0 (kk')^\theta}{T_k}. \quad (2.46)$$

The diffusion rate  $p$  is still the same in each location but now individuals move on each connection in a proportion dependent on the actual traffic connection. The proper normalization constant of the diffusion matrix is  $T_k = Ak^{(1+\theta)}$  with  $A = \langle k \rangle^{1+\theta} w_0 / \langle k \rangle$ . By plugging the heterogeneous  $d_{kk'}$  into expression 2.44, the dynamical equations read

$$\frac{\partial N_k(t)}{\partial t} = -pN_k(t) + pk^{(1+\theta)} \frac{w_0}{A\langle k \rangle} \sum_{k'} P(k') N_{k'}(t). \quad (2.47)$$

Here it has been assumed that we are working with uncorrelated networks and thus  $P(k'|k) = \frac{kP(k)}{\langle k \rangle}$ . The explicit stationary solution can be easily obtained by setting  $\partial N_k(t)/\partial t \equiv 0$ :

$$N_k = \frac{k^{(1+\theta)}}{\langle k^{(1+\theta)} \rangle} \bar{N}. \quad (2.48)$$

We see then that the population of each node scales with the node degree in the stationary limit. This is the result of the diffusion process bringing a large number of individuals into well-connected locations, thus showing the impact of the network's topological (through  $k$ ) and traffic (through  $\theta$ ) fluctuations on the location size behavior. Note that when  $\theta = 0$ , we recover the homogeneous diffusion case in which  $d_{kk'} = d_k = p/k$ , obtaining

$$N_k = \frac{k}{\langle k \rangle} \bar{N}. \quad (2.49)$$

Here, the location size is just fixed from topological fluctuations. It is worth remarking that the location size as a function of the degree is constrained by the diffusion process, a feature not to be expected in real systems where the population size of local patches can be considered as an independent variable. On the other hand, the degree dependence is close to those observed in real systems where in several cases it is possible to find a relation  $N_k \approx k^\phi$  with  $0.5 \leq \phi \leq 1.5$  [15].

### Population dependent mobility rates

When we have a metapopulation based on a real setting, quite possibly we will want to feed the population variables  $N_l$  or  $N_k$  with real data, and thus it is important to have them as independent variables. Here, then, the number of people traveling from one location to the other in a unitary time scale is a defined number  $w_{ij}$  and the number of traveling individuals is independent of the population size  $N_i$ . Now, each individual location has a diffusion rate  $\sum_j w_{ij} = N_i$ , where  $\sum_j w_{ij}$  is the total number of people traveling out of location  $i$  in the unitary time scale. In other words, the diffusion rate of each individual is inversely proportional to the population size. Empirical evidence obtained from the study of the worldwide air transportation network [157] showed the symmetry of the network both in the directionality and in the travel fluxes. We can state that the condition  $w_{ij} = w_{ji}$  is satisfied in large-scale real transportation networks and therefore

$$\frac{\partial N_i(t)}{\partial t} = \sum_j (w_{ji} - w_{ij}) = 0, \quad (2.50)$$

so that any initial conditions for the population size satisfy the stationary state. Returning to the degree block variable representation, we can recover the above condition by considering a diffusion rate of the form  $p_k = T_k/N_k$ . The diffusion rate on any given edge from a location of degree  $k$  to a location of degree  $k'$  is therefore given by

$$d_{kk'} = \frac{w_0(kk')^\theta}{N_k}, \quad (2.51)$$

and the degree block diffusion equations read in the case of the uncorrelated network as

$$\frac{\partial N_k(t)}{\partial t} = -T_k + k^{(1+\theta)} w_0 \frac{\langle k^{1+\theta} \rangle}{\langle k \rangle}. \quad (2.52)$$

By normalization we have  $T_k = k^{(1+\theta)} w_0 \langle k^{1+\theta} \rangle / \langle k \rangle$ , and thus we recover the solution  $\partial N_k(t) / \partial t = 0$  that allows any stationary value distribution  $N_k$ . Here, equation 2.50 shows that a population-dependent diffusion process does not fix the location size, at odds with the traffic-dependent mobility rate case, and thus it can be given as a free parameter of the model.

### Recurrent mobility

The heterogeneous mobility model previously presented surely captures some macroscopic features of human mobility, but also lacks an important component of everyday human mobility: recurrence. The question of spreading under recurrent mobility patterns was looked at in some of the first metapopulation models in the classical epidemiology community [201, 202]. Later, Balcan and Vespignani [213, 219], extended the idea to complex settings in which subpopulations exhibit heterogeneous demographic and mobility properties. In this approach, a time-scale separation is assumed, and the analysis is performed in a regime in which the commuting is much faster than the infection dynamics.

The mobility models previously shown, being macroscopic and Markovian, do not resolve individual human behavior, and thus neither preferentially attach individuals or groups of them to specific locations, like if they simulated households or workplaces. Actually, one does not need to make an agent-based model to approach the problem, but as a first approximation suffices to sort groups of individuals by assigning them a home or origin location to which to return. In every node  $i$ , we group its members according to the location in which they are actually present at a given time  $t$ ,  $N_{ii}(t)$  and  $N_{ij}(t)$ , where  $j \in v(i)$  can be any neighboring location of  $i$ . Sure, this increases the complexity of the analytical treatment and also the computational implementation. It is assumed that individuals in subpopulation  $i$  will visit any one of the connected subpopulations with a per capita diffusion rate  $\sigma_i$ . The return trip for the individuals displaced out of their origin node  $i$  is simply simulated with a per capita diffusion rate  $\tau_i$ .

The details of the mean-field dynamical equations for this type of mobility, either using the conventional node-based approach and also the degree-based heterogeneous approximation, can be found in [220].

### Other macroscopic mobility models

The models previously shown represent human mobility flows using a node's degree as a proxy to its socioeconomic relevance or *attractiveness*. However, other approaches take into account explicitly the population of the areas that constitute the nodes of the metapopulation system, like in Balcan and Vespignani [219].

In addition, an important dimension or factor is being left or at least is not being explicitly represented in these models: spatial distance. The gravity model, which predates all these models, dates back to 1946, introduced by George K. Zipf [418], and obviously inspired by Newton's law of gravity. This model has been paradigmatic in spatial economics and transportation studies [419, 420]. It comes as no surprise that in the gravity model, the rate of going from location  $i$  to location  $j$  is given by:

$$d_{ij} = K \frac{A_i A_j}{r_{ij}^2}, \quad (2.53)$$

where  $A$  represents a location's attractiveness, and  $r_{ij}$  is the Euclidean distance between location  $i$  and  $j$ . Some generalized gravity models show the attractiveness powered to some exponent or indeed consider different functional forms involving the distance  $r_{ij}$  between locations, like power-laws (other than with exponent 2) or exponentially decaying functions [421].

Following physical analogies, we also have the radiation model of human mobility [422, 423]. This model has been specifically developed to reproduce commuter movements and has the additional desirable feature of being parameter-free. The traveling probability from location  $i$  to  $j$  is given by

$$w_{ij} = \frac{N_i N_j}{(N_i + P_{ij})(N_i + N_j + P_{ij})} \sum_{i \neq j} w_{ij}, \quad (2.54)$$

with  $N_i$  and  $N_j$  being the populations of origin and destination,  $P_{ij}$  the total population living between location  $i$  and location  $j$  (computed as the total population living in a circle of radius  $r_{ij}$  centered at  $i$ , excluding the populations of origin and destination locations), and  $\sum_{j \neq i} w_{ij}$  the total number of commuters daily leaving their home in location  $i$ .

For a comprehensive review of human mobility modeling and applications, we refer the reader to the work of Barbosa et al. [424]. In an upcoming section, we will take a quick peek at some of the advances done in individual-based or microscopic mobility models.

#### 2.3.3 The global invasion threshold

Single-population epidemiological models show a threshold behavior established by the basic reproduction number  $R_0$ , separating the free-disease phase ( $R_0 < 1$ ) and the

endemic phase ( $R_0 > 1$ ). In [216], Colizza and Vespignani showed the existence of a global invasion threshold. This threshold, denoted as  $R_*$ , sets the condition for the infection of a macroscopic fraction of the metapopulation system. The main factors controlling the feasibility of a solid invasion other than  $R_0$ , are the diffusion or mobility rate and the topological properties of the system.

By considering the invasion dynamics as a basic branching process, an explicit analytical expression for the global invasion threshold quantity can be derived for certain epidemic metapopulation dynamics carrying out a series of simplified but reasonable assumptions. These assumptions involve that the system is very close to the epidemic threshold, and also that at the early stage of the epidemic, the fraction of invaded districts is also very small. It can be shown then [216, 217]:

$$R_* = (R_0 - 1) \frac{p\bar{N}\alpha}{\mu} \chi > 1. \quad (2.55)$$

This expression comprises different factors at play in the spatial propagation of an epidemic. The first factor,  $(R_0 - 1)$ , relates to the classical single-population threshold behavior of a spreading process. The second factor,  $\frac{p\bar{N}\alpha}{\mu}$ , is related to the probability of exporting the disease out of a population given a certain mobility regime and a particular epidemic spreading model. This last thing is set by the factor  $\alpha$ . For the SIR, and  $R_0$  close to 1,  $\alpha$  can be approximated to  $\alpha \approx 2(R_0 - 1)/R_0^2$ . Finally, the third factor,  $\chi$ , differs depending on the network topology. For homogeneous networks it is  $\chi = \frac{\langle k \rangle}{\langle k \rangle - 1}$ , whereas for heterogeneous (uncorrelated) networks it is  $\chi = \frac{\langle k^{2+2\theta} \rangle - \langle k^{1+2\theta} \rangle}{\langle k^{1+\theta} \rangle^2}$ . This ratio is extremely small in heavy-tailed networks and it is vanishing in the limit of infinite network size. This implies that the heterogeneity of the metapopulation network favors the global spread of epidemics by lowering the global invasion threshold. Topological fluctuations favor the subpopulation invasion and suppress the phase transition in the infinite size limit.

As  $R_*^c \equiv 1$  sets the critical behavior in the invasion process, we can find the associated threshold condition for the mobility rate:

$$p_c \bar{N} = \frac{\mu}{\alpha} (R_0 - 1)^{-1} \frac{1}{\chi}. \quad (2.56)$$

Below  $p_c$ , the diffusion dynamics is sluggish enough to be able to invade the system. Indeed, we see that the closer  $R_0$  is to 1, the larger  $p$  needs to be to overcome the global invasion threshold. It is important to stress that when  $R_0 - 1 \ll 1$  no longer applies, a closed form for  $R_*$  is unachievable.

These global invasion and mobility threshold findings have provided a theoretical framework and rationale for the evidence concerning the inefficacy of travel restrictions in the containment of global epidemics [209, 425, 16]. It is easy to see, by using the

expressions above with actual figures for modern transportation networks, population sizes, and realistic disease parameters, that not even a reduction of 1 order of magnitude of the mobility parameter is enough to bring the system below the invasion threshold. Unfortunately, this was overlooked by policymakers in charge during the COVID-19 pandemic at different scales, either national, regional, or urban. This will be the topic of Chapter 3 in this thesis.

### 2.3.4 Simulations

Computer simulations are an indispensable tool for scientists studying complex systems. Often, developing a comprehensive theory that can be fully expressed and solved analytically poses a significant challenge. Even when feasible, analytical solutions might be limited or trivial due to the necessary assumptions and simplifications, potentially overlooking unexpected or nonlinear behaviors. Moreover, in fields such as epidemiology, where contagion processes are typically non-reproducible and undesirable in real-life settings, research must rely on *in silico* studies using computer-simulated models of the dynamical processes of interest.

Numerical or computational solutions involve discretization, both of the system's state variables and of the time parameter. However, when choosing the simulation technique to represent a dynamical system, one of the first decisions to adopt is whether we are going to use a continuous-time or a discrete-time simulation algorithm. When we are writing down a series of mean-field equations we commented that these represent average values from a continuous-time Markovian stochastic process, whose probability distribution can be obtained by solving the corresponding master equation. There exists a stochastic algorithm (and several versions of it) to obtain exact solutions of the dynamics given by this continuous-time process, famously known as the Gillespie algorithm [173, 174], which has been also extensively used in network epidemiology [426]. In case we adopt a discrete-time scheme, the procedure is to simulate the dynamics with what is known as Monte Carlo mechanistic simulations. This also involves the generation of random outcomes for the elementary dynamical process occurring. Without entering into technicalities, the main difference in procedure between these two schemes lies in the following. The Gillespie algorithm samples both the most probable individual-based reaction given some parameters and the current state of the system, and the exact time at which the reaction takes place. Being a Markovian continuous-time algorithm, the reaction times are sampled from an exponential distribution. On the contrary, the discrete-time simulations occur at discrete time steps  $\Delta t$ , typically of unitary length  $\Delta t = 1$ . During each of these time steps, all the elementary dynamical processes for all the system constituents occur.

When either  $\Delta t$  or the reaction rates are small enough, both these approaches yield the same dynamical evolution, but one has to be cautious about comparisons between them [348]. Moreover, if one wants to build a mean-field set of differential equations related to the discrete-time simulations, special care must be taken [347].

In the following, we explain how to implement the discrete-time Monte Carlo mechanistic simulations for epidemic processes running on a metapopulation. This is the type of approach that has been used for instance in Chapters 3 and 4 (with a more complex mobility model in this case). We distinguish the main following ingredients: (i) Network generation, (ii) metapopulation generalities, (iii) the diffusion process, and (iv) the epidemic process.

## Network generation

The network structure establishes which set of location pairs will be effectively connected. Thinking of realistic settings, if we are dealing with the air transportation network, we should expect a scale-free degree distribution. If, on the contrary, we are modeling an urban environment at the district level, we could expect simpler structures like a regular lattice or even a complete network.

For abstract or synthetic settings, there exists a variety of generative network models [427]. A standard and versatile algorithm to generate synthetic uncorrelated networks is the so-called configuration model, which needs only the degree sequence of the network, and then randomly chooses any pair of nodes, takes one stub for each of them (half a connection), and proceeds through iteration until all the stubs are exhausted.

## Metapopulation generalities

When implementing our metapopulation model, one elemental thing to consider is the level of description we want or need to achieve. The framework is usually introduced as one where the agents are homogeneous and indistinguishable. If that is the case, the description is simple, and we do not need to keep track of the status and events related to every agent in the system. It suffices to count populations by health status and the corresponding flows moving between locations. If, by any chance, we need a more detailed description of the dynamics, we would move into the agent-based territory, possibly suffering higher memory and performance costs.

Once we have come to terms with the description needed, we have the parameters needed for the computations, and the problem initial conditions are provided (i.e. initial number of infected individuals, where the first infected are seeded, how many people occupy every location, etc.), we enter the dynamical loop where the stochastic

elementary processes occur. In this loop, we advance in steps of time  $\Delta t$ . For every time step, we loop over every node/location in the system and we compute the epidemic reactions and the diffusion. By the end of the time step, the system is properly updated and we move on to the next step, until either we run out of time steps or the dynamics reaches an absorbing state. Finally, we extract all the relevant information for post-processing and analysis. Since these are stochastic simulations, it is convenient to execute several realizations of the dynamics under the same set of parameters to obtain statistically reliable results.

### Contagion process

The contagion process normally assumes that individuals in each location mix homogeneously and have a finite number of contacts, so that the probability of a susceptible individual contracting the disease from an infected is proportional to the transmission rate  $\beta$  and normalized to the location size,  $\beta/N_l$ . Then, each susceptible is turned into an infected individual with probability:

$$P(\mathcal{S} \rightarrow \mathcal{I}) = 1 - \left(1 - \frac{\beta \Delta t}{N_l}\right)^{I_l}, \quad (2.57)$$

where  $I_l$  is the total population of infected individuals at location  $l$  (at the corresponding time step). Similarly, each infected individual spontaneously experiences a recovery or removal process with probability

$$P(\mathcal{I} \rightarrow \mathcal{R}) = \mu \Delta t. \quad (2.58)$$

Notice that  $\beta$  and  $\mu$  are rates and thus have inverse time units. It is a formality if  $\Delta t = 1$ , but this multiplication ensures that we are working with probability instead of rates.

These processes can be executed individual by individual or, more efficiently, when there is no need to distinguish individuals, through binomial samplings. Then, the total number of new infected cases at a location and time step are specifically generated from:

$$\Delta I_l(t) = \text{Binomial}(S_l(t), P(\mathcal{S} \rightarrow \mathcal{I})). \quad (2.59)$$

As for the individuals that leave the infected state, similarly, we generate the total number of newly removed cases:

$$\Delta R_l(t) = \text{Binomial}(I_l(t), P(\mathcal{I} \rightarrow \mathcal{R})). \quad (2.60)$$

The updating of the local populations by health status:  $S_l(t) \leftarrow S_l(t) - \Delta I_l(t)$ ,  $I_l(t) \leftarrow I_l(t) + \Delta I_l(t) - \Delta R_l(t)$ , and  $R_l(t) \leftarrow R_l(t) + \Delta I_l(t)$ , proceeds before the diffusion process at every location.

## Diffusion process

Independently of the particular macroscopic mobility model used, this is encoded in the matrix  $d_{ij}$  of rates between locations. Technically, again, since these quantities are rates, we should work with  $d_{ij}\Delta t$ , even though since normally  $\Delta t = 1$ , there is no need to. For any location or origin  $i$ , we must ensure that  $1 = \sum_j d_{ij}\Delta t$ . That is, this behaves as a probability and thus all the available options of displacement add up to 1.

Then, at every time step, each individual in any compartment in location  $i$  moves to a neighboring location  $j$  with rate  $d_{ij}$ . This is performed by a proportional or weighted sampling against  $d_{ij}$ , based on the origin location  $i$  of every individual. This again can be speeded up if there is no need for microscopic resolution. Now, since generally there are more than two options for the destination, we work with multinomial distributions. Then, for every compartment, the number of travelers from location  $i$  traveling to every neighboring location at time  $t$  is sampled from a multinomial distribution:

$$\mathbf{N}_{\text{travelers}} = \text{Multinomial}(X_i(t), \mathbf{d}_i\Delta t), \quad (2.61)$$

where  $N_{\text{travelers}}$  is the array of travellers from location  $i$  to every corresponding neighboring location,  $X_i(t)$  is the total population of the  $X$  health status at location  $i$  and time  $t$  (with  $X = S, I, R$ ), and  $\mathbf{d}_i\Delta t$  is the array of probabilities of traveling from location  $i$  to every neighboring location. In this context, the set of neighboring locations always includes the origin location.

As a cautionary note, the updating of the traveling population must proceed in parallel and not sequentially in the location loop. All the epidemic and diffusion processes are happening within the time step  $\Delta t$ , but the complete new configuration is not seen after all changes have happened at the beginning of the next time step. Otherwise, we would have people who experienced the epidemic reactions in their location, travel to another location, and could again be subjected to new epidemic reactions there.

Finally, it is worth mentioning that, unless we introduce some type of intervention to the mobility process,  $d_{ij}$  is an object that can be pre-computed at the beginning of the whole dynamical evolution, saving a lot of computational operations, especially if the system is very large.

## 2.4 Microscopic human mobility models

In section 2.3, when introducing the topic of spatial epidemics, we saw that mobility is an essential part of the metapopulation framework. Additionally, we saw that, typically, the way mobility is introduced is through a macroscopic (population-based) Markovian approach. That is to say, it does not resolve at the individual scale and thus can potentially wash away relevant heterogeneities and realistic microscopic features at play.

In a separate but not isolated area, that of human mobility analysis, important discoveries have been unveiled in the last decade, advancing the understanding of how we, humans, move. These advances have been possible in recent years thanks to the mass availability of mobile phone records, global-positioning data, and other data sets. As in many other fields that benefited from the Big Data era, this pushed the field empirically and allowed in turn to advance the theoretical models at hand.

These unveiled regularities are interesting and valuable in themselves but also have the potential to advance the understanding of complex phenomena in other areas, as is the case of spatial epidemics and the interplay of spreading phenomena with human mobility.

In this section, we take a quick tour of selected microscopic human mobility models to illustrate their motivation and main features. One of them, the d-EPR model, will be fundamental to the work developed in Chapter 4.

### 2.4.1 Explorers and returners: The EPR family of models

#### The EPR model

Song et al. [428] found that human trajectories follow several highly reproducible scaling laws. Yet, many of those laws were not explained by, at the time, standard mobility models such as the continuous-time random walk (CTRW) model or the Lévy-flight model. Worse than that, the findings were in direct contradiction with the theoretical predictions. The research was based on the inspection of two datasets: i) a one-year period of time-resolved trajectories of three million anonymized mobile phone users, and ii) an anonymized location record of 1,000 users who signed up for a location-based service. Preliminarily, computation of the displacement  $\Delta r$  at hourly intervals followed  $P(\Delta r) \sim |\Delta r|^{-1-\alpha}$ , with  $\alpha = \pm 0.05$  and an expected cutoff at  $\Delta r \approx 100\text{km}$ . And the waiting time  $\Delta t$ , defined as the time a user spent at one location, followed  $P(\Delta t) \sim |\Delta t|^{-1-\beta}$  with  $\beta = 0.8 \pm 0.1$ , and a cutoff of  $\Delta t = 17\text{h}$ . These fat-tailed distributions suggest that humans follow a CTRW during their daily mobility. However, three empirical observations were also uncovered:

1. **Exploration exhaustion.** The number of distinct locations  $S(t)$  visited by a randomly moving object is expected to follow

$$S(t) \sim t^\mu, \quad (2.62)$$

where  $\mu = 1$  is for Lévy flights and  $\mu = \beta$  for CTRW. Measurements in [428], though, indicated  $\mu = 0.6 \pm 0.02$ , smaller than the CTRW prediction above. There is a decreasing tendency of the user to visit previously unvisited locations.

2. **Heterogeneous visitation frequency.** The probability  $f$  of a user visiting a given location is expected to be asymptotically uniform everywhere for both Lévy flights and CTRWs. In contrast, the visitation patterns of humans are rather uneven, so that the frequency  $f$  of the  $k$ th most visited location follows Zipf's law

$$f_k \sim k^{-\zeta}, \quad (2.63)$$

where  $\zeta \approx 1.2 \pm 0.1$ , suggesting that the visitation frequency distribution follows  $P(f) \sim f^{-1+1/\zeta}$ .

3. **Ultraslow diffusion.** The CTRW model predicts that the mean square displacement (MSD) asymptotically follows  $\langle \Delta x^2(t) \rangle \sim t^\nu$  with  $\nu = 2\beta/\alpha \approx 3.1$ . As both  $P(\Delta r)$  and  $P(\Delta t)$  have cutoffs, asymptotically the MSD should converge to a Brownian behavior with  $\nu = 1$ . However, this convergence is too slow to be relevant in the observational time frame available. Either way, CTRW predicts that the longer a human trajectory is followed, the further it will drift from its initial position. Yet, humans show a tendency to return home on a daily basis, suggesting that simple diffusive processes, which are not recurrent in two dimensions, do not offer a suitable description of human mobility.

Therefore, on the one hand, these summarized findings indicate that individual human mobility does follow reproducible scaling laws. However, they also show systematic deviations from the predictions of standard null models.

In order to explain the observed features, two principles that should govern human mobility were introduced:

- i) **Exploration.** Random-walk models assume that the next diffusive step is independent of the previously visited locations. However one of the scaling laws found indicated that the tendency to explore additional locations decreases with time.

- ii) **Preferential return.** In contrast with the random-walk-based model for which the visitation probability is random and uniform in space, humans show a significant propensity to return to the locations they visited frequently before (i.e. their home or workplace).

Accordingly, they formulated the EPR model for individual mobility (see Figure 2.5), which incorporated the ingredients above, and allowed them to explain the apparent anomalies found.

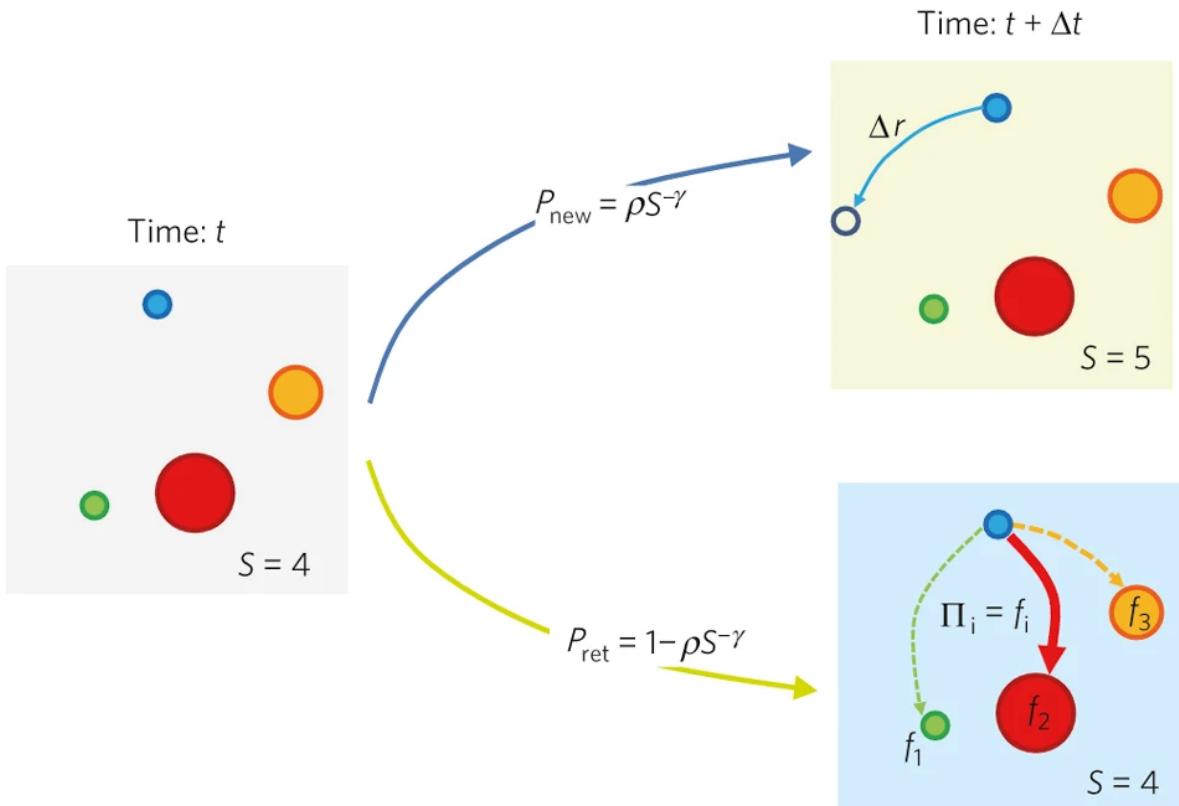


Figure 2.5: **Schematic description of the individual-mobility model.** Starting at time  $t$  from the configuration shown in the left panel, indicating that the agent visited previously  $S = 4$  locations with frequency  $f_i$  that is proportional to the size of circles drawn at each location, at time  $t + \Delta t$  (with  $\Delta t$  drawn from the  $P(\Delta t)$  fat-tailed distribution) the user can either visit a new location at distance  $\Delta r$  from his/her present location, where the distance is chosen from the  $P(\Delta r)$  fat-tailed distribution (exploration; upper panel) or return to a previously visited location with probability  $P_{ret} = 1 - \rho S^{-\gamma}$ , where the next location will be chosen with probability  $\Pi_i = f_i$  (preferential return; lower panel). Reference: Image obtained from Song et al. [428] (<https://www.nature.com/articles/nphys1760>).

The model intends to describe the trajectory of an individual, assuming that at time  $t = 0$  the individual is at some preferred location. After a waiting time  $\Delta t$  chosen from the  $P(\Delta t)$  distribution, the individual will change their location. Thus, it is assumed

that the individual has two choices. Either explores a new location with probability:

$$P_{\text{exp}} = \rho S^{-\gamma}, \quad (2.64)$$

or returns to an already visited location with the complementary:

$$P_{\text{ret}} = 1 - \rho S^{-\gamma}. \quad (2.65)$$

Parameter  $\rho$  characterizes the individual tendency to either explore or return, and it is directly related to the rate of different visits performed during a certain time interval. The number of different visits performed until a particular instant of time is represented with  $S$ , and decays with parameter  $\gamma$ . In the exploration phase, the new location is visited by performing a jump of distance  $\Delta r$ , chosen from the  $P(\Delta r)$  distribution, while their direction is uniformly randomly chosen. Naturally, as the individual visits their new position, the number of previously visited locations increases from  $S$  to  $S + 1$ . In the preferential return phase, the probability  $\Pi_i$  to visit an already visited location  $i$  is chosen to be proportional to the number of visits the individual had to that location. That is, there is a preferential attachment mechanism at play:

$$\Pi_i = f_i. \quad (2.66)$$

When comparing the newly proposed individual model with the empirical data, the authors found that  $P_{\text{exp}} \propto S^{-\gamma}$ , with  $\gamma = 0.21 \pm 0.02$ , confirming the exploration hypothesis. Contrasting the ratio  $P_{\text{new}}$  with  $S^{-\gamma}$ , the  $\rho$  parameter was measured. By plotting  $P(\rho)$  for their dataset, they found that  $\rho$  follows approximately a normal distribution with mean at  $\langle \rho \rangle \approx 0.6$ . Moreover, they also found that  $\Pi = f$ , confirming the validity of the preferential return.

### d-EPR model

Following the tracks by [428], Pappalardo et al. [429] discovered the emergence of two distinct mobility profiles - explorers and returners- by analyzing datasets on call records and private vehicle GPS tracks. It is claimed that these two distinct mobility profiles are well-separated and show distinctively characteristic movement attributes.

The total radius of gyration  $r_g$  metric, defined as:

$$r_g = \sqrt{\frac{1}{N} \sum_{i \in L} n_i (\mathbf{r}_i - \mathbf{r}_{\text{cm}})^2}, \quad (2.67)$$

is introduced to characterize the typical distance traveled by an individual. Here,  $L$  is the set of locations visited by the individual,  $\mathbf{r}_i$  is a 2-dimensional vector describing the geographic coordinates of location  $i$ ;  $n_i$  is the visitation frequency or the total time

spent by the individual in location  $i$ ;  $N = \sum_{i \in L} n_i$  is the total number of visits or time spent, and  $\mathbf{r}_{\text{cm}}$  is the center of mass of the individual. Additionally, one can define the  $k$ -radius of gyration  $r_g^{(k)}$  over the  $k$ -th most frequented locations  $L_1, \dots, L_k$  as:

$$r_g^{(k)} = \sqrt{\frac{1}{N_k} \sum_{i \in L}^k n_i (\mathbf{r}_i - \mathbf{r}_{\text{cm}})^2}. \quad (2.68)$$

With this metric, we can understand how the  $k$ -th most frequented locations of an individual determine the characteristic distance traveled by them. Now,  $\mathbf{r}_{\text{cm}}^{(k)}$  is the center of mass computed on the  $k$ -th most frequented locations;  $N_k$  is the sum of the weights assigned to the  $k$ -th most frequented locations ( $r_g^{(k)} = r_g$  if  $k \geq N$ ). Thus,  $r_g^{(k)}$  represents the mobility range restricted to the  $k$ -th most frequented locations. Then, for instance, if an individual's  $r_g^{(2)} \approx r_g$ , their characteristic traveled distance is dominated by the two most frequented locations. Conversely, if  $r_g^{(2)} \ll r_g$ , then clearly the individual's mobility cannot be satisfactorily characterized by the top two locations.

The authors in [429] characterized the mobility data in their data sets by quantifying the degree of similarity between overall ( $r_g$ ) and recurrent mobility ( $r_g^{(k)}$ ). One way of looking at this is by computing the ratio  $s_k = r_g^{(k)}/r_g$  for every individual and drawing the probability distribution  $P(s_k)$ . It was found that this distribution shows a certain bimodal regime for low values of  $k$ . One of the peaks sits around  $s_k = 0$ , whereas the other one does it around  $s_k = 1$ . A value of  $s_k \rightarrow 0$  tells us that, for those individuals, the  $k$ -th most visited locations are but a rather negligible contribution of their total mobility. Thus, we can state that those individuals are explorers: during a certain time span, they visit a high number of locations. The other mode observed,  $s_k \rightarrow 1$ , tells us that, for those individuals, the  $k$ -th most visited locations greatly comprise their total mobility. Then, we can say that those individuals are returners: given the same time span, they concentrate their mobility in a very reduced number of locations, thus showing a high level of recurrence.

The empirical findings here were precisely tested against the above-introduced EPR model developed in [428], and the authors found that failed to properly account for this bimodality in real human mobility trajectories. The empirically observed split into returners and explorers was absent from the model-generated trajectories, and the model overestimated by more than an order of magnitude the number of locations needed to accurately estimate the total radius of gyration. At odds with the real data, for the EPR model, it was found no significant correlation between  $r_g$  and the sum of the distances for the  $k$ -th most visited locations, neither for  $k$ -returners nor for  $k$ -explorers. It was suspected that a potential source of mismatching could arise

from the fact that in the EPR model individuals can travel arbitrarily large distances, increasing their  $r_g$  with each jump. Then, the authors in [429] were led to correct this by proposing a modification to the original EPR model, the d-EPR model [430]. Here, the exploration phase is driven by the gravity model of human mobility, and, therefore, visiting a new location depends on its distance from the location of origin and also its relevance or attractiveness. Taking into account these spatial attributes, it was found that the d-EPR model generates trajectories that are in much better agreement with the empirical data.

### EPR extensions

The formulation of the original EPR model sparked renovated research on the microscopic properties of human mobility and generated a new wave of models with the exploration and preferential dynamics as an underlying mechanism. Each of these iterations tried to add more realistic features in order to explain in a more detailed and accurate way the observed patterns in the data and overcome the limitations of previous models. Thus, this has left us with a landscape of EPR models. Even though human mobility is not the main focus of this dissertation, we outline some of these models for the reader to have a broader context of the field.

**Recency** The concept of recency was introduced to solve discrepancies related to the standard preferential return mechanism. The issue is that, as conceived, earlier discovered locations have a cumulative advantage over later visited locations. In turn, this precludes people from changing location preferences, which is something that contradicts observed data. In effect, using mobility data (CDRs) and location-based check-ins produced by thousands of users, Barbosa et al. [431] uncover a strong tendency of individuals to return to recently visited locations, and show that such tendency is not conditioned to the previous visitation frequencies. The authors then proposed a model in which the return phase in human movements also considers recently-visited locations and not solely frequently-visited locations. Their approach is based on the empirical evidence that the longer the time since the last visit to a location, the lower the probability of observing a user at this location [51].

The return stage then consists of two exclusive processes to select the location to return to. With probability  $\alpha$ , a preferential return based on the visitation frequency rank acts. In this case, the location  $i$  is selected with probability  $\Pi_i \propto k_f(i)^{-1-\gamma}$ , where  $k_f(i)$  is the frequency rank of location  $i$ . With the complementary  $1 - \alpha$ , the novel rank-based return occurs. Here, the  $i$ -th last visited location is selected from a Zipf distribution with probability  $p(i) \propto k_s(i)^{-\eta}$ , where  $k_s(i)$  is the recency-based rank of the

location  $i$ . Notice that when  $\alpha = 1$ , the original preferential return of the EPR model is recovered, while if  $\alpha = 0$ , visitation returns will be based exclusively on the recency. Authors tested that when  $\alpha = 0$ , the heavy tail of the visitation frequency disappears, while if  $\alpha = 1$ , the power law of the recency distribution vanishes. This suggests that both mechanisms must be present in order to reproduce those two observed features in their study.

**Memory** If one runs any of the standard EPR mobility models, one will soon discover that even though the rate of exploration can significantly decay, given a long enough time span, individuals *never* end up discovering new locations. This is surely in stark contrast with everyday human behavior which even though heterogeneous, is rigidly marked by some space and time constraints, and the exploitation of familiar places. In Alessandretti et al. [432], they look precisely at these opposite trends in human mobility. By analyzing high-resolution multi-year traces of big data sets of human trajectories, evidence is found of a conserved quantity: the number of familiar locations an individual visits at any point has a typical size of  $\sim 25$ . This discovery is used to improve the EPR model through the introduction of a memory mechanism. In this modification, agents obey the same exploration strategy as in the EPR model but dispose of a limited memory  $M$ . Hence, the return probability to a given location  $i$  only accounts for the total number of visited locations occurring at most  $M$  days before any exploration step. The introduction of this simple modification memory qualitatively reproduces all the observations exposed in the work, including the long-term evolution of the activity set, improving previous models.

**Preferential Exploration** This model was introduced in [433], once again to overcome some limitations that the original EPR model presented. The empirical analysis done by Schläpfer et al. revealed that:

1. Spatio-temporal population flows to locations follow a highly reproducible scaling law of the form

$$\rho_i(r, f) = \frac{\mu_i}{(rf)^\eta}, \quad (2.69)$$

with scaling exponent  $\eta \approx 2$ . Here,  $\rho_i(r, f)$  is the spectral flow of location  $i$ ,  $r$  is the travel distance, and  $f$  is the visitation frequency.

2. Although the magnitudes of these flows vary substantially across locations, they show a systematic spatial clustering.

The EPR model effectively generates the distance-frequency scaling of the flows to individual locations with a scaling exponent  $\eta \approx 2$ , showing an excellent agreement

with the empirical observations. However, the model is unable to reproduce the heterogeneity in the attractiveness of locations and their systematic spatial clustering. Actually, the trajectories of individuals are not independent but spatially coupled through common attraction points, the so-called points of interest: people tend to go to popular places, and those places are popular because other people go there in turn. The problem lies in that the EPR model ignores the coupling of agents' motion and thus generates attractiveness values for locations that are rather homogeneous and uniform across space.

To resolve this flaw, the authors introduced the preferential exploration and preferential return (PEPR) model. The PEPR not only generates the distance-frequency scaling of the spectral flows with the correct scaling exponent but also leads to the formation of clear spatial clusters that follow an area distribution that is quantitatively consistent with the data.

Regarding the technical aspects of the model, what differs from the EPR model is the exploration part. Here, the radial jump distance  $\Delta r$  is still sampled from the same distribution  $P(\Delta r) \sim |\Delta r|^{-1-\alpha}$  as in the original EPR model. But now, the direction  $\theta$  of the motion is no longer sampled uniformly from  $[0, 2\pi)$ . Instead, a mechanism to preferentially select direction towards regions of high visitation is introduced. Assuming a regular lattice as the spatial substrate, let, for a given cell  $i$ ,  $\tilde{d}_i(\theta; R)$  be the effective distance traveled by all agents to all cells within distance  $R$  and between angles  $\theta$  and  $\theta + d\theta$ . Then agents starting from cell  $i$  sample  $\theta$  from the distribution  $P(\theta; R, v) \sim \tilde{d}_i(\theta; R)^\nu$  with parameter  $\nu \geq 0$ .

**Bursty visitation** In [434], the authors propose a modification to the preferential return step in the original EPR model to reproduce the observed bursty behavior of location visitation in human mobility. Bursty behavior in the number of visits to a location could be attributed to dynamical aspects like sociocultural events occurring at a particular date, or to exogenous factors such as a location gaining popularity due to social media promotion.

Even though the EPR model satisfactorily reproduces empirical behaviors in human mobility including the power-law growth of the number of visited locations with time and power-law decay of visitation frequency of ranked locations, does not deal with that kind of bursts in the visitation frequency to a location. The extension proposed to accommodate this fact is the exploration and rank-shift return (ERSR) model, which is able to capture how unpopular locations may change into bursty visited ones. The authors also find that replacing the preferential return with the ranking return scheme still preserves the original features of the EPR model.

In the ERSR model, the rank-shift return mechanism introduces an additional parameter  $\theta$ , which controls the probability of a certain location moving up to a higher ranking position in the rank-based preferential return selection. Starting at each time window, the individual's locations are ranked by visitation traffic in the first place. Then, every location may move up to a new position towards the front of the list with probability  $\theta$ . The new rank is randomly chosen with a uniform probability between 1 and its current rank  $j$ . During the time window, the individual returns to a location visited with probability  $P_i = r_i^{-\delta}$ , where  $r$  is the rank in the rearranged list and  $\delta > 0$  is a free parameter with a typical value of 1.

**Social dimension** As we appreciate in this model review, models tend to focus on the spatial and temporal dimensions of mobility only, while the social dimension is often neglected. On the other hand, models with some social mechanisms embedded tend to have rather trivial or unrealistic spatial and temporal mechanisms. This gap is what Cornacchia and Pappalardo try to fill in [435] with the STS-EPR model.

The STS-EPR combines (i) a mechanism that takes into account the spatial distance between locations and their collective relevance; (ii) a temporal mechanism to capture the individual's tendency to follow a circadian rhythm; (iii) a mechanism that models the social dimension of human mobility. Also, it includes an action-correction mechanism that deals with borderline cases during the simulation.

More schematically, the main steps of the model comprise the following processes:

- Exploration, with  $P_{\text{exp}} = \rho S^{-\gamma}$ .
- Social, with probability  $\alpha$ . The individual selects another agent among their social contacts. The probability of being selected is proportional to a mobility similarity metric. After the contact is chosen, the candidate location to explore is an unvisited location for the focal agent that was visited by their contact.
- Individual, with probability  $1 - \alpha$ . The focal agent chooses a new location to explore following the gravity model for human mobility.
- Return, with  $1 - P_{\text{exp}}$ .
  - Social, with probability  $\alpha$ . Similarly, a contact for the focal agent is chosen, and a common already visited location among the two is selected following the frequency-based preferential attachment scheme.
  - Individual, with probability  $1 - \alpha$ . The individual chooses a location to return to based on the conventional preferential return scheme.

The authors assessed the realism of this sophisticated model by computing the distribution of a series of standard mobility metrics like the jump length  $\Delta r$ , radius of gyration  $r_g$ , location frequency  $f(r_i)$ , or the waiting time  $\Delta t$ , and compared its performance against other models like GeoSim [436] and DITRAS [437]. Results showed that the STS-EPR produced more realistic trajectories than the models lacking either the social, spatial, or temporal dimensions.

### 2.4.2 Scales in human mobility: The container model

The models showcased in the previous section all share the exploration and preferential return mechanism as the underlying principle. This modeling effort has successfully managed to explain features of human mobility that previous standard models struggled to do it or whose predictions were in stark contradiction with the empirical evidence. The model in this section departs from the EPR essence but still addresses recent findings in the human mobility literature. We refer to the lack of evidence for characteristic spatial scales in how people travel [438, 51, 428, 439]. As we have explicitly seen at the beginning of this section, studies have shown that the distribution of displacement lengths  $\Delta r$  traveled by an individual has a power-law (scale-free) tail  $P(\Delta r) \sim \Delta r^{-\beta}$  over several orders of magnitude. However, Alessandretti et al. [440] pose the question: How is it possible that our intuitive conception of space is clearly hierarchical and characterized by typical scales, whereas an increasing number of diverse datasets suggest that human mobility is scale-free? The container model of human mobility arises when trying to answer this question.

Each typical scale of human mobility can be conceived as a container of certain mobility behavior. Naturally, these containers present a characteristic length scale. Containers may, in principle range from ridiculously tiny spaces, such as a room, to our practically maximum range, the whole planet, and all the scales in between (neighborhoods, cities, countries, *et cetera*). The observed power law arises when we aggregate mobility behavior within containers and mobility that transports a person between containers. Specifically, it is assumed that for each individual, physical space is organized as a nested structure of containers. This structure is surely influenced by the organization of the transportation system and the concrete structure of our built environment. Since the nested container structure drives and constrains how individuals move, the authors claim that could be inferred from raw mobility data.

We devote now some lines to explain the mechanistic aspects of the model. For each individual, the physical space is conceived as a hierarchy of  $L$  levels, ordered from the smallest to largest (for instance, individual locations to countries). At any level  $l$ , space is partitioned into topologically compact containers, with a characteristic

size. For  $l < L$ , a container is fully included within a single parent container (for example, each neighborhood is part of a single city). Hence, each geographical location  $k$  can be identified as a sequence of containers,  $k = (k_1, \dots, k_l, \dots, k_L)$ , where container  $k_l$  is included in  $k_{l+1}$ . Then, each container  $k_l$  is characterized by its probability to be selected within its parent container, and its attractiveness value  $a(k_l)$ . We define the level distance  $d(j, k)$  between locations  $j$  and  $k$  as the highest index at which the two sequences of containers describing  $j$  and  $k$  differ. The traces can be modeled individually; each yielding a unique hierarchical structure.

Based on the assumption that the amount of time spent in a container depends on its place in the hierarchy, we design a model of trajectories, where the probability of transitioning from location  $j$  to location  $k$  depends on the level distance between them. For an agent located in  $j$ , the probability of moving to  $k$  is modeled as the product of two factors:

$$P(j \rightarrow k) = p_{d(j,k), d(j,h)} \Pi_{l \leq d(j,k)} a(k_l). \quad (2.70)$$

The first factor,  $p_{d(j,k), d(j,h)}$ , represents the probability of traveling at level distance  $d(j, k)$ , given that the current location  $j$  is at level distance  $d(j, h)$  from the individual home location,  $h$ . This probability follows a multinomial distribution, which must depend on level distance from home to account for the fact that higher-level transitions are more likely when individuals are not in the home container; for example, one is typically more likely to transition at the country scale, when not in the home country. The second factor  $\Pi_{l \leq d(j,k)} a(k_l)$  is the probability of choosing a specific location  $k$  at that level distance, where  $a(k_l)$  is the attractiveness of a container at level  $l$  including location  $k$ .

The container model provided better results and characterization of empirical data when tested against the EPR model and other six state-of-the-art micro-mobility models in key features of human mobility. Additionally, the model was further validated by aggregating users based on demographics and contextual features: gender, level of urbanization, and walkability score, with promising results in line with the literature, allowing for a better understanding of mobility behavior. Despite this, it is also true that presents some limitations in scope. For instance, some important features are neglected, including temporal visitation patterns, exploration, and the structural connectedness of geographical networks. It is also true that some of these aspects could be incorporated in future revisions of the model.



## **Part II**

# **Metapopulation models**



# Chapter 3

## Assessing the effectiveness of perimeter lockdowns at the urban scale: the case of Madrid

*Madrid es España dentro de España. ¿Qué es Madrid si no es España?*

— Isabel Díaz de Ayuso

*Si cada español hablase de lo que entiende, y de nada más, habría un gran silencio que podríamos aprovechar para el estudio.*

— Manuel Azaña (atribuida también a Antonio Machado)

### 3.1 Introduction

During the COVID-19 pandemic, we have become acquainted with a myriad of measures to halt or at least control the spreading of the novel coronavirus SARS-CoV-2 at different scales. Measures that are different from vaccination and medical treatments are called non-pharmaceutical interventions (NPIs). These may include travel bans, lockdowns of different geographical extensions, curfews, restrictions on occupancy in public closed spaces, or self-protection measures such as wearing masks or sanitation (see [24] for a comprehensive review of NPIs within the context of COVID-19) [16, 441, 442, 443, 444]. The objective of most NPIs is to reduce the number of contacts between individuals, with the ultimate goal of breaking transmission chains and, subsequently, controlling the spread and minimizing hospitalizations and deaths. However, certain NPIs may bring severe economic downturns [445, 446, 447], dysfunctional supply chains [448, 449, 450], and social backlash [451, 452, 453]. Therefore, competent authorities should be properly informed of the benefits and costs of the possible interventions, especially if these are enforced.

Given the negative impact that global lockdowns and quarantines inflict on societies, it is important to look for alternative strategies in successive COVID-19 waves. In this

regard, a proposed alternative in the context of urban settlements is that of localized mobility restrictions, which is a more fine-grained restriction that acts only on the neighborhoods that are especially affected by an epidemic outbreak, instead of acting indiscriminately on the whole city. In Spain, the city of Madrid stands as a unique and paradigmatic example of pursuing the so-called perimeter lockdowns (PLs) at the level of Basic Health Zones (BHZ, subareas of the city defined by public health criteria). The main characteristic of these lockdowns was that they were implemented in highly integrated urban areas. As such, they restricted mobility in and out of the BHZ but allowed residents to go to work, attend academic activities, or for other essential purposes in the rest of the city, for which public transportation was available. Similarly, small businesses such as restaurants and shops were allowed to open, although at 50% maximum capacity. Hence, the lockdown was much more permeable than those implemented at larger scales, in which mobility is completely restricted between cities or regions. The case of Madrid reached the literature through a commentary article by members of Madrid's regional public health counseling [454]. But the commentary received several responses that pointed to inaccuracies and lack of self-criticism given the severe impact of the epidemic in the region [455, 456, 457, 458]. Two different studies approached the question through statistical analyses and concluded its lack of effectiveness at least in the real-life setting under consideration [459, 460]. Independently, perimeter lockdowns within a large city were also implemented during the first wave of COVID-19 in Santiago de Chile. In [461], the authors similarly conclude, through causal inference methods, that this kind of strategy is ineffective in the context of highly integrated human settlements, such as cities.

Due to the rarity of this strategy, to the best of our knowledge, it has not been explored using mechanistic models. In contrast to studies that rely on the statistical analysis of the collected data, mechanistic models can provide important insights into the effects of any strategy regardless of the peculiarities of the situation under analysis. This is important to properly gauge if a strategy will be effective in other places and under different conditions. Thus, in this work, we aim to assess the effectiveness of PLs at the urban scale using these models. To do so, we first take a brief look at the data collected during the implementation of the strategy in Madrid. This case serves as an example of how this strategy can be implemented in practice. Then, we build a simplified model of COVID-19 transmission to test the effect of PLs at the urban scale using a metapopulation framework. It is worth stressing that, even though we use the case of Madrid as an inspiration to conceive and guide the study, the framework developed is general enough to be extended to other locations. As such, we do not intend to replicate Madrid's epidemiological trajectory but rather design, based on

epidemiological first principles, a minimal model that captures the essential features of the spreading at the urban level and that can inform public health policy on the outcome of potential scenarios.

## 3.2 Material and methods

### Madrid's surveillance data

To test the effectiveness of the measures in the particular case of Madrid, we have explored the official epidemiological data reported by the government [462] and collected the time series for the 14-day cumulative incidence rate (14d CIR) for all the BHZs belonging to the city of Madrid. The 14d CIR is a standard in the field of public health [463] and is measured as:

$$14\text{d CIR} = \frac{\sum_{i=1}^{14} \text{new cases}_i}{N} \times 10^5 \text{ inhabitants}, \quad (3.1)$$

where  $N$  is the total population assuming a fully susceptible population at the beginning of the outbreak, and the summation of new cases is done over the previous 14 days. The value is given for every 100,000 inhabitants. Therefore, for the sake of brevity, all references to this variable will omit this factor.

In Madrid, this indicator was used to determine whether or not a particular BHZ had to be confined, together with other auxiliary information such as its trend or the presence of community transmission. The threshold value upon which a PL was implemented varied in time. Consequently, for a more complete picture, we also reconstructed the 14d CIR threshold, or activation threshold, by looking at official bulletins [464]. The confined status of a BHZ was held for two weeks and then reviewed according to its epidemiological situation.

### Metapopulation model

Metapopulation models have been applied to global-scale situations, such as the global spread of influenza-like diseases [213, 281], COVID-19 [465, 466, 467], and to regional and national scales [468, 280]. The metapopulation framework [217, 242] is versatile and allows the description of a system on varying spatial scales, and thus can also be conceived on the urban scale.

In this work, we create a metapopulation system composed of 21 subpopulations representing each of the 21 administrative districts of the city of Madrid, Spain. Within each subpopulation, the epidemic dynamics are governed by the classical SIR model under the well-mixed approximation. It must be noted that, even though this model

cannot capture some of the particular characteristics of COVID-19, it can still provide useful insights for settings in which those characteristics do not play a major role, as in our analysis [469, 470]. In cases in which case detection or the latency period might be relevant, it would be necessary to use an SEIR model or one of its variants.

## Epidemic modeling

At the district level, we implement a well-mixed SIR compartmental model. In this model, individuals are classified according to their health status: susceptible ( $\mathcal{S}$ ) if they are susceptible to catching the disease, infected ( $\mathcal{I}$ ) if they have been infected and can infect others, and removed ( $\mathcal{R}$ ) when they are either recovered or deceased. Within each district, the transition between compartments results from the following rules, iterated at each time step, corresponding to  $\Delta t = 1$  day [239]. The contagion interaction reads:

$$\mathcal{S} + \mathcal{I} \xrightarrow{R_0/(T_I N)} \mathcal{I} + \mathcal{I} \quad (3.2)$$

Susceptible individuals in district  $i$  are infected with the following expression for the probability:

$$P_i(\mathcal{S} \rightarrow \mathcal{I}) = 1 - \left(1 - \frac{R_0 \Delta t}{T_I N_i}\right)^{I_i}, \quad (3.3)$$

where  $R_0$  is the basic reproduction number of the disease,  $T_I$  is the mean infectious time,  $N_i$  represents the number of people in district  $i$  and  $I_i$  represents the number of infected individuals in the said area. This probability is based on a frequency-dependent force of infection assumption.

Recovery/removal is modeled as a spontaneous transition:

$$\mathcal{I} \xrightarrow{1/T_I} \mathcal{R}. \quad (3.4)$$

Infected individuals become removed at a rate inversely proportional to the mean infectious period,  $T_I$ . The probability for an infected individual to transition to the removed state is thus simply  $P(\mathcal{I} \rightarrow \mathcal{R}) = \Delta t/T_I$ .

The new infected and removed individuals are generated stochastically by sampling from binomial distributions using the above probabilities, respectively [239, 219]<sup>1</sup>.

It must be clarified that we are assuming a situation of general disease awareness in society and thus with generalized mitigation measures in place like wearing face masks, extra hygiene measures, some capacity limits, avoidance of crowded situations,

---

<sup>1</sup>Although we are talking of “individuals” here, the model is not agent-based, that is, does not resolve at the microscopic or individual level. Instead we work with populations (and flows of them) as compartmentalized by health-status. A microscopic implementation is indeed feasible and not hard to do, but given the scope of the work is not required, and we avoid incurring in additional computational costs.

etc. This awareness situation is more similar to the one experienced in Madrid at the time of the onset of the perimeter lockdown strategy. In this context, it would be more appropriate to model the spreading using the effective reproductive number  $R_t$  rather than the basic reproductive number  $R_0$ , the latter being defined only for a fully susceptible population without any mitigation strategies in place. For simplicity, we set  $R_0 \approx R_t$  to include directly all these factors without explicitly modeling them [470]. We set  $R_0 = 1.25$  which is in the range of the estimations of the effective reproduction number performed by the Spanish Ministry of Health in August 2020, when the second wave of infections started [471]. Finally, we use an infectious period of  $T_I = 4.5$  days. All simulations are seeded with 5 infected individuals in the central district *Centro*. The simulations end when the total number of infected individuals in the system reaches zero.

## Mobility

To estimate realistic mobility patterns we use the data from a mobility survey carried out by the Spanish Ministry of Transport, Mobility and Urban Agenda [472]. This survey provides the estimated number of individuals going from one district to another every day from February 2020 to May 2021. We use the data from the pre-pandemic period to build a baseline mobility matrix that is not influenced by any changes in social interactions produced by the pandemic. Indeed, as described above, we use an effective value of the basic reproduction number that already incorporates all these effects. Thus, if we were to use the real data from the period under study, we would be counting certain interventions twice.

Each element of the mobility matrix,  $D_{ij}$ , represents the rate at which individuals move from district  $i$  to district  $j$ . This can be estimated from the origin-destination (OD) matrix provided by the Spanish government,  $M$ , in which  $M_{ij}$  represents the number of individuals traveling from district  $i$  to district  $j$ . In particular, the elements of  $D$  are defined as:

$$D_{ij} = \begin{cases} \kappa \frac{M_{ij}}{\sum_j M_{ij}} & \text{if } i \neq j \\ 1 - \sum_{k \neq i} D_{ik} & \text{if } i = j \end{cases} \quad (3.5)$$

where  $D_{ij}$  is the rate at which individuals move from district  $i$  to district  $j$  and  $\kappa$  is a general mobility parameter. The baseline values of  $D_{ij}$  are obtained considering  $\kappa = 1$ . When  $\kappa \rightarrow 0$ , we see that  $D_{ii} \rightarrow 1$ , so that the rate of individuals staying at their origin district approaches one. The particular values of  $D_{ij}$  are calculated using data from the Ministry's survey from the pre-pandemic period. Then, at each step,  $\Delta t$ , the number of travelers from district  $i$  to every destination  $j$  is sampled from a multinomial process with probabilities  $D_{ij}\Delta t$ .

### Perimeter lockdown strategy

The response against the spreading process is assumed to be the best-case scenario, with perfect information and neither exceptions nor violations of the policies enacted by the authorities. The observable to monitor the effectiveness in every district is the same as in the real-life setting, the 14d CIR. We set a risk threshold  $\Theta$  so that whenever  $14d\text{ CIR}_i \geq \Theta$  for district  $i$ , we say that this district is at risk and thus it is set under a perimeter lockdown. Albeit this parameter changed many times during the implementation of this strategy in Madrid (see Figure 3.1), we fixed it for each simulation.

The implications for a district under this type of lockdown are twofold:

1. Travel bans. Flows of travelers from and to the district at risk are completely suppressed. Note that this intervention is independent of the value of  $\kappa$ , which controls the overall mobility in the system even in the absence of restrictions. It is also important to remark that this is again a best-case scenario since in reality trips involving affected areas were allowed for a variety of circumstances.
2. Transmissibility reduction. Localized lockdowns usually come with additional measures to try to control the spreading within the affected area. This may include a rise in disease awareness through mass media and authorities' advertisements and actions like occupancy limits and mandatory wearing of face masks. The combination of such actions would translate into a reduction of the transmissibility, its magnitude being dependent on people's compliance and the intensity and effectiveness of those same measures. To model this, we set a local transmissibility rate for each closed district  $i$ ,  $\beta_i = \chi\beta$ , where  $\beta$  is the general transmissibility rate related to the basic reproductive number of the disease  $R_0 = \beta T_I$ , and  $\chi$  is the fraction of the reduction in transmissibility. For normal districts,  $\chi = 1$ , but when a district is under lockdown, this quantity is  $\chi < 1$ .

Once these measures are in place for a district at risk, the restriction remains until the outbreak ends in the whole system. That is when the global incidence is  $I = 0$ . In reality, however, once the affected areas stopped fulfilling the lockdown requirements, these restrictions were lifted. Since we are not seeking a replication of the Madrid epidemiological trajectory, but a general exploration and understanding of the basic dynamics under these types of lockdowns, we limit ourselves to studying what happens in this extreme scenario.

## Population and mobility data

The metapopulation model is informed by real population and mobility data from Madrid. Madrid is the capital city of Spain with a population of  $N = 3,312,310$  inhabitants in 2020. The prime administrative divisions of the city are  $V = 21$  *distritos* (districts). The model does not resolve at the basic health zone (BHZ) level, which is the one that was actually considered for the PLs, due to the lack of more fine-grained mobility data. In table 3.1, we collect the population size at the district level  $N_i$  (note that a BHZ generally encompasses between 5,000 and 25,000 inhabitants). Regarding mobility data, we obtained interdistrict mobility flows provided by the Spanish Ministry of Transport as part of a survey aimed at analyzing changes in mobility patterns within Spain amid the COVID-19 crisis and at evaluating the effect of mobility restrictions [472]. This survey is based on cell phone data provided by cell phone carriers.

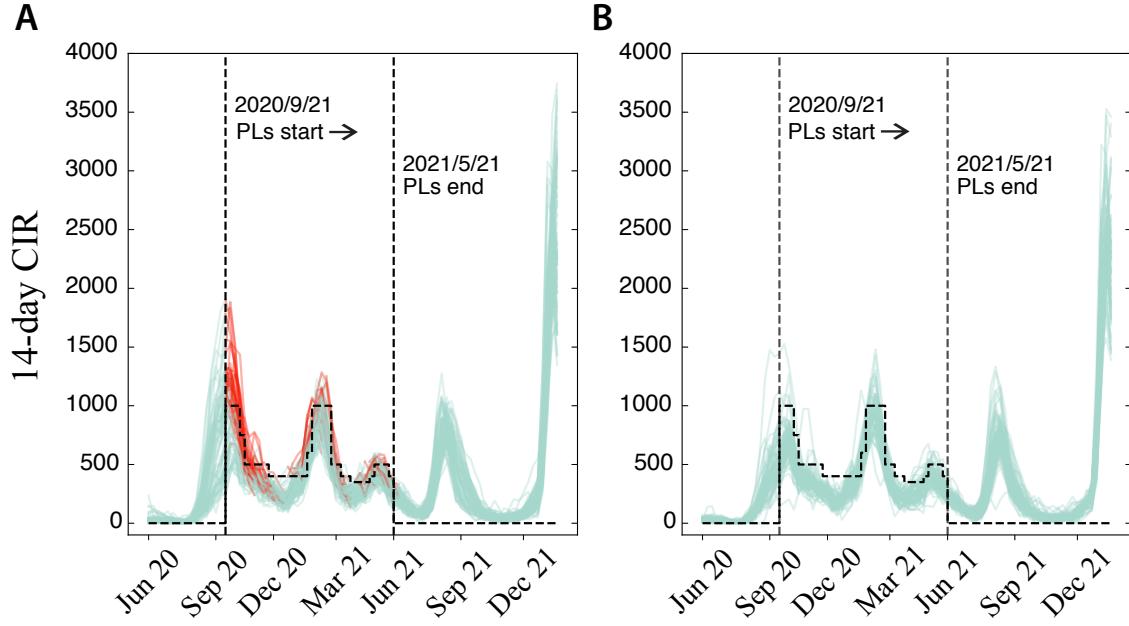
The data of interest spans a period ranging from March 2020 until May 2021. There is also a week of data in February 2020, from day 14 to day 21, used as a baseline reference for comparisons with the COVID-19 period. We refer to this period as the regular mobility scenario. We took only this reference period since we wanted to inform our urban model of Madrid with regular or unperturbed data. These data are used to build the origin-destination matrix  $\mathbf{M}$ . To do so, we average the mobility flows between  $i$  and  $j$  for the whole week so that  $\mathbf{M}$  is time-independent.

District	$N_i$	District	$N_i$	District	$N_i$
Centro	141,236	Fuencarral	247,692	Ciudad Lineal	216,818
Arganzuela	154,243	Moncloa	120,834	Hortaleza	193,228
Retiro	118,557	Latina	240,155	Villaverde	154,808
Salamanca	146,016	Carabanchel	258,633	Villa de Vallecas	114,733
Chamartín	145,700	Usera	142,454	Vicálvaro	75,485
Tetuán	159,849	Puente de Vallecas	239,057	San Blas	160,258
Chamberí	138,667	Moratalaz	93,810	Barajas	50,077

Table 3.1: Population data (2020) of the 21 districts in which Madrid is subdivided.

## 3.3 Results and discussion

First, we take a brief look at the real data from Madrid during the period in which the PL strategy was in place. Then, we show the results from the simulations of our metapopulation model with PLs applied to an idealized scenario inspired by Madrid.



**Figure 3.1: Madrid’s epidemiological trajectory.** Real 14-day cumulative incidence rate time series for the basic health zones (BHJs) in Madrid city. Panel A: Trajectories for BHJs that during some time period experienced a perimeter lockdown (PL). In red, the period in which they were under a PL. Panel B: Trajectories for BHJs that were not confined (as extracted from the official bulletins). Vertical dashed lines mark the beginning and the end of the perimeter lockdown strategy in Madrid and step-wise horizontal dashed lines signal the risk threshold considered by the authorities to activate the lockdowns.

## Analysis of Madrid’s epidemiological data

In general, perimeter lockdowns were enforced in areas in which the 14-day cumulative incidence rate (14d CIR) was above a certain threshold, although this information could be complemented by its trend and the, broadly defined, presence of community transmission in the area. In Figure 3.1, the dashed horizontal lines indicate the threshold set by the authorities which, as we can see, was not static. Rather, it mostly followed the epidemic waves. This strategy was in place from late September 2020 to May 2021 (vertical dashed lines in the Figure), a period that comprises the second and third waves of the epidemic in Spain (the first one is not shown in the Figures). Given the typical time scale of the spreading, it has been shown that the change in trend after the implementation of the strategy cannot be attributed to it [459, 460]. Note also that there were several BHJs with a 14d CIR above the threshold that were not confined (Figure 3.1 B). Lastly, we observe high synchronization in the spreading for all the BHJ time series, whether under lockdown or not. Thus, either the application of the strategy in a few areas had effects on the whole system, or it had limited consequences.

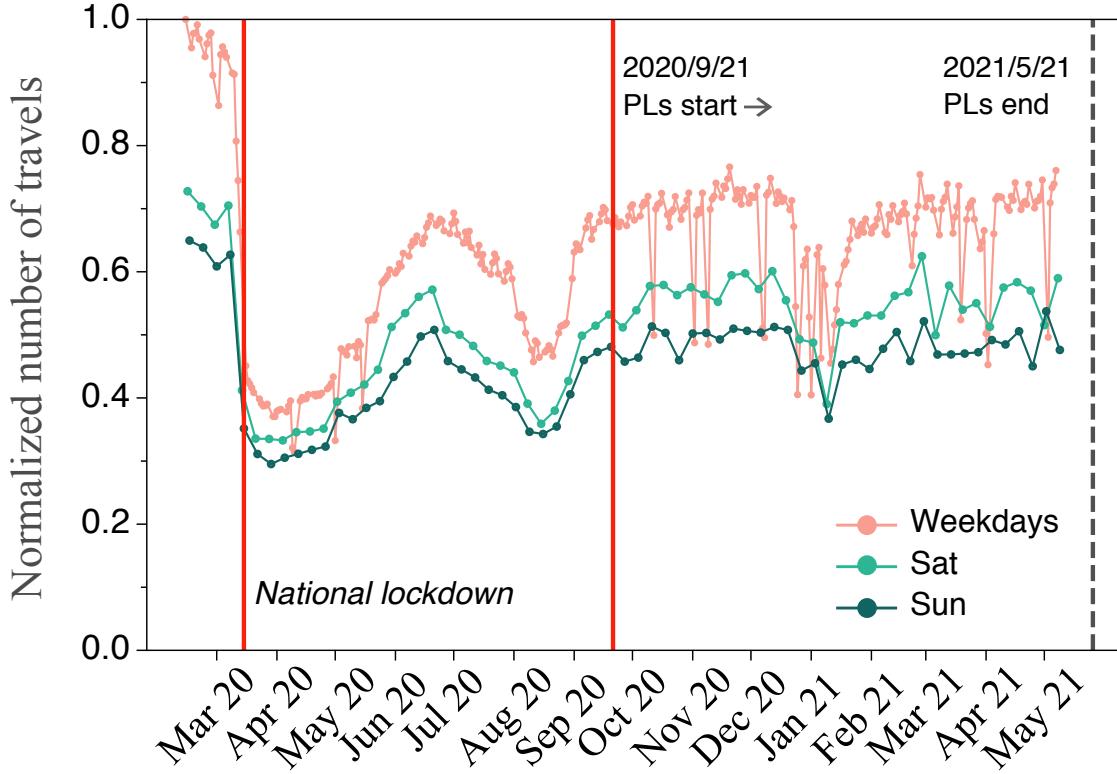
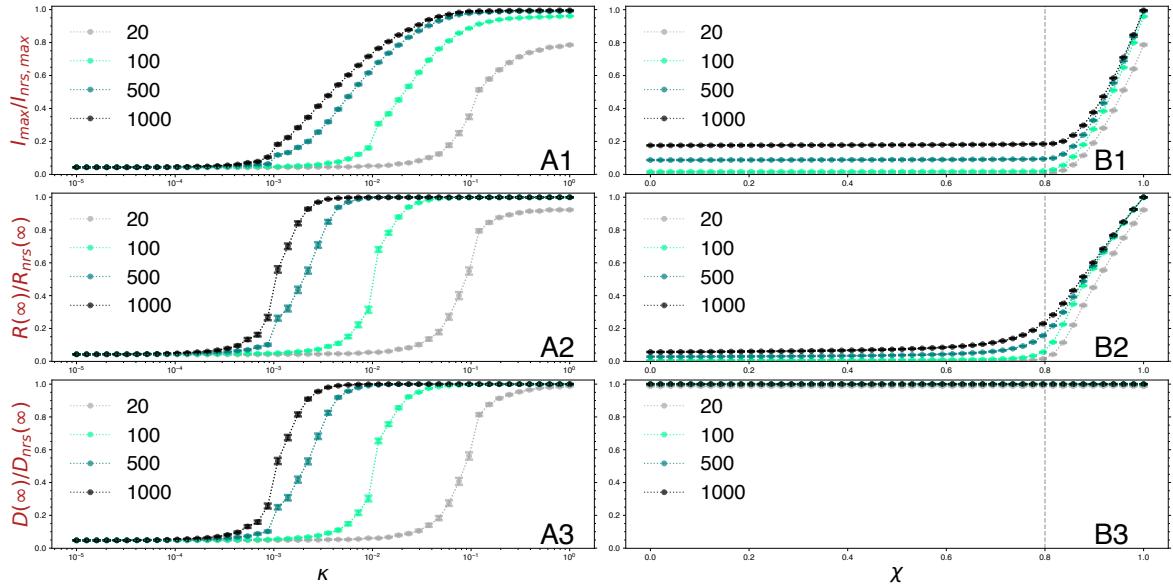


Figure 3.2: **Madrid’s mobility patterns.** Normalized total number of travels (with respect to the maximum value in the data) within the city of Madrid throughout the considered period. Weekend days are distinguished for the sake of clarity. There was a national lockdown from March 2020 to June 2020, corresponding to the first wave of the epidemic in Spain. The PL strategy was implemented in Madrid in late September 2020. Note that the Spanish Government stopped collecting mobility data two weeks before the end of the PL strategy in Madrid.

Next, we look at the overall mobility from the beginning of the pandemic to the end of the PL strategy, Figure 3.2. The imposition of the national lockdown (first vertical red line) reduced mobility to 40% of the usual values. Then, it slowly increased following the progressive lifting of the restrictions during the summer of 2020, reaching mobility values of 70% of the baseline. The value is reduced once again to 50% of the baseline mobility during August 2020, which is the period in which, typically, people go on vacation in Spain. By the time the PL strategy was put in place, mobility was close to 70% on weekdays. In this period, we observe that the mobility levels tend to stabilize, except for the occasional public holidays, around that 70% figure. Averaging over the full period of PLs, mobility levels remain at 0.63 [0.62, 0.64], and, excluding weekends, increase to 0.67 [0.66, 0.68]. Thus, the impact of the PLs on the overall mobility within the city was negligible.

## Effects of perimeter lockdowns

To study the effect of PLs, we define three epidemiological observables: global peak incidence,  $I_{\max}$ , global final prevalence,  $R(\infty)$ , and the final fraction of districts whose 14d CIR exceeded a certain risk threshold,  $D(\infty)/V$ . For brevity, we refer to these observables simply as the peak incidence, prevalence, and locked districts, respectively. Furthermore, in all cases the values are normalized over the ones obtained in the non-response scenario (subscripted as  $nrs$ ), corresponding to a fully unmitigated outbreak. As such, when the normalized observables are less than unity, it signals that the intervention had some positive effect on containing the epidemic. In the following, we explore how these observables change when we vary the three free parameters in our model: the general mobility parameter,  $\kappa$ , the risk threshold,  $\Theta$ , and the transmissibility reduction,  $\chi$ .



**Figure 3.3: Epidemic impact curves for varying control parameter values.** Panels A1, A2, and A3 show the peak incidence, prevalence, and locked districts fraction, respectively, for varying values of mobility parameter  $\kappa$  and different threshold  $\Theta$  values, without transmissibility reductions in the locked areas ( $\chi = 1$ ). Panels B1, B2, and B3 show the same observables for varying values of transmissibility reduction,  $\chi$ , and different  $\Theta$  values, with baseline mobility ( $\kappa = 1$ ). Quantity values are normalized with respect to a no-response scenario. The vertical dashed line in the B panels marks the point where, by reducing  $\chi$ , the global reproduction number turns  $R^* = 1$ , and thus the threshold under which the spread is under control [217].

Panels A1 (peak incidence), A2 (prevalence), and A3 (locked districts) of Figure 3.3 show a series of curves for several threshold risk values  $\Theta = 20, 100, 500$ , and  $1000$ , without transmissibility reduction in locked areas ( $\chi = 1$ ) and varying  $\kappa$  as control parameter. Every point in the curves depicts a situation where mobility is lower than

the baseline scenario ( $\kappa = 1$ ). Note that, since  $\chi = 1$ , the only effect of PLs is to cut down mobility in and out the affected districts. These curves then allow us to disentangle the effects of general mobility on the system. Clearly, all the observables show that reducing the general mobility parameter  $\kappa$  helps to reduce the impact of the epidemic with respect to the no-response scenario. The problem, though, lies in the magnitude of the decrease. Madrid's PLs pursued risk threshold values mainly between  $\Theta = 500$  and  $1000$  most of the time. Under these conditions, even with lower mobility than the one observed during the national lockdown ( $\kappa \sim 0.4$ , see Figure 3.2) the effects are mostly negligible. To achieve, for instance, a reduction of 20% in peak incidence, mobility should be one order of magnitude lower than during the national lockdown ( $\kappa \sim 0.01$ ). And this still would not impact the prevalence or the fraction of locked districts. Only when  $\kappa = 10^{-3}$  (99.9% reduction), these quantities are reduced by 50%. With a much more strict threshold for PL activation,  $\Theta = 20$ , and under the regular mobility scenario, we can obtain a reduction in peak incidence  $> 20\%$ , and around 10% in prevalence, but this does not keep the system from undergoing a *de facto* general lockdown unless mobility is also reduced by 90% or more.

In Panels B1 (peak incidence), B2 (prevalence), B2 (locked districts) of Figure 3.3, instead, we fix the general mobility to the baseline,  $\kappa = 1$ , and vary  $\chi$  in the districts under PLs. In this situation, the intervention restricts the in and out mobility in the selected districts and also reduces the transmissibility within the area by a factor  $\chi$ . Here we readily see that PLs, when accompanied with additional mitigation measures inside the affected area, can effectively reduce the impact of the disease on the system. Given the proposed  $R_0 = 1.25$  and  $T_I = 4.5$  days, reducing baseline transmissibility by 20% ( $\chi = 0.8$ ) translates into an effective reproduction number for a quarantined district that is on the verge of criticality, that is,  $R^* = 1$  [217]. Note, however, that the value of  $R_0$  is already very low as we assume that multiple non-pharmaceutical interventions are in place. Thus, achieving further reductions might not be possible. In any case, when  $\chi = 0.8$  and  $\Theta = 500$ , we obtain normalized values for peak incidence of 0.093 [0.093, 0.093] and prevalence of 0.166 [0.166, 0.167]. This is a reduction of more than 90% and 80% with respect to the no-response scenario, respectively. However, we can see that all districts in the urban system have outbreaks above the risk threshold, for any  $\Theta$  and  $\chi$  shown in the figure. Thus, under these conditions, perimeter lockdowns can greatly reduce the impact on the population but would not protect parts of the city from the most affected ones, even under mild epidemiological conditions.

Lastly, we explore in more detail the space of parameters  $(\kappa, \chi)$  for two selected  $\Theta$  scenarios. We consider a rather proactive strategy, with  $\Theta = 20$  (Panels A1, A2, A3 in Figure 3.4), and a more reactive strategy, with  $\Theta = 500$  (Panels B1, B2, B3 in

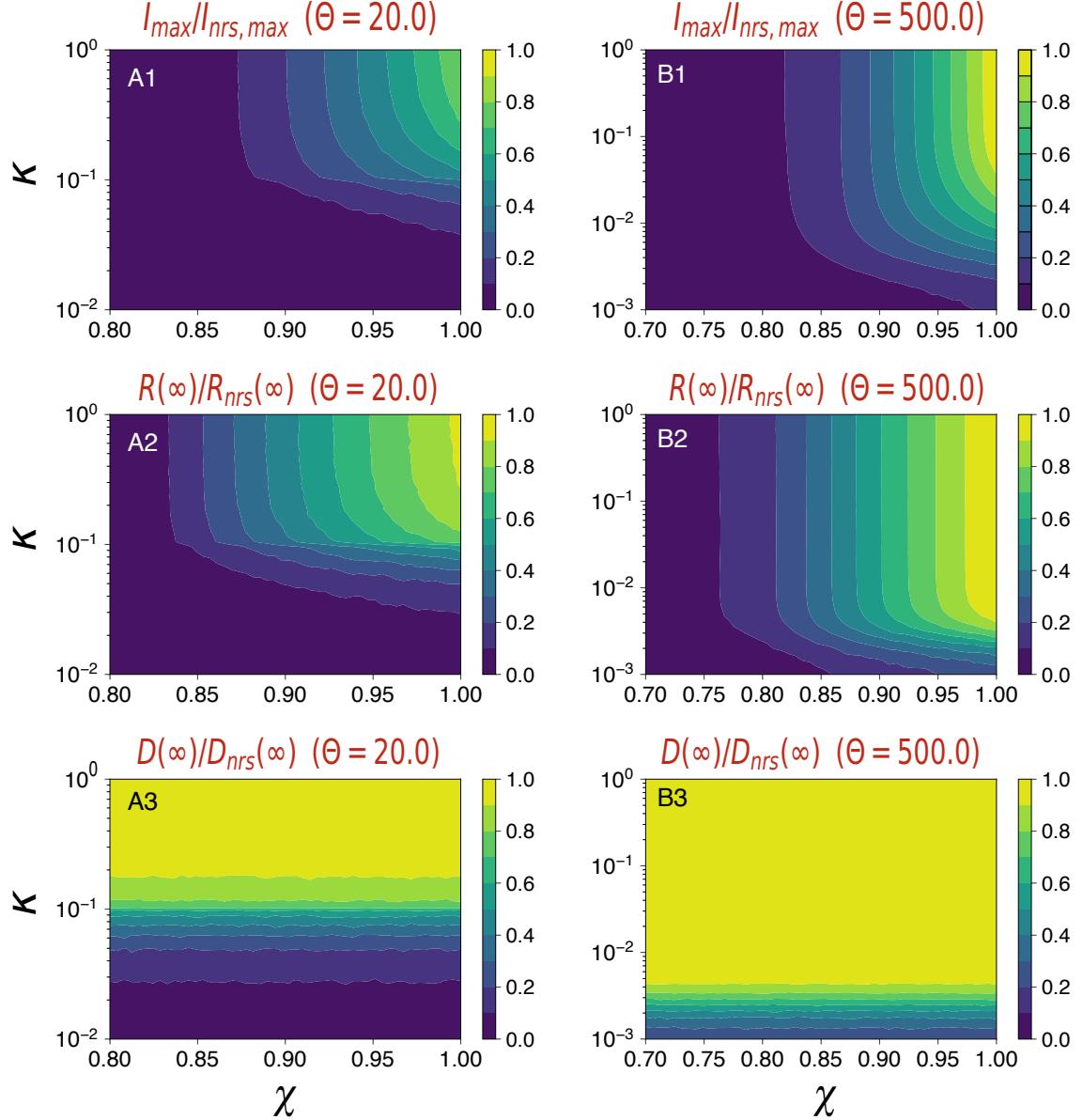


Figure 3.4: **Epidemic impact when varying simultaneously  $\kappa$  and  $\chi$  for selected  $\Theta$  scenarios.** Panels A1, A2, and A3 show the peak incidence, prevalence, and locked districts fraction, respectively, in  $(\chi, \kappa)$ -space for the threshold  $\Theta = 20$ . Panels B1, B2, and B3 show the same observables for  $\Theta = 500$ . All values are normalized with respect to a no-response scenario.

Figure 3.4). With this analysis, we can better appreciate the limited role of general mobility  $\kappa$  in both the reduction of the peak incidence and the final global prevalence when  $\kappa$  decreases from unity down to sensible but really strict values (around  $\kappa = 0.1$ ). Now, moving from  $\chi = 1$  to  $\chi = 0.8$  we can obtain an important reduction of the impact, both for  $\Theta = 20$  and also  $\Theta = 500$ . See tables 3.2 and 3.3 for an outline of the results given a selection of  $(\kappa, \chi)$  scenarios. Finally, with respect to locked districts, with  $\Theta = 20$  it is possible to avoid the invasion of the entire system, but it requires

a strong reduction of mobility, going well below  $\kappa = 0.5$  (see table 3.2). Note the difficulty of doing this: PLs have to be activated very soon and, at the same time, it is essential to achieve a strong reduction of global mobility in the system to effectively save a few districts from experiencing an outbreak. With  $\Theta = 500$ , the full invasion is unavoidable, except for extremely low values of mobility.

$\Theta = 20$				
	$\chi = 0.8$	$\chi = 0.9$	$\chi = 1$	
$\kappa = 1$	0.01 [0.01, 0.01]	0.43 [0.42, 0.43]	1.69 [1.68, 1.70]	$I_{\max} \%$
$\kappa = 0.5$	0.01 [0.01, 0.01]	0.42 [0.42, 0.43]	1.63 [1.62, 1.64]	
$\kappa = 0.1$	0.01 [0.01, 0.01]	0.31 [0.30, 0.32]	0.97 [0.94, 1.00]	
$\kappa = 1$	0.67 [0.65, 0.68]	17.16 [17.05, 17.28]	34.24 [34.12, 34.37]	$R(\infty) \%$
$\kappa = 0.5$	0.65 [0.63, 0.66]	17.13 [17.02, 17.24]	34.07 [33.92, 34.22]	
$\kappa = 0.1$	0.52 [0.51, 0.54]	13.80 [13.40, 14.20]	27.04 [26.35, 27.74]	
$\kappa = 1$	98.72 [98.57, 98.87]	98.72 [98.57, 98.88]	98.80 [98.63, 98.96]	$D(\infty) \%$
$\kappa = 0.5$	97.33 [97.08, 97.57]	97.44 [97.22, 97.67]	97.19 [96.89, 97.49]	
$\kappa = 0.1$	71.20 [69.92, 72.47]	75.71 [73.82, 77.61]	74.44 [72.59, 76.29]	

Table 3.2: **Epidemic impact with the threshold  $\Theta = 20$ .** Average values as a percentage of the total population  $N$  for  $I_{\max}$  and  $R(\infty)$ , and as a percentage of system size  $V$  for locked districts  $D(\infty)$ , together with a 95% confidence interval in brackets for selected scenarios of  $(\kappa, \chi)$  values under  $\Theta = 20$ . NRS gives  $I_{\max} = 2.15$  [2.15, 2.15],  $R(\infty) = 37.14$  [37.12, 37.16] and full invasion.

$\Theta = 500$				
	$\chi = 0.8$	$\chi = 0.9$	$\chi = 1$	
$\kappa = 1$	0.20 [0.20, 0.20]	0.73 [0.73, 0.73]	2.13 [2.13, 2.14]	$I_{\max} \%$
$\kappa = 0.5$	0.20 [0.20, 0.20]	0.73 [0.73, 0.73]	2.13 [2.13, 2.14]	
$\kappa = 0.1$	0.20 [0.20, 0.20]	0.73 [0.73, 0.73]	2.11 [2.11, 2.11]	
$\kappa = 1$	6.19 [6.17, 6.20]	21.98 [21.97, 22.00]	37.13 [37.11, 37.15]	$R(\infty) \%$
$\kappa = 0.5$	6.19 [6.17, 6.20]	21.98 [21.96, 21.99]	37.13 [37.12, 37.14]	
$\kappa = 0.1$	6.17 [6.16, 6.19]	21.99 [21.97, 22.00]	37.12 [37.11, 37.14]	
$\kappa = 1$	100 [100, 100]	100 [100, 100]	100 [100, 100]	$D(\infty) \%$
$\kappa = 0.5$	100 [100, 100]	100 [100, 100]	100 [100, 100]	
$\kappa = 0.1$	100 [100, 100]	100 [100, 100]	100 [100, 100]	

Table 3.3: **Epidemic impact with the threshold  $\Theta = 500$ .** Average values as a percentage of the total population  $N$  for  $I_{\max}$  and  $R(\infty)$ , and as a percentage of the size of the system  $V$  for the locked districts  $D(\infty)$ , together with 95% confidence interval in brackets for selected scenarios of values of  $(\kappa, \chi)$  under  $\Theta = 500$ . NRS gives  $I_{\max} = 2.15$  [2.15, 2.15],  $R(\infty) = 37.14$  [37.12, 37.16] and full invasion.

Examples of perimeter lockdowns of highly dense-integrated suburban areas are scarce. The inspection of the real epidemiological and mobility data from Madrid casts a lot of doubts about its proper implementation and its effectiveness. As we referenced

previously, this is something that the few works devoted to the experience of Madrid have formally confirmed [459, 460]. In [459], through a joint point trend analysis, the authors found that the decrease in the epidemic curve, both in the entire city and the BHZ affected, started before the impact of the perimeter lockdown could have been reflected, and also found that the strategy did not increase the speed at which cases were decreasing. The authors offer several reasons for the strategy’s failure: 1) the curve started to decrease before measures were taken. 2) Perimeter lockdowns were focused on mobility rather than preventing high-risk situations. 3) Even though the focus lied on mobility, it was allowed for essential activities such as working. And 4), BHZ boundaries were unknown to many residents since these are used for public healthcare administration. Regarding our particular experiment, points 2) and 3) apply and align well with our conclusions: mobility is not very much effective in this kind of setting, and the focus should be on preventing high-risk situations like workplace or community mixing. In [460], through the use of a generalized additive models technique, they reach a similar conclusion: the perimeter confinements did not have a significant impact on the 14-day cumulative incidence.

To further understand the mechanisms underpinning the spreading dynamics in this strategy, we conceived a minimal model using a metapopulation framework parameterized using real mobility data. Our model results show how difficult it is to control an outbreak in an urban environment under the PLs. The reduction is critically marked by how successful the transmissibility reduction is within a district, with mobility playing a minor role. Note that Madrid initially set its lockdown risk threshold at  $\Theta = 1000$  and then changed it dynamically up and down, being  $\Theta = 350$  the lowest documented. In our model, even when the threshold was  $\Theta = 20$ , a very high effort on reducing general mobility was needed to avoid invasion of the full system. This mobility reduction was not even achieved during the national lockdown that closed most businesses and industries (Figure 3.2). Furthermore, during a PL in Madrid, it was still allowed to go to your workplace even if it was in another district.

Localized lockdowns were also implemented in Santiago de Chile during its first COVID-19 wave. In [461], Li et al. explored their performance, obtaining mixed results. The authors state that localized lockdowns on their own are insufficient to control pandemic growth in the presence of indirect effects from contiguous neighboring areas that do not have lockdowns. As we did, they found that localized lockdowns can help contain the transmission of the virus but their effectiveness is dependent on its duration and indirect effects from neighboring areas with high social interconnectivity. Moreover, their estimates showed that in Greater Santiago the epidemic is only controlled when generalized lockdowns are in place. In contrast, these measures showed

promising results in municipalities that were rather isolated from affected neighboring areas. They concluded that the growth of disease transmission is reversed only when lockdowns are implemented in a coordinated fashion across interconnected geographic areas.

Urban systems for the most part are typically very well interconnected systems where it is easy to move from one part to another in time scales significantly lower than 1 day. Large portions of the population in these systems mix at different sites during the day: whether at workplaces, at schools, at the public transportation system, or during leisure activities. Therefore, a high level of synchronization in the spreading can be expected. Hence, when an outbreak emerges in one area, it will quickly spread through the system unless mobility is extremely reduced. If perimeter lockdowns are accompanied by significant transmissibility reductions, the measures can make a difference to flatten the curves. But unless the action is performed soon enough, the most probable course of action for the spreading is to quickly invade the full system. If the goal of the PLs is to protect some parts of the system so that citizens inside can normally live and perform their daily activities, this seems unachievable under reasonable assumptions and regular behavior. Our results from a sound theoretical model which idealized several critical features and thus, overall, offered a best-case scenario, seem to confirm this thesis. Moreover, the real epidemiological data for Madrid clearly shows the synchronization phenomena among different areas. Responses that aim to partially isolate areas of a system that are so spatially interconnected and temporally synchronized are likely to be insufficient. As results have shown, the emphasis should not be on mobility (either general or among particular areas) but on general measures that bring transmission down where the mixing or contact is effectively produced.

## 3.4 Conclusions

Our model explains the qualitative features of the real epidemic spreading observed in Madrid under the PLs and sets a general model to approach this type of situation. Nonetheless, it has its limitations. As we have already stated, our model is minimal and is intended to capture only the most basic features of the spreading together with the control strategy, but it is not intended to reproduce the real COVID-19 trajectories that took place during the period studied. COVID-19 is a complex disease with a long latency period, an important amount of pre-symptomatic infectivity and even asymptomatic transmission. Moreover, the disease can present high hospitalization and death rates which are highly age-dependent. These are elements worth taking

into account when aiming for accurate COVID-19 forecasting. However, regarding the assessment of non-pharmaceutical interventions as the one presented here, the SIR model stands as a simple but best-case scenario modeling choice. The rationale is that even though we neglect key elements that would hinder surveillance and disease control, such as a latency period and asymptomatic disease progression, the outcome is still unfavorable for the success of the strategy. Another limitation is that, at the subpopulation level, a homogeneous well-mixing scheme is considered. This is a standard assumption within the metapopulation framework but indeed relevant heterogeneous features may exist both intra-district and inter-district, such as age differential mixing. We also remark that our implementation of PLs assumed perfect information and absolute compliance. Moreover, travel bans were totally cut down, which is at odds with the permeability allowed in the real-life setting. Thus, even though our model is limited in many aspects, it can be, in general, regarded as a best-case scenario. The fact that the PLs did not achieve high effectiveness under these circumstances signals that in a real setting the situation would be worse.

Even though our minimal model indicates that the effectiveness of perimeter lockdowns at the urban scale is low, more detailed modeling could unveil subtleties at lower scales. Future works could aim for increased heterogeneity in human behavior. One could use micro-mobility models based on points of interests at the subpopulation level to break the homogeneous mixing assumption. In terms of the population, the addition of the socioeconomic characteristics of the population may help to disentangle the differential impact of non-pharmaceutical strategies on the diverse population strata. Indeed, even though at the aggregated level perimeter lockdowns appear to be inefficient, their impact on the most vulnerable parts of the population is yet unknown and is open for further research. From a theoretical point of view, an interesting research line would be to determine the amount of spatial cohesion at which these systems tend to be highly synchronized and for which mobility plays only a minor role beyond the initial stages. This could facilitate the analysis of strategies similar to the one presented here at multiple scales, such as metropolitan areas or urban-rural systems.

# Chapter 4

## Epidemic spreading in an urban environment under the d-EPR model of microscopic human mobility

*When the streetlights fade. Warm rain like judgement descends.  
Their voice numbs me. Speaking words in a dead tongue.  
I have walked a road that lead me back to you.  
From a window our glances met. My true colours I cannot hide.  
The landscape has changed. You don't recognise me.  
These pictures slowly fade. Memories wither, they are all gone.  
Further down the steps get steeper. You haunt me in my dreams.  
I let go and fall deeper. This will be the end of me.*  
— Cult of Luna - Dark City, Dead Man

### 4.1 Introduction

Metapopulation models are a well-established framework to analyze the spatial propagation patterns of infectious diseases across spatially extended areas [130, 131, 473, 206, 216, 239, 228, 235, 474, 225]. Differently from single-population models, metapopulations enable the inclusion of human mobility and thus account for a more realistic representation of the contagion dynamics. This allows us to understand the interplay between human mobility and the epidemic processes, and conceive potential interventions related to mobility control, in order to mitigate the impact of an outbreak on the population. Metapopulations are represented by networks, where the nodes represent the geographic units of interest and the links represent the connection between these units through human mobility flows [217]. This framework has been used most

extensively to analyze disease propagation at country or regional levels [468, 466, 467], or globally [45, 211, 16], but they are versatile enough to be applied to finer spatial scales, such as in urban settlements [339].

Independently of this, there are some pervading assumptions in this framework that are rarely trespassed. In particular, here we are concerned with those affecting the way in which mobility is modeled in metapopulations. As it happens with standard compartmental models, the description of the system stays at the population level and does not resolve the microscopic or individual scale. In other words, metapopulations deal with indistinguishable agents. There are some exceptions to this [228, 224], but in those models, heterogeneity traits are considered in relation to the mixing and contagion process, and not with mobility. Moreover, another standard assumption is the use of Markovian random walk-like models, where the displacement to the next location just depends on the immediate previous spatial state. Again, exceptions exist where, for instance, recurrent mobility is introduced to account for commuting, which is an integral part of everyday human mobility [280, 219, 220, 221, 225, 224]. Still, these models make at most a distinction by grouping individuals related to their origin location and their destination but far from acknowledging a more fine-grained heterogeneity. While, on the one hand, this set of assumptions is understandable due to analytical tractability and avoidance of increasingly high computational costs, on the other hand, this hampers the understanding of a more detailed and realistic picture of the system and subsequent tailored interventions aimed at mitigating the epidemic impact and the associated socioeconomic impact of coarse-grained control policies.

Moreover, these mobility models are in stark contrast with the discoveries and understanding of human mobility patterns in recent years [438, 51, 428, 429] thanks to the ever-increasing availability of high-resolution mobility data from sources such as GPS or mobile-phone carriers. It has been found that human trajectories show a high degree of temporal and spatial regularity, each individual being characterized by a time-independent characteristic travel distance and a significant probability to return to a few highly frequented locations, at odds with the trajectories that yield Lévy flights and random walk models [438]. In [428], it is shown how predictions of continuous-time random walks are in systematic conflict with empirical results, and two principles are proposed that govern human trajectories, that properly account for the empirically observed laws. These principles are (i) exploration of new locations and (ii) preferential return to already visited locations, and were crystalized in the seminal EPR model [428], that spurred further research in the topic [431, 429, 475, 433]. Independently, Pappalardo et al. [429], found through analysis of several real mobility datasets, that indeed two distinct mobility profiles emerge: explorers and returners.

They went on to propose a model, the d-EPR model [430], that captures these findings more adequately.

The literature on human mobility naturally motivates their results on the basis of potential impacts of phenomena driven by human mobility, including our topic of interest here, the spread of infectious diseases, but we miss a thorough exploration of the topic, with a few recent exceptions like [476] or [477]. Thus, in this work, we aim to contribute to filling this gap and offer a more detailed account of the modeling of an infectious disease spreading subjected to a more realistic microscopic mobility model. In particular, we run the d-EPR model of human mobility in a realistic setting by informing the attractiveness of locations with anonymized high-resolution real mobility trajectories in the Greater Boston Area, MA, USA. We want to assess how the epidemic spreading is affected by this type of exploration and preferential return mobility and, in turn, how are the different mobility groups affected by the epidemic. We run stochastic simulations of a SIR process under the d-EPR model with a homogeneous and heterogeneous distribution of the parameter  $\rho$  that characterizes the exploration probability. We also compare the d-EPR with simplified baseline scenarios that remove some relevant features of the original model. Mainly, we find that only the d-EPR model with a heterogeneous distribution is able to show noticeable effects across mobility groups of explorers and returners. In general, explorers, contribute to the spread of the disease across the whole system, suffer also a higher disease prevalence, and contribute more to the generation of new infected cases than returners. While the impact on the invasion dynamics is large, the differential effects across  $\rho$  groups for the disease prevalence and generation of new cases are somewhat small. The fact that even for low  $\rho$  values, an important fraction of time is spent in high attractiveness locations, which act as contagion centers, greatly contributes to the spread of the disease and blurs some potential effects of the heterogeneous mobility profiles.

## 4.2 Material and methods

### d-EPR model of human mobility

The d-EPR model is a microscopic or agent-based mobility model of human mobility. Originally, the model conceives heterogeneity both spatially, through the agents' motion, and temporally, through the distribution of sojourn times at every location. Since we are adopting a discrete-time approach for the simulations, we take a unitary time step  $\Delta t = 1$ . Then, at every time step, the  $N$  agents in the system face two options, either explore a new location with probability  $P_{\text{exp}} = \rho S^{-\gamma}$  or return to an already visited location with the complementary  $P_{\text{ret}} = 1 - P_{\text{exp}}$ . Here,  $S$  is the

number of different visits to locations performed by the agent up until time  $t$ , and  $\rho$  and  $\gamma$  are scaling parameters related to the exploration law whose values are  $\rho = 0.6$  and  $\gamma = 0.21$  as obtained from the analysis of data sets in [428]. The parameter  $\rho$  is related to the ratio of the number of different visits and the total number of visits performed by an individual. This parameter is typically considered constant for every individual in the model, but we consider that this imposition washes away the genuine heterogeneity related to them. Instead, agents and their associated mobility profile will be characterized by  $\rho$ , which will be sampled from several distributions, to explore the role of the agents' mobility heterogeneity. In particular, we show results for a Gaussian distribution with parameters  $\rho = 1/2$  and  $\sigma = 0.1$ , and for a beta distribution with parameters  $a = 2$  and  $b = 2$ . This choice of parameters makes both distributions symmetrical and with  $\langle \rho \rangle = \sum_a^N \rho_a = 1/2$ . Thus, there is the same quantity of agents with  $\rho < 1/2$  and  $\rho > 1/2$ . Along the text, we will refer to this last setting as the heterogeneous or beta setting, and to the former as the homogeneous or Gaussian one.

Now, in the case of exploration, the selection of the new location follows the gravity law, so that the probability of going to a location  $j$  from location  $i$  is given by:

$$p_{ij} = C \frac{A_i A_j}{r_{ij}^2}, \quad (4.1)$$

where  $A_j$  is the attractiveness of the destination location  $j$ , and  $r_{ij}$  is the Euclidean distance between origin  $i$  and destination  $j$ .  $C$  is a normalization constant given by  $C^{-1} = \sum_{k \neq i} \frac{A_i A_k}{r_{ik}^2}$  ensuring that  $p_{ij}$  behaves like a probability. The attractiveness value  $A$  of a location refers to the relevance of it within the system. It can be represented by the population in the location or by some other socioeconomic proxy. In an upcoming section, we give the details of the attractiveness field used in our simulations.

In the case of a return, an already visited location  $j$  is selected following a preferential attachment scheme based on the proportion of visits or visit frequency, related to that location. Then the probability of going to that location  $j$  is given by:

$$p_j = \frac{f_j}{\sum_k f_k}, \quad (4.2)$$

where  $f_j$  is the visitation frequency to that location  $j$  up until the current time steps  $t$ , and it is simply the ratio of the total number of visits performed to that place divided by the total number of visits performed until then.

For the synthetic trajectories generated under this model, we assume that the initial location of every agent  $a$ , that is,  $x_a(t = 0)$ , is their actual home location. While in real life, regular people have their home as a highly preferential location to return to on a daily or weekly basis, the d-EPR does not confer any special role on homes. As a consequence, the initial condition may end up commonly unnoticeable as

a special location for an agent's trajectory. Without affecting the general conclusions of this work, we confer every agent's initial position a very high visitation frequency (corresponding to an initial number of visits  $S_{home}(0) = 25$ ), so that whenever an agent chooses to return, the probability of returning to their ad-hoc home will be significantly higher than to any other previously visited location.

For the sake of comparing the results obtained under the d-EPR mobility model, we propose some baseline scenarios to be taken as reference. The first baseline scenario, which will be referred to as *plain* baseline, consists of: i) making all the places equivalent in terms of their attractiveness value, thus operating only the inverse square distance law when choosing a new location to explore, ii) removing the non-Markovian part of the d-EPR dynamics, and iii) making all the agents indistinguishable by setting  $\rho = 1/2$  for all. Removing the non-Markovian feature implies that  $P_{exp} = \rho$ , but also that the set of locations to return is reduced to the current location where the agent is before performing their next move, and lastly, already explored locations can be explored again in case of an exploration step. The second baseline we propose is removing the aforementioned non-Markovian features found in the d-EPR model, both when exploring and returning, but retaining the original attractiveness distribution of the Greater Boston Area; this second baseline will be referred to as *memory-less*. Eventually, the third baseline scenario will assume the regular d-EPR dynamics but with a uniform attractiveness distribution for locations, this scenario will be referred to as *uniform* baseline.

## Epidemic model

We work with the compartmental SIR model, which is the most basic model to analyze influenza-like illnesses. Thus individuals can be either susceptible, infected, or removed/recovered. Since we do not aim to analyze the spreading of a particular disease, the simplicity of the SIR model allows us to gain general insights regarding the spreading process, facilitates implementation, and allows us to focus on the mobility aspect.

As in a regular metapopulation model, contagion occurs at locations. Here we assume the standard homogeneous mixing approach, and thus the probability for an individual of getting the infection at a location  $\ell$  and a certain time step  $t$ , can be written as:

$$P(\mathcal{S} \rightarrow \mathcal{I}) = 1 - \left(1 - \frac{\beta \Delta t}{N_\ell(t)}\right)^{I_\ell(t)}. \quad (4.3)$$

Here,  $\beta$  is the disease's transmission rate,  $I_\ell(t)$  is the total number of infected individuals, and  $N_\ell(t)$  is the total population at location  $\ell$  and time  $t$ , respectively.

As for the individual probability of recovery, this is simply given as

$$P(\mathcal{I} \rightarrow \mathcal{R}) = \mu \Delta t, \quad (4.4)$$

where  $\mu$  is the recovery or removal rate, which is just the inverse of the disease's infectious period  $T_I = 1/\mu$ . Under the well-mixing assumption, we can relate the basic reproduction number of the disease to the infection and removal rates as  $R_0 = \beta/\mu$ . We work here with  $R_0 = 1.2$  and  $\mu = 0.1$ .

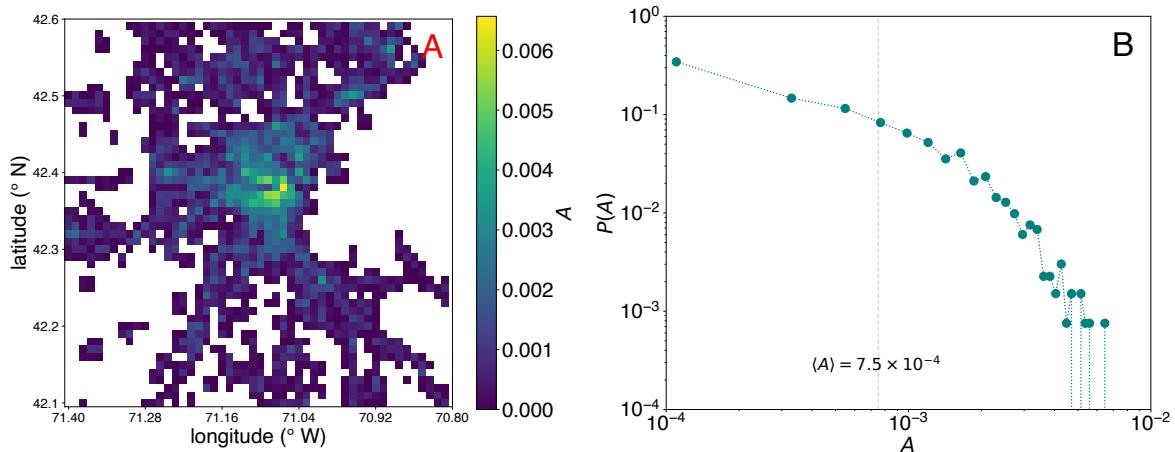
In terms of computational implementation, we can extract the newly infected cases and the newly removed cases from binomial samplings based, respectively on the individual probabilities for each process  $P(\mathcal{S} \rightarrow \mathcal{I})$  and  $P(\mathcal{I} \rightarrow \mathcal{R})$ . Moreover, later, in the analysis of results, we will look at the contribution of agents to the generation of newly infected cases in contagion events. In individual-based models of contagion, like epidemics on networks or epidemiological ABMs, the infection process is a pair-wise interaction where with a certain probability or rate infected agents may infect their neighbors or the people around them in a certain environment. Whereas our model is individual-based regarding the mobility model, at the level of location, we are modeling infection as a homogeneous mixing process. Therefore, infections are not produced through pairwise interactions in a network, but through the mean-field model described above, where all the infected individuals at a location effectively contribute to the force of infection and thus to the generation of new infected cases. Then, to attribute infections to a certain mobility group, we will randomly sample an infected individual from the infected population where the contagion event happened, which will be referred to as the infector. Statistically, over the course of the dynamics and simulations, we can expect that this sampling will faithfully reflect the representative infected agent.

## Mobility data and attractiveness field

The gravity law of mobility implies that every location  $\ell$  in the system has an attractiveness value associated,  $A_\ell$ . Here, we use the number of stays of individuals at every location to assign the corresponding attractiveness value to each of them. These quantities come from anonymized high-resolution trajectories obtained from mobile phone data.

We took this data from the Greater Boston Area and processed it in the following way. First, we build a square lattice representing the spatial substrate where the mobility and contagion dynamics take place. In spatial coordinates, the lattice ranges in longitude from  $70.8^\circ$  W to  $71.4^\circ$ W, and in latitude from  $42.1^\circ$ N to  $42.8^\circ$ N. This translates into a total area of roughly  $50 \times 50 \text{ km}^2$ , being each cell in the lattice given

by an area of  $1\text{km}^2$ , we refer to each of these cells in the lattice as locations. In order to compute the distance between locations needed for the gravity law, we just take the geometric centers of the locations. As for the attractiveness, we collect real high-resolution trajectories of individuals and relate their positions in time with the grid built. Then, all the individual positions that fell into a particular cell in the lattice, are added to account for the attractiveness of the corresponding location. Whereas in a lattice of the given dimensions the total number of locations is  $V = 2500$ , here, the effective number of locations for the dynamics is  $V_{\text{eff}} = 1325$ . This is a result of the specific stays of the mobility data used which, among other factors, is logically constrained by the geographical accidents of the area.



**Figure 4.1: Greater Boston Area attractiveness field and distribution.** In this Figure, Panel A depicts the attractiveness field, that is, its spatial distribution, as obtained by the curation of the high-resolution trajectories of individuals. Panel B shows the log-log plot of the distribution of locations' attractiveness values, remarking the average value  $\langle A \rangle = 7.5 \times 10^{-4}$ .

In Figure 4.1 we depict the attractiveness spatial field of the Greater Boston Area together with the attractiveness distribution (in logarithmic scale). Without affecting any result, attractiveness values have been rescaled so that  $A_\ell \leftarrow A_\ell / \sum_\ell^V A_\ell$ . The average value of attractiveness is  $\langle A \rangle = 7.5 \times 10^{-4}$ , whereas the maximum value is  $A_{\max} = 6.57 \times 10^{-3}$ , one order of magnitude higher. Clearly, the Boston conurbation shows a mono-centric structure around Boston downtown, with the attractiveness distribution behaving in a power-law manner, where the majority of locations show a very small attractiveness value and only a few have remarkably high values.

### 4.3 Results and discussion

We perform extensive stochastic simulations of the spatial propagation of an influenza-like illness epidemic under the d-EPR mobility in the Greater Boston Area

and analyze relevant features of the epidemiological impact on the population. We will explore the invasion process, the disease prevalence, and the contagion events by the agents' mobility profile characterized by the  $\rho$  parameter.

## Disease invasion

Let us focus first on the invasion process. We define an invasion as the first occurrence of a secondary case in a location other than the epicenter, i.e. the initial seed of the spreading. Thus, by construction, every location can be invaded once and only once, if at all, during the whole dynamics. All the simulations have the most attractive location in the system as the place where the first infected individuals are introduced, the spreading *epicenter*. Figure 4.2 A1 and B1 show, for the Gaussian and the beta settings, respectively, the average values over the ensemble of simulations of the invasion share per  $\rho$  profile,  $\langle N_{\text{inv},\rho}/V_{\text{inv}} \rangle$ ; that is, the average number of invaders in group  $\rho$  divided by the total number of invaded locations. The question here is, whether explorer profiles ( $\rho > 1/2$ ) tend to proportionally lead the invasion since they have a higher exploration probability, and we see that this is clearly the case for the beta setting, but not for the Gaussian one. The reference to acknowledge this fact is the dashed line, which is equal to  $\langle N_\rho \rangle/N$ . This quantity can be taken as the null case where the invasion would occur proportionally to the population of agents in every  $\rho$  group. In other words, as  $\langle N_{\text{inv},\rho}/V_{\text{inv}} \rangle \rightarrow \langle N_\rho \rangle/N$  per every  $\rho$  group, it means that  $\rho$  has no special role in the invasion process, being thus a purely random process. On the contrary, the separation of these values shows that an effect exists. In the beta setting, both for the regular d-EPR and the uniform baseline, we see that the invasion profile is shifted to the right, meaning that returners tend to underperform with respect to the null hypothesis. It must be noted that top explorers (those with  $\rho \rightarrow 1$ ) are not as much overrepresented as middle-high explorers with  $\rho \in (0.5, 0.8)$ . The reason for this is that locations are finite and can be invaded only once, and naturally from the beta distribution of agents, those in the middle-high range of  $\rho$  are more populous than those in the extreme. Moreover, these lower  $\rho$  explorers have in the beginning a high enough exploration probability to easily bring the disease to new locations.

Importantly, when removing recurrence, which is what characterizes both the memory-less and the plain baselines, the invasion occurs randomly across  $\rho$  groups. On the contrary, making the space homogeneous with respect to attractiveness, does not change the outcome. It is important to note that even though all the locations have the same attractiveness, the inverse square distance still applies in the exploration stage for location selection.

Apart from explorers taking the leading role in delivering the disease across the

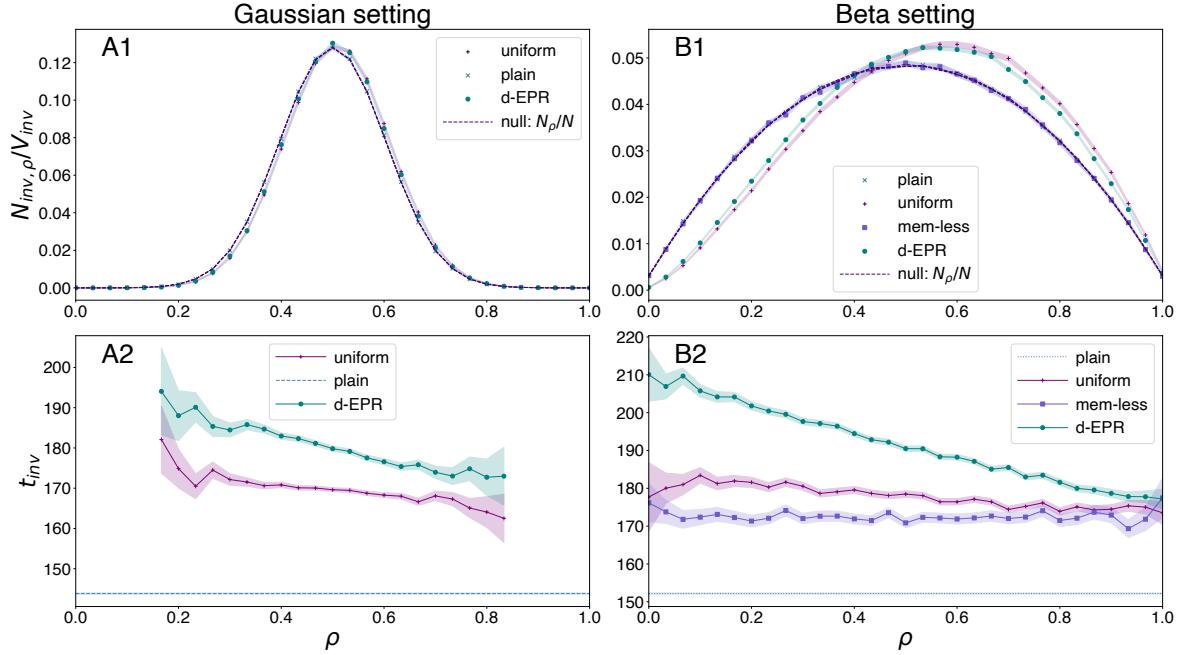
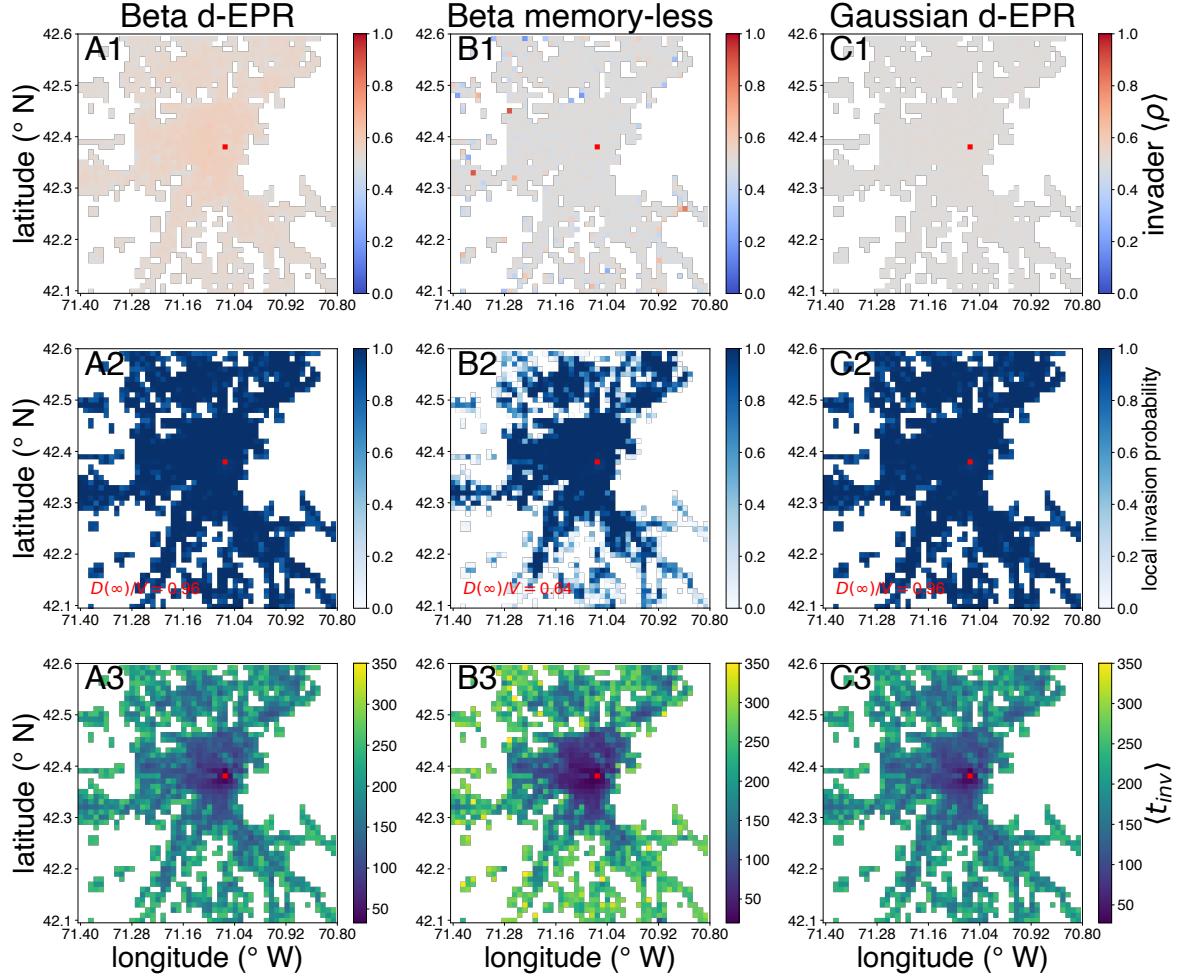


Figure 4.2: **Invasion events.** The top panels show the share of invasions per mobility group  $\rho$  for the Gaussian (Panel A1) and the beta distribution (Panel B1) settings. We compare the d-EPR model against the memory-less, the uniform, and the plain baselines, and also against a null reference where the invasion would occur randomly independently of  $\rho$ . Panels A2 and B2 depict the invasion times per mobility group, respectively for the Gaussian and the beta  $\rho$  distributions. Comparisons of the d-EPR results are also done against the previous baselines. Dots and dashed lines represent average values, while shaded areas cover the 95% confidence interval.

system, they also invade faster than returners on average in the d-EPR model for both the homogeneous (Figure 4.2 A2) and heterogeneous settings (Figure 4.2 B2). We also appreciate that the invasion occurs slower under the d-EPR mobility than with respect to the rest of the baseline scenarios. Clearly, the faster average invasion times occur for the plain scenario. Then, by a margin, the memory-less and the uniform baselines show similar global invasion times. The lack of the non-Markovian features in the memory-less baselines facilitates the discovery of new places and thus it can be understood that the invasion comes faster. Also, the fact that in the uniform baseline, invasion times are also faster than in the d-EPR mobility could be understood by the real broad attractiveness distribution. The mono-centric nature of the Greater Boston Area generally makes the furthest locations less attractive with respect to the center. This hampers the arrival of the disease, slowing the spreading dynamics. In the uniform baseline, even though distance still matters, the attractiveness *handicap* disappears, making more plausible a faster disease invasion across the whole system.

To conclude this analysis of the invasion process, we can take a look at a series of maps of the Greater Boston Area that bring spatial insights into the phenomenon

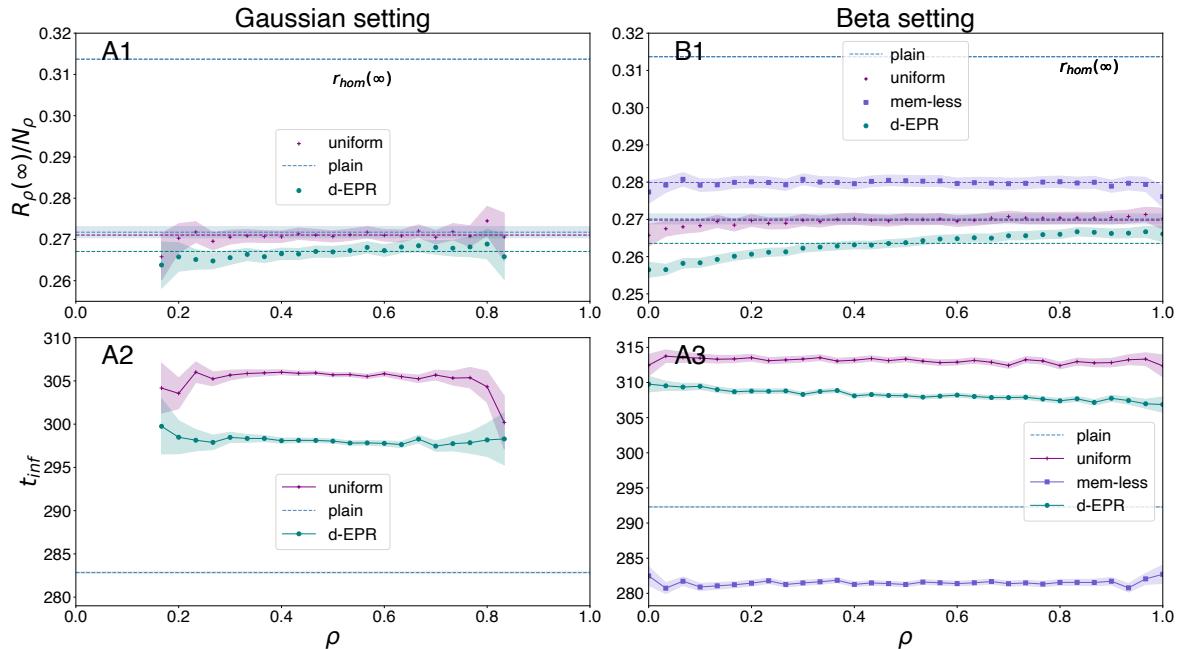


**Figure 4.3: Invasion spatial patterns.** Invader’s  $\langle \rho \rangle$  profile, local probability of invasion, and average invasion time maps, respectively for the beta d-EPR (Panels A1, A2, and A3), the beta memory-less (Panels B1, B2, and B3), and the Gaussian d-EPR scenarios (Panels C1, C2, and C3) on the Greater Boston Area. The epicenter cell in the lattice is shown in bright red.

(Figure 4.3). We have selected the beta d-EPR scenario (Panels A1, A2, A3), the beta memory-less scenario (Panels B1, B2, B3), and the Gaussian d-EPR scenario (Panels C1, C2, C3). The top panels (A1, B1, and C1) focus on the spatial representation of the invasion process, depicting the  $\rho$  profile of a typical invader. Under the same d-EPR mobility, the contrast between the beta (Panel A1) and Gaussian settings (Panel C1) is striking. While on the first case, the map is tainted reddish, signaling that on average invasion is led by an agent with  $\rho > 1/2$ , only slightly descending into  $\langle \rho \rangle < 1/2$  in the less attractive and furthest from the center locations, this pattern blurs in the Gaussian setting, where the typical invader is rather characterized by  $\langle \rho \rangle \approx 1/2$  at any location. Removing the non-Markovian component of the model (Panel B1), leaves with a spatial pattern rather similar to the Gaussian case, with some punctuated exceptions where a more marked value of  $\langle \rho \rangle$  appears. When looking at the local probability

of being invaded (Panel B2), these locations have a rather low probability of being invaded, accounting for a reduced sample for the statistics. These locations have a very low attractiveness, which reduces the probability of being explored and invaded. But, differently in this setting, as the recurrent mobility is removed, there is no home attachment by the agents that started their trajectories here, further complicating returning to these locations with the disease. Finally, bottom panels (A3, B3, and C3) in Figure 4.3, show local average invasion times, respectively for the scenarios discussed. In any scenario, we can appreciate a clear spatial structure where, predominantly, as we move away from the unique big center in the system, invasions occur later on average.

## Disease prevalence



**Figure 4.4: Overall infections.** The top panels represent the prevalence per mobility group for the Gaussian (Panel A1) and the beta distribution (Panel B1) settings. We compare the d-EPR against the proposed baselines and also against the classical single-population SIR solution  $r_{hom}(\infty)$ . Bottom panels (A2 and B2) represent the infection times per mobility group  $\rho$ . Dots and dashed lines represent average values and the shaded areas represent the 95% confidence interval.

Similarly, as we did for the invasion process, we look at the overall infection share per  $\rho$  group, i.e. the disease prevalence  $r_\rho(\infty) = \langle R_\rho(\infty)/N_\rho \rangle$ , and the infection times (Figure 4.4), together with the spatial patterns of the contagion process (Figure 4.5). Under the Gaussian distribution for the mobility profile  $\rho$  (Figure 4.4 A1), we appreciate no differential impact across  $\rho$  groups, and the average global prevalence  $r(\infty) = \langle \sum_\rho r_\rho(\infty) N_\rho / N \rangle$  for any setting is essentially the same. Note also that the

solution for the classical single-population SIR model  $r_{\text{hom}}(\infty)$  has been added as a further reference. All of the outcomes are separated from this solution, signaling we are working far enough from the homogeneous mixing assumption. Moving to the beta setting in Figure 4.4 B1, differences emerge. Both the plain and memory-less baselines show a higher impact than the uniform baseline and the d-EPR model. As we have already hinted, the lack of structure contributes to this higher impact, even though the differences are not large. When looking at differences across the  $\rho$  profile, only the beta d-EPR model shows a noticeable differential impact in prevalence. The impact across groups increases as  $\rho \rightarrow 1$ . The separation from the global average is higher in the lower regime of  $\rho$ . We can see then that explorers tend to be more affected by the disease spreading than returners. As we will soon see, this outcome is supported by the result in Figure 4.5 (Panel A2), where higher attractiveness correlates with higher local new cases generation, and moreover, with an explorer as the typical infected agent. As for the infection times, in either setting (Figure 4.4 A2 and B2), there are no meaningful differences, and within each scenario, on average, all mobility groups experience the infection at around the same time. This hints at a very synchronized dynamics, once the disease has been spread across the system, which contributes to partially explaining the small differences seen in prevalence across groups.

The spatial analysis of the contagion process helps to understand these previous results and support them (Figure 4.5). Panels A1, B1, and C1 in Figure 4.5 show the typical infected agent  $\rho$  profile at every location for the aforementioned scenarios. The differences in prevalence across  $\rho$  groups found in the beta d-EPR scenario can be appreciated spatially here. The closest and most attractive locations are dominated by a pale red pattern, signaling  $\langle \rho \rangle$  slightly above  $1/2$ , which turns into a pale blue as we approach the spatial boundaries of the system. Thus, there is a tendency for explorers' profiles to be infected in the most attractive locations and for returners to be infected in the least attractive ones. Differences in prevalence by  $\rho$  profile can be understood then by looking at these spatial patterns. Panels A2, B2, and C2 go on to show the correlation between location attractiveness and the average local number of cases generated during the whole epidemic. The most attractive locations have a higher total generation of cases, which is something to be expected given that these concentrate, individually, more visits than lower attractive locations.

The correlation is especially high in the beta and Gaussian d-EPR scenarios, and high but remarkably lower in the memory-less one. In this scenario, the lack of recurrent mobility would tend to overpopulate the most attractive locations and underpopulate the least attractive ones, with respect to the initial values. This provokes a superlinear case production in the former ones, leaving a high quantity of locations with very small

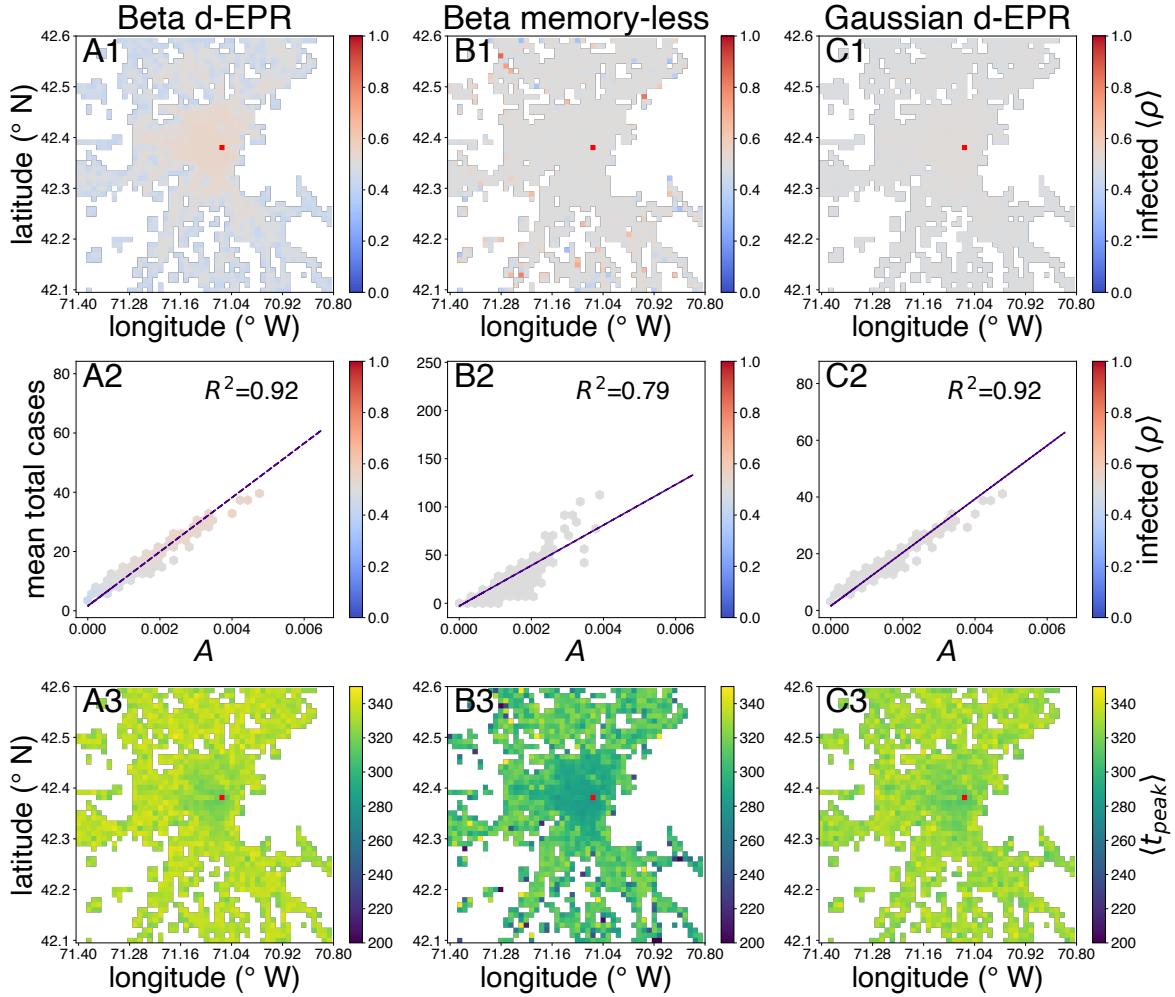
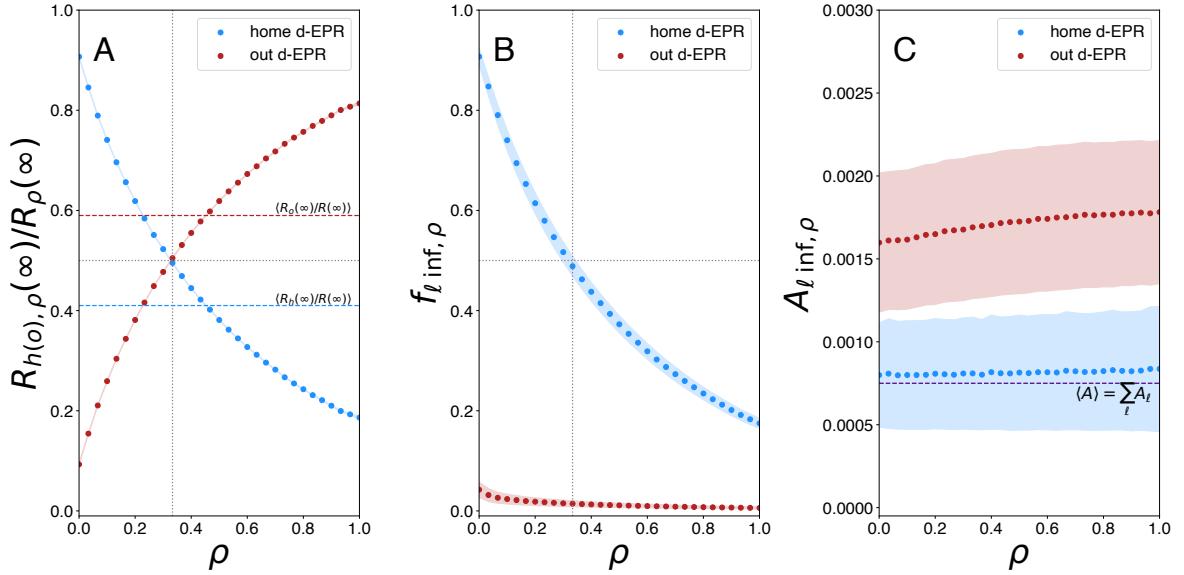


Figure 4.5: **Contagion spatial patterns.** Infected's  $\langle \rho \rangle$  profile map, average local case generation versus attractiveness, and average peak time maps, respectively for the beta d-EPR (Panels A1, A2, and A3), the beta memory-less (Panels B1, B2, and B3), and the Gaussian d-EPR scenarios (Panels C1, C2, and C3) on the Greater Boston Area. The epicenter cell in the lattice is shown in bright red.

to medium attractiveness values with very few cases compared to the other scenarios. Finally, panels A3, B3, and C3 in Figure 4.5 depict, respectively the average local peak times for the studied scenarios (beta d-EPR, beta memory-less, and Gaussian d-EPR). The main takeaway here is that, whereas invasion times spanned over the breadth time interval of the order of 200-300 time steps, peak times show a much shorter span, concentrated mainly in less than 50 time steps, or even shorter for the majority of locations. Therefore, as we stated previously, the local dynamics are highly synchronized by this stage of the spreading, and since the majority of infections occur around the peak time, this helps us to understand why there are no meaningful differences in infection times across  $\rho$  groups.

To finish the analysis of the disease prevalence, let us delve into more detail on the geography of infections. In this model, by construction, home locations are the most

visited locations with respect to any other location, by every agent, even for the top explorers. Consequently, we want to characterize the following issues: i) Where are the agents getting infected, preferentially at home or outside? ii) What is the visitation frequency  $f_{\ell \text{ inf}}$  of the location where the infection happened? And iii) What is the attractiveness of this location  $A_{\ell \text{ inf}}$ ? For simplicity, we make the distinction between home locations and outside home locations. As before, we analyze these quantities by mobility group  $\rho$  (see Figure 4.6).



**Figure 4.6: Infections at home locations and outside.** All results are shown for the d-EPR model under beta-distributed  $\rho$  values. Panel A represents the share of infected individuals at home (outside). Global values for the prevalence at home and outside are computed and shown in colored horizontal dashed lines. The point where half of the infected population has been infected either at home or outside intersect is visually guided by dotted lines, and occurs at  $\rho = 0.33$ . Panel B depicts the visitation frequency to the location (home or outside) where the agents were infected. Dotted lines are also added to signal the point where agents spent half of their time at home ( $\rho = 0.33$ ). Panel C represents the average attractiveness of that location per mobility group  $\rho$ . Symbols represent the average values over the ensemble of simulations and shaded areas represent the 95% confidence interval.

We focus only on this scenario because, as posed, it does not apply to the others, as either there is no sense of *home* (memory-less and plain) or attractiveness variation (uniform and plain). Moreover, the Gaussian setting shows no noticeable differences with respect to the beta one. Then, Panel A represents the fraction of agents in every  $\rho$  group that was infected either at their home location or outside of it. In this particular system, we see that the majority of infections, around 60% occur outside the infected agent's location, whereas the 40% of them occur at the agents' home location. When looking at the share of infections per  $\rho$  profile, for low  $\rho$ , infections at home dominate

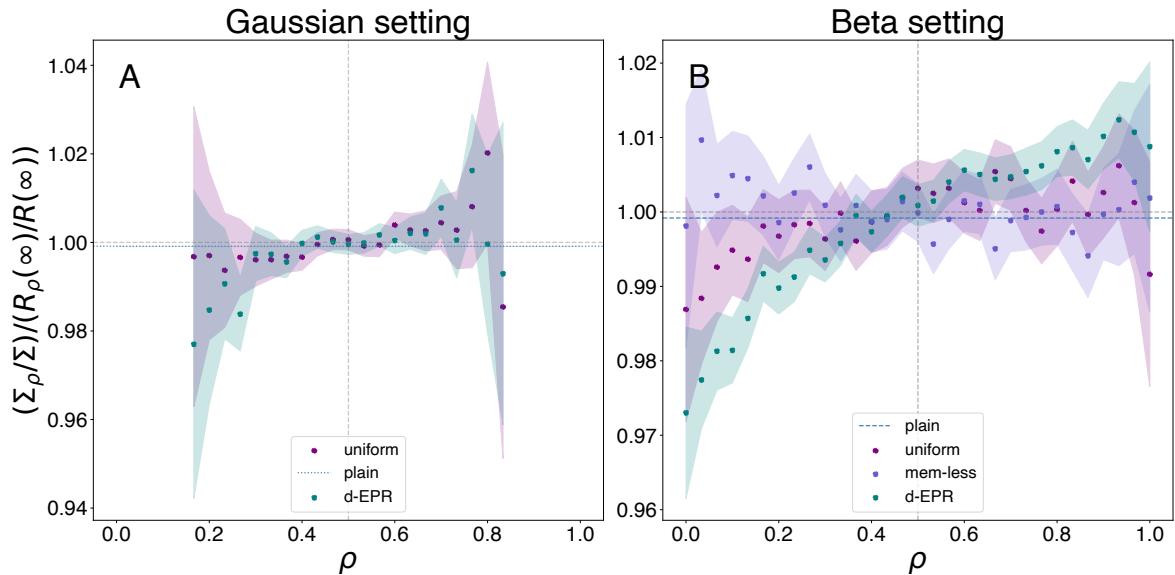
over infections outside, whereas for high  $\rho$ , the dominance is inverted. The crossover point occurs at  $\rho = 0.33$ . Panel B now checks the visitation frequency during the whole trajectory of the locations where agents suffered the infection. As expected, the visitation frequency of the agents infected at home is high, even though it decays steeply as  $\rho \rightarrow 1$ , accounting for less  $f_{\ell \text{ inf}} < 0.2$  for top explorers. For a certain value of  $\rho$ , the share of time spent at the home location falls below 50% of the time. This value is precisely  $\rho = 0.33$ . The results of infections at home and outside now come as no surprise: the change of trend, where infections outside dominate over those happening at home, occurs at the point where the probability of being outside the home location surpasses 1/2.

Now, for agents infected outside, we observe that the visitation frequency of that location is extremely low, only showing a slight uptick for top returners. Thus, given this low visitation frequency, were infections outside the home location just a bad lucky strike? Well, no for two reasons. For one, possibly obvious, is that every agent has only one home location, whereas there are other  $V - 1$  locations to potentially get the infection. Whereas as  $\rho \rightarrow 0$ , agents can spend a very high amount of time at their home location, for higher  $\rho$  groups, the visitation frequency quickly decays and the agents are able to sample several other locations. This leads to the second motive which is shown in Panel C. On average, the attractiveness of the infection location of the agents infected outside is remarkably higher than the attractiveness of agents infected at home. Knowing from before that higher  $A$  leads to higher new case generation, if these agents, when leaving home, are on average sampling higher frequency locations, getting infected should come as no surprise even though they do not visit that location again. Note also how we can appreciate here that this average attractiveness of the location of the infection is higher for higher  $\rho$  groups. On the other hand, since home locations are assigned stochastically in a homogeneous way, we see a flat average attractiveness with  $\rho$  when the infection occurs at home. We check, in fact, that this line is very close to the average value of the attractiveness distribution in the system  $\langle A \rangle = 7.5 \times 10^{-4}$  (dashed horizontal line in Figure 4.6).

## Contribution to infections and contagion events

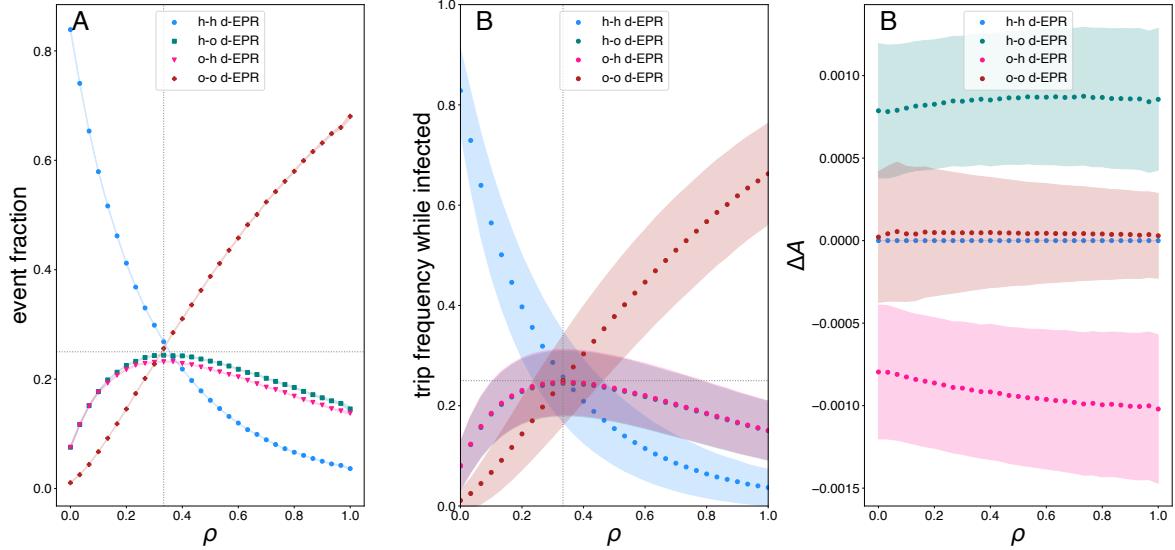
Previously we looked at the role of the exploration and preferential return mobility in delivering the disease across the system and on being affected by the disease. To conclude this work, we look at the infection process through the eyes of the *infector*, not the infected, and ask ourselves about the role of the mobility profile in the contribution to the generation of newly infected cases at any contagion event. To assess this issue by  $\rho$  mobility group, we count the number of infected cases generated in every contagion

event by infectors belonging to the corresponding  $\rho$  group, let  $\Sigma_\rho$  be this quantity, and let  $\Sigma$  be the total number of cases generated by adding over all the  $\rho$  groups. By rescaling  $\Sigma_\rho$  with  $\Sigma^{-1}$ , we get a normalized account of generated new cases per  $\rho$  group. The quantity  $\sigma_\rho = \Sigma_\rho/\Sigma$  is then compared against  $r_\rho(\infty) \equiv R_\rho(\infty)/R(\infty)$ . A constant quotient  $\sigma_\rho/r_\rho(\infty)$  across  $\rho$  groups is expected assuming no special role of mobility when triggering contagion events. Moreover, if  $\sigma_\rho/r_\rho(\infty) = 1$  for any group  $\rho$ , proportions  $\sigma_\rho$  and  $r_\rho(\infty)$  are both identical, meaning that they are conserved and that infectors on those groups tended, on average, to trigger contagion events of size 1 (one new infected case). If  $\sigma_\rho/r_\rho(\infty) > 1$ , for some  $\rho$ , then the corresponding group  $\rho$  contributed to infections with more than one case, and conversely if the sign is reversed ( $< 1$ ). Then, results (Figure 4.7) show a clear enough picture for both settings and all scenarios within. Even though the d-EPR model under the beta setting is the only model showing a more consistent tendency toward higher contribution to newly infected cases as  $\rho \rightarrow 1$ , the separation from the unit horizontal line is small. For the top returner category,  $\rho \rightarrow 0$ , the average number of new cases generated is within  $[0.961, 0.985]$ , whereas for the top explorer category,  $\rho \rightarrow 1$ , this quantity is within  $[1.00, 1.02]$ . Thus, even in the extremes, the contribution is very close or within unity.



**Figure 4.7: Contribution to infections.** Panel A represents the re-scaled contribution to infections per  $\rho$  group,  $(\Sigma_\rho/\Sigma)/(R_\rho(\infty)/R(\infty))$ , for the Gaussian distribution setting and tested baseline scenarios. Panel B shows the same quantity under the beta distribution setting and corresponding scenarios. Dots and dashed lines represent average values and the shaded areas represent the 95% confidence interval.

From all the analysis of this work, we know that explorer profiles (and even from  $\rho \approx 0.33$  on) tend to spend more time outside their home location, which makes them more prone to be infected outside. This feature still holds during the infectious



**Figure 4.8: Event and trip decomposition by origin and destination of infections.** Under the beta setting, per every  $\rho$  group Panel A represents the fraction events where the infected agent contributed to the infection and where the agent was infected. Panel B represents the frequency of trips by origin-destination while the agent was infected. Panel C represents the changes in the attractiveness of those trips. Locations considered can be either the home location or outside of it. Dots and dashed lines represent average values and the shaded areas represent the 95% confidence interval.

period of the agents and, therefore, explorer-infected agents should be expected to contribute more to triggering infections outside their home location. This is indeed clearly observed in Figure 4.8 A, where we have collected all the successful contagion events in which all infected agents (designated as infectors or not) participated. We look at the origin location where the agent was infected, and at the destination location where the infected agent participated in a successful contagion event. We decompose these origin-destination situations by home-home (h-h), home-outside (h-o), outside-home (o-h), and outside-outside (o-o). For  $\rho \in (0, 0.33)$ , we see that the dominant contribution comes from h-h situations. It is at  $\rho \approx 0.33$  where each of the four situations contributes similarly by 1/4. That is, agents characterized by that mobility parameter, can be infected at home/outside and infect outside/at home with equal probability. Noticeably, during most of the  $\rho$  range, there is a far from negligible contribution of h-o/o-h situations. Panel B in Figure 4.8 shows a similar picture, where now h-h, h-o, o-h, o-o, refer to the trips between locations during the infected trajectory of agents. Thus, we can see that the same dynamic occurs when we look at a more general picture during this time period. The higher uncertainty is due to the fact that trajectories of agents while infected are obviously conditioned on their infectious period duration, which is sampled from an exponential distribution and which on average is

$T_I = 10$  time steps. Finally, Panel C in 4.8 shows the jumps in attractiveness when switching between the different situations considered during their infectious period. Trivially, h-h trips have  $\Delta A = 0$ , but o-o trips also show a very small change in attractiveness, meaning that those types of situations tend to involve locations very close in  $A$  value. Then, h-o trips show a noticeable change in  $A$ , supporting the claim that these kinds of trips involve trips from homes toward the core of the system, where the most attractive places are located. Correspondingly, o-h trips show  $\Delta A < 0$ , as they represent backward trips from o-h situations.

Therefore, as explorer-infected agents tend to contribute to infections outside the home location and, on average, at locations with higher attractiveness, which in turn concentrate more visits, we could expect that these types of agents trigger events with a higher number of infected cases with respect to those triggered by returners, which tend to spend more time at home and, on average, in lower attractiveness locations. One reason, though, to see a small effect in the contribution to infections across  $\rho$  groups can be due to the exponentially distributed infectious period, for which an important part of the infected population can be short enough to blur the differences in location sampling by explorers and returners.

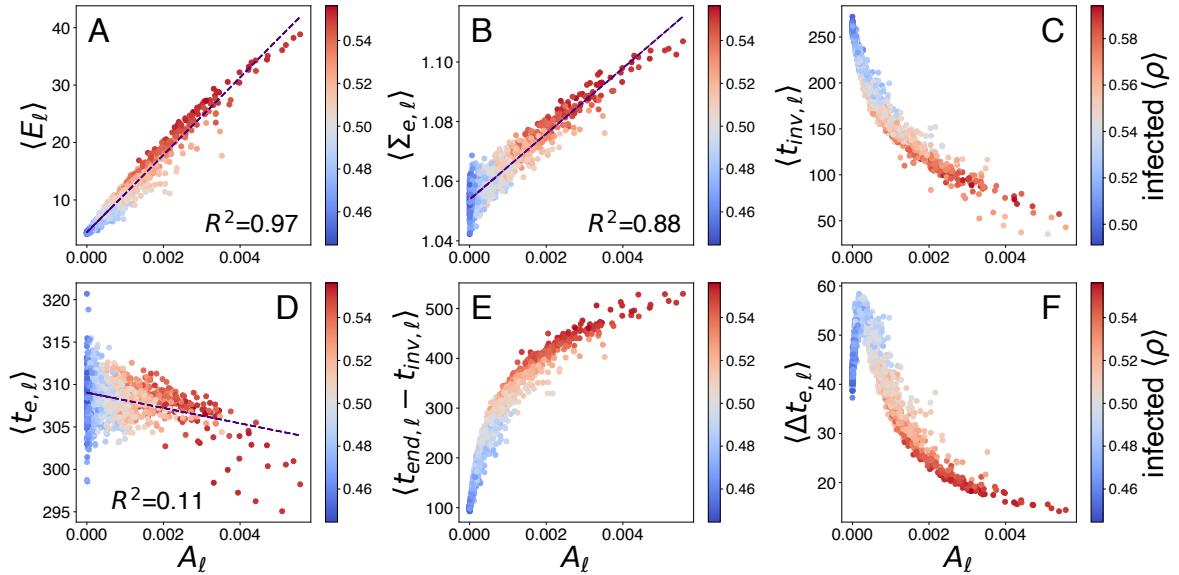


Figure 4.9: **Contagion events and locations attractiveness.** Scatter plots of locations  $\ell$  as characterized by their attractiveness value  $A_\ell$  against some features of the contagion events experienced in such location  $\ell$ . Panel A: Number of events ( $R^2 = 0.97$ ). Panel B: Size of events ( $R^2 = 0.88$ ). Panel C: Invasion times. Panel D: Average event times ( $R^2 = 0.11$ ). Panel E: Time window between the occurrence of the first event (invasion time) and the last event. Panel F: Inter-event times. All points show average values over the ensemble of simulations. Color bars represent the  $\rho$  value of the typical infected agent.

To gain some more insights about this and conclude the analysis, Figure 4.9 delves

into some features of contagion events related to locations' attractiveness. First (Panel A), we find that attractiveness  $A_\ell$  correlates ( $R^2 = 0.97$ ) with the average number of events  $\langle E_\ell \rangle$  happening in a location  $\ell$ , which in turn also correlates with the  $\rho$  parameter of the infector triggering the event. We next find that attractiveness correlates with the average size of the event triggered ( $R^2 = 0.88$ ) but, importantly, even though the correlation is clear, the range of sizes is very tight,  $\langle \Sigma_{e,\ell} \rangle \in (1.04, 1.12)$ . The implication is that the most attractive locations, by the fact of having a larger number of visits and population, are neither greatly nor disproportionately contributing to the generation of new infected cases with respect to the least attractive locations. What makes them centers of contagion (as seen in Figure 4.5 A2, B2, and C2) is the fact that the probability of there occurring a successful event is higher, as given by Figure 4.9 A. Then, although, the most attractive locations are prone to experience more contagion events, their average sizes do not differ greatly from those of events happening in the least attractive locations. This fact could be another factor to diminish the impact of exploration in the generation of new cases.

Now, Panels C, D, E, and F look specifically at the relationship of a location attractiveness  $A_\ell$  to the time of certain event-related milestones happening there. Panels C and D represent average invasion times and infection times (whose spatial perspective was offered previously in Figures 4.3 and 4.5). The trend is clear: the most attractive locations are impacted first and as  $A$  decreases, invasion times are delayed in a nonlinear way, being the invasion especially late in the least attractive locations. The time window of all the invasion events spans more than 200 time steps. However, the average infection times are barely correlated with the attractiveness values and are strongly concentrated in a short time window of approximately 30 time steps. As already stated, this is signaling a strong synchronization of contagion events across the system. Finally, Panels E and F focus on the time window from the first case generated at  $\ell$  (invasion event) to the last one locally, and on the inter-event times. According to Panel E, the least attractive locations experience their local outbreaks during a shorter amount of time, whereas the most attractive locations sustain the epidemic during significantly greater times. Panel F shows complementarily that inter-event times are smaller in the most attractive locations, and that this time span grows as  $A$  decreases, but up to a point where a peak is reached and the least attractive locations see a reduction in inter-event times. Rebuilding the full picture given by Figure 4.9, we can state that, on average, the most attractive locations are invaded first, the time window of contagion events is largely broader than for the least attractive locations, and during this time span, more contagion events occur with a shorter time separation. The least attractive locations experience their first cases substantially later and show fewer events

during a much shorter time window, also these few events are more separated in time. In these locations, the epidemic is then short-lived and tends to be focused around the time that the majority of cases are happening.

## 4.4 Conclusions

Mobility modeling within the metapopulation framework has typically followed the use of naive assumptions regarding how humans move and has been disconnected from the recent, and not so recent, advances in the understanding of human mobility. In this work, we have contributed to exploring this venue, by coupling the spreading of an epidemic with a more realistic microscopic model of human mobility, based on the EPR family, the d-EPR model. We have tried to give an exhaustive characterization of the epidemic under this type of mobility and to assess the role of exploration and preferential return in relevant epidemiological observables. To better frame the model capabilities, we have explored the epidemic impact under different distributions of the  $\rho$  parameter characterizing agents' mobility, homogeneous (Gaussian) and heterogeneous (beta), and also under different variants of the model, that removed essential features (like the non-Markovian traits in both exploration and return and the heterogeneity in the attractiveness distribution).

Overall, we find that in the setting where the population shows heterogeneity in their mobility profile as given by  $\rho$  and all the original features of the d-EPR model are conserved, differential impact across  $\rho$  groups is observed, and we can state that this type of mobility has an impact on the spreading and, in turn, mobility groups play a different role and are affected differently. The invasion dynamics is the most evident explored phenomenon where the d-EPR shows a clear departure from more naive baselines or homogeneous settings. Explorers drive the disease invasion and therefore help to deliver the disease across the system, they do it proportionally more than returner profiles, and significantly faster. Regarding the disease prevalence in the population, returner profiles tend to be clearly less affected by the disease with respect to the global average value, whereas explorers tend to be more impacted. As it is found that the most attractive places show both more contagion event occurrence and accumulate more cases generated, exploration favors the chances of getting the disease. Admittedly, even though the differential impact across  $\rho$  groups is evident, the deviations are not notoriously large. In fact, overall infection times show no meaningful differences across  $\rho$  values. This points to a highly synchronized stage, as it is indeed found, in which the majority of cases are concentrated within a rather short time window, the disease is widely spread and then there is little difference in the mobility

profile of the infected agents. This, in turn, is facilitated by the fact that from rather low  $\rho$  values ( $\rho \approx 0.33$ ), agents tend to spend more time outside their home location, visiting locations that on average tend to be more attractive. Looking at the impact from the point of view of provoking the infections, we saw also a similar tendency, i.e. explorers contribute more to the generation of new infected cases than returners. This effect can be also explained by the same argument that they are more prone to being infected, but it must be noted that the differences in contribution are rather small, even for extreme values of  $\rho$ . From the point of view of locations, we find that the most attractive locations tend to be impacted first by the disease. The invasion process shows a rather broad range of times, which are then shortened in the exponential phase of the spreading, being local peak times concentrated in a noticeably narrower time window. The most attractive locations experience a high number of contagion events, spanning in longer time windows and thus showing shorter inter-event times, therefore helping to globally sustain the spreading.

In summary, the coupling of the d-EPR model of microscopic human mobility with an epidemiological model brings a new dimension to the interplay between mobility and the spreading of an epidemic. Under certain conditions, EPR-like mobility can indeed play a role in the spread and impact of a disease in a population, compared to more naive and traditional mobility models traditionally used within the metapopulation framework. Accounting for this type of behavior and heterogeneity in mobility attitudes can indeed help to devise more fine-grained strategies to contain an epidemic in environments where this type of behavior is relevant. However, even though the d-EPR model incorporates some realistic behavior as unveiled by the scientific literature in the field of human mobility [428, 429], it is far from a definitive and faithful account of complex human dynamics, and some limitations should be acknowledged.

Regarding our implementation, we can think of some potentially important ones. One of them is the coarse spatial resolution used, due in part to the expensive cost of simulating and then analyzing an agent-based model. A higher spatial resolution could reveal subtleties in the dynamics and the effects observed. Another one is related to the time scale resolution. The original EPR as well as the d-EPR consider heterogeneous sojourn-times at locations following a power-law distribution. However, we followed the standard procedure in the metapopulation framework of performing discrete-time simulations with unitary time steps. Accounting for this finer temporal structure could impact the time and size of the gatherings towards favoring the heterogeneity brought by explorers and returners profiles. Finally, related to the time scale problem, the infectious periods in our epidemiological model are exponentially distributed. This translates into an important part of the infected population decaying

to recovered/removed health status in a very short time span. Given the coarse time resolution of the mobility dynamics, this means that for an important part of the population, it does not really matter their mobility attitude during their infectious period, thus contributing to vanishing potential differences between mobility profiles. Albeit important, the exploration of the impact of these modeling options lies out of the scope of this work and should be considered as future work.

As for the limitations of the d-EPR model, some of these have been already presented in Section 2.4, when introducing the family of EPR models for microscopic mobility. To recapitulate, the d-EPR model lacks recency [431], which gives earlier discovered locations a cumulative advantage over later visited locations, precluding people from changing location preferences. The model also lacks memory in the sense that agents can potentially add newly discovered locations to the set of already visited locations, which is also in contrast with the observed behavior [432]. Considering these more realistic features could impact the way the epidemic is spread across the population. Regarding the exploration stage, the d-EPR model uses the gravity model, but other extensions consider different mechanisms like the PEPR [433], based on preferential exploration, or the STS-EPR [437], introducing a social dimension. Moreover, here, the attractiveness values of locations are static and reflect a mean-field configuration. The first feature does not take into account things like circadian rhythms or event-based dynamics (such as weekends versus weekdays, concerts, special gatherings, and so on). The second feature refers to the fact that location attractiveness values are based on collective flows of agents and disregards the heterogeneity in agent preferences: not necessarily all the locations may appear as the same relevance for all the agents, since this may depend on personal circumstances.

As it can be appreciated, there exist several venues to extend the model in order to acknowledge the complexities of real human dynamics. However, we believe that the next step for future work before further complicating the model or trying a different approach is to test it against running the epidemic with mobility based on real trajectories, something that nowadays can be done thanks to the availability of massive high-resolution data on human mobility. Complementarily, here we have studied an epidemic in the Greater Boston Area as the spatial substrate, therefore it could be interesting to explore other urban environments with different spatial organization and locations' attractiveness distribution.

# **Part III**

## **Single-population behavioral structured models**



# Chapter 5

## Exploring the impact of vaccine hesitancy in secondary outbreaks of COVID-19 in the US

*We have it totally under control. It's one person coming in from China.*

*We have it under control. It's going to be just fine.*

— Donald J. Trump on January 22, 2020

*We're using the full power of the federal government to defeat the virus, and that's what we've been doing.*

— Donald J. Trump on March 14, 2020

*And I think we're doing very well on the vaccines but, with or without a vaccine, it's going to pass, and we're going to be back to normal.*

— Donald J. Trump on April 5, 2020

*Even without the vaccine, the pandemic's going to end. It's gonna run its course. It's gonna end. They'll go crazy. He said 'without the vaccine' — watch, it'll be a headline tomorrow. These people are crazy. No, it's running its course.*

— Donald J. Trump on October 16, 2020<sup>1</sup>

### 5.1 Introduction

The COVID-19 outbreak, first reported by the Chinese authorities on December 31, 2019, and declared as a pandemic in February 2020 by the World Health Organization, has been the worst global health crisis in at least a century.

---

<sup>1</sup>For more, please see <https://edition.cnn.com/interactive/2020/10/politics/covid-disappearing-trump-comment-tracker/>.

Multiple aspects of our life were severely affected at various scales: psychological [478, 479] and social [480]; human-related systems and infra-structures [481, 482]; supply chains [483, 450]; and the economy in general [484, 485]. To manage the disease, it was mandatory to adopt a plethora of measures aimed at reducing the mixing and interaction among individuals in order to mitigate SARS-CoV-2 transmission and propagation. Lockdowns [486, 487, 488, 489, 490], curfews and mobility restrictions [491, 16, 492, 493], social distancing [494, 495, 496], personal protection [497, 498] are now part of the new normalcy across the world. This “new normal” [499, 500, 501], being its impact as critical as the virus itself, was conceived and promised as something that should be ephemeral, a toll to pay, just until the ultimate solution arrives: the vaccines.

A rapid and massive scientific effort [502, 503] to develop a vaccine against SARS-CoV-2 was deployed and successfully achieved in less than a year; another unprecedented fact. In December 2020, just a year after the onset of the pandemic, several nations started their vaccination campaigns in the pursuit of herd immunity to control the pandemic. But again, further problems proliferate. From lack of confidence due to the relatively short time for vaccine development and approval to the typical fears of suffering serious side effects or due to outlandish conspiracy theories, some people hesitate or are reluctant to vaccination [504, 505, 506, 507]. For good or bad, this phenomenon is neither exclusive nor new [508, 504]. To be vaccinated (or not) is, in most countries, a choice of the individual, even though the consequences of such a choice go beyond the self and affect the social sphere. Hesitancy poses an ethical problem since if a critical fraction of individuals declines vaccine uptake for any disease, resurgence is to be expected [509, 510, 511]. This already happened in the UK, which was declared measles-free in 2017 but lost this status just 2 years later due to sub-optimal vaccination uptake [512].

In this work, our aim is to quantify, via modeling, the effects of vaccine hesitancy in the US during the COVID-19 pandemic. In particular, given the large heterogeneity found across the US, we perform our analysis on each state, although our analysis and conclusions could be extrapolated elsewhere provided relevant data is at our disposal. We make use of an age-structured SIR model to simulate the spreading dynamics, which is fed with real and up-to-date data of the US age-distributed population and contact matrices, as well as with survey-based seroprevalence estimations [513]. Despite these elements of realism, it is worth noting that we do not intend to replicate the real trajectory of the COVID-19 pandemic in the United States until now, nor do we aim to accurately forecast the unfolding of future outbreaks and epidemic sizes. Rather, we try to answer the following question: what would happen if a certain fraction of the

population hesitates or declines vaccine uptakes and normalcy is re-established (as is now the case in many places, not only the US) or new aggressive variants emerge? We propose a hypothetical scenario in which COVID-19 outbreaks emerge in each state, independently, with a mitigated propagation due to the presence of some restrictions, while there is an ongoing vaccination campaign designed following the information obtained from public surveys [514]. Once the vaccination and this first outbreak end, we assume a *back to normal* situation, where all restrictions are lifted disease awareness vanishes, and a new outbreak sets in. To make our simulations more realistic, we assume that these successive outbreaks happen for a more transmissible variant of the virus, mimicking in this way the evolutionary path of the SARS-CoV-2 variants of concern that have evolved towards more transmissible forms of the virus. We focus on looking at correlations between variables related to hesitancy and the impact on the population, i.e., attack rates and deaths, for every US state. We also explore the role that age structure may play in conditioning the outcomes and estimate potentially averted deaths if a 1% point more of the hesitant fraction of the population would change its attitude. Finally, we compare the model output from our hypothetical scenario with the epidemic impact caused by the COVID-19 Delta wave during July–November 2021.

## 5.2 Material and methods

### Epidemic model

Given the utmost relevance of age in the effects of COVID-19, it is compulsory to introduce the age distribution of the population and the specific interaction between age groups to adequately model the dynamics of the disease [515, 516, 397]. We use the estimated age-contact matrices provided by Mistry et al. [360] updated to the population structure of 2019 [517, 404]. Then, we build an age-structured SIR model defined by this set of equations [360]:

$$\begin{aligned} \frac{dS_a}{dt} &= -\lambda_a S_a, \\ \frac{dI_a}{dt} &= \lambda_a S_a - \gamma I_a, \\ \frac{dR_a}{dt} &= \gamma I_a, \end{aligned} \tag{5.1}$$

where  $S_a$  is the number of susceptible individuals of age  $a$ ,  $I_a$  is the number of infected individuals of age  $a$ ,  $R_a$  is the number of removed individuals of age  $a$ , and  $\gamma^{-1}$  is the infectious period, which is assumed to be the same for all age classes and equal to  $\gamma^{-1} = 4.5$  days. COVID-19 is a disease with a more complex natural history than a SIR

model can account for, being required to add some pre-symptomatic or asymptomatic compartments, as well as a latency period, for certain applications. Nonetheless, it has been shown that SIR models can correctly describe the overall evolution of the disease [518], which is enough for the scope of this paper. Lastly,  $\lambda_a$  is the force of infection for individuals of age  $a$  and it is expressed as

$$\lambda_a = \beta \chi_a \sum_{a'} M_{aa'} \frac{I_{a'}}{N_{a'}}, \quad (5.2)$$

where  $\beta$  is the transmissibility of the virus,  $N_a$  is the total number of individuals of age  $a$ , and  $M_{aa'}$  measures the average number of contacts of an individual of age  $a$  with individuals of age  $a'$ . Finally,  $\chi_a$  is an age-dependent susceptibility factor accounting for the lower susceptibility of children to the disease, i.e.  $\chi_a = 0.56$  if  $a \leq 19$  and 1 otherwise [519].

The basic reproductive number  $R_0$  is defined in this model as

$$R_0 = \frac{\beta}{\gamma} \rho(\chi M), \quad (5.3)$$

where  $\rho(\chi M)$  is the spectral radius, or largest eigenvalue, of the age-contact matrix (in this case also incorporating the susceptibility factor) [117].

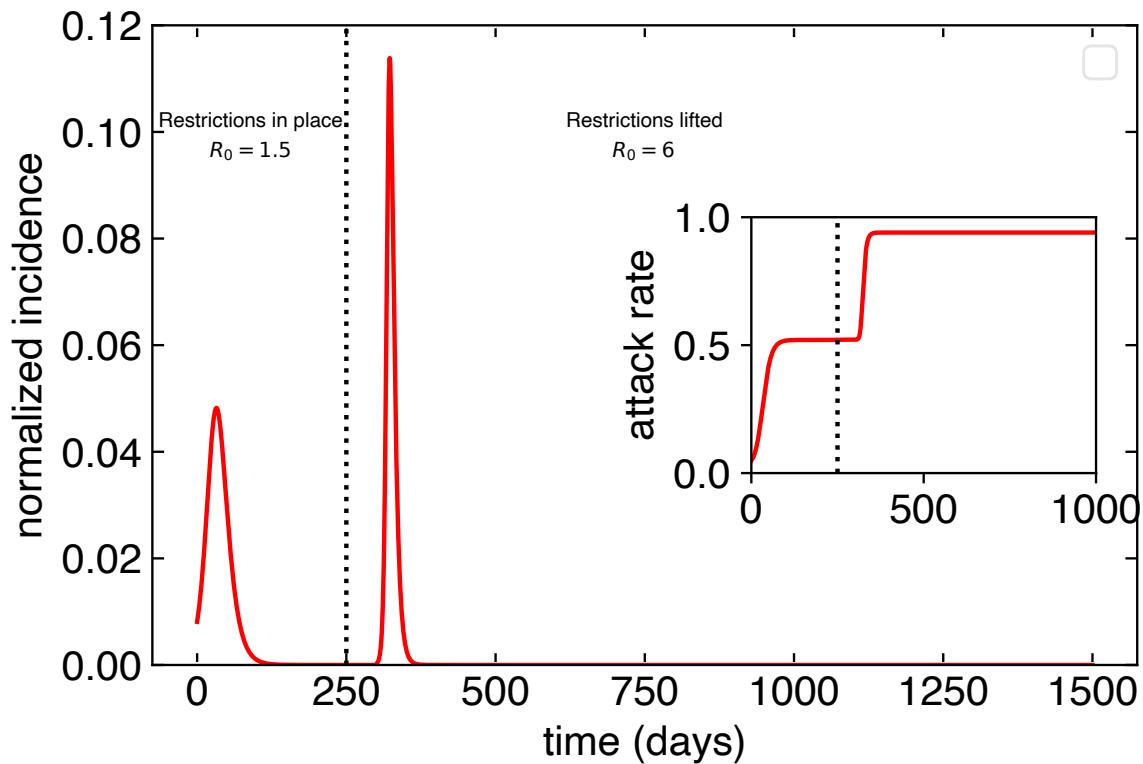
## Scenario

First, we collect the seroprevalence data measured in September 2020 for each age group and US state. We set the corresponding fraction of the population in each state into the removed compartment. Second, we simulate an initial outbreak with an  $R_0 = 1.5$ . This basic reproductive number is below the estimated  $R_0$  value for unmitigated transmission of the original variant of SARS-CoV-2 which is around 2.5-3 [520]. With this choice, we mimic a scenario in which there are some restrictions, social distancing, and other protective measures in place, yielding a smaller effective  $R_0$ . During this outbreak, we implemented a vaccination campaign (described below). By the end of the campaign, all individuals who have not refused vaccine uptake will have been vaccinated.

Once the vaccination campaign is completed, we assume that societies have returned to normalcy, i.e., any kind of restrictions and precautionary measures are lifted. Then, a new outbreak is seeded in each state, emulating a spillover from other states in the country or importation from other countries. In this second outbreak, we set  $R_0 = 6$  which is closer to the dominant variant of concern (delta variant) [521], predominant in the US since mid-2021 until the emergence of the Omicron strain by the end of 2021 [522, 523], being this last variant of concern more highly transmissible with

respect to Delta even though with lower pathogenicity [524]. The reason to propose these secondary outbreaks in each US state with an  $R_0$  more in line with Delta is the availability of data to compare our hypothetical scenario with reality. Note that we assume that no awareness or other non-pharmaceutical interventions are in place during this outbreak. Thus, it can be thought of as the worst-case scenario of resurgence after a vaccination campaign.

As a visual example of the proposed scenario, in Figure 5.1 we show how the incidence would evolve at the level of state for the full epidemics if no vaccination campaign were deployed during the first outbreak. When an aggressive variant sets in, secondary outbreaks may still cause havoc. The inset depicts the evolution of the prevalence, which can reach almost the whole population for large enough  $R_0$ . In section 5.3, we explore how the vaccination efforts modify this baseline scenario.



**Figure 5.1: Proposed baseline scenario.** Following the first wave of the epidemic, part of the population acquires natural immunity. Then, we simulate the propagation of a mitigated outbreak due to the presence of some restrictions, social distancing, and prophylaxis measures, leading to a slower propagation of the original variant of the disease ( $R_0 = 1.5$ ). After the outbreak is extinguished a back-to-normal situation is assumed and all prevention measures are lifted. Then, an outbreak is seeded again with a higher basic reproductive number,  $R_0 = 6$ . On top of this baseline scenario, we will introduce a vaccination campaign during the first outbreak and explore the impact of vaccination hesitancy on the second outbreak.

## Vaccination

We use data from The COVID States Project (<https://covidstates.org>), in particular the surveys in Report #43: COVID-19 vaccine rates and attitudes among Americans [514]. These surveys provide information on vaccination acceptance/hesitancy by age at the state level. Therein, several degrees of predisposition toward vaccines are reported. The following categories are distinguished: individuals who are “already vaccinated”, individuals who are inclined to be vaccinated “as soon as possible”, “after at least some people I know”, “after most people I know”, and finally people who “would not get the COVID-19 vaccine”. The shares of people in each category are given at a national level for different age groups. The data shows an important amount of heterogeneity in each of those categories by age group. However, at the level of state, the data is not disaggregated by age groups, only the share of people in each vaccine acceptance category is shown. More specifically, we are looking for the coefficients  $g_{c,a}^{\text{state}}$ , which represent the share of people for every acceptance category,  $c$ , and age-class,  $a$  in every US state. These coefficients satisfy  $N_c^{\text{state}} = \sum_a g_{c,a}^{\text{state}} N_a^{\text{state}}$ , where  $N_a^{\text{state}}$  is the population of the state in the age class  $a$  and  $N_c^{\text{state}}$  is the population of the state in the acceptance category  $c$ . These  $N_c^{\text{state}}$  values are provided in appendix A within the report, but the information is not disaggregated by age at the level of the state [514].

The report offers information at a national level about how people are distributed within acceptance categories by age group. We refer to the shares shown in the report as  $h_{c,a}^{\text{national}}$ , which are normalized by age class, that is,  $1 = \sum_a h_{c,a}^{\text{national}}$  for a particular category  $c$ . The quantity  $\sum_a h_{c,a}^{\text{national}} N_a^{\text{state}}$  would be the number of people if national coefficients apply for a certain state and vaccine acceptance category  $c$ . We relate these coefficients  $h_{c,a}^{\text{national}}$  to coefficients  $g_{c,a}^{\text{state}}$  through a linear transformation:

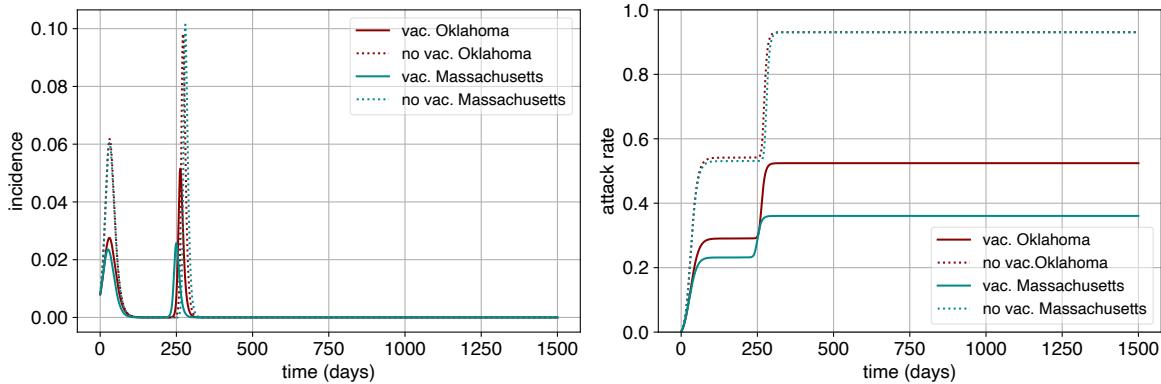
$$g_{c,a}^{\text{state}} = h_{c,a}^{\text{national}} \frac{N_c^{\text{state}}}{\sum_a h_{c,a}^{\text{national}} N_a^{\text{state}}}. \quad (5.4)$$

This transformation preserves the shares of people in a certain vaccine acceptance category  $c$  in every state and also allows for the introduction of age heterogeneity adapted from the national-level data.

Vaccination campaigns are complex and depend highly on several properties of the population: age, risk groups, professions, supplies, infrastructure, etc. Since we are mostly interested in the aftermath after the vaccination campaign, we adopt a simple scheme. From the aforementioned surveys, we extract the fraction of the population within each age group and state that is willing to be vaccinated,  $V_a^{\text{state}}$ . We set the length of the vaccination campaign to be  $\Delta t_v = 150$  days and assume that the fraction

of population vaccinated per unit of time is constant and equals to  $V_a^{\text{state}}/\Delta t_v$ . Both susceptible and recovered individuals can be vaccinated. For simplicity, the vaccine is assumed to be 100% effective in preventing the infection.

### 5.3 Results and discussion



**Figure 5.2: Comparison of spreading dynamics.** Comparison of peak incidences and final epidemic sizes for the states of Oklahoma (OK), which has the highest vaccine hesitancy, and Massachusetts (MA), where the vaccine hesitancy is the lowest according to surveys [514]. Continuous trajectories (blue and red) represent the simulation with the vaccination campaign, whereas dotted trajectories represent the simulation without introducing the vaccination campaign. All simulations started with a fully susceptible population.

In Figure 5.2, we show how the incidence and prevalence of the disease changes from the baseline scenario depicted in Figure 5.1 when vaccination is in place. In particular, we consider the state with the highest vaccine hesitancy, Oklahoma (OK), and the state with the lowest one, Massachusetts (MA). Additionally, for a fairer comparison, the simulations were started with a null initial condition for prevalence ( $R_a(t = 0) = 0$ ) (i.e. considering that the whole population is in the susceptible state). The dotted lines in the figure show the case without vaccination. We can see that the impact, in each isolated outbreak and for the full epidemic, is more or less the same for both states, with differences owing to the population's internal structure. When vaccination is introduced in the model (continuous lines), we can appreciate the reduction of peak incidence and epidemic final sizes for both states during the first outbreak. However, when we simulate the second outbreak, the state with the lowest vaccine hesitancy shows a remarkably lower impact, while the other state experiences a sizable second outbreak. The peak of the outbreaks is similar in both outbreaks for Massachusetts, while in Oklahoma, the secondary outbreak is around twice as large as the first outbreak.

Next, we focus on the overall effect of vaccination on the spreading. We explore the relationship between the final attack rate (total fraction of the population that was infected) and the fraction of non-vaccinated individuals in each state. At the time of the surveys vaccine uptake on underage people was not being considered and there was no data regarding the attitudes of this age group. Thus, this set of individuals is composed by both underage people and adults who manifested vaccine hesitancy in the aforementioned surveys [514].

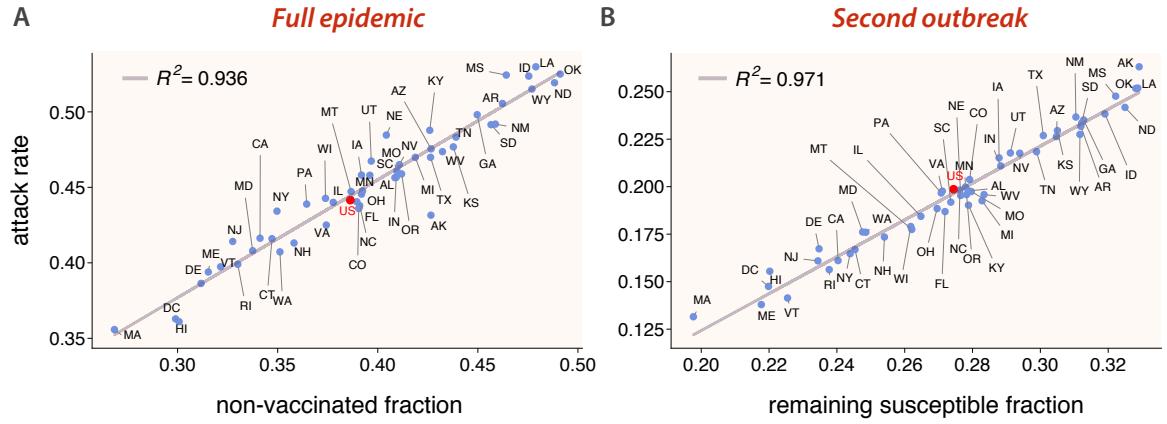
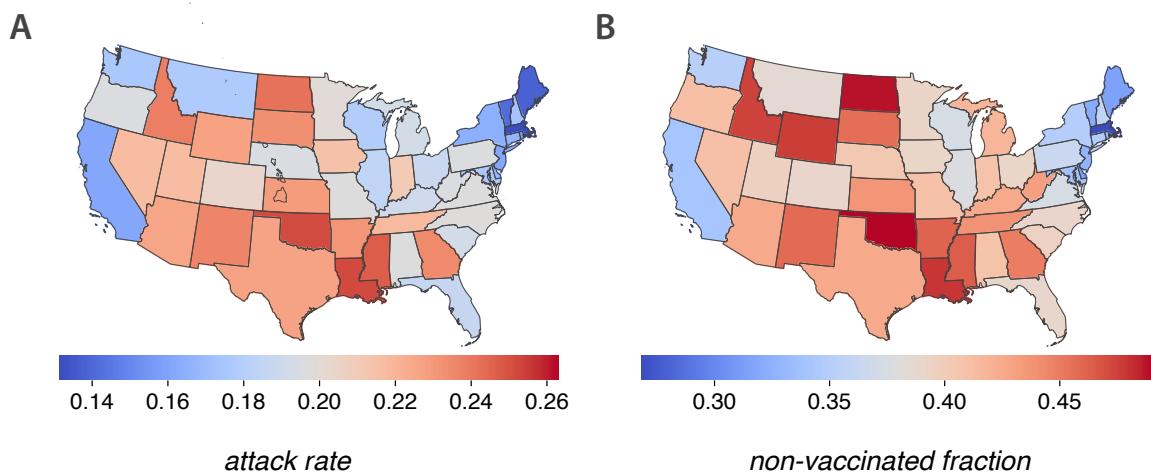


Figure 5.3: **Attack rates scatter plots.** Scatter plot of attack rates after the full epidemic (first outbreak with  $R_0 = 1.5$  and the second one with  $R_0 = 6$ ) versus the non-vaccinated fraction of individuals (A), and attack rates of the second outbreak ( $R_0 = 6$ ) versus the remaining susceptible fraction after the first outbreak (B) for every US state. The red dot corresponds to a simulation of a population representing the whole country. It is clearly seen that higher hesitancy translates into higher attack rates.

In order to look for a possible correlation between state-level attack rates and the fraction of non-vaccinated individuals, we performed a linear regression. Figure 5.3 (A) depicts a scatter plot of the attack rates versus the fraction of non-vaccinated individuals for the simulated full epidemic unfolding in every state. The correlation coefficient,  $R^2 = 0.936$ , shows a clear relationship between attack rates and vaccine hesitancy for the full period. Note that we have added a simulation for the whole country (the red dot in the scatter plots) with the age structure from the whole population. Figure 5.3 (B) shows a scatter plot of the attack rate of the second outbreak, versus the fraction of remaining susceptible individuals at the end of the first outbreak. Here, the correlation coefficient is also very high,  $R^2 = 0.971$ . Note that the use of the remaining susceptible fraction rather than directly the fraction of non-vaccinated individuals owes to the fact that once the first outbreak and the vaccination campaign have ended, the demographic structure of the pool of susceptible individuals has changed dramatically. This pool is all comprised of individuals who

either declined vaccination or are underage. Since, according to data, hesitancy rates are low in older people, there is a predominance now of younger susceptible individuals. Additionally, Figure 5.4 represents the very same data as Figure 5.3 (A) on the map of the United States. We observe some geographical clustering, even though we are treating each state as a completely isolated population. The states with higher attack rates or, similarly, the states with a higher fraction of vaccine hesitancy, are concentrated mainly in the interior of the country (inner Pacific west, Intermountain, ranging from north (Midwest) to south (inner Southeast)).



**Figure 5.4: US map of attack rates and vaccine hesitancy.** Representation on the US map of the attack rates of every state after the end of the epidemic trajectory proposed in this paper (Panel A), and the fraction of non-vaccinated individuals (Panel B). Some spatial clustering can be appreciated along the country, even though in the simulations all states are completely isolated.

Let us next try to get a deeper understanding of what is happening during our simulated second outbreak. Looking at some particular extreme examples, we can appreciate that the state of Massachusetts (MA), with the lowest vaccine hesitancy (9% of the adult population), had the lowest epidemic size during the first outbreak and also during the second outbreak. On the other hand, Alaska (AK) shows one of the lowest attack rates in the first outbreak, but the highest one in the second outbreak, together with the highest fraction of remaining susceptible at the end of the first outbreak, whereas its hesitancy amounts to 23% of the adult population, way behind the most reluctant states. Interestingly, there are other states with a relatively low hesitancy rate that also show a sizable second outbreak. This is the case of the state of Utah (UT), with a hesitancy of about 15% among the adult population but nevertheless ranking high in the size of the second outbreak. One could hypothesize that these two states should have a similar number of deaths during the second outbreak. But, remarkably,

as we show below, there is indeed more than a simple extrapolation of the correlation between the outbreak size and the number of non-vaccinated/susceptible individuals when it comes to forecasting mortality. The reason is that the age of non-vaccinated and/or remaining susceptible matters, not only because it usually determines behavior (and risk of infection) but also because the infection fatality rate heavily depends on it. As [515] emphasizes, considering transmission through the lens of (age-based) contact patterns is fundamental to understanding which population groups are driving disease transmission. Several reports, at least for the US, point to the fact that transmission dynamics shifted from older adults in the first stages of the pandemic to younger groups later [515, 525, 526]. This is understandable since once the harshest lockdowns were lifted, naturally younger groups are more socially active and thus can act as the main drivers while elders are less active and as the disease awareness exists due to the ongoing epidemic, one should expect that they mix more carefully. Regarding mortality, it has been well documented an increasing risk of suffering severe disease and death for the oldest age groups, especially beyond 65 years old [516, 527, 528, 529, 530, 531].

In Figure 5.5, we show a scatter plot of deaths per million individuals in the second outbreak versus the fraction of non-vaccinated individuals at the end of the first outbreak. We estimate the number of deaths in each age group by applying the corresponding infection fatality rate (IFR) [527], so that:

$$D_a(\infty) = \text{IFR}_a \times R_a(\infty), \quad (5.5)$$

where  $R_a(\infty)$  and  $D_a(\infty)$  are, respectively, the prevalence and the number of deceased individuals at the end of a particular outbreak.

Even though a high correlation coefficient is obtained, its explanatory power is smaller than for the attack rate, which suggests that there are other factors playing a relevant role. Certainly, we can appreciate that higher proportions of deaths tend to occur in those states with higher hesitancy. Bringing back the case of Alaska (AK), and contrary to what could be naively expected, we see that it has been overtaken by several states. Even more striking is the case of Utah (UT), which is in the lower part of the ranking. This clearly reveals that apart from vaccine hesitancy, the age structure is playing a key role in the disease dynamics and COVID-related fatalities [527, 529, 530].

To understand better these interdependencies, we next look at the attack rates during the second outbreak by coarse-graining the 85 age groups resolved in our model into four main relevant categories for the sake of the analysis. In Figure 5.6, we show results for 0-18 (A), 18-45 (B), 45-65 (C), and more than 65-year-old age groups (D). For each one, the attack rates during the second outbreak are computed as  $R_a(\infty)/R(\infty)$ ,

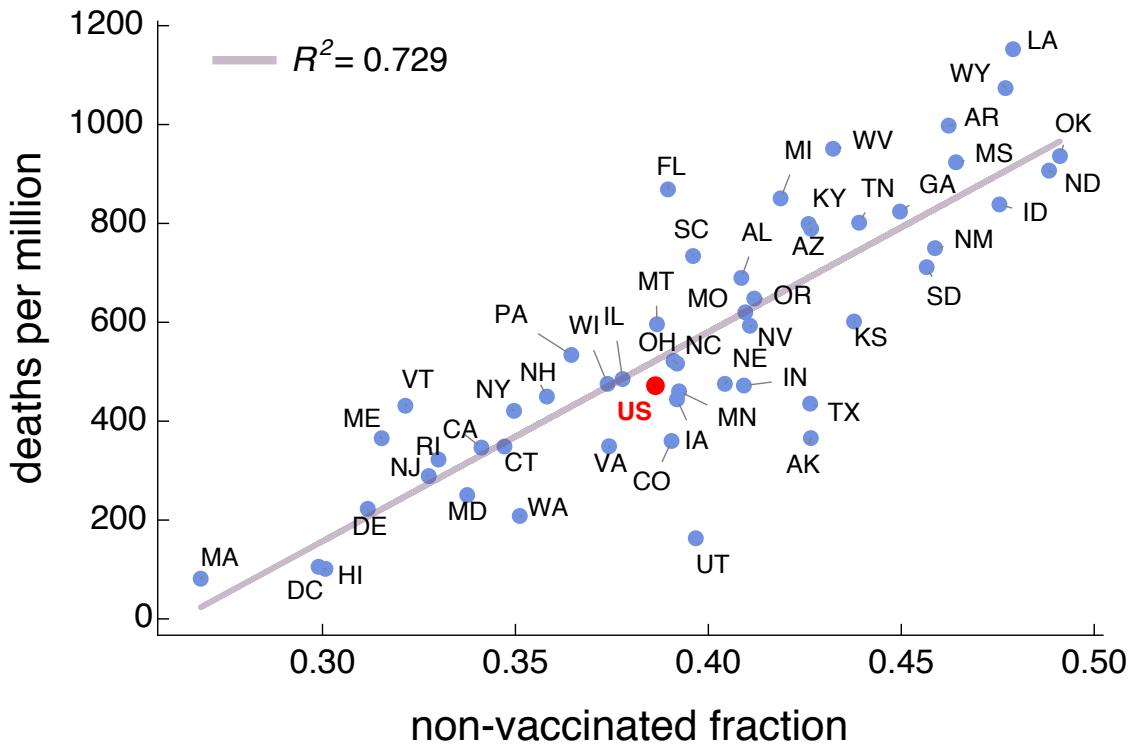
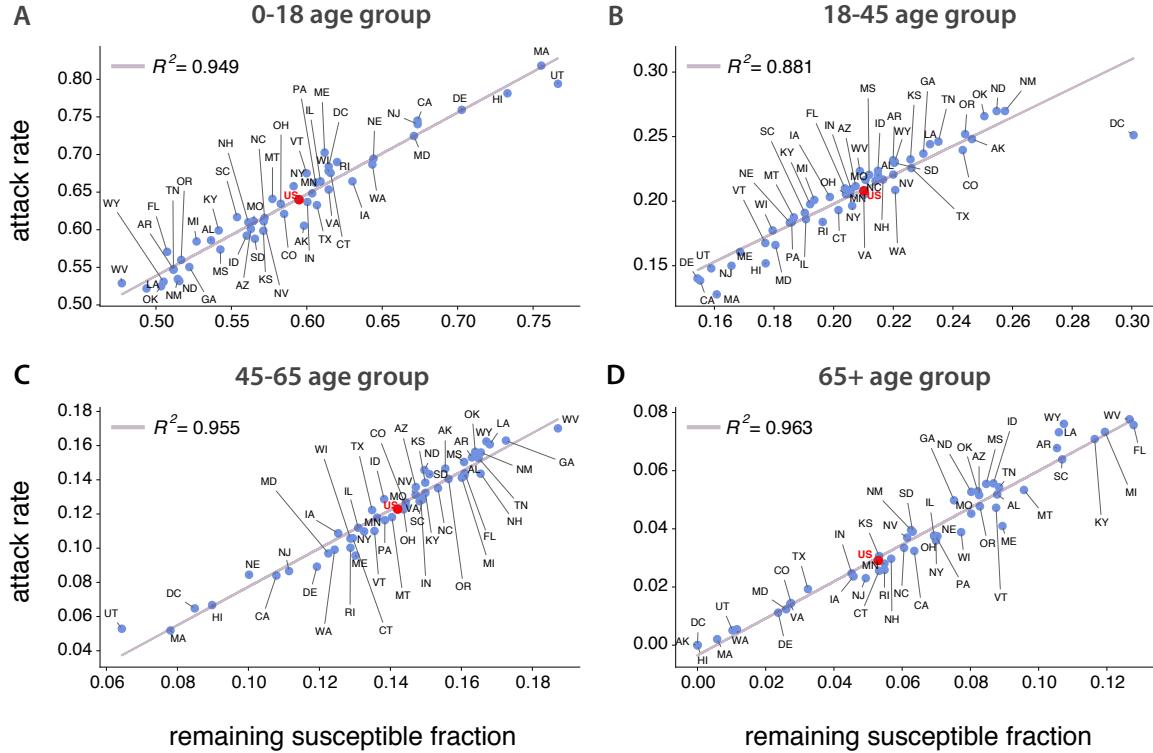


Figure 5.5: **Death scatter plots.** Scatter plot of deaths per million during the second outbreak versus the non-vaccinated fraction at the end of the first outbreak for every US state. Results are also shown for a simulation of the epidemic for the whole country as if it were a single age-structured population (red dot). The model does not include deaths as part of the dynamics, but they can be estimated by applying the infection fatality rate to the final fraction of individuals in the removed compartment for each age class (eq. (5.5)).

while the fractions of remaining susceptible individuals at the end of the first outbreak are computed as  $S_{1a}/S_1$ , where  $R(\infty)$  is the final attack rate, and  $S_1$  is the total fraction of remaining susceptible subjects. Thus, these figures tell us the share of people in each group  $a$  within the susceptible and removed pools.

The results by age groups exhibit a very high correlation for the linear fittings, which indicates the relevance of age structure in the transmission of the disease. For every age group, states with higher hesitancy tend to experience larger epidemic sizes. Regarding the cases mentioned before, namely, Utah (UT) and Alaska (AK), one can see that their fractions of remaining susceptible individuals are large in the youngest age groups and rather small (null for Alaska) in the 65+ age strata. This ultimately explains why these two states undergo large second outbreaks that are not translated into a higher number of deaths. Finally, we also note that the fraction of remaining susceptible individuals is the highest for roughly every state in the two youngest age brackets (around or higher than 50%), which means that the younger age groups will



**Figure 5.6: Attack rate scatter plots by age.** Scatter plot of attack rates during the second outbreak versus the remaining susceptible fraction for every US state. Top-left (A): 0-18 years old group. Top-right (B): 18-45 years old group. Bottom-left (C): 45-65 years old group. Bottom-right (D): over 65 years old group. Results are also shown for a simulation of the epidemic for the whole country as if it were a single age-structured population (red dot). These high correlations show also the relevant role of age structure in disease propagation.

be the driving group of the second outbreak.

At this point in the analysis of results, we can appreciate the expected relevance of vaccine hesitancy but also that of age heterogeneity in order to project the impact of the epidemic spreading on a territory. Now, we look for an estimation of how many deaths could potentially be averted just by reducing the fraction of individuals in the “would not get the COVID-19 vaccine” category by one percentage point. It may occur that for states with an important share of younger population and not very high hesitancy, an extra effort does not pay off. Conversely, in states with an older population and for those with high hesitancy, such an additional increase in the percentage of vaccinated may represent an important benefit. We believe these are important considerations for public health policy-making. Table 5.1 shows the number of averted deaths per million people if vaccine hesitancy is reduced by one percent during the vaccination campaign in every state. A first look at the table would lead us to believe that, overall, states with higher hesitancy will tend to avert more deaths by improving their vaccine

rollout. But if we look for correlations between the total number of averted deaths in both outbreaks and the fraction of non-vaccinated individuals, we obtain a not-so-high Pearson coefficient of  $\rho = 0.61$ , signaling correlations but not quite strongly. We have learned throughout the discussion and related literature review the severe impact that disease has on the eldest groups within a population and especially in this work, the importance of having a low pool of remaining susceptible individuals in the oldest age groups (65+) for having lower deaths rates in secondary outbreaks. If then we correlate this quantity with the total number of averted deaths, we obtain a Pearson coefficient of  $\rho = 0.92$  and thus a very high correlation and greater explanatory power.

State	Averted deaths (1st outbreak)	Averted deaths (2nd outbreak)	State	Averted deaths (1st outbreak)	Averted deaths (2nd outbreak)
AK	5.12 [2.83- 10.57]	16.78 [9.37- 35.14]	MT	12.03 [6.62- 23.86]	34.49 [18.93- 67.57]
AL	8.62 [4.73- 17.13]	39.18 [21.49- 76.94]	NC	8.08 [4.44- 16.11]	34.3 [18.83- 67.7]
AR	10.23 [5.61- 20.29]	42.38 [23.17- 82.64]	ND	10.83 [5.92- 21.52]	35.56 [19.34- 69.23]
AZ	9.66 [5.29- 19.12]	39.74 [21.73- 77.58]	NE	9.75 [5.34- 19.35]	27.7 [15.11- 54.23]
CA	8.79 [4.81- 17.45]	24.96 [13.62- 49.08]	NH	12.1 [6.67- 24.09]	33.04 [18.19- 65.35]
CO	7.66 [4.21- 15.43]	26.57 [14.62- 53.06]	NJ	4.49 [2.47- 8.91]	27.55 [15.08- 54.42]
CT	10.43 [5.72- 20.71]	28.18 [15.43- 55.57]	NM	10.73 [5.9- 21.38]	37.28 [20.45- 73.19]
DC	4.09 [2.22- 8.28]	9.43 [5.16- 20.01]	NV	7.57 [4.16- 15.1]	34.65 [19.04- 68.39]
DE	8.34 [4.59- 16.6]	26.71 [14.71- 52.98]	NY	4.18 [2.29- 8.29]	31.78 [17.35- 62.54]
FL	10.3 [5.62- 20.2]	48.74 [26.54- 94.48]	OH	10.55 [5.79- 20.98]	32.53 [17.83- 63.96]
GA	5.89 [3.24- 11.8]	38.9 [21.38- 76.83]	OK	9.74 [5.34- 19.41]	38.24 [20.93- 74.83]
HI	9.31 [5.09- 18.44]	10.87 [6.06- 22.76]	OR	11.05 [6.07- 21.97]	35.75 [19.61- 70.23]
IA	7.31 [4.0- 14.56]	31.54 [17.25- 62.05]	PA	7.84 [4.3- 15.2]	36.33 [19.88- 71.2]
ID	10.18 [5.59- 20.29]	35.19 [19.29- 68.9]	RI	10.26 [5.63- 20.38]	26.66 [14.59- 52.55]
IL	9.28 [5.08- 18.44]	30.61 [16.73- 60.12]	SC	9.66 [5.32- 19.19]	39.87 [21.91- 78.17]
IN	9.00 [4.95- 18.04]	30.06 [16.51- 59.43]	SD	10.53 [5.79- 21.03]	35.13 [19.24- 68.87]
KS	9.08 [4.99- 18.16]	33.38 [18.29- 65.69]	TN	9.54 [5.24- 19.0]	39.14 [21.47- 76.84]
KY	11.88 [6.52- 23.52]	37.12 [20.32- 72.48]	TX	5.99 [3.29- 12.05]	27.89 [15.3- 55.49]
LA	7.68 [4.22- 15.25]	45.27 [24.79- 88.51]	UT	5.68 [7.38- 26.53]	17.02 [9.31- 34.05]
MA	8.37 [4.59- 16.67]	16.94 [9.29- 33.83]	VA	8.31 [4.57- 16.66]	27.93 [15.37- 55.55]
MD	5.64 [3.1- 11.27]	25.95 [14.27- 51.71]	VT	13.40 [7.38- 26.53]	31.36 [17.21- 61.56]
ME	14.13 [7.77- 27.89]	29.7 [16.29- 58.21]	WA	8.14 [4.48- 16.35]	22.54 [12.4- 45.11]
MI	12.03 [6.6- 23.82]	40.84 [22.34- 79.66]	WI	10.70 [5.88- 21.28]	30.5 [16.71- 59.9]
MN	7.42 [4.08- 14.82]	31.34 [17.18- 61.9]	WV	13.25 [7.28- 26.2]	44.72 [24.5- 87.12]
MO	10.60 [5.81- 21.04]	34.9 [19.09- 68.37]	WY	12.39 [6.82- 24.64]	39.78 [21.81- 77.73]
MS	8.78 [4.82- 17.48]	38.48 [21.09- 75.36]	US	7.95 [4.36- 15.85]	32.15 [17.62- 63.43]

Table 5.1: **Averted deaths.** Average number of averted deaths per million (95% CI in brackets), separately in the 1st and 2nd outbreak, due to reducing vaccine hesitation by one percent point. Results shown for every state and for the whole country (US).

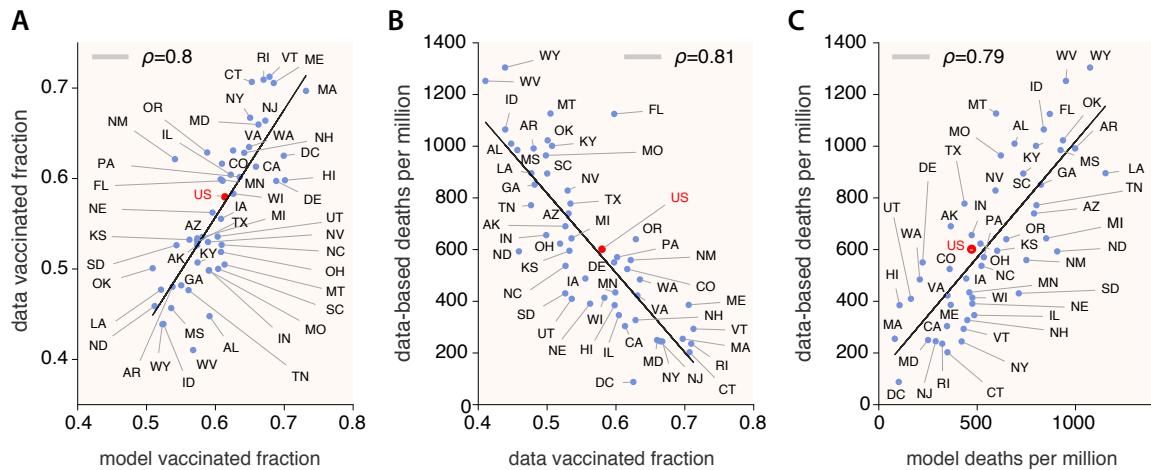
The results shown in this work are based on a standard and sound epidemiological approach based on compartmental ODE modeling with a heavily-based data-driven input for several aspects: population age structure, mixing patterns by age, and vaccine uptake attitudes. The proposed scenario on which the model is run, however, does not map exactly any real situation experienced in the US, and thus our aim was not to reproduce or forecast accurately realistic COVID-19 trajectories. The

particular outcomes brought about should be regarded as what-if scenarios or hypothetical outcomes of what to expect overall given the premises hold. If, for instance, vaccine acceptance fractions should differ, as well as the transmissibility of a given virus strain, the specific figures could change dramatically, but not so the underlying conclusions. Additionally, from the beginning, the model confection and the devised scenario were not conceived to realistically simulate or reproduce the myriad of complexities and heterogeneities involved in COVID-19 spreading country-wise. However, we can still check to a certain point how the model performed with respect to reality and thus gain confidence in the conclusions derived from it or rather discard the approach as insufficient or unsatisfactory.

As explained above, we proposed a hypothetical situation in which the epidemic spreading was ongoing but under rather mild transmission conditions due to disease awareness and general restrictions, while at the same time, a mass vaccination campaign was deployed. After vaccination was completed and the epidemic wave was rather halted, we simulated that societies turned back to normal from very low daily incidence, but the virus was still there and new secondary outbreaks emerged.

In reality, vaccine rollout in the US took off at the beginning of 2021 among concerns about vaccine hesitancy, a few months before we posed our research questions and designed the aforementioned hypothetical scenario. By mid-February, the highest epidemic wave experienced by the country until that moment was ending and reaching a plateau of rather low incidence. Overall, the situation stayed under control except for a very slight peak around mid-April and then a relaxed decrease until reaching the lowest national incidence levels by June 2021 since the beginning of the pandemic. Thus, we could draw some parallelisms here with the first wave in our experiment: a rather controlled and decaying progression with vaccination going on. Then, restrictions were overall lifted up and then a new wave started to build up by the beginning of July 2021. This wave was mainly driven by the at the time designated new variant of concern, the Delta variant, peaking in the first days of September and reaching a higher plateau than before by early November 2021, immediately followed by the huge Omicron wave. Given the history of the spreading dynamics in the US, we consider that our hypothesized secondary outbreaks in every state, describing a rather unmitigated scenario with a more aggressive strain, could match reasonably well the Delta wave that took place across the described period. Therefore, it is informative to check how the model output relates to the real impact of the epidemic during the aforementioned period. In Figure 5.7, we looked at the following pairs of observables to check for correlations: real data [532] vs. model/survey vaccination fractions (Figure 5.7 A), data-based deaths [533] vs. real vaccination fractions (5.7 B), and data-based deaths

vs. model deaths ( $5.7^{\circ}\text{C}$ ). We can see that all confronted observables show a high correlation. First, in panel A, comparing data and model vaccination fractions we obtain a Pearson correlation coefficient of  $\rho = 0.8$ . Overall, we can say that the used surveys on attitudes towards vaccination were reality meaningful and thus strengthened subsequent model results. Second, in panel B we obtain a value of  $\rho = 0.81$  when correlating what happened in reality regarding deaths and vaccination. As expected, the higher the vaccination fraction in a state, the lower the deaths that took place due to COVID-19. When comparing in the paper the model deaths per million with the non-vaccinated fraction (for the matter, this is equivalent to comparing it to the vaccinated fraction), we obtained an  $R^2 = 0.729$  and therefore a Pearson coefficient of  $\rho = 0.85$ , a very high signal of correlation. We can then see that our model projections captured the real trend quite satisfactorily and this emphasizes the dominant role of the vaccine in mitigating the impact on the population. Finally, in panel C we obtained a Pearson coefficient  $\rho = 0.79$  when comparing this time the real data-based estimation of deaths and the model-based estimation of deaths. Thus we find again a high correlation figure between our hypothesized scenario of secondary outbreaks and the Delta wave.



**Figure 5.7: Comparison of model output and real data for the Delta wave in the US.** Correlation analysis (Pearson correlation coefficient) for data and model observables. Left (A): Data vaccinated fraction until 31/10/2021 versus model/survey vaccinated fraction. Center (B): Data-based deaths per million versus real vaccinated fraction. Right (C): Data-based deaths vs. model deaths per million. High correlations are obtained between the model output and real data.

## 5.4 Conclusions

In this work, we have explored SARS-CoV-2 transmission dynamics in a population that is partially vaccinated and is seeded again with the virus when restrictions are fully lifted. We explored, in particular, to what extent vaccine hesitancy may still drive sizable outbreaks in a context where a more transmissible SARS-CoV-2 variant of concern is dominant. We used data from vaccination acceptance surveys, together with up-to-date age-distributed populations and contact matrices in the US to inform an age-structured SIR model.

Our results show a clear correlation between the size of experienced outbreaks, once all kinds of measures are lifted, and the fraction of vaccine hesitancy or, similarly, the fraction of remaining susceptible individuals at the onset of a second outbreak. Higher vaccine hesitancy ratios expose the population to larger outbreaks and, inversely, higher vaccine acceptance ratios can mitigate the impact to the point of negligible secondary waves due to the immunity of the population.

We have also inspected in detail the role of the age structure of the population in both the attack rate and the mortality of secondary outbreaks. Our findings reveal that the prevalence is highly correlated with the fraction of remaining susceptible individuals by age class, with the youngest contributing the most to the attack rate. It is however not immediate to project such a correlation to the expected number of deaths, as here too age plays a role, though in the opposite direction, e.g., the younger the population, the lower the mortality.

Lastly, we estimated the number of potentially averted deaths during the course of the simulated epidemic if the number of people reluctant to vaccine uptake was reduced by one percentage point. Results show again the relevance of age structure in transmission since not all the states with higher hesitancy rank highest in averting deaths. It is very relevant to the fraction of hesitant individuals in the older groups.

To round up the analysis, we investigated how the model fared when comparing real-life data. Even though the devised experiment here was not intended to accurately replicate or forecast real COVID-19 trajectories, the data-driven approach and sound modeling offered very high correlations when comparing survey/model vaccination against real vaccination, and model death estimation against real data-based deaths during the Delta wave in the US.

We acknowledge that our model has several limitations. One is at the core of its compartmental structure, not including a more detailed progression of the natural history of the disease, which might affect our estimation of deaths, and does not consider hospitalizations of any kind. The vaccination campaign could be implemented

in a more realistic way owing to each state's idiosyncrasy but, more importantly, vaccines seem to be not sterilizing and thus not fully preventing transmission, and, on top of that, immunity decays with time. These facts do not affect the overall dynamics explored in this paper but should be incorporated to provide reliable estimations on the exact amount of expected infections or deaths. Additionally, the behavioral responses are not completely accounted for. All these factors open important challenges for future work. Finally, our analysis represents hypothetical scenarios that can unveil mechanisms and correlations rather than producing accurate forecasts. Yet, these scenarios can be helpful for policy-making and providing quantitative arguments to the public debate about the role of vaccination in mitigating and containing disease propagation.

To conclude, the most important implications of the results reported here include: (i) data on vaccination by age is important to accurately capture the evolution of mortality in secondary waves; (ii) surveys on vaccination attitude are a valuable proxy to estimate the hesitancy of the population; (iii) allocation of additional resources is more important in states with relatively high hesitancy rates but especially in states where the remaining susceptible population is older; (iv) reintroduction of restrictions could be needed in states with very high attack rates to reduce pressure over healthcare systems; and (v) incentives to vaccination directed towards the younger population will reduce the prevalence, while they will reduce the number of deaths if they focus on the older generations.



# Chapter 6

## Epidemic spreading in contact networks coupled to a threshold-based opinion dynamics on vaccine uptake

*Upon their heads  
He places crowns  
To witness entire towns  
The boy blesses whispers into words  
In the painted valleys they await rain.*

— Isis - Threshold of Transformation

### 6.1 Introduction

Epidemics and human behavior are inextricably intertwined. The history of humanity has indeed been frequently plagued by notable episodes of infectious diseases, with devastating impacts on individual health and entire human populations [534, 535]. Progress, understood as economic growth and technological development, has contributed to mitigating impacts, even to the point of eradicating diseases [536], but also, ironically, to facilitating their spread. Human behavior has indeed played a crucial role in generating, driving, and shaping the spread of infectious diseases through processes such as industrial farming and animal exploitation [537, 538, 539], urbanization [540, 541, 542, 543], and globalization [544, 545, 546], among others. These behaviors are reflected in contact [402, 233, 547] and mobility patterns [239, 548], driven by the way our societies function. Another fundamental aspect of human behavior is the reaction to the disease itself [549, 550, 551, 552]. Disease awareness often prompts individuals to adopt self-protective measures like general prophylaxis, social

distancing, quarantines, or getting vaccinated. Political and health authorities can also implement these strategies [24]. These measures can change the course of the epidemic, resulting in a feedback loop between epidemic and human dynamics [553, 554, 261, 52]. However, the human reaction can also push in the direction of vaccine hesitancy [555, 556, 509, 557], non-compliance with enforced non-pharmaceutical interventions [558, 559, 560, 561, 562], or simply carelessness and de-prioritization due to low perceived risk of the disease, or the absence of recent infections in a population [536]. Despite the undeniable centrality of human behavior in the spread of epidemics, modeling it remains a formidable challenge [563], mainly due to its complex nature and the lack of real-world, quantitative data on behavioral changes in populations affected by epidemic outbreaks [243, 244].

Focusing on vaccination as a paramount protective measure against infectious diseases, vaccination strategies have long been a topic within the epidemic modeling literature [53, 564, 536]. Vaccination modeling has been tackled either through non-behavioral or behavioral approaches. The former involves conceiving vaccination as a process that *automatically* acts on the population, such as continuous vaccination [565, 566], pulse vaccination [567, 568], age-specific vaccination [569, 570, 571], or, in networked populations, targeted vaccination relying on centrality metrics and vaccination without global knowledge of the network [536]. However, these strategies usually assume perfect compliance and totally disregard the human response- both individually and collectively- that can be triggered throughout unfolding outbreaks. On the other hand, approaches exist that try to incorporate more genuine behavioral dynamics under various schemes both in well-mixed [243, 244] and in complex networks [572, 573, 536], including also temporal [574, 573] and multilayer approaches [573, 575]. One of the standard frameworks to integrate human behavior into epidemiological models is game theory through the so-called vaccination games [272, 273]. Here, individuals decide whether to vaccinate or not depending on the assessment of the cost incurred by the options they face. This assessment although traditionally assumed to be perfectly rational and mostly isolated from the social environment, has been progressively relaxed to incorporate more realistic features that affect the human decision-making process [573, 576, 272, 273]. Another framework for integrating epidemic models and vaccination-related behavior, significantly less exploited than vaccination games, is that of opinion dynamics [36]. Here, individuals possess an opinion on a topic (such as vaccine uptake), which can change or be influenced through different mechanisms by peer interaction, as well as social and mass media exposure. Some references include [577, 578, 274, 275, 276, 579, 580], where rich complex behavior involving sudden transitions, bistability, or network segregation has been shown.

Among the most notable models of opinion dynamics is the seminal Watts-Granovetter threshold model [581, 582, 583], originally aimed to model phenomena such as riots or other social movements, was later generalized to the study of information cascades, stock market crashes, or cascading failures in infrastructure networks. These collective phenomena are typically referred to as *complex contagions* [584]. Unlike simple contagions, where transmission of infection occurs at a rate between individuals and a single infected neighbor is always sufficient to expose a susceptible node, in complex contagions the exposure of an individual is conditional on the decision of a fraction of its peers [584, 585, 586]. Thus, while biological processes such as epidemics may be seen as simple contagion processes, human behavior dynamics related to opinions and decisions on matters relative to self-protection and vaccine uptake might constitute complex contagions. Recent studies have indeed begun to explore the interaction of simple and complex contagions within the context of epidemics [587, 300, 588, 589, 590, 591, 579, 592]. Surveys on vaccination attitudes in the US suggest these attitudes can depend on the vaccination status of a certain portion of one’s social circle, aligning with the concept of a complex contagion phenomenon [514].

In this work, we draw on these studies to investigate the interplay between the spread of an epidemic on networks and the dynamics of opinion on vaccine uptake. We employ the standard SIR model, which we couple to a dynamic vaccination campaign. This campaign’s progression is, in turn, influenced by a threshold-based opinion dynamics process, effectively integrating models of simple and complex contagion. We obtain results for homogeneous networks (Erdős-Rényi) and heterogeneous networks (Barabási-Albert), characterizing the system’s prevalence in relation to the opinion dynamics threshold parameter (initially assumed homogeneous), the initial fraction of pro-vaccination individuals, and the vaccination rate. We then assess the effects of zealots, those with strong anti-vaccine attitudes, on the system’s prevalence and vaccination coverage. Lastly, we use vaccination surveys from the US to analyze the epidemiological consequences in scenarios featuring a heterogeneous spectrum of attitudes toward vaccination.

## 6.2 Material and methods

### Epidemic dynamics

The epidemiological model utilized is the standard and well-known SIR model running on a networked population of agents<sup>1</sup>. We consider discrete-time dynamics with unitary

---

<sup>1</sup>We will refer to the constituents of the network as either *agents* or *individuals* equivalently.

time step  $\Delta t = 1$ . Then, at every time step, susceptible agents can experience a pairwise contagion interaction and become infected with probability  $\beta\Delta t$ :

$$\mathcal{S} + \mathcal{I} \xrightarrow{\beta} \mathcal{I} + \mathcal{I}. \quad (6.1)$$

Note that since every agent can interact with any other of their first neighbors in the network, the total probability of becoming infected is  $P(\mathcal{S} \rightarrow \mathcal{I}) = 1 - (1 - \beta\Delta t)^{I_a(t)}$ , where  $I_a(t)$  denotes the number of infected neighbors of agent  $a$  at time  $t$ , and  $\beta$  is the disease's transmission rate. Unlike mass-action or well-mixed models, where a correction based on the total population  $N$  (typically as  $\beta/N$ ) is introduced, the quantity  $I_a(t)$  is bounded by the agent's degree  $k_a$ .

Here, all the dynamical processes occur on the same network, which is generated according to either the Erdős-Rényi (ER) or the Barabási-Albert (BA) model. Given that these models are characterized by markedly different degree distributions, we anticipate disparate epidemiological outcomes in each.

Lastly, recovery or removal from the dynamics is, as usual, a spontaneous transition modeled as a Markovian process:

$$\mathcal{I} \xrightarrow{\mu} \mathcal{R}. \quad (6.2)$$

Then, for every infected individual, the probability of decaying to the recovered/removed state is  $P(\mathcal{I} \rightarrow \mathcal{R}) = \mu\Delta t$  at every time step.

## Vaccination campaign and opinion dynamics

The SIR process on networks runs simultaneously with two other intimately related dynamical processes: the vaccination campaign and the opinion dynamics on vaccination. Vaccination by itself is modeled as a spontaneous Markovian process proceeding continuously as long as there are eligible agents to be vaccinated and without any preferential target. The catch is that in order to be eligible for vaccine uptake, the agent must be in a convinced or active state. Here is where the opinion formation process enters.

To model the opinion dynamics we consider the Watts-Granovetter threshold model of social contagion [582]. In this model, agents can be classified into two states: inactive and active. Within this context, we will refer to inactive agents as hesitant, being them denoted by  $\mathcal{H}$ , and will retain the active nomenclature to refer to pro-active agents regarding vaccine uptake, and will be denoted by  $\mathcal{A}$ . As in the original model, the state  $\mathcal{A}$  is irreversible, and therefore the dynamics is ruled by the transition:

$$\mathcal{H} \rightarrow \mathcal{A}. \quad (6.3)$$

Mapping the status  $\mathcal{H}$  to 0 and  $\mathcal{A}$  to 1, and by designating by  $o_a$  to agent's  $a$  opinion, the updating rule in the original proceeds in the following way:

$$o_a(t+1) = \begin{cases} 1 & \text{if } \sum_{j \in \Omega_a} \frac{o_j(t)}{k_a} \geq \theta, \\ 0 & \text{otherwise.} \end{cases} \quad (6.4)$$

Here, the summation extends to  $\Omega_a$ , which is the neighborhood of agent  $a$ ,  $o_j$  is the status of the neighbor  $j$ , and  $k_a$  is the degree or number of neighbors of agent  $a$ ; finally,  $\theta$  is the activation threshold ( $0 \leq \theta \leq 1$ ), which in this case is assumed to be homogeneous across the system. Due to its lack of symmetry, if the focal agent status is already at 1, nothing happens. Thus, as aforementioned, once the agents adopt the state designed by 1, it is conserved until the end of the dynamical process.

For the present work, however, we do not follow the rule described in equation 6.4 but a variation. Rather than adopting a proactive attitude towards vaccination based on the vaccine views of peers, agents will make the decision on vaccine uptake based on the neighbors' vaccination status (see Section 6.2 for further justification). Therefore, for an agent  $a$  hesitant to vaccination and thus with status  $o_a = 0$  we set the updating as:

$$o_a(t+1) = \begin{cases} 1 & \text{if } \frac{V_{\Omega_a}(t)}{k_a} \geq \theta, \\ 0 & \text{otherwise.} \end{cases} \quad (6.5)$$

Where  $V_{\Omega_a}(t)$  is the total number of vaccinated agents in the neighborhood of agent  $a$ . This change directly connects the success of the ongoing vaccination campaign to the opinion on vaccination, setting a co-evolving feedback loop between vaccination and opinion dynamics.

Finally, then, every convinced or active susceptible agent  $(\mathcal{S}, \mathcal{A})$  will be vaccinated:

$$\mathcal{S} \xrightarrow{\alpha} \mathcal{V}, \quad (6.6)$$

with probability  $P(\mathcal{S} \rightarrow \mathcal{V}) = \alpha \Delta t$ .

The system's behavior will be examined across varying control parameters such as the activation threshold  $\theta$ , the initial fraction of active agents  $n_A(0)$ , and the vaccination rate  $\alpha$ . The parameter space related to the opinion dynamics process,  $(\theta, n_A(0))$  is scrutinized in finer detail, and we select certain vaccination scenarios based on the rate  $\alpha$ .

To gain insight into these vaccination rates, the lowest one we investigate here is  $\alpha = 0.001$ , which amounts to vaccinating 0.1% of the eligible population. For completeness, we also consider the extreme and unrealistic case of  $\alpha = 1$ . For context, the highest daily vaccination record in the United States, according to [593], was 3.15 million doses on April 11, 2021, when the total population that had received two doses was

approximately 73 million, as reported by [594]. Given that the total US population is about 350 million, this represents a rate of roughly 0.012 on that record day (1.2% of the population). Nonetheless, actual vaccination rates are much more variable as compared to our simplified model of a constant rate. From a global perspective, during the COVID-19 pandemic, despite the urgent vaccination deployment efforts and increased disease awareness, the daily percentage of the population receiving a COVID-19 vaccine dose rarely exceeded 1%, equivalent to  $\alpha = 0.01$ , with occasional exceptions in countries like China and the UK. As reported in [595], these countries experienced peaks of 1.57% and 1.41% of their populations getting vaccinated in a single day, respectively. These figures are instrumental in contextualizing our vaccination rate scenarios.

## Vaccination surveys

In the final stage of this work, we inform our model with activation thresholds derived from real data collected from surveys. This dataset is the same that was utilized in Chapter 5; however, in this instance, we exploit the heterogeneity in attitudes towards the vaccine. To ensure the completeness of this discussion, we recall that the data originates from The COVID States Project (<https://covidstates.org>), specifically from the surveys in Report #43: COVID-19 vaccine rates and attitudes among Americans [514]. These surveys offer insights into vaccination acceptance and hesitancy, segmented by age at the state level. In this particular case, while we do not account for age within our model, it remains a critical element for consideration in future iterations of this research.

The surveys identify several degrees of vaccine predisposition, categorizing individuals as “already vaccinated”, inclined to be vaccinated “as soon as possible”, “after at least some people I know”, “after most people I know”, and those who “would not get the COVID-19 vaccine”. It is from this threshold-based perspective and its heterogeneous distribution—varying according to others’ vaccination statuses—that we draw inspiration for our coupled epidemic-vaccination-opinion dynamics model.

For the subsequent analysis in Section 6.3, we translate these categories into quantifiable activation thresholds, denoted by  $\theta$ . Individuals in the “already vaccinated” category are exempted from further opinion dynamics, effectively setting their activation threshold at zero. The same is presumed for individuals categorized as “as soon as possible”. We posit that these individuals are already convinced and merely awaiting their opportunity to be vaccinated; hence, we assign them a  $\theta = 0$ . For those who would get vaccinated “after at least someone I know”, we interpret this as requiring one known individual to be vaccinated, leading to a threshold of  $\theta = 1/k_a$ , where  $k_a$  is the agent’s number of first neighbors. Individuals in the “after most people

I know” group are assigned a simple majority criterion, corresponding to  $\theta = 0.5$ . It could be argued that this approach is somewhat lenient with the constraint, considering that the majority criterion might vary on an individual basis. Lastly, for the “would not get the COVID-19 vaccine” category, we assign a threshold of  $\theta = 1^+$ , signifying that their decision remains unchanged irrespective of the vaccination status of their entire neighborhood, indicative of anti-vaccine sentiment. Table 6.1 presents a summary of the proposed mappings.

survey category	$\theta$
<i>already vaccinated</i>	0
<i>as soon as possible</i>	0
<i>after at least someone I know</i>	$1/k_a$
<i>after most people I know</i>	0.5
<i>would not get the [COVID-19] vaccine</i>	$1^+$

Table 6.1: Mapping to inform the agents’ activation thresholds in our model based on the vaccination attitude categories as given by survey [514].

## US states’ contact networks

Although constructing realistic contact networks for the US states is beyond the scope of this work, we strive to introduce a degree of realism by building the ER and BA networks with an average degree  $\langle k \rangle$  estimated from synthetic contact matrices inferred by Mistry et al. [360]. To this end, we compile the age-structured contact matrices and population distributions by age for each state and compute the average degree  $\langle k \rangle$  as follows:

$$\langle k \rangle = \frac{1}{N} \sum_i N_i \sum_j M_{ij}, \quad (6.7)$$

where  $M_{ij}$  is the contact matrix, representing the average number of contacts per individual between age groups  $i$  and  $j$ ,  $N_i$  is the population in age group  $i$ , and  $N$  is the total population of the system.

The computed values of  $\langle k \rangle$  for every state are then used to construct the networks accordingly. For the ER networks, the connection probability is determined by  $p = \langle k \rangle / (N - 1)$ . For the BA networks, the parameter  $m$ , representing the number of edges of newly added nodes, is set as  $m = \langle k \rangle / 2$ .

However, it must be noted that the computed  $\langle k \rangle$  for all the US states fall within a very narrow range, approximately between 11.7 and 13.2. Consequently, within this level of approximation, we do not anticipate significant differences in outcomes attributable to network characteristics (within the same model).

## Simulations

Our results are obtained through extensive discrete-time Monte Carlo simulations. For the sake of clarification, we describe the algorithm followed to implement the coupled dynamics described above:

1. A network with  $N$  agents is sampled following either the Erdős-Rényi or Barabási-Albert generative algorithms [427]. An infected agent is assigned initially, as well as the initial fraction  $n_A(0)$  of pro-active agents. In both network settings, we assign the infected seed to the highest degree agent.
2. At every time step  $t$ , every agent  $i$  in the contact network is visited, and we apply the threshold model updating rule (Eq. 6.5). In practice, we keep a list of hesitant/inactive and susceptible agents since they are the only type of agents that could change their opinion status. Opinion updating follows a parallel scheme.
3. During the same time step, the transitions related to the SIR+V model occur. Susceptible and proactive individuals who were not vaccinated undertake a Bernoulli trial with each infected neighbor. Effective updating of health statuses takes place following a parallel scheme.
4. If the population of infected individuals drops to zero, the process terminates, otherwise, we move to the next time step  $t \rightarrow t + 1$  (and back to item 2).

In all the experiments performed, each result for a given set of control parameters has been derived by averaging over 25 network realizations and 25 dynamical process iterations for each network.

## 6.3 Results and discussion

### Homogeneous thresholds

First, we investigate the most basic setup of a homogeneous system where every agent is endorsed with the same activation threshold  $\theta$  for the opinion change on vaccination. The system size is  $N = 2 \times 10^4$ , and the disease parameters, which stay fixed for the whole analysis, are  $\beta = 0.03$  and  $\mu = 0.2$ . The average degree of both network models used, Erdős-Rényi (ER) and Barabási-Albert (BA), is kept fixed too at  $\langle k \rangle = 10$ .

The main observables to analyze are the total prevalence  $R(\infty)$ , vaccination coverage (VC)  $V(\infty)$ , and number of active agents  $N_A(\infty)$  at the absorbing state. By normalizing these quantities by the system's size  $N$ , we represent them as  $r(\infty) =$

$R(\infty)/N$ ,  $v(\infty) = V(\infty)/N$ , and  $n_A(\infty) = N_A(\infty)/N$ , respectively. We examine these quantities across varying control parameters such as the activation threshold  $\theta$ , the initial fraction of active agents  $n_A(0)$ , and the vaccination rate  $\alpha$ .

### Homogeneous networks

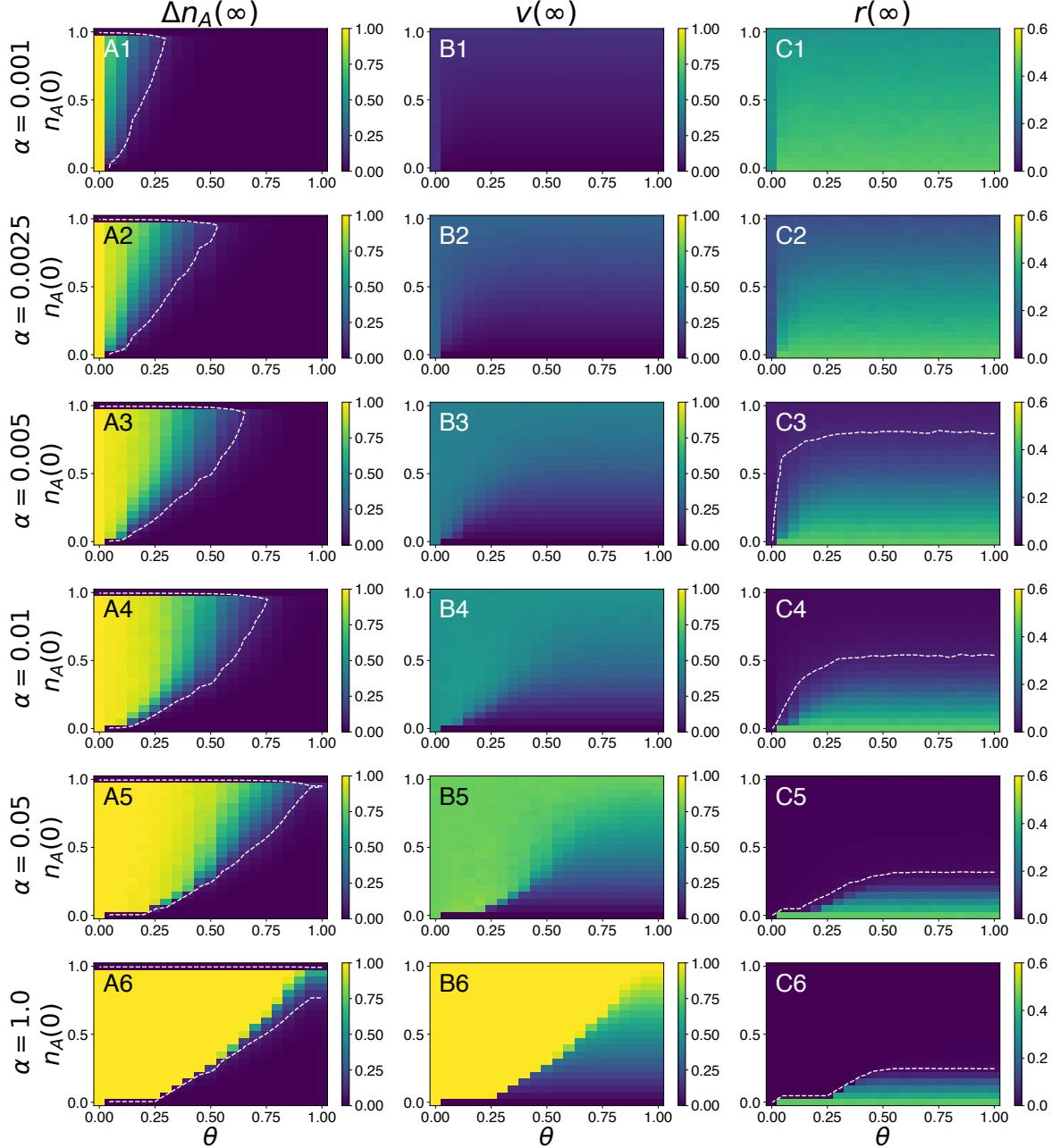
We run the coupled epidemic-opinion dynamics on Erdős-Rényi networks as specified in Section 6.2. The multi-panel in Figure 6.1 depicts the outcome at the absorbing state of the main observables for varying tuples of  $(\theta, n_A(0), \alpha)$  values. Note that since the heatmaps of  $n_A(\infty)$  and  $v(\infty)$  look very similar, due to being intimately related dynamical processes, we look at  $n_A(\infty)$  differently, through the quantity defined as  $\Delta n_A(\infty) \equiv [n_A(\infty) - n_A(0)]/[1 - n_A(0)]$ , which can be interpreted as a measure of the normalized final size of the cascade triggered given the initial condition  $n_A(0)$ . In other words,  $\Delta n_A(\infty)$  informs us about the change in vaccine support relative to the initial support  $n_A(0)$ .

Although not presented here, the trivial scenario with  $\alpha = 0$  triggers no cascading behavior, and subsequently, no vaccination at all (since  $v(0) = 0$ ). Therefore the prevalence is always the maximum expected according to the given disease and network characteristics. In this case, for the stochastic simulations on the ER network, we obtain an average value  $r_{ER}^{\max} \equiv r_{ER}(\infty) = 0.454$ , with  $[0.450, 0.460]$  as the 95% confidence interval (CI). Indeed, for the limit case  $n_A(0) \rightarrow 0$  and  $\theta > 0$ , with  $\theta > n_A(0)$ , vaccination is trivially null for any  $\alpha$  and the system's prevalence approaches the case with  $\alpha = 0$ ,  $r_{ER}(\infty) \rightarrow r_{ER}^{\max}(\infty)$ . As a further reference, given the values being used as epidemiological parameters,  $\beta = 0.03$  for the transmission rate and  $\mu = 0.2$  for the removal rate, and the social interaction term  $\langle k \rangle = 10$ , under a homogeneous mixing approximation we would expect  $R_0 = \beta \langle k \rangle / \mu = 1.5$ , and thus  $r_{hom}(\infty) = 0.58$  as computed from the classical SIR solution given by the transcendental equation  $r(\infty) = 1 - e^{-R_0 r(\infty)}$ . Despite ER networks being regarded as homogeneous systems that can resemble more a standard well-mixing approach than, for instance, scale-free networks where the degree distribution has a power-law form, the relatively low  $\langle k \rangle$  results in a significant contact saturation that leads to a reduced epidemiological impact<sup>2</sup>.

For the lowest proposed vaccination rate scenario,  $\alpha = 0.001$ , we observe that for the most part of the  $(\theta, n_A(0))$ -space, there is no growing support for vaccine adoption, that is,  $\Delta n_A(\infty) = 0$  (Figure 6.1 A1). We refer to the space where  $\Delta n_A(\infty) = 0$  to the *impassive* region. Consequently, vaccination coverage is extremely low (Figure 6.1 B1), and sizeable outbreaks emerge, ranging roughly between  $r(\infty) \in (0.3, 0.46)$

---

<sup>2</sup>It can be shown that as  $\langle k \rangle$  grows (while preserving the product  $\beta \langle k \rangle$ ), the simulation outcomes approach the homogeneous mixing assumption, and then  $r_{ER}(\infty) \rightarrow r_{hom}(\infty)$ .



**Figure 6.1: Homogeneous thresholds in Erdős-Rényi networks.** Results for normalized activation change  $\Delta n_A(\infty)$ , vaccination coverage  $v(\infty)$ , and prevalence  $r(\infty)$ . In every panel, outcomes are explored in  $(\theta, n_A(0))$  space. Every row shows the solution for a different vaccination rate  $\alpha$ . Every point in the diagrams amounts to 25 network realizations times 25 dynamical realizations. White dashed line contours mark an approximate (due to finite size effects) separation between  $\Delta n_A(\infty) = 0$  or  $r(\infty) = 0$  and non-null values of the respective observables.

(Figure 6.1 C1). The region of the parameter space where adherence to vaccine uptake is the maximum possible,  $\theta = 0$  with  $n_A(0) \in (0, 1)$ , shows a slightly larger vaccination coverage and a lower prevalence. In any case, all the parameter space is characterized by an endemic solution ( $r(\infty) \neq 0$ ).

As the vaccination rate  $\alpha$  increases, there is a growing region of  $(\theta, n_A(0))$ , where hesitancy recedes and fades away, and consequently vaccine uptake emerges and the disease prevalence decreases. For  $\alpha = 0.0025$ , the maximum VC attained is at  $v(\infty) = 0.324$  [0.318, 0.327] (Figure 6.1 B2), and the corresponding prevalence is reduced to  $r(\infty) = 0.145$  [0.140, 0.149] (Figure 6.1 C2). The impact has been notably reduced in some regions of the parameter space, but the vaccination rates are still slow enough to avoid the emergence of epidemic outbreaks. Doubling  $\alpha$ ,  $\alpha = 0.05$ , clearly, the solution landscape changes. We attain a vaccination coverage around  $v(\infty) = 0.425$  [0.417, 0.432] (Figure 6.1 B3), which provokes the beginning of an emergent disease-free region (Figure 6.1 C3). Still increasing  $\alpha$  values translates into successful positive feedback between the vaccination adoption opinion dynamics and the vaccination campaign, which in turn increases the parameter space region where a disease-free solution reigns and, consequently, the endemic solution is bounded to domains with high adoption threshold  $\theta$  and low to very low initial support  $n_A(0)$  (Figure 6.1 C4, C5, and C6, progressively).

Overall, for a fixed value of  $\theta$ , moving from  $n_A(0) = 0$  to  $n_A(0) = 1$ , means that hesitancy loses ground, vaccination coverage increases and disease prevalence tends to zero. In high enough  $\alpha$  scenarios, for low  $\theta$ , the transitions occur in a more abrupt way, whereas for medium to high  $\theta$ , the transitions are smoother. Varying  $\theta$  with fixed  $n_A(0)$  does not have a notable effect except where a phase separation exists. If the initial vaccine acceptance  $n_A(0)$  is high enough, even  $\theta \rightarrow 1$  has no effect on vaccination and prevalence. A giant component of vaccinated agents can emerge fast enough to avoid a sizeable outbreak. On the other hand, as  $n_A(0)$  decreases (for a fixed  $\alpha$ ), critical values of  $\theta$  appear that, if surpassed, can bring the system from a disease-free phase to an endemic phase. This critical threshold  $\theta$ , however, can be pushed toward higher values if the vaccination campaign proceeds at faster rates.

The extreme and unrealistic case of  $\alpha = 1$ , depicts an abrupt transition when looking at the behavior of  $\Delta n_A(\infty)$  in  $(\theta, n_A(0))$ -space (Figure 6.1 A6). A rather marked boundary separates the region where the maximum size of cascading behavior occurs and the region where factoring in the initial support  $n_A(0)$ , there is no further change induced. This *impassive* region precludes the system from reaching total VC there. However, for the largest part of it (roughly when  $n_A(0) > 0.25$ ), VC, as being propelled by a high vaccination rate, is large enough to drive the system to the free-disease phase. Finally, it is noteworthy that the solution landscape for  $r(\infty)$  is very similar across the vaccination scenarios with  $\alpha = 0.05$  and  $\alpha = 1$  (Figure 6.1 C5 and C6, respectively). This observation suggests the existence of diminishing returns to the vaccination process as determined by  $\alpha$ , possibly constrained by the underlying

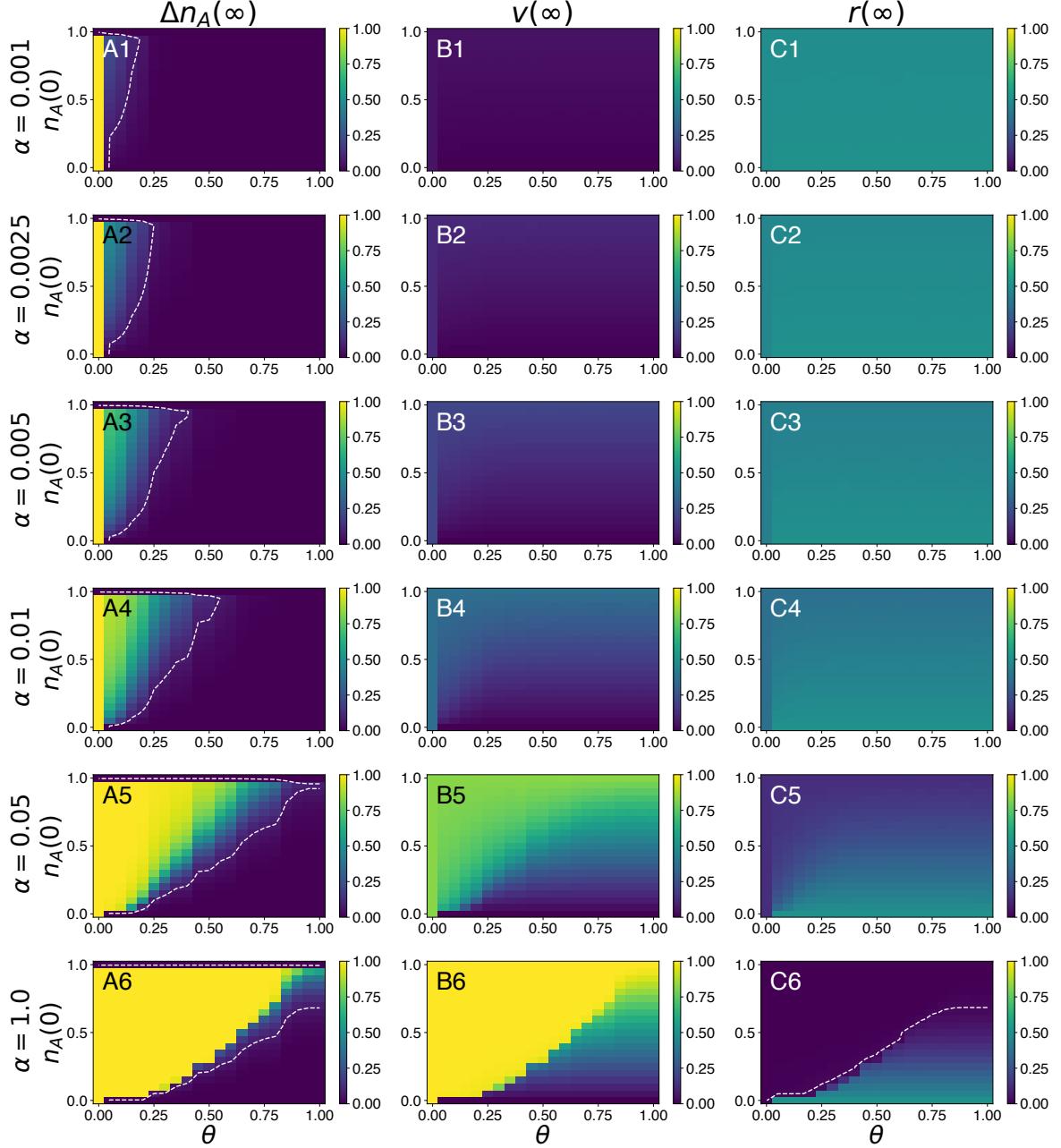
network topology. Therefore, a moderate vaccination may be sufficient to efficiently mitigate the epidemic impact of the disease.

### Heterogeneous networks

Now we run the same coupled epidemic-opinion dynamics on Barabási-Albert networks with the same average degree  $\langle k \rangle = 10$  as in the previous case. Similarly as before, the multi-panel in Figure 6.2 depicts the results for the main observables and control parameters scanned.

For  $\alpha = 0.001$ , the situation is qualitatively similar to the coupled dynamics running on homogeneous networks. Vaccination rates are too low to trigger a cascading behavior and change agents' hesitant opinions towards vaccine proactive agents. Virtually, VC is null across the  $(\theta, n_A(0))$  (Figure 6.2 B1), and prevalence is maximal for the BA network under the current epidemiological parameters, being  $r_{BA}(\infty) = 0.504$  [0.503, 0.505] (Figure 6.2 C1). However, at odds with the ER network case, continuing to increase  $\alpha$  has hardly an effect. Adherence to vaccination advances very slowly (Figure 6.2 A1 to A4), and consequently VC is sluggish. Even at  $\alpha = 0.01$ , disease prevalence continues to be very high across all the  $(\theta, n_A(0))$  space. Still, even at  $\alpha = 0.05$ , an important region of the parameter space suffers a high epidemic impact, and no disease-free solution emerges. Lastly, for  $\alpha = 1$ , the situation of  $\Delta n_A(\infty)$  (Figure 6.2 A6) and  $v(\infty)$  (B6) now resembles more closely the results for the ER network (Figure 6.1 A6 and B6, respectively). However, in the impassive region ( $\Delta n_A(\infty) = 0$ ) where VC remains very low to null, the epidemic impact is larger than in homogeneous networks.

In BA networks, the competition between disease contagion and the opinion-vaccination dynamics is overwhelmingly won by the former. This outcome is expected, considering the well-documented fact that the topology of heterogeneous networks facilitates a rapid spread of disease [154]. In this coupled dynamical system under study, disease transmission is a simple contagion process, requiring only a single contact between an infected and a susceptible individual to propagate, whereas opinion dynamics—a complex contagion—demands a critical mass of individuals who are not only persuaded but also vaccinated, to initiate a cascading effect of behavioral change. In contrast to the dynamics on ER networks, the heavy-tailed degree distribution found in BA networks means that highly connected nodes play a significant role and necessitate a larger number of influenced neighbors to shift their stance toward a pro-vaccination viewpoint, thereby impeding or delaying widespread vaccination coverage.



**Figure 6.2: Homogeneous thresholds in Barabási-Albert networks.** Results for normalized activation change  $\Delta n_A(\infty)$ , vaccination coverage  $v(\infty)$ , and prevalence  $r(\infty)$ . In every panel, outcomes are explored in  $(\theta, n_A(0))$  space. Every row shows the solution for a different vaccination rate  $\alpha$ . Every point in the diagrams amounts to 25 network realizations times 25 dynamical realizations. White dashed line contours mark an approximate (due to finite size effects) separation between  $\Delta n_A(\infty) = 0$  or  $r(\infty) = 0$  and non-null values of the respective observables.

### Zealots (anti-vaccine attitudes)

In the subsequent stage of our analysis, we factor in the presence of zealots. In opinion dynamics jargon, this refers to agents that will not change their opinion by any means. Actually, a zealot could embody the archetypal anti-vaccine individual, as well as

an agent committed to vaccination regardless of circumstances. For the purposes of this study, however, we focus solely on the former category. Hence, we use the terms “zealots” and “anti-vaccines” interchangeably. The incorporation of such agents into the model introduces a distinct activation threshold, designated as  $\theta_Z = 1^+$  for anti-vaccine agents. By definition, this threshold is insurmountable. Consequently, a key question we address is the extent to which anti-vaccine sentiments could influence the outcomes of the vaccination campaign and, as a result, the prevalence of the disease.

Now, first of all, without needing to recur to simulations, the closed population closure establishes that,  $n_A(0) + n_H(0) + n_Z = 1$ , where  $n_H(0)$  is the initial fraction of hesitant (or inactive) agents. Increasing  $n_Z$ , while preserving  $n_A(0)$ , the pool of agents amenable to persuasion toward vaccination diminishes. Maximally, we can reach  $n_A(0) + n_Z = 1$ . If  $n_A(0)$  is sufficiently large, the presence of zealots might have minimal impact. Conversely, if  $n_A(0)$  is low, further increasing  $n_Z$  will significantly contract the initial support base, potentially leading to a marked rise in disease prevalence. This behavior is indeed independent of the topology where the dynamical processes occur.

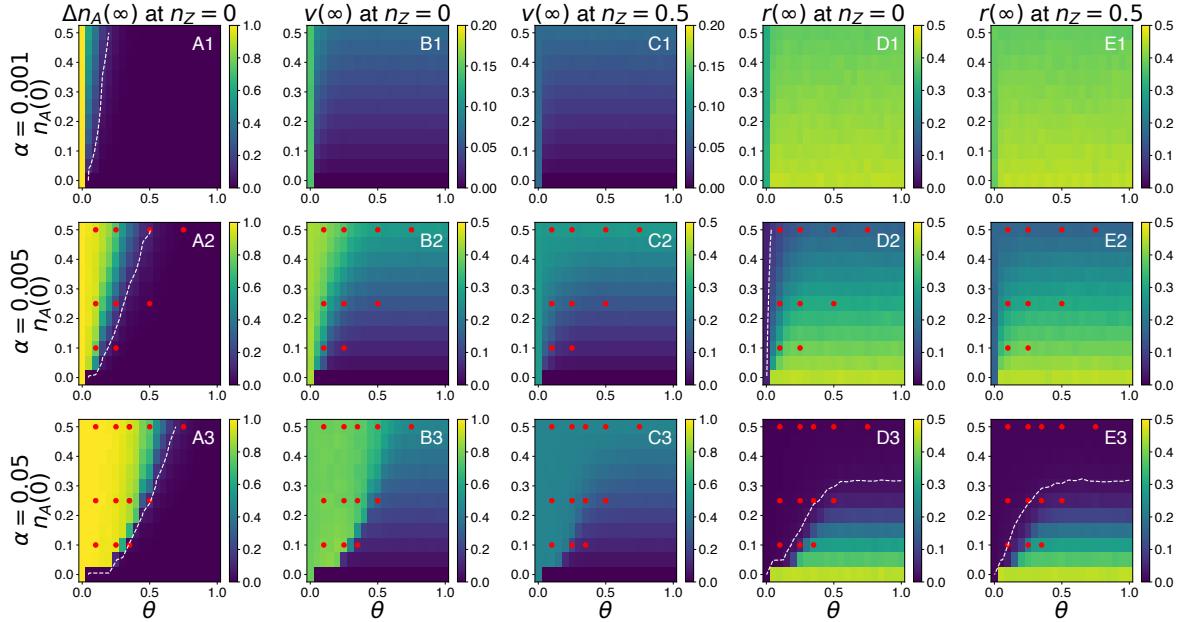
### **Homogeneous networks**

Turning to stochastic simulations, for Erdős-Rényi networks, we observe the behavior of the system for selected vaccination scenarios,  $\alpha = 0.001, 0.05$ , and  $0.05$ , under  $n_Z = 0$  and a (hopefully) extreme case where half of the system’s population has an anti-vaccine attitude (Figure 6.3).

Panels A in Figure 6.3 show  $\Delta n_A(0)$  at  $n_Z = 0$  (already shown in Figure 6.1) to aid the discussion, and similarly, panels B and D for  $v(\infty)$  and  $r(\infty)$ , respectively. Selected points in  $(\theta, n_A(0))$  are highlighted, which will be further analyzed subsequently in Figure 6.3.

For the lowest vaccination scenario  $\alpha = 0.001$ , the introduction of zealots has no noticeable impact on VC, and subsequently, neither on prevalence, except for the trivial case of  $\theta = 0$  (see the top Panels in Figure 6.3). This outcome is expected since vaccination proceeds too slowly to trigger a sizeable opinion-vaccination cascade that could mitigate the epidemic.

As we increase  $\alpha$  to  $\alpha = 0.005$ , in the absence of zealots, pro-vaccine activation and subsequent vaccination begin to take effect, and a disease-free region starts to appear for very low  $\theta$  and  $n_A(0) \geq 0$ . With  $n_Z = 0.5$ , we observe a notable decrease in VC with respect to the case  $n_Z = 0$  (compare Panels B2 and C2). Regarding disease prevalence, the small disease-free region has vanished, even though the magnitude of changes across  $(\theta, n_A(0))$  is not readily apparent: vaccination coverage remains high enough in the region of space where the introduction of zealots has had an impact.



**Figure 6.3: Comparison of impact at  $n_Z = 0$  and  $n_Z = 0.5$  on Erdős-Rényi networks.** Results for normalized vaccination coverage  $v(\infty)$ , and prevalence  $r(\infty)$  in  $(\theta, n_A(0))$ -space for  $n_Z = 0$  and  $n_Z = 0.5$ , and several vaccination scenarios  $\alpha$ . The size of the active cascade at  $n_Z = 0$  is shown for visual support. Highlighted points for  $\alpha = 0.005$  and  $\alpha = 0.05$  are analyzed later in detail in Figure 6.4. Every point in the heatmaps amounts to 25 network realizations times 25 dynamical realizations. White dashed line contours mark an approximate (due to finite size effects) separation between  $\Delta n_A(\infty) = 0$  or  $r(\infty) = 0$  and non-null values of the respective observables.

Despite the subtle changes in prevalence, it is worth noting—as mentioned earlier in the discussion—the valid domain of  $n_A(0)$  has dramatically shrunk with  $n_Z = 0.5$ , and now the disease-free phase is significantly reduced (refer to Figure 6.1 for a full appreciation of this effect). Increasing one order of magnitude to  $\alpha = 0.05$ , consensus towards pro-vaccine attitudes at  $n_Z = 0$  expands (as the  $\Delta n_A(\infty) = 0$  border shifts towards higher  $\theta$  from A2 to A3). This leads to higher VC and, subsequently, a larger disease-free phase in the  $(\theta, n_A(0))$  space. Upon moving to  $n_Z = 0.5$ , although VC is significantly reduced, this does not necessarily translate into higher prevalence. Upon closer examination, the  $r(\infty) = 0$  boundary is slightly shifted leftward from  $n_Z = 0$  to  $n_Z = 0.5$  (compare Figure 6.3 D3 and E3), affecting only specific regions in  $(\theta, n_A(0))$ .

In Figure 6.4, we take a closer look at selected cases (the colored points in the middle row panels of Figure 6.3) for  $\alpha = 0.005$  and  $0.05$ . Panels A, B, and C examine how  $\Delta n_A(0)$ ,  $v(\infty)$ , and  $r(\infty)$  respond for varying fractions of anti-vaccine agents ( $n_Z$ ).

In analyzing the selected cases, we can discern two distinct situations when evaluating  $\Delta n_A(\infty)$  as a function of  $n_Z$ , for both values of  $\alpha$  shown in Figure 6.4. Some points exhibit constant value lines with  $\Delta n_A(\infty) \approx 0$ , while others demonstrate a decline in  $\Delta n_A(\infty)$  toward 0 as  $n_Z$  approaches 1. The horizontal lines correspond to

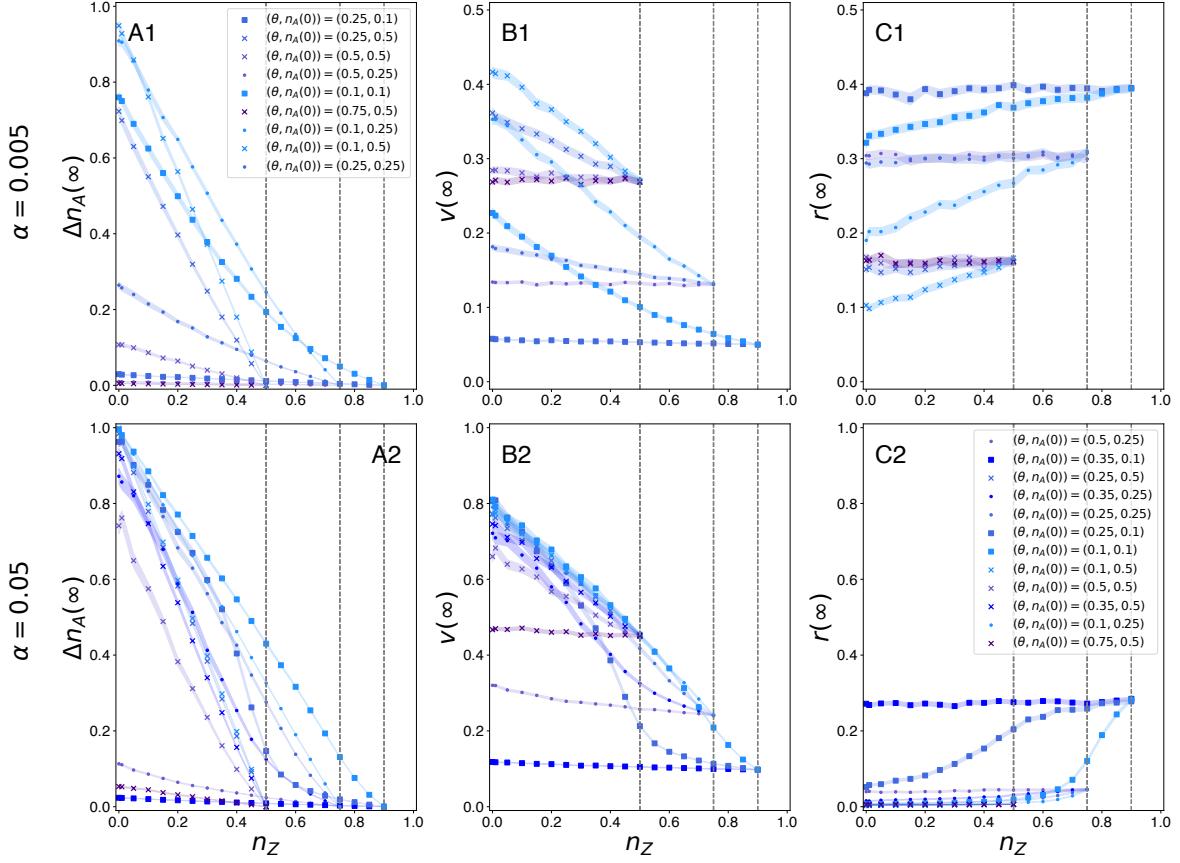


Figure 6.4: **Varying anti-vaccine attitude ( $n_Z$ ) curves on Erdős-Rényi networks.** Results obtained under  $\alpha = 0.005$  (Panels A1, B1, and C1) and  $\alpha = 0.05$  (Panels A2, B2, and C2) for normalized activation change  $\Delta n_A(\infty)$ , normalized vaccination coverage  $v(\infty)$ , and prevalence  $r(\infty)$  against the fraction of anti-vaccine agents  $n_Z$  in the system. Dashed lines mark the  $n_Z$  value at which  $n_A(0) + n_Z = 1$  for a given  $n_A(0)$  value. Solid lines represent average values and shaded areas the 95% CI. Results were obtained over 25 network realizations and 25 dynamical realizations.

curves that maintain constant or nearly constant values of  $v(\infty)$  (Panels B1 and B2) and  $r(\infty)$  (Panels C1 and C2).

Specifically, in situations such as  $(\theta, n_A(0)) = (0.25, 0.1)$  or  $(0.5, 0.5)$  for  $\alpha = 0.005$ , the system exhibits no change even as  $n_Z$  increases. This is because  $\Delta n_A(\infty) = 0$  is already achieved at  $n_Z = 0$ , indicating no further shift toward pro-vaccine opinion, and thus the system is in the inactive phase where further increases in anti-vaccine sentiment have no effect.

Conversely, in scenarios where  $\Delta n_A(\infty)$  decreases with increasing  $n_Z$ , there is a corresponding decrease in  $v(\infty)$  and an increase in  $r(\infty)$ . A convergent behavior is observed among the family of curves as identified by their respective  $n_A(0)$  values. Remarkably, the convergence point aligns with the population closure  $n_Z + n_A(0) = 1$ . At this juncture, the initial pool of hesitant agents is depleted, with  $n_H(0) = 0$ , rendering the threshold value  $\theta$  inconsequential as no more hesitant or inactive agents

remain to be influenced. This explains why, for the same  $n_A(0)$ , different  $\theta$  curves tend to approach the same behavior as  $n_Z \rightarrow 1$ .

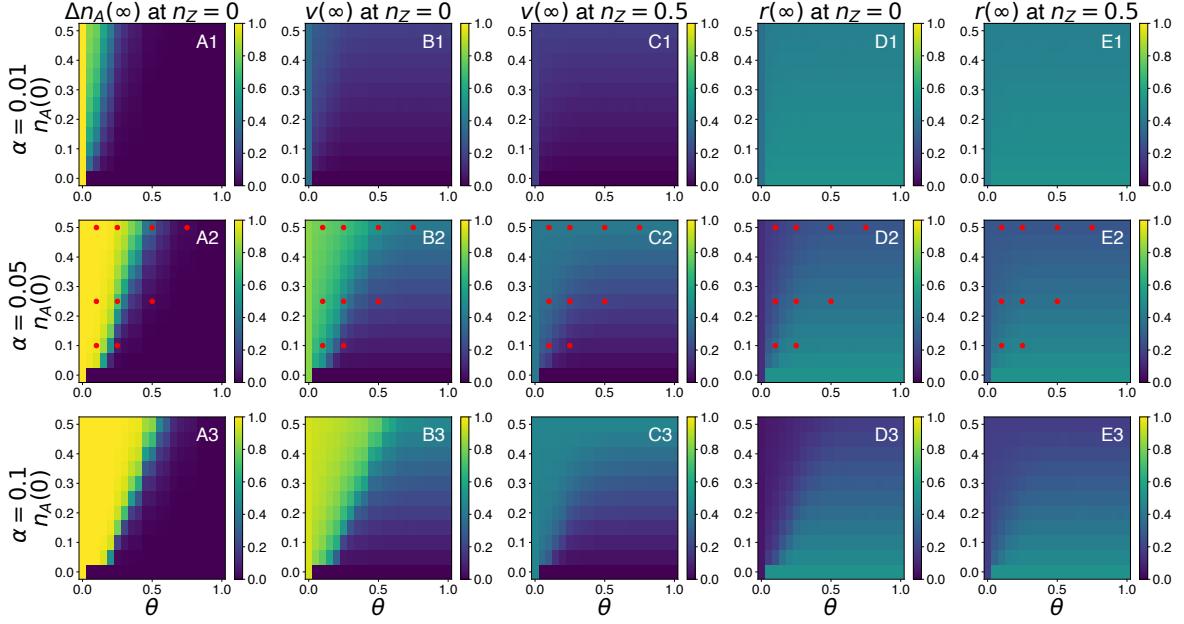
Within a family of curves sharing the same  $n_A(0)$  value, curves with higher  $\theta$  tend to flatten, as they are at or near the  $\Delta n_A(\infty) = 0$  region when  $n_Z = 0$ . In contrast, curves with lower  $\theta$  values experience the steepest changes, indicating that systems with lower  $\theta$  are less resilient to an increase in  $n_Z$ . For  $\alpha = 0.005$ , changes in  $v(\infty)$  and  $r(\infty)$  are predominantly linear, whereas for  $\alpha = 0.05$ , they can exhibit strong nonlinearities.

Take, for instance, the points  $(0.1, 0.1)$ ,  $(0.25, 0.1)$ , and  $(0.35, 0.1)$  for  $\alpha = 0.05$ . The point  $(0.35, 0.1)$ , which is on the cusp of  $\Delta n_A(\infty) \approx 0$ , shows little variation in VC and prevalence as  $n_Z$  increases. Moving to  $\theta = 0.25$ , there is a dramatic change:  $\Delta n_A(\infty)$  and  $v(\infty)$  significantly decrease. While  $r(\infty)$  does increase, it requires a substantial presence of anti-vaccine agents to see marked changes. Further reducing  $\theta$  to 0.1, we observe a similar trend for  $\Delta n_A(\infty)$  and  $v(\infty)$ , but the prevalence remains nearly unchanged (and in the disease-free phase  $r(\infty) \approx 0$ ) over a large range of  $n_Z$  values until it suddenly surges. For the rest of the points examined under  $\alpha = 0.05$ , despite strong decreases in VC, prevalence remains approximately bounded to the case  $n_Z = 0$ . From this, we can infer that under high vaccination rates, such as  $\alpha = 0.05$ , the system exhibits more resilience to the influence of anti-vaccine sentiment.

### Heterogeneous networks

For heterogeneous Barabási-Albert networks, Figure 6.5 represents the behavior of the system for selected vaccination scenarios,  $\alpha = 0.01, 0.05$  and  $0.1$ . Again, we compare the case  $n_Z = 0$  against the case  $n_Z = 0.5$ . Note, regarding vaccination scenarios, that given how the coupled dynamical processes behave on these networks, we have explored higher  $\alpha$  values.

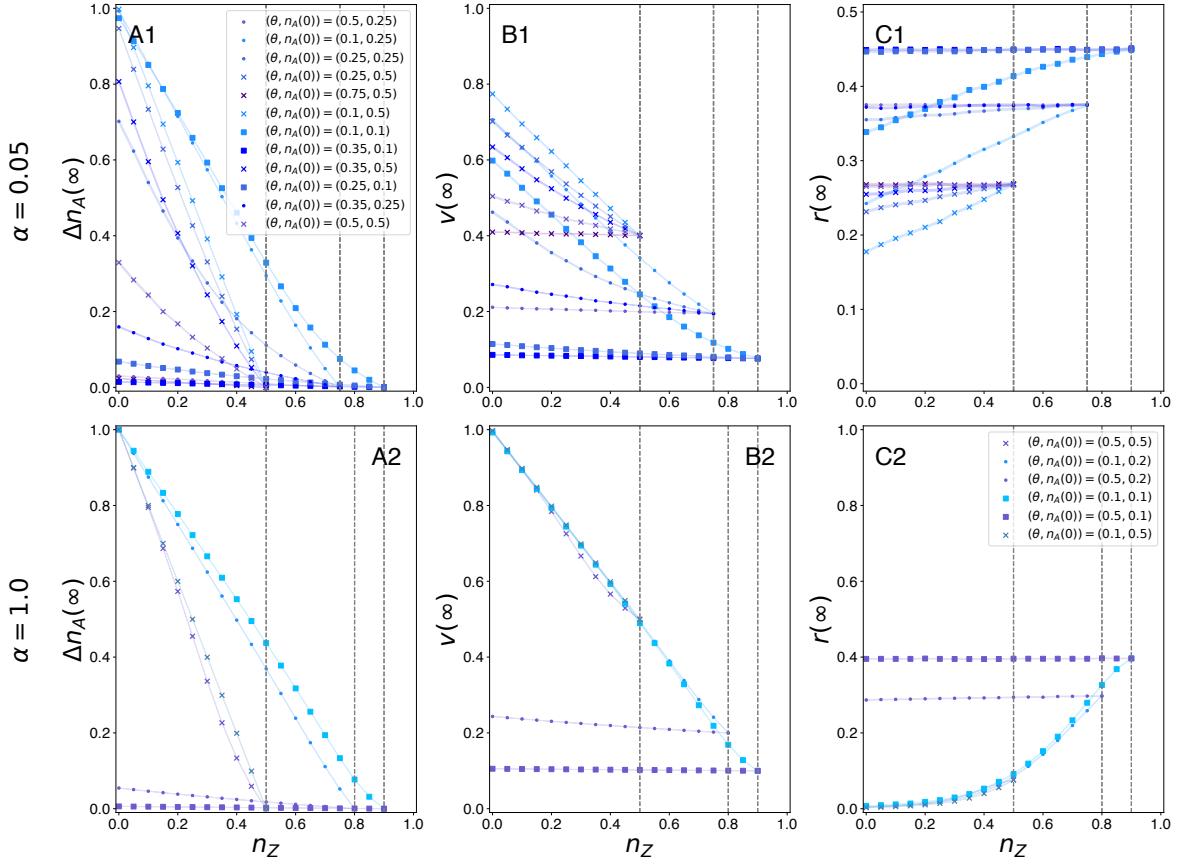
The bulk of our prior analysis remains applicable to BA networks. Within the realistic range of vaccination rates, we observed that for BA networks, vaccination coverage was insufficiently low, and high disease prevalence was typical, even when  $n_Z = 0$ . Consequently, increasing  $n_Z$  does not exacerbate the situation further, other than by reducing the valid domain for  $n_A(0)$ . However, as we consider higher vaccination rates, such as  $\alpha = 0.05$ , we can visually discern that a decrease in vaccination coverage, particularly in the low  $\theta$  region (from Figure 6.5 B2 to C2), drives an increase in disease prevalence (from Figure 6.5 D2 to E2). This pattern persists even for  $\alpha = 0.1$ . In this scenario, the disease-free region (Panel D3) at  $n_Z = 0$  contracts at  $n_Z = 0.5$ . It is noteworthy that even at this vaccination rate, which is unprecedented as seen during the COVID-19 era [595], the majority of the  $(\theta, n_A(0))$



**Figure 6.5: Comparison of impact at  $n_Z = 0$  and  $n_Z = 0.5$  on Barabási-Albert networks.** Results for normalized vaccination coverage  $v(\infty)$ , and prevalence  $r(\infty)$  in  $(\theta, n_A(0))$ -space for  $n_Z = 0$  and  $n_Z = 0.5$ , and several vaccination scenarios  $\alpha$ . The size of the active cascade at  $n_Z = 0$  is shown for visual support. Colored points for  $\alpha = 0.05$  are analyzed in detail later in Figure 6.6. Every point in the heatmaps amounts to 25 network realizations times 25 dynamical realizations. White dashed line contours mark an approximate (due to finite size effects) separation between  $\Delta n_A(\infty) = 0$  or  $r(\infty) = 0$  and non-null values of the respective observables.

space is susceptible to outbreaks with high prevalence.

As with the analysis of homogeneous networks, we refocus on specific points in the  $(\theta, n_A(0))$  space and vaccination scenarios to investigate the system's behavior with increasing  $n_Z$  (see Figure 6.6). The principal findings are reaffirmed: points where  $\Delta n_A(\infty)$  is already zero at  $n_Z = 0$  are unaffected by subsequent increases in  $n_Z$ . Conversely, scenarios where  $\Delta n_A(\infty) \neq 0$  at  $n_Z = 0$  demonstrate a trend towards diminished pro-vaccination sentiment, reduced vaccination coverage, and heightened prevalence with rising  $n_Z$ . Within a given family of curves characterized by an  $n_A(0)$  value, those with lower  $\theta$  values exhibit the most pronounced changes, suggesting that increasing  $n_Z$  effectively raises the system's activation threshold. The convergence within a curve family occurs when the condition  $n_A(0) + n_Z = 1$  is fulfilled. Although purely hypothetical, we conclude by presenting the case of  $\alpha = 1$ , to underscore that for BA networks, even an exhaustive vaccination campaign cannot completely counteract the potential for significant epidemic outbreaks as increasing anti-vaccine sentiment grows.

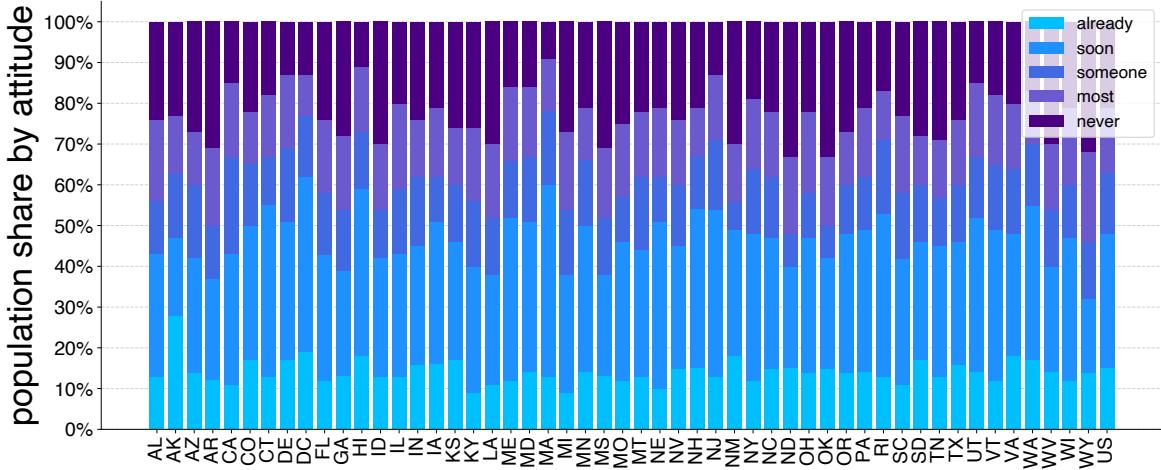


**Figure 6.6: Varying anti-vaccine attitude ( $n_Z$ ) curves on Barabási-Albert networks.** Results obtained under  $\alpha = 0.05$  (Panels A1, B1, and C1) and  $\alpha = 1.0$  (Panels A2, B2, and C2) for normalized activation change  $\Delta n_A(\infty)$ , normalized vaccination coverage  $v(\infty)$ , and prevalence  $r(\infty)$  against the fraction of anti-vaccine agents  $n_Z$  in the system. Dashed lines mark the  $n_Z$  value at which  $n_A(0) + n_Z = 1$  for a given  $n_A(0)$  value. Solid lines represent average values and shaded areas the 95% CI. Results were obtained over 25 network realizations and 25 dynamical realizations.

## Survey-based thresholds

To conclude the analysis of our model, we introduce heterogeneous threshold values based on US survey data on vaccine hesitancy [514]. In Table 6.1 we offered a mapping between the nominal categories of vaccination attitude and the interpreted activation threshold  $\theta$ . Now, in Figure 6.7 we present the population share belonging to these categories by US state.

Since the control parameters  $\theta$ ,  $n_A(0)$ , and  $n_Z$ , previously used are now fixed for every state, we focus on  $\alpha$  to explore the outcomes of the coupled dynamics. As before, simulations are conducted on both ER and BA networks to assess their impact on significantly different topologies. The average degree value,  $\langle k \rangle$ , for each US state is computed as previously specified. It must be noted that the assignation of the different vaccination attitudes and corresponding thresholds to the agents in the network is

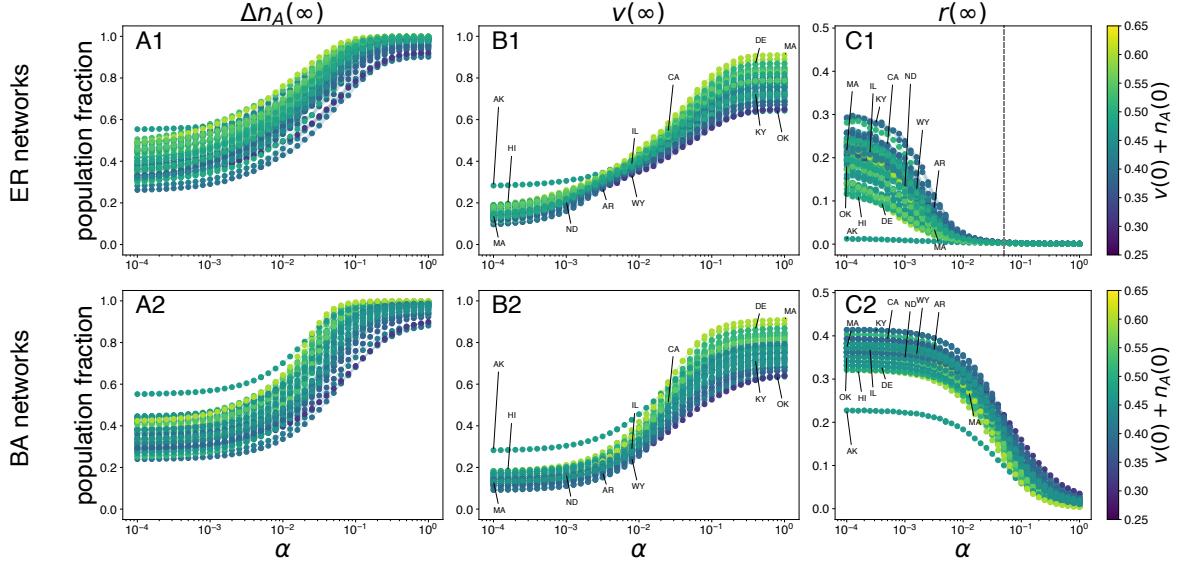


**Figure 6.7: US states’ population fractions by attitude towards vaccination.** From the bottom to the top, for each state the bar segments represent the following attitudes towards vaccination: “already vaccinated”, “as soon as possible”, “after at least some people I know”, “after most people I know”, “I would never take the [COVID-19] vaccine”, as obtained from Lazer et al. [514].

random.

We then examine how  $\Delta n_A(\infty)$ ,  $v(\infty)$ , and  $r(\infty)$  behave across varying  $\alpha$  values in the curves for both network models (see Figure 6.8). Panels A1 and A2 present  $\Delta n_A(\infty)$  as a function of  $\alpha$  in ER and BA networks, respectively. In both cases, we observe that, across states, the quantity remains between 0.2 and 0.6 for the lowest  $\alpha$  scenarios. This separation diminishes as  $\alpha$  increases, aligning within the range of 0.8 to 1.0 for scenarios with the highest vaccination rate  $\alpha$ . Notably, no state exhibits  $\Delta n_A(\infty) = 0$  in any scenario, implying, as per previous sections’ analysis, that each state would be impacted by an increase in their respective fraction of zealots  $n_Z$ .

In Panels B1 and B2 of Figure 6.8, we focus on vaccination coverage. In low  $\alpha$  regimes, VC is primarily between 0.1 and 0.2, with the exception of Alaska (AK). For intermediate  $\alpha$  values, representing more realistic vaccine uptake rates, the disparity in VC across states reduces but widens again as  $\alpha \rightarrow 1$ . Due to inherent differences in each state’s  $\langle k \rangle$  and, more importantly, their  $\theta$  distribution, the curves intersect at various points. These varying gaps and curve crossings for different  $\alpha$  values reflect the diversity in vaccine attitudes across states. As a specific example, the case of Massachusetts (MA) is interesting. In the lowest  $\alpha$  regimes, MA occupies a median rank among states, with 13% of its population having already received the vaccine. However, with increasing  $\alpha$ , the vaccination-opinion feedback loop is further stimulated. This facilitates the progressive vaccination of individuals in subsequent categories, specifically “soon” at 47%, and “someone” at 18%. Ultimately, this leads MA to emerge as one of the states with the highest vaccination coverage.



**Figure 6.8: Epidemiological impact across US states for varying vaccination rates  $\alpha$ .** Top panels: Results across US states on ER networks. The vertical dashed line represents an approximate separation between the endemic phase and the disease-free phase. Bottom panels: BA networks. Panels A1 and A2:  $\Delta n_A(\infty)$  vs.  $\alpha$ . Panels B1 and B2:  $v(\infty)$  vs.  $\alpha$ . Panels C1 and C2:  $r(\infty)$  vs.  $\alpha$ . Color bar represents the initial fraction of support for the vaccine: “already vaccinated”  $v(0)$  plus “as soon as possible”  $n_A(0)$ . Some state labels are shown for reference.

Regarding the final disease prevalence  $r(\infty)$ , Panels C1 and C2 display results for ER and BA networks, respectively. The most notable difference between the network types aligns with previous analyses. For ER networks (Figure 6.8 C1), in very low  $\alpha$  regimes, prevalence across US states ranges from 0.1 to 0.3. These values decrease rapidly with increasing vaccination deployment, reaching a disease-free state beyond  $\alpha \approx 0.05$ . Alaska is correspondingly an exception here too; its prevalence is negligible across all vaccination rate scenarios, as represented by  $\alpha$ . With an initial “already vaccinated” fraction of  $v(0) = 0.29$ , Alaska’s outcomes suggest that for very low  $\alpha$ , the opinion-vaccination cascades do not initiate, resulting in  $v(\infty) \approx v(0)$ . This indicates that such an initial  $v(0)$  is close to the herd immunity threshold  $p_c$ , preventing sizeable outbreaks. For a well-mixed population, akin to an ER network with a basic reproduction number  $R_0 = 1.5$ , the herd immunity threshold is  $p_c = 1 - 1/R_0 \approx 0.33$ . Alaska’s negligible impact is thus understandable in light of its initial condition  $v(0)$ <sup>3</sup>. Conversely, in BA networks we saw the impossibility of attaining a disease-free solution for most  $\alpha$  values. Now, even Alaska experiences significant outbreaks for most  $\alpha$  values. In the case of Massachusetts, previously highlighted for its vaccination coverage

<sup>3</sup>It should be noted that  $v(0)$  represents an initial condition in a population otherwise naive to the disease. In reality, however, the “already vaccinated” category in surveys preceded the second wave of COVID-19 in the US.

(VC) progression with varying  $\alpha$  in both ER and BA networks, a similar trend is evident in its prevalence data. It begins at a mid-rank position but becomes one of the states least impacted when  $\alpha$  reaches high (yet realistic) values.

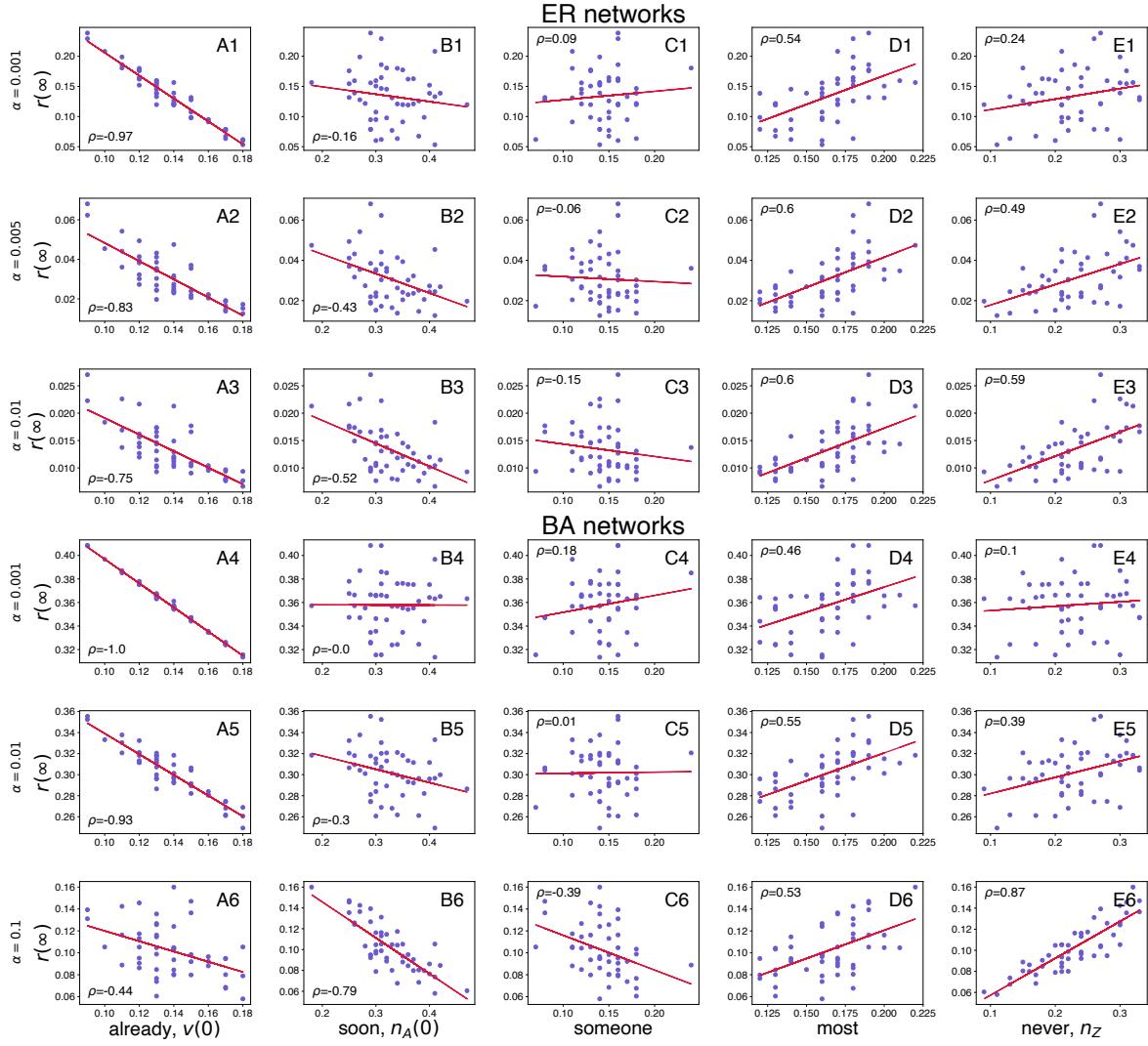
Finally, all curves incorporate additional information represented by a colormap for the sum of initial conditions for vaccinated and pro-vaccine individuals,  $v(0) + n_A(0)$ . While the relationship is not strictly monotonic, it is evident that larger  $v(0) + n_A(0)$  values correspond to higher VC and lower prevalence, and vice versa, underscoring the significance of initial vaccine support conditions in influencing the epidemiological impact.

To gain a clearer understanding of how diverse vaccine attitudes affect the system's disease prevalence, we perform a correlation analysis based on pairwise scatter plots of the prevalence with various vaccination attitude categories across US states.

In both networks tested, we observe a clear correlation as indicated by Pearson's correlation coefficient for low and moderate  $\alpha$  scenarios of  $r(\infty)$  and  $v(0)$  ("already vaccinated") (refer to Figure 6.9 A1 to A6) and a relatively small or very poor correlation with respect to the "soon" category (Figure 6.9 B1 to B6), the only exception being  $\alpha = 0.1$  for the BA network case (Panels A6 and B6). As expected, the higher  $v(0)$  is, the lower  $r(\infty)$ . Particularly in the lowest vaccination rate scenarios, the correlation between  $r(\infty)$  and  $v(0)$  is very high (Panel A1 with  $\rho = -0.97$ ) or even absolute (Panel A4 with  $\rho = -1$ ), in contrast to  $r(\infty)$  vs.  $n_A(0)$ . In these scenarios, where the vaccination campaign is irrelevant, it is evident that the emphasis is on the already vaccinated portion of the population rather than those pro-vaccine but not yet vaccinated. We observe that as  $\alpha$  increases, the Pearson coefficient's explanatory power for  $v(0)$  diminishes, while it grows (in absolute value) for  $n_A(0)$ .

For the attitude "someone" (I will take the vaccine after someone I know), we observe mixed results. At the lowest  $\alpha$ , it indeed correlates positively with prevalence, and as  $\alpha$  increases, the sign of the correlation changes, and  $\rho$  grows in magnitude. However, correlations are generally weak, suggesting that these individuals are not key in shaping the spread. Only at high  $\alpha$ , particularly in BA networks (see Figure 6.9 C6), does this vaccination attitude begin to gain some (albeit small) explanatory power. The presence of hubs, signifying individuals with a very high degree  $k \gg \langle k \rangle$ , makes it more likely that some of their neighbors have been vaccinated, thereby enhancing their role in the opinion dynamics compared to ER networks.

The "most" attitude (I will take the vaccine after most people I know) shows moderate but distinct correlations. In states where this fraction of individuals is larger, larger outbreaks tend to occur. Across different  $\alpha$  scenarios and networks tested, the correlation coefficient  $\rho$  varies within a small range. One reason these individuals



**Figure 6.9: Pairwise scatter plots of prevalence and vaccine attitudes.** Panels with letters A, B, C, D, and E show, respectively, the scatter plots of prevalence (y-axis) against vaccine attitudes (x-axis): “already vaccinated”, “as soon as possible”, “after at least some people I know”, “after most people I know”, “I would never take the vaccine”. Panels numbered 1, 2, and 3 show results for ER networks with  $\alpha = 0.001$ ,  $0.005$ , and  $0.01$ , respectively. Panels numbered 4, 5, and 6 show results for BA networks with  $\alpha = 0.001$ ,  $0.01$ , and  $0.1$ , respectively.

are not more influential is that if a cascade of pro-vaccine individuals grows quickly enough, those with the “most” attitude will also likely become pro-vaccine, thus helping to reduce the final epidemic size. On the other hand, the presence of such individuals would be highly detrimental if it were to be accompanied by either high anti-vaccine (zealot) sentiment or low pro-vaccine attitudes.

Finally, we consider the anti-vaccine attitude population fraction (“never”), i.e. the zealots denoted by  $n_Z$ . A higher  $n_Z$  correlates with a higher  $r(\infty)$ , although their statistical correlation varies depending on the vaccination scenario. Similar to other attitudes, the fraction of zealots or anti-vaccine individuals increases its explanatory

power as the vaccination campaign gains traction. Indeed, when vaccination becomes less relevant, it is logical to expect that individuals refusing vaccines under any circumstances will not have a significant impact. Their influence becomes particularly notable in BA networks for high to very high vaccination rates.

To conclude the discussion, let us revisit the population distribution by vaccination attitudes as shown in Figure 6.7. Two observations complement our understanding of the results in this section. First, in several states, the combined fractions of the “most” and “never” attitudes amount to less than 0.5. This indicates that in scenarios with rapid vaccination, individuals with a “most” attitude are likely to find most of their neighborhood already vaccinated, prompting their shift to a pro-vaccine stance. Second, in most states, the fraction of the population already vaccinated or willing to be vaccinated as soon as possible,  $v(0) + n_A(0)$ , is more than 0.4, and often exceeds 0.5. Including the “someone” category, which requires just one vaccinated neighbor per individual, this vaccinated/pro-vaccine segment increases to over 0.5 in most cases. This suggests that, from moderate  $\alpha$  scenarios onwards, vaccination-opinion cascades can readily grow, leading to widespread vaccination coverage and significantly mitigated outbreaks. Although not directly comparable, assuming a uniform threshold setting in ER networks with initial vaccination support between 0.4 and 0.6, we see in Figure 6.1 (Panel C5) that we are within the disease-free phase of the diagram. Similarly, Figure 6.8 (Panel C1) shows that all states reach a disease-free solution at approximately  $\alpha = 0.05$ . This suggests that the initial conditions across states are favorable enough to effectively yield reduced attack rates, provided vaccination progresses swiftly. Conversely, for BA networks, consistently with the simplified homogeneous setting previously explored, even such a robust initial support is insufficient to significantly lower the system’s prevalence.

In reality, however, the structure of how people interact is more complex than the contact network models used here allow for. In the context of this work, it is crucial to consider realistic elements such as homophily and community structure. Individuals with closed-minded attitudes may prefer interactions within their own group, while actively avoiding opposing views, as has been previously documented in relation to vaccine hesitancy and anti-vaccine sentiments [596, 597, 598]. This phenomenon does not only happen in virtual spaces, such as the creation of echo chambers in online social networks [599, 600], but also occurs geographically [596]. Echo chambers can facilitate the diffusion of vaccine hesitancy online and reinforce the views of closed-minded individuals. In turn, physical clusters of hesitant individuals can dangerously become hotspots for the resurgence and transmission of infectious diseases.

In addition to these aspects, for influenza-like illnesses, age plays a crucial role in

transmission patterns and disease prevalence [402, 400, 601]. Incorporating age into the network structure would render this analysis more realistic and could bring new insights. These and other considerations will be addressed in future work.

## 6.4 Conclusions

In this work, we have examined the role of human behavior in the spread of epidemics, and in particular, the impact of vaccine hesitancy as influenced by social interaction. We propose a model that couples: i) the spread of an epidemic of an influenza-like illness (SIR model), ii) a vaccination campaign, and iii) opinion dynamics on vaccine uptake, which is modeled through the Watts-Granovetter threshold model. Unlike the original model, our agents do not consider their neighbors' (positive) opinion on vaccination, but rather their actual vaccination status, thereby tightly coupling vaccination and opinion dynamics. Consequently, agents in our system observe their peers' actions rather than their assertions when deciding to participate in the vaccination campaign. This modeling choice was inspired by surveys on attitudes toward vaccination in the US, which indicated that an individual's decision to vaccinate was often predicated on the vaccination status of their peers.

We explored the dynamics on Erdős-Rényi (ER) and Barabási-Albert (BA) networks, scanning the parameter space given by the activation threshold (to switch to pro-vaccine attitude)  $\theta$ , the initial fraction of activated (pro-vaccine) agents  $n_A(0)$ , and the vaccination rate  $\alpha$ . First, we assumed homogeneous thresholds  $\theta$  in the population. In ER networks, we find that as  $\alpha$  increases a successful feedback loop between the cascading behavior of the opinion threshold model and the vaccine uptake drives disease prevalence down. For high enough values - but within reasonable capabilities- a disease-free solution emerges in an increasingly large part of  $(\theta, n_A(0))$  space. Moreover, further increasing  $\alpha$  (as explored here beyond  $\alpha = 0.05$ ), the endemic solution does not disappear for low enough initial support (roughly  $n_A(0) < 0.25$ ) and an important range of  $\theta$  values, signaling to diminishing returns in vaccination. In BA networks, qualitatively the picture is similar but, critically, high and unrealistic vaccination efforts are needed for the disease-free phase in  $(\theta, n(0)_A)$  to emerge. Overall, disease prevalence is higher in BA than in ER networks for any  $\alpha$  value.

We then shifted our focus to the impact of zealotry (anti-vaccine behavior) on the system's dynamics. Qualitatively, both ER and BA networks exhibit similar responses. When  $\alpha$  is low, anti-vaccine sentiment has a negligible effect as vaccine hesitancy remains minimal; the system fails to propagate opinion-vaccine cascades effectively. As  $\alpha$  increases, two distinct scenarios emerge: one where vaccine coverage and disease

prevalence remain unchanged due to  $\Delta n_A(\infty) = 0$  at  $n_Z = 0$ , and another where a decline in VC corresponds to an upsurge in prevalence. It is also found that curves with identical  $n_A(0)$  values, albeit with different  $\theta$  values, tend to converge with increasing  $n_Z$ , specifically at the juncture where  $n_Z + n_A(0) = 1$ . Within the same  $n_A(0)$  family, curves marked by a lower  $\theta$  exhibit more pronounced changes in VC and prevalence, while those with a higher  $\theta$  reach a null gradient more promptly. As the pool of hesitant agents dwindles, the system's behavior becomes predominantly determined by  $n_A(0)$ , rendering the activation threshold  $\theta$  irrelevant. Furthermore, an increase in  $\alpha$  may precipitate pronounced nonlinearities in the evolution of  $v(\infty)$  and  $r(\infty)$  as a function of  $n_Z$ .

Finally, we applied the proposed framework of coupled dynamics to characterize the epidemic's impact across the United States, using heterogeneous activation thresholds derived from vaccination surveys. In ER networks, a disease-free solution can be attained with moderate vaccination efforts in every state. In contrast, in BA networks, significant outbreaks occur even at high vaccination rates. Regarding different vaccine attitudes, we observed mixed behavior depending on the vaccination scenario. Generally, and as expected, individuals who are already vaccinated play a key role in reducing the epidemic's final size. Conversely, those who significantly delay vaccine uptake or directly refuse to vaccinate have a negative impact on the disease's prevalence. Fortunately, given that initial support for vaccine uptake is sufficiently high across states, the influence of anti-vaccine zealots is mitigated.

Compared to the homogeneous vaccination threshold setting studied in the beginning, the introduction of heterogeneity in the vaccination activation thresholds leads to a more nuanced evolution in the adoption of vaccination attitudes. In a homogeneous threshold scenario, every local neighborhood must surpass the same threshold  $\theta$  for the focal individual to adopt a pro-vaccine stance (“soon”). The feasibility of initiating a widespread vaccine-opinion cascade depends heavily on the initial level of vaccine support  $n_A(0)$ . This could make it challenging to trigger such a cascade. In contrast, with heterogeneous thresholds, initiating a pro-vaccine cascade might be simpler. It could start within a few clusters of individuals already supportive of vaccines and gradually gain momentum, eventually influencing more vaccine-hesitant individuals. However, this is also very dependent on the global and local population share of vaccine attitudes. Indeed, the presence of tightly-knit groups of highly vaccine-hesitant individuals could hinder achieving widespread vaccination coverage. Therefore, further research is necessary to draw a comprehensive comparison between scenarios with uniform and varied vaccination thresholds.

Our work has endeavored to delve into the complex interplay between coupled

dynamical processes: epidemic spreading, vaccination, and opinion dynamics. Despite incorporating elements of realism, our study presents several limitations, thus paving the way for future research.

In relation to the simplest version of our model and in a theoretical fashion, it would be beneficial to develop a mean-field theory for comparison with simulations and to derive an analytical expression for the epidemic threshold as a function of relevant control parameters. This would provide a clearer picture of the conditions separating the endemic and disease-free phases.

Furthermore, for practical applications and policy-making, additional layers of realism and complexity are necessary. Primarily, accounting for the role of community structure and homophily in the distribution of vaccine attitudes is critical for understanding the reemergence and successful transmission of infectious diseases [596]. This consideration could have a more significant impact on the spreading dynamics than when a random distribution of vaccine attitudes is assumed. Additionally, given the significance of age as a factor driving disease transmission and its disproportionate effects on different population groups, and considering the availability of high-resolution age-structured contact matrices [360], a logical next step is to incorporate age into the model. This would involve constructing a multilayer network [601] that accurately represents these age-structured contact patterns.

As for the vaccination dynamics, we have assumed a Markovian process without any strategy that prioritizes certain population groups, nor we have assumed heterogeneous vaccination rates when analyzing the impact on US states. A more nuanced portrayal of a vaccination campaign would be beneficial for making accurate projections. Finally, although the threshold model proposed is grounded in responses from real surveys, human decision-making is multifaceted, and additional mechanisms undoubtedly influence the choice to vaccinate. These mechanisms might include rational action, as is typically considered in vaccination game approaches [573, 576], where individuals weigh the costs of different options. Additionally, the influence of mass media and social networks is significant, where figures of authority may exert more influence than peers or acquaintances [602, 603, 604, 600]. Moreover, these platforms can also facilitate the rapid spread of misinformation and fake news, further complicating the landscape of decision-making. Furthermore, communities evolve and adapt, often leading to fragmentation between opposing viewpoints and the creation of echo chambers [605, 606, 599]. These factors are increasingly pivotal in shaping attitudes toward vital issues and warrant consideration in future studies.



## **Part IV**

### **Closure**



# Chapter 7

## Closure

*I wonder where I'll go now. The net is vast and infinite.*

— Motoko Kusanagi

### Outline and main remarks

This dissertation comprises several works on the modeling of the spreading of epidemics in human populations and the analysis of their subsequent impact. In *Part I: Overture*, we present an extensive introduction to the field in Chapter 1 and cover the fundamentals—and some advanced aspects—of the framework used throughout this dissertation in Chapter 2. Then, the original contributions are structured in two parts. *Part II: Metapopulation models*, containing Chapters 3 and 4, focuses on the spatial spread of diseases and the role of mobility. *Part III: Single-population behavioral structured models* includes Chapters 5 and 6, and concentrates on the integration of behavioral features into models with either age or contact structure. Although the works were discussed and concluding remarks were offered in each corresponding chapter, we will summarize them again and offer some general closing thoughts.

We begin *Part II* with Chapter 3, where we examine a real-life situation motivated by the COVID-19 pandemic. As has been already explained, the COVID-19 pandemic, along with the ensuing health and socioeconomic crisis, caught all authorities off guard. The absence of competent preventive plans for emerging infectious disease outbreaks, as was the case in Spain, resulted in authorities scrambling to catch up with the spread and being compelled to implement highly aggressive control measures to bend the curve and avoid a catastrophic impact on the population. The enforcement of perimeter lockdowns in the Comunidad de Madrid, as well as in the city of Madrid at the level of Basic Health Zones, can be seen as one of those measures that were both disruptive and of limited effectiveness in containing the spread. By developing a metapopulation model of the city of Madrid at the district level, informed by mobility flows from actual

data provided by the Spanish Ministry of Transport, Mobility, and Urban Agenda, we assess the potential success of the perimeter lockdown strategy in containing the disease's spread and mitigating its impact on the population. Although the model is parsimonious and is not designed to forecast future scenarios or retrodict the actual experience of Madrid, it sufficiently demonstrates that in highly interconnected urban systems, the disease spreads easily and swiftly throughout the entire system. Local outbreaks tend to synchronize, making it virtually impossible to contain the disease in localized areas while expecting the rest of the system to function normally. It must be noted that our model assumes perfect information from the surveillance system, absolute compliance with the rules, and no exceptional permeability. Even with these significant simplifications, which should theoretically favor this strategy, the results indicate the ineffectiveness of such a strategy in an urban context.

Considerable emphasis has been placed on mobility restrictions at different spatial scales before and during this pandemic. However, our study contributes to the established evidence that mobility reductions, at best, only allow for a modest delay in the epidemic's progression. This reduction must be unrealistically large in magnitude, thus significantly disrupting normal mobility patterns, and should be enacted almost immediately from the first detected cases. Consequently, we argue that the focus should instead be placed on the contact and mixing patterns of the population, which ultimately drive transmission of influenza-like illnesses.

Moving to Chapter 4, this work resonates closely with the previous one, as it also deals with the spatial spreading of an epidemic and the role of mobility. Driven by advancements in human mobility modeling and insights such as the exploration and preferential return mechanisms, and recognizing that epidemic metapopulation literature has yet to explore the implementation of more realistic microscopic mobility models, we identified an opportunity to contribute something novel. Among the family of EPR models, we selected the d-EPR model, aiming for a balance between realism and ease of implementation. As the model's exploration stage follows the gravity model of human mobility, the locations where the population mixes were assigned a value reflecting their attractiveness or relevance. This assignment is informed by real data from the Greater Boston Area, courtesy of the abundant high-resolution mobility data from mobile phone records. Furthermore, to adequately evaluate the d-EPR model's strengths and weaknesses, we investigated the spatial spreading of the disease under various settings and scenarios. The former pertains to the distribution of mobility patterns within the population, ranging from homogeneous to heterogeneous, while the latter involves variations of the original d-EPR model where key features are modified. We discovered that only the d-EPR model, with a

heterogeneous distribution of the parameter characterizing agents' mobility, can exhibit a discernible differential impact on epidemiological outcomes across mobility groups. Specifically, the non-Markovian nature of the exploration and preferential return mechanisms appears vital in demonstrating that this type of mobility is influential in shaping the epidemic's spread. Explorers accelerate the disease invasion, thereby facilitating its dissemination across the system, contributing proportionally more—and significantly faster—than returner profiles. As for disease prevalence, returner profiles are consistently less affected than explorers. However, the average infection times per mobility group show negligible differences. A closer examination of the infection dynamics reveals that agents are more frequently infected outside their home locations than within. This is simply due to the fact that agents, even those with low to medium mobility profiles, tend to spend more time outside their home location and frequent spots that are generally more attractive. Since more attractive locations are likelier to host contagion events, it is clear that exploration enhances disease diffusion. The ease with which even low-medium mobility agents can reach highly attractive locations leads to rapid and widespread dissemination of the disease throughout the system.

In conclusion, integrating the d-EPR model with an epidemiological model provides fresh perspectives on the interplay between mobility and the spread of an epidemic. Nonetheless, it must be acknowledged that our model and its implementation have limitations that should be addressed in future work, such as testing how the model compares with actual mobility trajectories and considering models that incorporate more realistic mobility features. For now, this represents a modest initial step towards acknowledging the empirically observed heterogeneity and complexities of human mobility and evaluating their implications in epidemic dynamics.

As we begin with *Part III*, Chapter 5 examines the effects of vaccine hesitancy in age-structured populations. This investigation spurred by the COVID-19 pandemic, coincides with the period when vaccines were being deployed following an unprecedented scientific endeavor that yielded effective vaccines in record time. Paradoxically, the rapid development of these vaccines, along with the pervasive *infodemics* across social networks and the Internet, sparked concerns among certain sectors of the population. This resulted in the dissemination of misinformation and outright falsehoods regarding the safety and efficacy of the COVID-19 vaccines. Utilizing surveys on vaccine acceptance conducted in the US, alongside detailed age-structured contact matrices for US states, we explored the potential impact of secondary COVID-19 outbreaks. The central inquiry was to ascertain the extent to which vaccine hesitancy could fuel significant outbreaks, especially in an environment dominated by a more transmissible SARS-CoV-2 variant of concern. This question

can be fully understood by recalling that after half a year of the beginning of the vaccination campaign, the highly transmissible Delta variant wave was building up across the US and the rest of the world. Our approach involved a SIR model segmented into 85 age brackets. Although the SIR model's simplicity might not fully capture the complexities of COVID-19, SARS-CoV-2 transmission dynamics, and the vaccination process, it is nevertheless informed by state-specific data on prevalence, demographics, vaccine acceptance surveys, and high-resolution age-structured contact patterns.

The analysis yielded a clear link between the magnitude of secondary outbreaks in each state and the level of vaccine hesitancy present. Delving into the influence of age distribution, we discovered a strong association between outbreak prevalence and the proportion of remaining susceptible individuals (after the initial wave) within each age group, with younger populations contributing most significantly to the transmission rates. However, caution is advised when projecting these correlations onto expected mortality figures, as the impact of age on death rates tends to be inversely related. Our model's projections suggested that reducing vaccine hesitancy by even one percentage point could potentially prevent numerous deaths, underlining the importance of considering the age structure in transmission dynamics. Notably, not all states with high hesitancy levels topped the list for preventable deaths; instead, the percentage of hesitant individuals among older populations emerged as a crucial factor. Comparing our model's outcomes with actual data revealed high correlations for both predicted vaccination rates against actual figures and estimated deaths against recorded fatalities during the Delta variant's surge in the US. To sum up, despite our model's relative simplicity, the use of granular data for key processes such as demographics, contact patterns, and vaccine sentiment surveys enabled the development of a fairly precise depiction of the repercussions of vaccine hesitancy during the ongoing COVID-19 pandemic. As part of the scientific discourse, this work contributes valuable data-backed insights and modeling to support the widespread adoption of vaccines.

We conclude in Chapter 6 with a second look at the role of vaccine hesitancy, this time with populations structured on contact networks, and explicitly modeling the interplay between vaccination and behavioral attitudes toward it. We present a model that intricately links the epidemiology of an influenza-like illness (using the SIR model), the dynamics of a vaccination campaign, and the social opinion dynamics regarding vaccine uptake. This latter aspect is informed by the Watts-Granovetter threshold model, with a novel twist: our model accounts for individuals' vaccination actions rather than merely their expressed opinions, drawing on U.S. survey data that suggests peer vaccination status significantly influences personal vaccination decisions.

Our analysis proceeds on two types of networks, Erdős-Rényi (ER) and Barabási-Albert (BA), examining the outcomes across a range of parameters including the pro-vaccine attitude activation threshold, the initial proportion of pro-vaccine individuals, and the vaccination rate.

Our findings reveal distinct dynamics within ER and BA networks. In ER networks, increasing the vaccination rate can foster a positive feedback loop between opinion dynamics and actual vaccine uptake, effectively reducing disease prevalence and potentially leading to disease eradication within a realistic parameter space. However, the emergence of a disease-free state in BA networks requires higher and less practical vaccination efforts, with overall higher disease prevalence compared to ER networks for any given vaccination rate. Additionally, the introduction of anti-vaccine individuals  $n_Z$  into the model demonstrates varied impacts on vaccination coverage and disease prevalence, dependent on the system's position in parameter space. The presence of anti-vaccine sentiment is negligible at low vaccination rates but becomes influential as vaccination rates increase. For a given pro-vaccine initial support  $n_A(0)$ , settings with different activation thresholds converge as the fraction of anti-vaccine individuals increases, yielding the threshold irrelevant at  $n_A(0) + n_Z = 1$  due to depletion of the hesitant individual pool. Furthermore, nonlinearities in vaccination coverage and prevalence emerge for increased vaccination ratio values.

Ultimately, the model is applied to the US context, incorporating regionally diverse vaccination thresholds obtained from survey data to assess the epidemic's impact at a national scale.

To bolster realism and practical impact, future directions of this work should focus on integrating age demographics into contact networks, and a more detailed representation of decision-making that accounts for cost-benefit considerations and the effects of social networks, such as highly influential individuals, misinformation, polarization, and the creation of echo chambers.

\* \* \*

This concludes the summary and key observations from the main work developed during my pre-doctoral training. While the studies address various questions and issues within epidemiology and epidemic spread modeling, they are not narrowly focused on any singular problem but rather span a breadth of topics. Nevertheless, certain fundamental principles are woven throughout this body of work. One such principle is the commitment to mechanistic methods that model the elementary processes leading to the observed complex dynamics from first principles. Given this complexity,

numerical and computational methods have been the primary and preferred means of implementing and resolving these models, rather than relying solely on analytical approaches. Although simulations of dynamical processes using mechanistic approaches can be computationally intensive and demanding in terms of memory usage, they afford the modeler the opportunity to create intricate microcosms that emulate real-life processes with a reasonably high level of detail. This approach enables the exploration of various configurations, adaptations, and hypothetical scenarios, providing a more lucid understanding of the dynamics at play.

Another thread linking these works is the integration of real-world data to inform our models wherever possible, thereby offering a more grounded portrayal of the phenomena under scrutiny within the scope of the models' limitations. While theoretical models can illuminate the foundational mechanisms of complex dynamics and the exploration of parameters can reveal the full spectrum of potential behaviors within a model, such models are tethered to reality only by empirical reference points. At times, models may become so complex and parameter-laden that what should be an exhilarating journey of discovery turns into an overwhelming challenge. In these cases, real-world data is invaluable for parameterizing key aspects such as mobility patterns, social mixing, or the disease's progression, ensuring that our efforts remain directed and our analyses sharp. Furthermore, to achieve relevance in our research and possibly influence policy-making or contribute constructively to public discourse, robust and high-quality data is indispensable.

## Perspectives

As the reader may have noticed through the journey that began in Chapter 1, the field of mathematical epidemiology and epidemic modeling has significantly evolved since its inception in the 19th and early 20th centuries. It has grown exponentially, much like the phenomena it aims to model, and has undergone various waves of discovery and innovation, with the COVID-19 pandemic being the latest catalyst. The pandemic has spurred an extensive body of work, both theoretical and empirical, within which the research presented here humbly takes its place. Yet, as we well understand, scientific progress is often a gradual process, with small steps building upon the shared knowledge and insights of the past. Over time, these contributions can accumulate and lead to breakthroughs, opening new avenues and triggering a burst of discoveries until that particular area of study matures and stabilizes, awaiting fresh questions and paradigms to reignite the cycle of innovation.

In a field thus so vast, interdisciplinary, and mature in many key aspects, seems

difficult, for barely a beginner in this field, to foresee what to expect or what venues are worth pursuing in the years to come. However, I will offer, from my short experience, a brief and general commentary on what I feel we need to push the field forward.

Reflecting on the evolution of this field—or indeed, any field of study—hypotheses and models often start from a rather basic and unrefined state. Initially grounded in evidence and observation, they can be quite rudimentary, but over years and even decades, these models undergo refinements or experience paradigm shifts that drive the field’s progression and sophistication. Take, for example, the modeling of the force of infection, which inherently involves modeling the process of contact or mixing among individuals. This concept began with the assumption of mass-action interaction and only after a century evolved with the advent of network epidemiology. Even as we stand on the cusp of a potential revolution in the study of higher-order interactions, the assumption of well-mixing or homogeneous mixing continues to hold sway in much of the literature and certain areas within mathematical epidemiology. Our own work presented here is not exempt from this reliance on the well-mixing assumption. In defense of this assumption and our work, it remains a justifiable simplification under specific circumstances, which explains its persistence. Nevertheless, there is no doubt that exploring theoretical and empirical alternatives to this assumption—those that are less computationally and data-intensive than network approaches—would be intriguing. Such alternatives could provide a more accurate representation of the contagion process.

Closely related to the well-mixing approach, the assumption of homogeneous populations extends not only to social interactions but also to other potentially relevant attributes affecting epidemic spread, such as sex, age, mobility patterns, or socioeconomic status. Decades after the foundational works in mathematical epidemiology, the endeavor to measure and incorporate heterogeneity, as well as to understand its impact, began—and it continues to this day. Recognizing population heterogeneity is not just an acknowledgment of diversity; it is a critical requirement. Heterogeneity significantly influences the dynamics of disease transmission and prevalence, offering a more nuanced and accurate depiction of epidemic spread and its repercussions. This understanding allows for more effective resource allocation and the formulation of efficient control or mitigation strategies.

Consider the international response to the COVID-19 pandemic as an example. Broad and indiscriminate actions were implemented in many regions, likely preventing overwhelming loss of life. Yet, the profound psychological, social, and economic turmoil caused by these measures still echoes throughout societies. Hopefully, a more nuanced understanding of how complex, heterogeneous societies interact with the spread of epidemics could have guided us in developing balanced strategies. Such strategies

would aim to contain or mitigate the spread while minimizing collateral damage to other aspects of people’s lives.

The underlying illness affecting all these issues raised that, again also invades other fields, is the insatiable demand for data. As noted earlier, access to high-quality, fine-grained data is essential to develop models that can accurately forecast the unfolding and progression of an epidemic in real-life settings or inform projections and hypothetical scenarios where, for instance, mobility restrictions or vaccination campaigns are deployed.

In Spain, as in many other places around the world, researchers and modelers often encounter substantial obstacles in accessing necessary data. These challenges encompass more than just data availability; they include concerns over data quality, the need for digitization, standardization discrepancies across regions, and the granularity of the data. Regrettably, it has become all too common for researchers to embark on arduous quests for this valuable data to feed their models. Compounding this issue is the dependency on data held by private corporations, which often (albeit understandably) limit access to the research community and the public. Data pertaining to mobility and human behavior, for instance, tend to be controlled by private entities, which hinders research efforts. Furthermore, the data needed to adequately parameterize models may not exist, usually due to a lack of comprehensive field studies and surveys—this is particularly true for understanding population mixing patterns.

It is crucial to acknowledge that these data shortfalls are not just the result of reluctance from involved parties; they also stem from a broader cultural issue: a lack of recognition of the essential role that science and research play in tackling and mitigating the most significant challenges faced by society. While researchers alone cannot solve this problem, we bear the responsibility to stress the importance of data accessibility to both authorities and private companies. Collaborative efforts are necessary to champion for improved data management that benefits the common good and equips us to meet societal challenges. Enhanced access to data is not merely a convenience; it is a vital requirement for making informed decisions, implementing effective public health measures, and, ultimately, contributing to our welfare.

# Chapter 8

## Bibliography

- [1] Jocelyne Piret and Guy Boivin. Pandemics throughout history. *Frontiers in microbiology*, 11:631736, 2021.
- [2] James W LeDuc and M Anita Barry. Sars, the first pandemic of the 21st century. *Emerging Infectious Diseases*, 10(11):e26, 2004.
- [3] JS Malik Peiris, Wen-wei Tu, and Hui-ling Yen. A novel h1n1 virus causes the first pandemic of the 21st century. *European journal of immunology*, 39(11):2946–2954, 2009.
- [4] Jean-Jacques Muyembe-Tamfum, S Mulangu, Justin Masumu, JM Kayembe, A Kemp, and Janusz T Paweska. Ebola virus outbreaks in africa: past and present. *Onderstepoort Journal of Veterinary Research*, 79(2):06–13, 2012.
- [5] World Health Organization et al. *World malaria report 2022*. World Health Organization, 2022.
- [6] Didier Raoult, Theodore Woodward, and J Stephen Dumler. The history of epidemic typhus. *Infectious Disease Clinics*, 18(1):127–140, 2004.
- [7] Divya Ganesan, Sanjukta Sen Gupta, and Dominique Legros. Cholera surveillance and estimation of burden of cholera. *Vaccine*, 38:A13–A17, 2020.
- [8] Michael JA Reid, Nimalan Arinaminpathy, Amy Bloom, Barry R Bloom, Catharina Boehme, Richard Chaisson, Daniel P Chin, Gavin Churchyard, Helen Cox, Lucica Ditiu, et al. Building a tuberculosis-free world: The lancet commission on tuberculosis. *The Lancet*, 393(10178):1331–1384, 2019.
- [9] Anthony S Fauci. Emerging and reemerging infectious diseases: the perpetual challenge. *Academic Medicine*, 80(12):1079–1085, 2005.
- [10] Bill Gates. Responding to covid-19—a once-in-a-century pandemic? *New england Journal of medicine*, 382(18):1677–1679, 2020.
- [11] Chaolin Huang, Yeming Wang, Xingwang Li, Lili Ren, Jianping Zhao, Yi Hu, Li Zhang, Guohui Fan, Jiuyang Xu, Xiaoying Gu, et al. Clinical features of patients infected with 2019 novel coronavirus in wuhan, china. *The lancet*, 395(10223):497–506, 2020.
- [12] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy SM Leung, Eric HY Lau, Jessica Y Wong, et al. Early transmission dynamics in wuhan, china, of novel coronavirus–infected pneumonia. *New England journal of medicine*, 382(13):1199–1207, 2020.
- [13] Xiaorong Wang, Qiong Zhou, Yukun He, Lingbo Liu, Xinqian Ma, Xiaoshan Wei, Nanchuan Jiang, Limei Liang, Yali Zheng, Ling Ma, et al. Nosocomial outbreak of covid-19 pneumonia in wuhan, china. *European Respiratory Journal*, 55(6), 2020.
- [14] Tanu Singhal. A review of coronavirus disease-2019 (covid-19). *The indian journal of pediatrics*, 87(4):281–286, 2020.
- [15] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The role of the airline transportation network in the prediction and predictability of global epidemics. *Proceedings of the National Academy of Sciences*, 103(7):2015–2020, 2006.

- [16] Matteo Chinazzi, Jessica T Davis, Marco Ajelli, Corrado Gioannini, Maria Litvinova, Stefano Merler, Ana Pastore y Piontti, Kunpeng Mu, Luca Rossi, Kaiyuan Sun, et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (covid-19) outbreak. *Science*, 368(6489):395–400, 2020.
- [17] P Sam. Redefining vulnerability in the era of covid-19. *Lancet*, 395(10230):1089, 2020.
- [18] Jogender Kumar and Praveen Kumar. Covid-19 pandemic and health-care disruptions: count the most vulnerable. *The Lancet Global Health*, 9(6):e722–e723, 2021.
- [19] Orestis Delardas, Konstantinos S Kechagias, Pantelis N Pontikos, and Panagiotis Giannos. Socio-economic impacts and challenges of the coronavirus pandemic (covid-19): an updated review. *Sustainability*, 14(15):9699, 2022.
- [20] Mario Draghi. We face a war against coronavirus and must mobilise accordingly. *Financial Times*, 25(3), 2020.
- [21] Reed Wood, Gina Yannitell Reinhardt, Babak RezaeeDaryakenari, and Leah C Windsor. Resisting lockdown: The influence of covid-19 restrictions on social unrest. *International Studies Quarterly*, 66(2):sqac015, 2022.
- [22] Javid Moosavi, Amir M Fathollahi-Fard, and Maxim A Dulebenets. Supply chain disruption during the covid-19 pandemic: Recognizing potential disruption management strategies. *International Journal of Disaster Risk Reduction*, 75:102983, 2022.
- [23] Emeline Han, Melisa Mei Jin Tan, Eva Turk, Devi Sridhar, Gabriel M Leung, Kenji Shibuya, Nima Asgari, Juhwan Oh, Alberto L García-Basteiro, Johanna Hanefeld, et al. Lessons learnt from easing covid-19 restrictions: an analysis of countries and regions in asia pacific and europe. *The Lancet*, 396(10261):1525–1534, 2020.
- [24] Nicola Perra. Non-pharmaceutical interventions during the covid-19 pandemic: A review. *Physics Reports*, 913:1–52, 2021.
- [25] Claudio Borio. The covid-19 economic crisis: Dangerously unique. *Business Economics*, 55:181–190, 2020.
- [26] David L Blustein, Ryan Duffy, Joaquim A Ferreira, Valerie Cohen-Scali, Rachel Gali Cinamon, and Blake A Allan. Unemployment in the time of covid-19: A research agenda. *Journal of vocational behavior*, 119:103436, 2020.
- [27] SreyRam Kuy, Raymond Tsai, Jay Bhatt, Quyen D Chu, Pritesh Gandhi, Rohit Gupta, Reshma Gupta, Michael K Hole, Benson S Hsu, Lauren S Hughes, et al. Focusing on vulnerable populations during covid-19. *Academic Medicine*, 95(11):e2–e3, 2020.
- [28] Neeta Kantamneni. The impact of the covid-19 pandemic on marginalized populations in the united states: A research agenda. *Journal of vocational behavior*, 119:103439, 2020.
- [29] John D Sterman. Learning in and about complex systems. *System dynamics review*, 10(2-3):291–330, 1994.
- [30] Sunny Y Auyang. *Foundations of complex-system theories: in economics, evolutionary biology, and statistical physics*. Cambridge University Press, 1998.
- [31] Sandra D Mitchell. *Unsimple truths: Science, complexity, and policy*. University of Chicago Press, 2009.
- [32] Albert-László Barabási. The network takeover. *Nature Physics*, 8(1):14–16, 2012.
- [33] Danielle S Bassett and Michael S Gazzaniga. Understanding complexity in the human brain. *Trends in cognitive sciences*, 15(5):200–209, 2011.
- [34] Sven E Jørgensen, Bernard C Patten, and Milan Straškraba. Ecosystems emerging: toward an ecology of complex systems in a complex future. *Ecological Modelling*, 62(1-3):1–27, 1992.
- [35] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Scale-free characteristics of random networks: the topology of the world-wide web. *Physica A: statistical mechanics and its applications*, 281(1-4):69–77, 2000.

- [36] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. Statistical physics of social dynamics. *Reviews of modern physics*, 81(2):591, 2009.
- [37] Michael Batty. Cities as complex systems: Scaling, interaction, networks, dynamics and urban morphologies. 2009.
- [38] Luís MA Bettencourt. Cities as complex systems. *Modeling complex systems for public policies*, pages 217–236, 2015.
- [39] Roger White, Guy Engelen, and Inge Uljee. *Modeling cities and regions as complex systems: From theory to planning applications*. MIT Press, 2015.
- [40] Didier Sornette. *Why stock markets crash: critical events in complex financial systems*. Princeton university press, 2009.
- [41] Alan Kirman. The economy as a complex system. *Economic Foundations for Social Complexity Science: Theory, Sentiments, and Empirical Laws*, pages 1–16, 2017.
- [42] Philip W Anderson. *The economy as an evolving complex system*. CRC Press, 2018.
- [43] W Brian Arthur. *The economy as an evolving complex system II*. CRC Press, 2018.
- [44] Ernesto Estrada. What is a complex system, after all? *Foundations of Science*, pages 1–28, 2023.
- [45] Vittoria Colizza, Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. The modeling of global epidemics: Stochastic dynamics and predictability. *Bulletin of mathematical biology*, 68(8):1893–1921, 2006.
- [46] George David Birkhoff. *Dynamical systems*, volume 9. American Mathematical Soc., 1927.
- [47] Leo P Kadanoff. More is the same; phase transitions and mean field theories. *Journal of Statistical Physics*, 137:777–797, 2009.
- [48] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [49] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4-5):175–308, 2006.
- [50] Nicolaas Godfried Van Kampen. *Stochastic processes in physics and chemistry*, volume 1. Elsevier, 1992.
- [51] Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779–782, 2008.
- [52] Kelly R Moran, Geoffrey Fairchild, Nicholas Generous, Kyle Hickmann, Dave Osthuis, Reid Piedhorsky, James Hyman, and Sara Y Del Valle. Epidemic forecasting is messier than weather forecasting: The role of human behavior and internet data streams in epidemic forecast. *The Journal of infectious diseases*, 214(suppl\_4):S404–S408, 2016.
- [53] Daniel Bernoulli and Dominique Chapelle. Essai d'une nouvelle analyse de la mortalité causée par la petite vérole, et des avantages de l'inoculation pour la prévenir. 1760.
- [54] Klaus Dietz and JAP Heesterbeek. Bernoulli was ahead of modern epidemiology. *Nature*, 408(6812):513–514, 2000.
- [55] Klaus Dietz and JAP Heesterbeek. Daniel bernoulli's epidemiological model revisited. *Mathematical biosciences*, 180(1-2):1–21, 2002.
- [56] Robert E Serfling. Historical review of epidemic theory. *Human biology*, 24(3):145–166, 1952.
- [57] Klaus Dietz. The first epidemic model: a historical note on pd en'ko. *Australian Journal of Statistics*, 30(1):56–65, 1988.
- [58] PD En'ko. The epidemic course of some infectious diseases (in russian). *Vrac'*, 10:1008–1010, 1889.
- [59] Ronald Ross. The mathematics of malaria. *British medical journal*, 1(2626):1023, 1911.

- [60] Ronald Ross. An application of the theory of probabilities to the study of a priori pathometry.—part i. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 92(638):204–230, 1916.
- [61] Ronald Ross and Hilda P Hudson. An application of the theory of probabilities to the study of a priori pathometry.—part ii. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 93(650):212–224, 1917.
- [62] Ronald Ross and Hilda P Hudson. An application of the theory of probabilities to the study of a priori pathometry.—part iii. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 93(650):225–240, 1917.
- [63] William Heaton 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.
- [64] D Daley J Gani and D Daley. Epidemic modeling: an introduction. *Cambridge U. Press, Cambridge*, 1999.
- [65] Hans Heesterbeek. The law of mass-action in epidemiology: a historical perspective. *Ecological paradigms lost: routes of theory change*, pages 81–104, 2005.
- [66] Roy M Anderson. Discussion: the kermack-mckendrick epidemic threshold theorem. *Bulletin of mathematical biology*, 53:1–32, 1991.
- [67] William Ogilvy Kermack and Anderson G McKendrick. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, 115(772):700–721, 1927.
- [68] William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. ii.—the problem of endemicity. *Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character*, 138(834):55–83, 1932.
- [69] William Ogilvy Kermack and Anderson G McKendrick. Contributions to the mathematical theory of epidemics. iii.—further studies of the problem of endemicity. *Proceedings of the Royal Society of London. Series A, Containing papers of a mathematical and physical character*, 141(843):94–122, 1933.
- [70] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics iv. analysis of experimental epidemics of the virus disease mouse ectromelia. *Epidemiology & Infection*, 37(2):172–187, 1937.
- [71] WO Kermack and AG McKendrick. Contributions to the mathematical theory of epidemics: V. analysis of experimental epidemics of mouse-typoid; a bacterial disease conferring incomplete immunity. *Epidemiology & Infection*, 39(3):271–288, 1939.
- [72] Mirjam Kretzschmar and Jacco Wallinga. Mathematical models in infectious disease epidemiology. *Modern infectious disease epidemiology: Concepts, methods, mathematical models, and public health*, pages 209–221, 2010.
- [73] AG M'kendrick. Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 44:98–130, 1925.
- [74] Helen Abbey. An examination of the reed-frost theory of epidemics. *Human biology*, 24(3):201, 1952.
- [75] John A Jacquez. A note on chain-binomial models of epidemic spread: what is wrong with the reed-frost formulation? *Mathematical Biosciences*, 87(1):73–82, 1987.
- [76] Lisa Sattenspiel. Modeling the spread of infectious disease in human populations. *American Journal of Physical Anthropology*, 33(S11):245–276, 1990.
- [77] MS Bartlett. Some evolutionary stochastic processes. *Journal of the Royal Statistical Society. Series B (Methodological)*, 11(2):211–229, 1949.
- [78] Norman TJ Bailey. A simple stochastic epidemic. *Biometrika*, pages 193–202, 1950.

- [79] Norman TJ Bailey. The total size of a general stochastic epidemic. *Biometrika*, pages 177–185, 1953.
- [80] M Bartlett. Deterministic and stochastic models for recurrent epidemics, vol. 4. *Proceedings of the Third Berkeley*, 1956.
- [81] Norman TJ Bailey. The simple stochastic epidemic: a complete solution in terms of known functions. *Biometrika*, pages 235–240, 1963.
- [82] Norman TJ Bailey. Some stochastic models for small epidemics in large populations. *Applied Statistics*, pages 9–19, 1964.
- [83] RM Anderson and RM May. Age-related changes in the rate of disease transmission: implications for the design of vaccination programmes. *Epidemiology & Infection*, 94(3):365–436, 1985.
- [84] Klaus Dietz and Dieter Schenzle. Proportionate mixing models for age-dependent infection transmission. *Journal of mathematical biology*, 22:117–120, 1985.
- [85] DJ Nokes, RM Anderson, and MJ Anderson. Rubella epidemiology in south east england. *Epidemiology & Infection*, 96(2):291–304, 1986.
- [86] RM Anderson, JA Crombie, and BT Grenfell. The epidemiology of mumps in the uk: a preliminary study of virus transmission, herd immunity and the potential impact of immunization. *Epidemiology & Infection*, 99(1):65–84, 1987.
- [87] Bryan Thomas Grenfell and Roy Malcolm Anderson. Pertussis in england and wales: an investigation of transmission dynamics and control by mass vaccination. *Proceedings of the Royal Society of London. B. Biological Sciences*, 236(1284):213–252, 1989.
- [88] Carlos Castillo-Chavez, Herbert W Hethcote, Viggo Andreasen, Simon A Levin, and Wei M Liu. Epidemiological models with age structure, proportionate mixing, and cross-immunity. *Journal of mathematical biology*, 27:233–258, 1989.
- [89] Robert M May and Roy M Anderson. Commentary transmission dynamics of hiv infection. *Nature*, 326(137):10–1038, 1987.
- [90] James Koopman, Carl Simon, John Jacquez, Jill Joseph, Lisa Sattenspiel, and Taesung Park. Sexual partner selectiveness effects on homosexual hiv transmission dynamics. *Journal of Acquired Immune Deficiency Syndromes*, 1(5):486–504, 1988.
- [91] RM Anderson, TW Ng, MC Boily, and RM May. The influence of different sexual-contact patterns between age classes on the predicted demographic impact of aids in developing countries. *Annals of the New York Academy of Sciences*, 569:240–274, 1989.
- [92] Roy Malcolm Anderson, SP Blythe, Sunetra Gupta, and E Konings. The transmission dynamics of the human immunodeficiency virus type 1 in the male homosexual community in the united kingdom: the influence of changes in sexual behaviour. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 325(1226):45–98, 1989.
- [93] Carlos Castillo-Chavez and Stephen P Blythe. Mixing framework for social/sexual behavior. In *Mathematical and statistical approaches to AIDS epidemiology*, pages 275–288. Springer, 1989.
- [94] Herbert W Hethcote, Harlan W Stech, and Pauline Van Den Driessche. Nonlinear oscillations in epidemic models. *SIAM Journal on Applied Mathematics*, 40(1):1–9, 1981.
- [95] Wei-min Liu, Herbert W Hethcote, and Simon A Levin. Dynamical behavior of epidemiological models with nonlinear incidence rates. *Journal of mathematical biology*, 25:359–380, 1987.
- [96] Herbert W Hethcote and Pauline van den Driessche. Some epidemiological models with nonlinear incidence. *Journal of Mathematical Biology*, 29(3):271–287, 1991.
- [97] Joan L Aron and Ira B Schwartz. Seasonality and period-doubling bifurcations in an epidemic model. *Journal of theoretical biology*, 110(4):665–679, 1984.
- [98] Roy M Anderson, Bryan T Grenfell, and RM May. Oscillatory fluctuations in the incidence of infectious disease and the impact of vaccination: time series analysis. *Epidemiology & Infection*, 93(3):587–608, 1984.

- [99] Herbert W Hethcote. A model for hiv transmission and aids. In *Mathematical Approaches to Problems in Resource Management and Epidemiology: Proceedings of a Conference held at Ithaca, NY, Oct. 28–30, 1987*, pages 164–176. Springer, 1989.
- [100] Robert M May and Roy M Anderson. The transmission dynamics of human immunodeficiency virus (hiv). In *Applied Mathematical Ecology*, pages 263–311. Springer, 1989.
- [101] John A Jacquez, Carl P Simon, and James Koopman. Structured mixing: heterogeneous mixing by the definition of activity groups. In *Mathematical and statistical approaches to AIDS epidemiology*, pages 301–315. Springer, 1989.
- [102] Lisa Sattenspiel, James Koopman, Carl Simon, and John A Jacquez. The effects of population structure on the spread of the hiv infection. *American Journal of Physical Anthropology*, 82(4):421–429, 1990.
- [103] Peter Haggett. Contagious processes in a planar graph: an epidemiological application. *Medical Geography: Techniques and Field Studies*, pages 307–324, 1972.
- [104] Peter Haggett. Hybridizing alternative models of an epidemic diffusion process. *Economic Geography*, 52(2):136–146, 1976.
- [105] Andrew David Cliff. *Spatial diffusion: an historical geography of epidemics in an island community*, volume 14. CUP Archive, 1981.
- [106] AR McLean and Roy M Anderson. Measles in developing countries part i. epidemiological parameters and patterns. *Epidemiology & Infection*, 100(1):111–133, 1988.
- [107] AR McLean and RM Anderson. Measles in developing countries. part ii. the predicted impact of mass vaccination. *Epidemiology & Infection*, 100(3):419–442, 1988.
- [108] OV Baroyan, LA Rvachev, and Yu G Ivannikov. Modelling and prediction of influenza epidemics in the ussr. *Moscow, Gamelei Institute of Epidemiology and Microbiology*, 1977.
- [109] Leonid A Rvachev and Ira M Longini Jr. A mathematical model for the global spread of influenza. *Mathematical biosciences*, 75(1):3–22, 1985.
- [110] Ira M Longini Jr. A mathematical model for predicting the geographic spread of new infectious agents. *Mathematical Biosciences*, 90(1-2):367–383, 1988.
- [111] Curtis C Travis and Suzanne M Lenhart. Eradication of infectious diseases in heterogeneous populations. *Mathematical biosciences*, 83(2):191–198, 1987.
- [112] Lisa Sattenspiel. Population structure and the spread of disease. *Human Biology*, pages 411–438, 1987.
- [113] Lisa Sattenspiel and Carl P Simon. The spread and persistence of infectious diseases in structured populations. *Mathematical Biosciences*, 90(1-2):341–366, 1988.
- [114] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1991.
- [115] K Dietz and D Schenzle. Mathematical models for infectious disease statistics. *A Celebration of Statistics: The ISI Centenary Volume A Volume to Celebrate the Founding of the International Statistical Institute in 1885*, pages 167–204, 1985.
- [116] HW Hethcote. Modeling heterogeneous mixing in infectious disease dynamics. *Models for Infectious Human Diseases: Their Structure and Relation to Data*, 6:215, 1996.
- [117] Odo Diekmann, Johan Andre Peter Heesterbeek, and Johan AJ Metz. On the definition and the computation of the basic reproduction ratio  $r_0$  in models for infectious diseases in heterogeneous populations. *Journal of mathematical biology*, 28(4):365–382, 1990.
- [118] Klaus Dietz. The estimation of the basic reproduction number for infectious diseases. *Statistical methods in medical research*, 2(1):23–41, 1993.
- [119] JAP Heesterbeek and Klaus Dietz. The concept of  $r_0$  in epidemic theory. *Statistica neerlandica*, 50(1):89–110, 1996.

- [120] Niels Becker. A general chain binomial model for infectious diseases. *Biometrics*, pages 251–258, 1981.
- [121] Ira M Longini Jr, James S Koopman, Arnold S Monto, and John P Fox. Estimating household and community transmission parameters for influenza. *American journal of epidemiology*, 115(5):736–751, 1982.
- [122] Ira M Longini Jr and James S Koopman. Household and community transmission parameters from final distributions of infections in households. *Biometrics*, pages 115–126, 1982.
- [123] Michael Haber, Ira M Longini Jr, and George A Cotsonis. Models for the statistical analysis of infectious disease data. *Biometrics*, pages 163–173, 1988.
- [124] Frank Ball and Damian Clancy. The final size and severity of a generalised stochastic multitype epidemic model. *Advances in applied probability*, pages 721–736, 1993.
- [125] Frank Ball, Denis Mollison, and Gianpaolo Scalia-Tomba. Epidemics with two levels of mixing. *The Annals of Applied Probability*, pages 46–89, 1997.
- [126] Mart CM De Jong, Odo Diekmann, and Hans Heesterbeek. How does transmission of infection depend on population size. *Epidemic models: their structure and relation to data*, 84, 1995.
- [127] O Diekmann, MCM De Jong, AA De Koeijer, and P Reijnders. The force of infection in populations of varying size: a modelling problem. *Journal of Biological Systems*, 3(02):519–529, 1995.
- [128] Odo Diekmann, MCM De Jong, and Johan Anton Jacob Metz. A deterministic epidemic model taking account of repeated contacts between the same individuals. *Journal of Applied Probability*, 35(2):448–462, 1998.
- [129] Michael Begon, Malcolm Bennett, Roger G Bowers, Nigel P French, SM Hazel, and Joseph Turner. A clarification of transmission terms in host-microparasite models: numbers, densities and areas. *Epidemiology & Infection*, 129(1):147–153, 2002.
- [130] Lisa Sattenspiel and Christopher Powell. Geographic spread of measles on the island of dominica, west indies. *Human biology*, pages 107–129, 1993.
- [131] Benjamin Bolker and Bryan Thomas Grenfell. Space, persistence and dynamics of measles epidemics. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 348(1325):309–320, 1995.
- [132] BM Bolker and BT Grenfell. Impact of vaccination on the spatial correlation and persistence of measles dynamics. *Proceedings of the National Academy of Sciences*, 93(22):12648–12653, 1996.
- [133] Alun L Lloyd and Robert M May. Spatial heterogeneity in epidemic models. *Journal of theoretical biology*, 179(1):1–11, 1996.
- [134] Matthew J Keeling and Bryan T Grenfell. Disease extinction and community size: modeling the persistence of measles. *Science*, 275(5296):65–67, 1997.
- [135] Sunetra Gupta, Roy M Anderson, and Robert M May. Networks of sexual contacts: implications for the pattern of spread of hiv. *Aids*, 3(12):807–818, 1989.
- [136] Andrew Barbour and Denis Mollison. Epidemics and random graphs. In *Stochastic Processes in Epidemic Theory: Proceedings of a Conference held in Luminy, France, October 23–29, 1988*, pages 86–89. Springer, 1990.
- [137] Albert-László Barabási, Réka Albert, and Hawoong Jeong. Mean-field theory for scale-free random networks. *Physica A: Statistical Mechanics and its Applications*, 272(1-2):173–187, 1999.
- [138] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’networks. *nature*, 393(6684):440–442, 1998.
- [139] Cristopher Moore and Mark EJ Newman. Epidemics and percolation in small-world networks. *Physical Review E*, 61(5):5678, 2000.
- [140] Mark EJ Newman. Spread of epidemic disease on networks. *Physical review E*, 66(1):016128, 2002.

- [141] Matthew J Keeling. The effects of local spatial structure on epidemiological invasions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 266(1421):859–867, 1999.
- [142] Mark EJ Newman. Power laws, pareto distributions and zipf’s law. *Contemporary physics*, 46(5):323–351, 2005.
- [143] Sergei N Dorogovtsev and José FF Mendes. *Evolution of networks: From biological nets to the Internet and WWW*. Oxford university press, 2003.
- [144] Mark Newman, Albert-László Barabási, and Duncan J Watts. *The structure and dynamics of networks*. Princeton university press, 2011.
- [145] Mark Newman. *Networks*. Oxford university press, 2018.
- [146] Luciano da Fontoura Costa, Osvaldo N Oliveira Jr, Gonzalo Travieso, Francisco Aparecido Rodrigues, Paulino Ribeiro Villas Boas, Lucas Antiqueira, Matheus Palhares Viana, and Luis Enrique Correa Rocha. Analyzing and modeling real-world phenomena with complex networks: a survey of applications. *Advances in Physics*, 60(3):329–412, 2011.
- [147] Giulio Cimini, Tiziano Squartini, Fabio Saracco, Diego Garlaschelli, Andrea Gabrielli, and Guido Caldarelli. The statistical physics of real-world networks. *Nature Reviews Physics*, 1(1):58–71, 2019.
- [148] Béla Bollobás and Béla Bollobás. *Random graphs*. Springer, 1998.
- [149] Ken TD Eames and Jonathan M Read. Networks in epidemiology. In *Bio-Inspired Computing and Communication: First Workshop on Bio-Inspired Design of Networks, BIOWIRE 2007 Cambridge, UK, April 2-5, 2007 Revised Selected Papers*, pages 79–90. Springer, 2008.
- [150] Leon Danon, Ashley P Ford, Thomas House, Chris P Jewell, Matt J Keeling, Gareth O Roberts, Joshua V Ross, Matthew C Vernon, et al. Networks and the epidemiology of infectious disease. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.
- [151] Matt Keeling. The implications of network structure for epidemic dynamics. *Theoretical population biology*, 67(1):1–8, 2005.
- [152] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001.
- [153] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. *Physical review letters*, 86(14):3200, 2001.
- [154] Yamir Moreno, Romualdo Pastor-Satorras, and Alessandro Vespignani. Epidemic outbreaks in complex heterogeneous networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 26:521–529, 2002.
- [155] Yamir Moreno, Javier B Gómez, and Amalio F Pacheco. Epidemic incidence in correlated complex networks. *Physical Review E*, 68(3):035103, 2003.
- [156] Yamir Moreno and Alexei Vazquez. Disease spreading in structured scale-free networks. *The European Physical Journal B-Condensed Matter and Complex Systems*, 31:265–271, 2003.
- [157] Alain Barrat, Marc Barthelemy, Romualdo Pastor-Satorras, and Alessandro Vespignani. The architecture of complex weighted networks. *Proceedings of the national academy of sciences*, 101(11):3747–3752, 2004.
- [158] Marc Barthélémy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. *Journal of theoretical biology*, 235(2):275–288, 2005.
- [159] Pascal Crepey, Fabián P Alvarez, and Marc Barthélémy. Epidemic variability in complex networks. *Physical Review E*, 73(4):046131, 2006.
- [160] Albert-László Barabási and Eric Bonabeau. Scale-free networks. *Scientific american*, 288(5):60–69, 2003.
- [161] Romualdo Pastor-Satorras, Alessandro Vespignani, et al. Epidemics and immunization in scale-free networks. *Handbook of Graphs and Networks*, Wiley-VCH, Berlin, 2003.

- [162] Nilly Madar, Tomer Kalisky, Reuven Cohen, Daniel Ben-avraham, and Shlomo Havlin. Immunization and epidemic dynamics in complex networks. *The European Physical Journal B*, 38:269–276, 2004.
- [163] Guilherme Ferraz de Arruda, Francisco A Rodrigues, and Yamir Moreno. Fundamentals of spreading processes in single and multilayer complex networks. *Physics Reports*, 756:1–59, 2018.
- [164] István Z Kiss, Joel C Miller, Péter L Simon, et al. Mathematics of epidemics on networks. *Cham: Springer*, 598:31, 2017.
- [165] Kjell Hausken and John F Moxnes. A closure approximation technique for epidemic models. *Mathematical and Computer Modelling of Dynamical Systems*, 16(6):555–574, 2010.
- [166] Kieran J Sharkey. Deterministic epidemic models on contact networks: Correlations and unbiological terms. *Theoretical population biology*, 79(4):115–129, 2011.
- [167] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.
- [168] Piet Van Mieghem. The n-intertwined sis epidemic network model. *Computing*, 93(2-4):147–169, 2011.
- [169] Angélica S Mata and Silvio C Ferreira. Pair quenched mean-field theory for the susceptible-infected-susceptible model on complex networks. *Europhysics Letters*, 103(4):48003, 2013.
- [170] Silvio C Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. *Physical Review E*, 86(4):041125, 2012.
- [171] Raúl Toral and Pere Colet. *Stochastic numerical methods: an introduction for students and scientists*. John Wiley & Sons, 2014.
- [172] Hiroki Sayama. *Introduction to the modeling and analysis of complex systems*. Open SUNY Textbooks, 2015.
- [173] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of computational physics*, 22(4):403–434, 1976.
- [174] Daniel T Gillespie. Exact stochastic simulation of coupled chemical reactions. *The journal of physical chemistry*, 81(25):2340–2361, 1977.
- [175] Naoki Masuda and Renaud Lambiotte. *A guide to temporal networks*. World Scientific, 2016.
- [176] Manlio De Domenico, Albert Solé-Ribalta, Emanuele Cozzo, Mikko Kivelä, Yamir Moreno, Mason A Porter, Sergio Gómez, and Alex Arenas. Mathematical formulation of multilayer networks. *Physical Review X*, 3(4):041022, 2013.
- [177] Emanuele Cozzo, Guilherme Ferraz de Arruda, Francisco A Rodrigues, and Yamir Moreno. Multilayer networks: metrics and spectral properties. *Interconnected networks*, pages 17–35, 2016.
- [178] Alberto Aleta and Yamir Moreno. Multilayer networks in a nutshell. *Annual Review of Condensed Matter Physics*, 10:45–62, 2019.
- [179] Naoki Masuda and Petter Holme. *Introduction to temporal network epidemiology*. Springer, 2017.
- [180] Márton Karsai, Hang-Hyun Jo, Kimmo Kaski, et al. *Bursty human dynamics*. Springer, 2018.
- [181] Jukka-Pekka Onnela, Jari Saramäki, Jörkki Hyvönen, Gábor Szabó, M Argollo De Menezes, Kimmo Kaski, Albert-László Barabási, and János Kertész. Analysis of a large-scale weighted network of one-to-one human communication. *New journal of physics*, 9(6):179, 2007.
- [182] Ciro Cattuto, Wouter Van den Broeck, Alain Barrat, Vittoria Colizza, Jean-François Pinton, and Alessandro Vespignani. Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one*, 5(7):e11596, 2010.

- [183] Manlio De Domenico, Clara Granell, Mason A Porter, and Alex Arenas. The physics of spreading processes in multilayer networks. *Nature Physics*, 12(10):901–906, 2016.
- [184] Iacopo Iacopini, Giovanni Petri, Alain Barrat, and Vito Latora. Simplicial models of social contagion. *Nature communications*, 10(1):2485, 2019.
- [185] Federico Battiston, Giulia Cencetti, Iacopo Iacopini, Vito Latora, Maxime Lucas, Alice Patania, Jean-Gabriel Young, and Giovanni Petri. Networks beyond pairwise interactions: Structure and dynamics. *Physics Reports*, 874:1–92, 2020.
- [186] Guillaume St-Onge, Vincent Thibeault, Antoine Allard, Louis J Dubé, and Laurent Hébert-Dufresne. Master equation analysis of mesoscopic localization in contagion dynamics on higher-order networks. *Physical Review E*, 103(3):032301, 2021.
- [187] Guillaume St-Onge, Vincent Thibeault, Antoine Allard, Louis J Dubé, and Laurent Hébert-Dufresne. Social confinement and mesoscopic localization of epidemics on networks. *Physical review letters*, 126(9):098301, 2021.
- [188] Guillaume St-Onge, Hanlin Sun, Antoine Allard, Laurent Hébert-Dufresne, and Ginestra Bianconi. Universal nonlinear infection kernel from heterogeneous exposure on higher-order networks. *Physical review letters*, 127(15):158301, 2021.
- [189] Alain Barrat, Guilherme Ferraz de Arruda, Iacopo Iacopini, and Yamir Moreno. Social contagion on higher-order structures. In *Higher-Order Systems*, pages 329–346. Springer, 2022.
- [190] Federico Battiston, Enrico Amico, Alain Barrat, Ginestra Bianconi, Guilherme Ferraz de Arruda, Benedetta Franceschiello, Iacopo Iacopini, Sonia Kéfi, Vito Latora, Yamir Moreno, et al. The physics of higher-order interactions in complex systems. *Nature Physics*, 17(10):1093–1098, 2021.
- [191] James Dickson Murray and James Dickson Murray. *Mathematical Biology: II: Spatial Models and Biomedical Applications*, volume 3. Springer, 2003.
- [192] Vittoria Colizza, Marc Barthélémy, Alain Barrat, and Alessandro Vespignani. Epidemic modeling in complex realities. *Comptes rendus biologies*, 330(4):364–374, 2007.
- [193] Dimitrios Tsiotas and Vassilis Tselios. Understanding the uneven spread of covid-19 in the context of the global interconnected economy. *Scientific Reports*, 12(1):666, 2022.
- [194] Richard Levins. Some demographic and genetic consequences of environmental heterogeneity for biological control. *Bulletin of the ESA*, 15(3):237–240, 1969.
- [195] Rebecca F Grais, J Hugh Ellis, and Gregory E Glass. Assessing the impact of airline travel on the geographic spread of pandemic influenza. *European journal of epidemiology*, 18:1065–1072, 2003.
- [196] RF Grais, J Hugh Ellis, A Kress, and GE Glass. Modeling the spread of annual influenza epidemics in the us: The potential role of air travel. *Health care management science*, 7:127–134, 2004.
- [197] Antoine Flahault and Alain-Jacques Valleron. A method for assessing the global spread of hiv-1 infection based on air travel. *Mathematical Population Studies*, 3(3):161–171, 1992.
- [198] Gerardo Chowell, Paul W Fenimore, Melissa A Castillo-Garsow, and Carlos Castillo-Chavez. Sars outbreaks in ontario, hong kong and singapore: the role of diagnosis and isolation as a control mechanism. *Journal of theoretical biology*, 224(1):1–8, 2003.
- [199] Lars Hufnagel, Dirk Brockmann, and Theo Geisel. Forecast and control of epidemics in a globalized world. *Proceedings of the national academy of sciences*, 101(42):15124–15129, 2004.
- [200] Vittoria Colizza, Alain Barrat, Marc Barthélémy, and Alessandro Vespignani. Predictability and epidemic pathways in global outbreaks of infectious diseases: the sars case study. *BMC medicine*, 5:1–13, 2007.
- [201] Lisa Sattenspiel and Klaus Dietz. A structured epidemic model incorporating geographic mobility among regions. *Mathematical biosciences*, 128(1-2):71–91, 1995.

- [202] Matt J Keeling and Pejman Rohani. Estimating spatial coupling in epidemiological systems: a mechanistic approach. *Ecology letters*, 5(1):20–29, 2002.
- [203] Shigui Ruan, Wendi Wang, Simon A Levin, et al. The effect of global travel on the spread of sars. *Mathematical Biosciences and Engineering*, 3(1):205, 2006.
- [204] David JD Earn, Pejman Rohani, and Bryan T Grenfell. Persistence, chaos and synchrony in ecology and epidemiology. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 265(1390):7–10, 1998.
- [205] Pejman Rohani, David JD Earn, and Bryan T Grenfell. Opposite patterns of synchrony in sympatric disease metapopulations. *Science*, 286(5441):968–971, 1999.
- [206] Matt J Keeling. Metapopulation moments: coupling, stochasticity and persistence. *Journal of Animal Ecology*, 69(5):725–736, 2000.
- [207] Andrew W Park, Simon Gubbins, and Christopher A Gilligan. Extinction times for closed epidemics: the effects of host spatial structure. *Ecology Letters*, 5(6):747–755, 2002.
- [208] Dirk Brockmann. Human mobility and spatial disease dynamics. *Reviews of nonlinear dynamics and complexity*, 2:1–24, 2009.
- [209] Vittoria Colizza, Alain Barrat, Marc Barthelemy, Alain-Jacques Valleron, and Alessandro Vespignani. Modeling the worldwide spread of pandemic influenza: baseline case and containment interventions. *PLoS medicine*, 4(1):e13, 2007.
- [210] Paolo Bajardi, Chiara Poletto, Duygu Balcan, Hao Hu, Bruno Goncalves, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Modeling vaccination campaigns and the fall/winter 2009 activity of the new a (h1n1) influenza in the northern hemisphere. *Emerging health threats journal*, 2(1):7093, 2009.
- [211] Duygu Balcan, Vittoria Colizza, Bruno Gonçalves, Hao Hu, José J Ramasco, and Alessandro Vespignani. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proceedings of the national academy of sciences*, 106(51):21484–21489, 2009.
- [212] Duygu Balcan, Hao Hu, Bruno Goncalves, Paolo Bajardi, Chiara Poletto, Jose J Ramasco, Daniela Paolotti, Nicola Perra, Michele Tizzoni, Wouter Van den Broeck, et al. Seasonal transmission potential and activity peaks of the new influenza a (h1n1): a monte carlo likelihood analysis based on human mobility. *BMC medicine*, 7:1–12, 2009.
- [213] Duygu Balcan, Bruno Gonçalves, Hao Hu, José J Ramasco, Vittoria Colizza, and Alessandro Vespignani. Modeling the spatial spread of infectious diseases: The global epidemic and mobility computational model. *Journal of computational science*, 1(3):132–145, 2010.
- [214] Wouter Van den Broeck, Corrado Gioannini, Bruno Gonçalves, Marco Quaggiotto, Vittoria Colizza, and Alessandro Vespignani. The gleamviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC infectious diseases*, 11(1):1–14, 2011.
- [215] Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007.
- [216] Vittoria Colizza and Alessandro Vespignani. Invasion threshold in heterogeneous metapopulation networks. *Physical review letters*, 99(14):148701, 2007.
- [217] Vittoria Colizza and Alessandro Vespignani. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *Journal of theoretical biology*, 251(3):450–467, 2008.
- [218] Vitaly Belik, Theo Geisel, and Dirk Brockmann. Natural human mobility patterns and spatial spread of infectious diseases. *Physical Review X*, 1(1):011001, 2011.
- [219] Duygu Balcan and Alessandro Vespignani. Phase transitions in contagion processes mediated by recurrent mobility patterns. *Nature physics*, 7(7):581–586, 2011.
- [220] Duygu Balcan and Alessandro Vespignani. Invasion threshold in structured populations with recurrent mobility patterns. *Journal of theoretical biology*, 293:87–100, 2012.

- [221] Jesús Gómez-Gardeñes, David Soriano-Panos, and Alex Arenas. Critical regimes driven by recurrent mobility patterns of reaction–diffusion processes in networks. *Nature Physics*, 14(4):391–395, 2018.
- [222] Clara Granell and Peter J Mucha. Epidemic spreading in localized environments with recurrent mobility patterns. *Physical Review E*, 97(5):052302, 2018.
- [223] David Soriano-Panos, Gourab Ghoshal, Alex Arenas, and Jesús Gómez-Gardeñes. Impact of temporal scales and recurrent mobility patterns on the unfolding of epidemics. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(2):024006, 2020.
- [224] Wesley Cota, David Soriano-Paños, Alex Arenas, Silvio C Ferreira, and Jesús Gómez-Gardeñes. Infectious disease dynamics in metapopulations with heterogeneous transmission and recurrent mobility. *New Journal of Physics*, 23(7):073019, 2021.
- [225] David Soriano-Panos, L Lotero, Alex Arenas, and Jesús Gómez-Gardeñes. Spreading processes in multiplex metapopulations containing different mobility networks. *Physical Review X*, 8(3):031039, 2018.
- [226] Angélica S Mata, Silvio C Ferreira, and Romualdo Pastor-Satorras. Effects of local population structure in a reaction-diffusion model of a contact process on metapopulation networks. *Physical Review E*, 88(4):042820, 2013.
- [227] Andrea Apolloni, Chiara Poletto, and Vittoria Colizza. Age-specific contacts and travel patterns in the spatial spread of 2009 h1n1 influenza pandemic. *BMC infectious diseases*, 13(1):176, 2013.
- [228] Andrea Apolloni, Chiara Poletto, José J Ramasco, Pablo Jensen, and Vittoria Colizza. Metapopulation epidemic models with heterogeneous mixing and travel behaviour. *Theoretical Biology and Medical Modelling*, 11:1–26, 2014.
- [229] Zhongyuan Ruan, Ming Tang, Changgui Gu, and Jinshan Xu. Epidemic spreading between two coupled subpopulations with inner structures. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 27(10), 2017.
- [230] EJ Crighton, SJ Elliott, R Moineddin, P Kanaroglou, and REG Upshur. An exploratory spatial analysis of pneumonia and influenza hospitalizations in ontario by age and gender. *Epidemiology & Infection*, 135(2):253–261, 2007.
- [231] Justin Lessler, Derek AT Cummings, Jonathan M Read, Shuying Wang, Huachen Zhu, Gavin JD Smith, Yi Guan, Chao Qiang Jiang, and Steven Riley. Location-specific patterns of exposure to recent pre-pandemic strains of influenza a in southern china. *Nature communications*, 2(1):423, 2011.
- [232] Lin Wang, Yan Zhang, Zhen Wang, and Xiang Li. The impact of human location-specific contact pattern on the sir epidemic transmission between populations. *International Journal of Bifurcation and Chaos*, 23(05):1350095, 2013.
- [233] Lin Wang, Zhen Wang, Yan Zhang, and Xiang Li. How human location-specific contact patterns impact spatial transmission between populations? *Scientific reports*, 3(1):1468, 2013.
- [234] Halvor Lund, Ludvig Lizana, and Ingve Simonsen. Effects of city-size heterogeneity on epidemic spreading in a metapopulation: a reaction-diffusion approach. *Journal of statistical physics*, 151:367–382, 2013.
- [235] Michele Tizzoni, Kaiyuan Sun, Diego Benusiglio, Márton Karsai, and Nicola Perra. The scaling of human contacts and epidemic processes in metapopulation networks. *Scientific reports*, 5(1):15111, 2015.
- [236] Chiara Poletto, Michele Tizzoni, and Vittoria Colizza. Heterogeneous length of stay of hosts' movements and spatial epidemic spread. *Scientific reports*, 2(1):476, 2012.
- [237] Yong-Wang Gong, Yu-Rong Song, and Guo-Ping Jiang. Epidemic spreading in metapopulation networks with heterogeneous infection rates. *Physica A: Statistical Mechanics and its Applications*, 416:208–218, 2014.
- [238] Yongwang Gong and Michael Small. Modelling the effect of heterogeneous vaccination on metapopulation epidemic dynamics. *Physics Letters A*, 383(35):125996, 2019.

- [239] Sandro Meloni, Nicola Perra, Alex Arenas, Sergio Gómez, Yamir Moreno, and Alessandro Vespignani. Modeling human mobility responses to the large-scale spreading of infectious diseases. *Scientific reports*, 1(1):1–7, 2011.
- [240] Christos Nicolaides, Luis Cueto-Felgueroso, and Ruben Juanes. The price of anarchy in mobility-driven contagion dynamics. *Journal of The Royal Society Interface*, 10(87):20130495, 2013.
- [241] Bing Wang, Yuexing Han, and Gouhei Tanaka. Interplay between epidemic spread and information propagation on metapopulation networks. *Journal of Theoretical Biology*, 420:18–25, 2017.
- [242] Lin Wang and Xiang Li. Spatial epidemiology of networked metapopulation: An overview. *Chinese Science Bulletin*, 59(28):3511–3522, 2014.
- [243] Sebastian Funk, Marcel Salathé, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. *Journal of the Royal Society Interface*, 7(50):1247–1256, 2010.
- [244] Nicola Perra, Duygu Balcan, Bruno Gonçalves, and Alessandro Vespignani. Towards a characterization of behavior-disease models. *PloS one*, 6(8):e23084, 2011.
- [245] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen. The spread of awareness and its impact on epidemic outbreaks. *Proceedings of the National Academy of Sciences*, 106(16):6872–6877, 2009.
- [246] Qingchu Wu, Xinchu Fu, Michael Small, and Xin-Jian Xu. The impact of awareness on epidemic spreading in networks. *Chaos: an interdisciplinary journal of nonlinear science*, 22(1), 2012.
- [247] Hai-Feng Zhang, Jia-Rong Xie, Ming Tang, and Ying-Cheng Lai. Suppression of epidemic spreading in complex networks by local information based behavioral responses. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(4), 2014.
- [248] Clara Granell, Sergio Gómez, and Alex Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. *Physical review letters*, 111(12):128701, 2013.
- [249] Clara Granell, Sergio Gómez, and Alex Arenas. Competing spreading processes on multiplex networks: awareness and epidemics. *Physical review E*, 90(1):012808, 2014.
- [250] Quantong Guo, Xin Jiang, Yanjun Lei, Meng Li, Yifang Ma, and Zhiming Zheng. Two-stage effects of awareness cascade on epidemic spreading in multiplex networks. *Physical Review E*, 91(1):012822, 2015.
- [251] Jia-Qian Kan and Hai-Feng Zhang. Effects of awareness diffusion and self-initiated awareness behavior on epidemic spreading—an approach based on multiplex networks. *Communications in Nonlinear Science and Numerical Simulation*, 44:193–203, 2017.
- [252] Thilo Gross, Carlos J Dommar D'Lima, and Bernd Blasius. Epidemic dynamics on an adaptive network. *Physical review letters*, 96(20):208701, 2006.
- [253] Thilo Gross and Ioannis G Kevrekidis. Robust oscillations in sis epidemics on adaptive networks: Coarse graining by automated moment closure. *Europhysics Letters*, 82(3):38004, 2008.
- [254] Leah B Shaw and Ira B Schwartz. Fluctuating epidemics on adaptive networks. *Physical Review E*, 77(6):066101, 2008.
- [255] Leah B Shaw and Ira B Schwartz. Enhanced vaccine control of epidemics in adaptive networks. *Physical Review E*, 81(4):046120, 2010.
- [256] Hui Yang, Ming Tang, and Hai-Feng Zhang. Efficient community-based control strategies in adaptive networks. *New Journal of Physics*, 14(12):123017, 2012.
- [257] Pierre-Yves Geoffard and Tomas Philipson. Rational epidemics and their public control. *International economic review*, pages 603–624, 1996.
- [258] Michael Kremer. Integrating behavioral choice into epidemiological models of aids. *The Quarterly Journal of Economics*, 111(2):549–573, 1996.

- [259] M Christopher Auld. Choices, beliefs, and infectious disease dynamics. *Journal of health economics*, 22(3):361–377, 2003.
- [260] Frederick Chen, Miaohua Jiang, Scott Rabidoux, and Stephen Robinson. Public avoidance and epidemics: insights from an economic model. *Journal of theoretical biology*, 278(1):107–119, 2011.
- [261] Eli P Fenichel, Carlos Castillo-Chavez, M Graziano Coddia, Gerardo Chowell, Paula A Gonzalez Parra, Graham J Hickling, Garth Holloway, Richard Horan, Benjamin Morin, Charles Perrings, et al. Adaptive human behavior in epidemiological models. *Proceedings of the National Academy of Sciences*, 108(15):6306–6311, 2011.
- [262] Chris T Bauch, Alison P Galvani, and David JD Earn. Group interest versus self-interest in smallpox vaccination policy. *Proceedings of the National Academy of Sciences*, 100(18):10564–10567, 2003.
- [263] Chris T Bauch and David JD Earn. Vaccination and the theory of games. *Proceedings of the National Academy of Sciences*, 101(36):13391–13394, 2004.
- [264] Feng Fu, Daniel I Rosenbloom, Long Wang, and Martin A Nowak. Imitation dynamics of vaccination behaviour on social networks. *Proceedings of the Royal Society B: Biological Sciences*, 278(1702):42–49, 2011.
- [265] Martial L Ndeffo Mbah, Jingzhou Liu, Chris T Bauch, Yonas I Tekle, Jan Medlock, Lauren Ancel Meyers, and Alison P Galvani. The impact of imitation on vaccination behavior in social contact networks. *PLoS computational biology*, 8(4):e1002469, 2012.
- [266] Eriko Fukuda, Satoshi Kokubo, Jun Tanimoto, Zhen Wang, Aya Hagishima, and Naoki Ikegaya. Risk assessment for infectious disease and its impact on voluntary vaccination behavior in social networks. *Chaos, Solitons & Fractals*, 68:1–9, 2014.
- [267] Chris T Bauch. Imitation dynamics predict vaccinating behaviour. *Proceedings of the Royal Society B: Biological Sciences*, 272(1573):1669–1675, 2005.
- [268] Haifeng Zhang, Feng Fu, Wenyao Zhang, and Binghong Wang. Rational behavior is a ‘double-edged sword’ when considering voluntary vaccination. *Physica A: Statistical Mechanics and its Applications*, 391(20):4807–4815, 2012.
- [269] Alison P Galvani, Timothy C Reluga, and Gretchen B Chapman. Long-standing influenza vaccination policy is in accord with individual self-interest but not with the utilitarian optimum. *Proceedings of the National Academy of Sciences*, 104(13):5692–5697, 2007.
- [270] Jingzhou Liu, Beth F Kochin, Yonas I Tekle, and Alison P Galvani. Epidemiological game-theory dynamics of chickenpox vaccination in the usa and israel. *Journal of the Royal Society Interface*, 9(66):68–76, 2012.
- [271] Piero Poletti, Bruno Caprile, Marco Ajelli, Andrea Pugliese, and Stefano Merler. Spontaneous behavioural changes in response to epidemics. *Journal of theoretical biology*, 260(1):31–40, 2009.
- [272] Sheryl L Chang, Mahendra Piraveenan, Philippa Pattison, and Mikhail Prokopenko. Game theoretic modelling of infectious disease dynamics and intervention methods: a review. *Journal of Biological Dynamics*, 14(1):57–89, 2020.
- [273] Jun Tanimoto. *Sociophysics approach to epidemics*, volume 23. Springer, 2021.
- [274] Marcelo A Pires and Nuno Crokidakis. Dynamics of epidemic spreading with vaccination: impact of social pressure and engagement. *Physica A: Statistical Mechanics and its Applications*, 467:167–179, 2017.
- [275] Marcelo A Pires, André L Oestereich, and Nuno Crokidakis. Sudden transitions in coupled opinion and epidemic dynamics with vaccination. *Journal of Statistical Mechanics: Theory and Experiment*, 2018(5):053407, 2018.
- [276] Marcelo A Pires, Andre L Oestereich, Nuno Crokidakis, and Sílvio M Duarte Queirós. Antivax movement and epidemic spreading in the era of social networks: nonmonotonic effects, bistability, and network segregation. *Physical Review E*, 104(3):034302, 2021.

- [277] Kenneth J Bart, Walter A Orenstein, Alan R Hinman, and Robert W Amler. Measles and models. *International Journal of Epidemiology*, 12(3):263–266, 1983.
- [278] Neil M Ferguson, Matt J Keeling, W John Edmunds, Raymond Gani, Bryan T Grenfell, Roy M Anderson, and Steve Leach. Planning for smallpox outbreaks. *Nature*, 425(6959):681–685, 2003.
- [279] Jacco Wallinga and Peter Teunis. Different epidemic curves for severe acute respiratory syndrome reveal similar impacts of control measures. *American Journal of epidemiology*, 160(6):509–516, 2004.
- [280] Marco Ajelli, Bruno Gonçalves, Duygu Balcan, Vittoria Colizza, Hao Hu, José J Ramasco, Stefano Merler, and Alessandro Vespignani. Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models. *BMC infectious diseases*, 10(1):1–13, 2010.
- [281] Michele Tizzoni, Paolo Bajardi, Chiara Poletto, José J Ramasco, Duygu Balcan, Bruno Gonçalves, Nicola Perra, Vittoria Colizza, and Alessandro Vespignani. Real-time numerical forecast of global epidemic spreading: case study of 2009 a/h1n1pdm. *BMC medicine*, 10(1):1–31, 2012.
- [282] Dianbo Liu, Leonardo Clemente, Canelle Poirier, Xiyu Ding, Matteo Chinazzi, Jessica Davis, Alessandro Vespignani, and Mauricio Santillana. Real-time forecasting of the covid-19 outbreak in chinese provinces: Machine learning approach using novel digital data and estimates from mechanistic models. *Journal of medical Internet research*, 22(8):e20285, 2020.
- [283] Aniruddha Adiga, Devdatt Dubhashi, Bryan Lewis, Madhav Marathe, Srinivasan Venkatramanan, and Anil Vullikanti. Mathematical models for covid-19 pandemic: a comparative analysis. *Journal of the Indian Institute of Science*, 100(4):793–807, 2020.
- [284] Christopher Barrett, Richard Beckman, K Berkbigler, Keith Bisset, Bryan Bush, K Campbell, S Eubank, K Henson, J Hurford, D Kubicek, et al. Transims: Transportation analysis simulation system. *Los Alamos National Laboratory Unclassified Report*, 2001.
- [285] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.
- [286] Neil M Ferguson, Derek AT Cummings, Simon Cauchemez, Christophe Fraser, Steven Riley, Aronrag Meeyai, Sopon Iamsirithaworn, and Donald S Burke. Strategies for containing an emerging influenza pandemic in southeast asia. *Nature*, 437(7056):209–214, 2005.
- [287] Neil M Ferguson, Derek AT Cummings, Christophe Fraser, James C Cajka, Philip C Cooley, and Donald S Burke. Strategies for mitigating an influenza pandemic. *Nature*, 442(7101):448–452, 2006.
- [288] Timothy C Germann, Kai Kadau, Ira M Longini Jr, and Catherine A Macken. Mitigation strategies for pandemic influenza in the united states. *Proceedings of the National Academy of Sciences*, 103(15):5935–5940, 2006.
- [289] Dennis L Chao, M Elizabeth Halloran, Valerie J Obenchain, and Ira M Longini Jr. Flute, a publicly available stochastic influenza epidemic simulation model. *PLoS computational biology*, 6(1):e1000656, 2010.
- [290] Dennis L Chao, Laura Matrajt, Nicole E Basta, Jonathan D Sugimoto, Brandon Dean, Dee Ann Bagwell, Brit Oiulstad, M Elizabeth Halloran, and Ira M Longini Jr. Planning for the control of pandemic influenza a (h1n1) in los angeles county and the united states. *American journal of epidemiology*, 173(10):1121–1130, 2011.
- [291] John S Brownstein, Clark C Freifeld, and Lawrence C Madoff. Digital disease detection—harnessing the web for public health surveillance. *The New England journal of medicine*, 360(21):2153, 2009.
- [292] Marcel Salathe, Linus Bengtsson, Todd J Bodnar, Devon D Brewer, John S Brownstein, Caroline Buckee, Ellsworth M Campbell, Ciro Cattuto, Shashank Khandelwal, Patricia L Mabry, et al. Digital epidemiology. 2012.

- [293] Shweta Bansal, Gerardo Chowell, Lone Simonsen, Alessandro Vespignani, and Cécile Viboud. Big data for infectious disease surveillance and modeling. *The Journal of infectious diseases*, 214(suppl\_4):S375–S379, 2016.
- [294] Marcel Salathé. Digital epidemiology: what is it, and where is it going? *Life sciences, society and policy*, 14(1):1, 2018.
- [295] David Lazer, Ryan Kennedy, Gary King, and Alessandro Vespignani. The parable of google flu: traps in big data analysis. *science*, 343(6176):1203–1205, 2014.
- [296] Emily H Chan, Timothy F Brewer, Lawrence C Madoff, Marjorie P Pollack, Amy L Sonricker, Mikaela Keller, Clark C Freifeld, Michael Blench, Abla Mawudeku, and John S Brownstein. Global capacity for emerging infectious disease detection. *Proceedings of the National Academy of Sciences*, 107(50):21701–21706, 2010.
- [297] John S Brownstein and CC Freifeld. Healthmap: the development of automated real-time internet surveillance for epidemic intelligence. *Weekly releases (1997–2007)*, 12(48):3322, 2007.
- [298] Daniela Paolotti, Annasara Carnahan, Vittoria Colizza, Ken Eames, John Edmunds, Gabriel Gomes, C Koppeschaar, Moa Rehn, Ronald Smallenburg, Clément Turbelin, et al. Web-based participatory surveillance of infectious diseases: the influenzanet participatory surveillance experience. *Clinical Microbiology and Infection*, 20(1):17–21, 2014.
- [299] Outbreaks near me. <https://outbreaksnearme.org/us/en-US>. Accessed: August 22, 2023.
- [300] Marcel Salathé and Shashank Khandelwal. Assessing vaccination sentiments with online social media: implications for infectious disease dynamics and control. *PLoS computational biology*, 7(10):e1002199, 2011.
- [301] Michael Paul and Mark Dredze. You are what you tweet: Analyzing twitter for public health. In *Proceedings of the international AAAI conference on web and social media*, volume 5, pages 265–272, 2011.
- [302] Jeffrey Shaman and Alicia Karspeck. Forecasting seasonal outbreaks of influenza. *Proceedings of the National Academy of Sciences*, 109(50):20425–20430, 2012.
- [303] Qian Zhang, Nicola Perra, Daniela Perrotta, Michele Tizzoni, Daniela Paolotti, and Alessandro Vespignani. Forecasting seasonal influenza fusing digital indicators and a mechanistic disease model. In *Proceedings of the 26th international conference on world wide web*, pages 311–319, 2017.
- [304] Jeremy Ginsberg, Matthew H Mohebbi, Rajan S Patel, Lynnette Brammer, Mark S Smolinski, and Larry Brilliant. Detecting influenza epidemics using search engine query data. *Nature*, 457(7232):1012–1014, 2009.
- [305] Aron Culotta. Towards detecting influenza epidemics by analyzing twitter messages. In *Proceedings of the first workshop on social media analytics*, pages 115–122, 2010.
- [306] Alessio Signorini, Philip M Polgreen, Alberto Maria Segre, et al. Using twitter to estimate h1n1 influenza activity. In *9th Annual Conference of the International Society for Disease Surveillance*, 2010.
- [307] Todd Bodnar and Marcel Salathé. Validating models for disease detection using twitter. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 699–702, 2013.
- [308] David A Broniatowski, Michael J Paul, and Mark Dredze. National and local influenza surveillance through twitter: an analysis of the 2012-2013 influenza epidemic. *PloS one*, 8(12):e83672, 2013.
- [309] Nicholas Generous, Sara Y Del Valle, Geoffrey Fairchild, Reid Piedhorsky, and Alina Deshpande. Detecting epidemics using wikipedia article views: A demonstration of feasibility with language as location proxy. Technical report, 2014.
- [310] David J McIver and John S Brownstein. Wikipedia usage estimates prevalence of influenza-like illness in the united states in near real-time. *PLoS computational biology*, 10(4):e1003581, 2014.

- [311] Elaine O Nsoesie, David L Buckeridge, and John S Brownstein. Guess who's not coming to dinner? evaluating online restaurant reservations for disease surveillance. *Journal of medical Internet research*, 16(1):e2998, 2014.
- [312] Patrick Butler, Naren Ramakrishnan, Elaine O Nsoesie, and John S Brownstein. Satellite imagery analysis: What can hospital parking lots tell us about a disease outbreak? *Computer*, 47(04):94–97, 2014.
- [313] Oktawia P Wójcik, John S Brownstein, Rumi Chunara, and Michael A Johansson. Public health for the people: participatory infectious disease surveillance in the digital age. *Emerging themes in epidemiology*, 11(1):1–7, 2014.
- [314] Donald R Olson, Kevin J Konty, Marc Paladini, Cecile Viboud, and Lone Simonsen. Reassessing google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS computational biology*, 9(10):e1003256, 2013.
- [315] Hyeoun-Ae Park, Hyesil Jung, Jeongah On, Seul Ki Park, and Hannah Kang. Digital epidemiology: use of digital data collected for non-epidemiological purposes in epidemiological studies. *Healthcare informatics research*, 24(4):253–262, 2018.
- [316] Dirk Brockmann. Digital epidemiology. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz*, 63:166–175, 2020.
- [317] Intenso debate sobre medidas covid. <https://www.lasexta.com/programas/sexta-noche/noticias/intenso-debate-medidas-covid-lasexta-noche-falta-datos-verguenza-20211120619969e89ef0da0001e72b9a.html>. Accessed: August 22, 2023.
- [318] Matemáticas para salir de la cuarentena del coronavirus. <https://elpais.com/ciencia/2020-04-15/matematicas-para-salir-de-la-cuarentena-del-coronavirus.html>. Accessed: August 22, 2023.
- [319] Lo que hemos aprendido de la pandemia de covid-19. <https://www.lasexta.com/constantes-vitales/noticias/que-hemos-aprendido-pandemia-covid19-2023061364883ebbea11350001f5f6bc.html>. Accessed: August 22, 2023.
- [320] Objetivo: Frenar la curva. [https://www.antena3.com/noticias/salud/objetivo-frenar-la-curva\\_202003125e6ab96cbf55340001bbf2a5.html](https://www.antena3.com/noticias/salud/objetivo-frenar-la-curva_202003125e6ab96cbf55340001bbf2a5.html). Accessed: August 22, 2023.
- [321] Las matemáticas de la epidemia: Así se extiende el coronavirus. <https://www.newtral.es/las-matematicas-de-la-epidemia-asi-se-extiende-el-coronavirus/20200201/>. Accessed: August 22, 2023.
- [322] El misterio de los supercontagiadores del coronavirus. <https://elpais.com/ciencia/2020-05-23-el-misterio-de-los-supercontagiadores-del-coronavirus.html>. Accessed: August 22, 2023.
- [323] Stephan Eubank, I Eckstrand, B Lewis, Srinivasan Venkatramanan, Madhav Marathe, and CL Barrett. Commentary on ferguson, et al., “impact of non-pharmaceutical interventions (npis) to reduce covid-19 mortality and healthcare demand”. *Bulletin of mathematical biology*, 82:1–7, 2020.
- [324] Inga Holmdahl and Caroline Buckee. Wrong but useful—what covid-19 epidemiologic models can and cannot tell us. *New England Journal of Medicine*, 383(4):303–305, 2020.
- [325] Tim Rhodes, Kari Lancaster, Shelley Lees, and Melissa Parker. Modelling the pandemic: attuning models to their contexts. *BMJ global health*, 5(6):e002914, 2020.
- [326] David M Steinberg, Ran D Balicer, Yoav Benjamini, Hilla De-Leon, Doron Gazit, Hagai Rossman, and Eli Sprecher. The role of models in the covid-19 pandemic. *Israel Journal of Health Policy Research*, 11(1):36, 2022.
- [327] Emma S McBryde, Michael T Meehan, Oyelola A Adegboye, Adeshina I Adekunle, Jamie M Caldwell, Anton Pak, Diana P Rojas, Bridget M Williams, and James M Trauer. Role of modelling in covid-19 policy development. *Paediatric respiratory reviews*, 35:57–60, 2020.

- [328] Neil Ferguson, Daniel Laydon, Gemma Nedjati Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, Adhiratha Boonyasiri, ZULMA Cucunuba Perez, Gina Cuomo-Dannenburg, et al. Report 9: Impact of non-pharmaceutical interventions (npis) to reduce covid19 mortality and healthcare demand. 2020.
- [329] Pablo Martinez De Salazar, René Niehus, Aimee Taylor, Caroline O'Flaherty Buckee, and Marc Lipsitch. Identifying locations with possible undetected imported severe acute respiratory syndrome coronavirus 2 cases by using importation predictions. *Emerging infectious diseases*, 26(7):1465, 2020.
- [330] Freya M Shearer, James Walker, Nefel Tellioglu, James M McCaw, Jodie McVernon, Andrew Black, and Nic Geard. Assessing the risk of spread of covid-19 to the asia pacificregion. *medRxiv*, pages 2020–04, 2020.
- [331] Adeshina Adekunle, Michael Meehan, Diana Rojas-Alvarez, James Trauer, and Emma McBryde. Delaying the covid-19 epidemic in australia: evaluating the effectiveness of international travel bans. *Australian and New Zealand journal of public health*, 44(4):257–259, 2020.
- [332] Oyelola A Adegbeye, Adeshina I Adekunle, Anton Pak, Ezra Gayawan, Denis HY Leung, Diana P Rojas, Faiz Elfaki, Emma S McBryde, and Damon P Eisen. Change in outbreak epicentre and its impact on the importation risks of covid-19 progression: A modelling study. *Travel Medicine and Infectious Disease*, 40:101988, 2021.
- [333] Ying Liu, Albert A Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of covid-19 is higher compared to sars coronavirus. *Journal of travel medicine*, 2020.
- [334] Gregory J Fox, James M Trauer, and Emma McBryde. Modelling the impact of covid-19 on intensive care services in new south wales. *The Medical Journal of Australia*, 212(10):468, 2020.
- [335] Emma S McBryde, Michael T Meehan, and James M Trauer. Flattening the curve is not enough, we need to squash it: An explainer using a simple model. *MedRxiv*, pages 2020–03, 2020.
- [336] David Adam. Special report: The simulations driving the world's response to covid-19. *Nature*, 580(7802):316–319, 2020.
- [337] Gianluca Manzo. Complex social networks are missing in the dominant covid-19 epidemic models. *Sociologica*, 14(1):31–49, 2020.
- [338] Alfonso de Miguel-Arribas, Alberto Aleta, and Yamir Moreno. Impact of vaccine hesitancy on secondary covid-19 outbreaks in the us: an age-structured sir model. *BMC infectious diseases*, 22(1):1–12, 2022.
- [339] Alfonso de Miguel Arribas, Alberto Aleta, and Yamir Moreno. Assessing the effectiveness of perimeter lockdowns as a response to epidemics at the urban scale. *Scientific Reports*, 13(1):4474, 2023.
- [340] A de Miguel-Arribas, J Morón-Vidal, LM Floría, C Gracia-Lázaro, L Hernández, and Y Moreno. Contests in two fronts. *arXiv preprint arXiv:2303.18109*, 2023.
- [341] Michael Y Li. *An introduction to mathematical modeling of infectious diseases*, volume 2. Springer, 2018.
- [342] MJ Keeling and P Rohani. Modeling infectious diseases in humans and animals. 837 princeton university press, 2008.
- [343] Linda JS Allen. An introduction to stochastic epidemic models. In *Mathematical epidemiology*, pages 81–130. Springer, 2008.
- [344] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, volume 5. John Wiley & Sons, 2000.
- [345] Sergio Gómez, Alexandre Arenas, Javier Borge-Holthoefer, Sandro Meloni, and Yamir Moreno. Discrete-time markov chain approach to contact-based disease spreading in complex networks. *Europhysics Letters*, 89(3):38009, 2010.
- [346] Linda JS Allen. Some discrete-time si, sir, and sis epidemic models. *Mathematical biosciences*, 124(1):83–105, 1994.

- [347] Joan Saldaña. Continuous-time formulation of reaction-diffusion processes on heterogeneous metapopulations. *Physical Review E*, 78(1):012902, 2008.
- [348] Peter G Fennell, Sergey Melnik, and James P Gleeson. Limitations of discrete-time approaches to continuous-time contagion dynamics. *Physical Review E*, 94(5):052125, 2016.
- [349] Zesheng Chen. Discrete-time vs. continuous-time epidemic models in networks. *IEEE Access*, 7:127669–127677, 2019.
- [350] Odo Diekmann, Hans G Othmer, Robert Planqué, and Martin CJ Bootsma. The discrete-time kermack–mckendrick model: A versatile and computationally attractive framework for modeling epidemics. *Proceedings of the National Academy of Sciences*, 118(39):e2106332118, 2021.
- [351] Liang Zhang, Chao-Ran Cai, Ji-Qiang Zhang, Xu-Sheng Liu, Chong-Yang Wang, and Zhi-Xi Wu. Nonlinear mapping between continuous-and discrete-time dynamics. *Europhysics Letters*, 137(6):61002, 2022.
- [352] A Böckh. *Die Staatshaushaltung der Athener*, volume 1-2. 3 edition, 1886. Revised by M. Fränkel.
- [353] Francis R Sharpe and Alfred J Lotka. L. a problem in age-distribution. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 21(124):435–438, 1911.
- [354] Loots I Dublin and Alfred J Lotka. On the true rate of natural increase: As exemplified by the population of the united states, 1920. *Journal of the American statistical association*, 20(151):305–339, 1925.
- [355] Robert R Kuczynski. The world’s population. *Foreign Affairs*, 7(1):30–40, 1928.
- [356] George Macdonald. The analysis of equilibrium in malaria. *Tropical diseases bulletin*, 49(9):813–829, 1952.
- [357] Klaus Dietz. Transmission and control of arbovirus disease. In *in Proc. SIMS Conf. on Epidemiology (eds. D. Ludwig and KL Cooke)*, page 104, 1975.
- [358] Herbert W Hethcote. Mathematical models for the spread of infectious diseases. *Epidemiology*, pages 122–131, 1975.
- [359] Jane M Heffernan, Robert J Smith, and Lindi M Wahl. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*, 2(4):281–293, 2005.
- [360] Dina Mistry, Maria Litvinova, Ana Pastore y Piontti, Matteo Chinazzi, Laura Fumanelli, Marcelo FC Gomes, Syed A Haque, Quan-Hui Liu, Kunpeng Mu, Xinyue Xiong, et al. Inferring high-resolution human mixing patterns for disease modeling. *Nature communications*, 12(1):1–12, 2021.
- [361] Pauline Van den Driessche and James Watmough. Reproduction numbers and sub-threshold endemic equilibria for compartmental models of disease transmission. *Mathematical biosciences*, 180(1-2):29–48, 2002.
- [362] Paul L Delamater, Erica J Street, Timothy F Leslie, Y Tony Yang, and Kathryn H Jacobsen. Complexity of the basic reproduction number ( $r_0$ ). *Emerging infectious diseases*, 25(1):1, 2019.
- [363] Simon Cauchemez, Pierre-Yves Boëlle, Guy Thomas, and Alain-Jacques Valleron. Estimating in real time the efficacy of measures to control emerging communicable diseases. *American journal of epidemiology*, 164(6):591–597, 2006.
- [364] Simon Cauchemez, Pierre-Yves Boëlle, Christl A Donnelly, Neil M Ferguson, Guy Thomas, Gabriel M Leung, Anthony J Hedley, Roy M Anderson, and Alain-Jacques Valleron. Real-time estimates in early detection of sars. *Emerging infectious diseases*, 12(1):110, 2006.
- [365] Luís MA Bettencourt, Ruy M Ribeiro, Gerardo Chowell, Timothy Lant, and Carlos Castillo-Chavez. Towards real time epidemiology: data assimilation, modeling and anomaly detection of health surveillance data streams. In *Intelligence and Security Informatics: Biosurveillance: Second NSF Workshop, BioSurveillance 2007, New Brunswick, NJ, USA, May 22, 2007. Proceedings* 2, pages 79–90. Springer, 2007.

- [366] Gerardo Chowell, Hiroshi Nishiura, and Luis MA Bettencourt. Comparative estimation of the reproduction number for pandemic influenza from daily case notification data. *Journal of the Royal Society Interface*, 4(12):155–166, 2007.
- [367] Luis MA Bettencourt and Ruy M Ribeiro. Real time bayesian estimation of the epidemic potential of emerging infectious diseases. *PloS one*, 3(5):e2185, 2008.
- [368] Hiroshi Nishiura and Gerardo Chowell. The effective reproduction number as a prelude to statistical estimation of time-dependent epidemic trends. *Mathematical and statistical estimation approaches in epidemiology*, pages 103–121, 2009.
- [369] Robert M May, Sunetra Gupta, and Angela R McLean. Infectious disease dynamics: what characterizes a successful invader? *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, 356(1410):901–910, 2001.
- [370] Jing Li, Daniel Blakeley, et al. The failure of  $r_0$ . *Computational and mathematical methods in medicine*, 2011, 2011.
- [371] Benjamin Ridenhour, Jessica M Kowalik, and David K Shay. Unraveling  $r_0$ : Considerations for public health applications. *American journal of public health*, 104(2):e32–e41, 2014.
- [372] David Adam. The limits of  $r$ . *Nature*, 583(7816):346–348, 2020.
- [373] Antonello Maruotti, Massimo Ciccozzi, and Fabio Divino. On the misuse of the reproduction number in the covid-19 surveillance system in italy. *Journal of Medical Virology*, 93(5):2569, 2021.
- [374] MG Roberts. The pluses and minuses of  $r$ . *Journal of the Royal Society Interface*, 4(16):949–961, 2007.
- [375] Laurent Hébert-Dufresne, Benjamin M Althouse, Samuel V Scarpino, and Antoine Allard. Beyond  $r_0$ : heterogeneity in secondary infections and probabilistic epidemic forecasting. *Journal of the Royal Society Interface*, 17(172):20200393, 2020.
- [376] Romulus Breban, Raffaele Vardavas, and Sally Blower. Linking population-level models with growing networks: a class of epidemic models. *Physical Review E*, 72(4):046110, 2005.
- [377] Hamish McCallum, Nigel Barlow, and Jim Hone. How should pathogen transmission be modelled? *Trends in ecology & evolution*, 16(6):295–300, 2001.
- [378] Hao Hu, Karima Nigmatulina, and Philip Eckhoff. The scaling of contact rates with population density for the infectious disease models. *Mathematical biosciences*, 244(2):125–134, 2013.
- [379] K Dietz. Overall population patterns in the transmission cycle of infectious disease agents. In *Population Biology of Infectious Diseases: Report of the Dahlem Workshop on Population Biology of Infectious Disease Agents Berlin 1982, March 14–19*, pages 87–102. Springer, 1982.
- [380] Dieter Schenzle and Klaus Dietz. Critical population sizes for endemic virus transmission. *Raumliche Persistenz und Diffusion von Krankheiten*, 83:31–42, 1987.
- [381] CJ Rhodes and RM Anderson. Contact rate calculation for a basic epidemic model. *Mathematical biosciences*, 216(1):56–62, 2008.
- [382] Matthew J Ferrari, Sarah E Perkins, Laura W Pomeroy, Ottar N Bjørnstad, et al. Pathogens, social networks, and the paradox of transmission scaling. *Interdisciplinary perspectives on infectious diseases*, 2011, 2011.
- [383] Junling Ma, Pauline van den Driessche, and Frederick H Willeboordse. Effective degree household network disease model. *Journal of Mathematical Biology*, 66:75–94, 2013.
- [384] Greg Huber, Mason Kamb, Kyle Kawagoe, Lucy M Li, Boris Veytsman, David Yllanes, and Dan Zigmond. A minimal model for household effects in epidemics. *Physical Biology*, 17(6):065010, 2020.
- [385] Lorenzo Pellis, Frank Ball, and Pieter Trapman. Reproduction numbers for epidemic models with households and other social structures. i. definition and calculation of  $r_0$ . *Mathematical biosciences*, 235(1):85–97, 2012.

- [386] Nele Goeyvaerts, Eva Santermans, Gail Potter, Andrea Torneri, Kim Van Kerckhove, Lander Willem, Marc Aerts, Philippe Beutels, and Niel Hens. Household members do not contact each other at random: implications for infectious disease modelling. *Proceedings of the Royal Society B*, 285(1893):20182201, 2018.
- [387] Akira Endo, Mitsuo Uchida, Adam J Kucharski, and Sebastian Funk. Fine-scale family structure shapes influenza transmission risk in households: Insights from primary schools in matsumoto city, 2014/15. *PLoS computational biology*, 15(12):e1007589, 2019.
- [388] Fang Li, Yuan-Yuan Li, Ming-Jin Liu, Li-Qun Fang, Natalie E Dean, Gary WK Wong, Xiao-Bing Yang, Ira Longini, M Elizabeth Halloran, Huai-Ji Wang, et al. Household transmission of sars-cov-2 and risk factors for susceptibility and infectivity in wuhan: a retrospective observational study. *The Lancet Infectious Diseases*, 21(5):617–628, 2021.
- [389] Roy M Anderson and Robert M May. Vaccination and herd immunity to infectious diseases. *Nature*, 318(6044):323–329, 1985.
- [390] DJ Nokes and RM Anderson. The use of mathematical models in the epidemiological study of infectious diseases and in the design of mass immunization programmes. *Epidemiology & Infection*, 101(1):1–20, 1988.
- [391] Paul EM Fine. Herd immunity: history, theory, practice. *Epidemiologic reviews*, 15(2):265–302, 1993.
- [392] David S Stephens. Vaccines for the unvaccinated: protecting the herd, 2008.
- [393] Paul Fine, Ken Eames, and David L Heymann. “herd immunity”: a rough guide. *Clinical infectious diseases*, 52(7):911–916, 2011.
- [394] Tom Britton, Frank Ball, and Pieter Trapman. A mathematical model reveals the influence of population heterogeneity on herd immunity to sars-cov-2. *science*, 369(6505):846–849, 2020.
- [395] Haley E Randolph and Luis B Barreiro. Herd immunity: understanding covid-19. *Immunity*, 52(5):737–741, 2020.
- [396] Colin J Worby, Sandra S Chaves, Jacco Wallinga, Marc Lipsitch, Lyn Finelli, and Edward Goldstein. On the relative role of different age groups in influenza epidemics. *Epidemics*, 13:10–16, 2015.
- [397] Nicholas G Davies, Petra Klepac, Yang Liu, Kiesha Prem, Mark Jit, and Rosalind M Eggo. Age-dependent effects in the transmission and control of covid-19 epidemics. *Nature medicine*, 26(8):1205–1211, 2020.
- [398] Jacco Wallinga, Peter Teunis, and Mirjam Kretzschmar. Using data on social contacts to estimate age-specific transmission parameters for respiratory-spread infectious agents. *American journal of epidemiology*, 164(10):936–944, 2006.
- [399] Sara Y Del Valle, James M Hyman, Herbert W Hethcote, and Stephen G Eubank. Mixing patterns between age groups in social networks. *Social Networks*, 29(4):539–554, 2007.
- [400] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 5(3):e74, 2008.
- [401] Laura Fumanelli, Marco Ajelli, Piero Manfredi, Alessandro Vespignani, and Stefano Merler. Inferring the structure of social contacts from demographic data in the analysis of infectious diseases spread. *PLoS Comput Biol*, 8(9):e1002673, 2012.
- [402] Sara Y Del Valle, James M Hyman, and Nakul Chitnis. Mathematical models of contact patterns between age groups for predicting the spread of infectious diseases. *Mathematical biosciences and engineering: MBE*, 10:1475, 2013.
- [403] Kiesha Prem, Alex R Cook, and Mark Jit. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS computational biology*, 13(9):e1005697, 2017.
- [404] Sergio Arregui, Alberto Aleta, Joaquín Sanz, and Yamir Moreno. Projecting social contact matrices to different demographic structures. *PLoS computational biology*, 14(12):e1006638, 2018.

- [405] Thang Hoang, Pietro Coletti, Alessia Melegaro, Jacco Wallinga, Carlos G Grijalva, John W Edmunds, Philippe Beutels, and Niel Hens. A systematic review of social contact surveys to inform transmission models of close-contact infections. *Epidemiology (Cambridge, Mass.)*, 30(5):723, 2019.
- [406] Kiesha Prem, Kevin van Zandvoort, Petra Klepac, Rosalind M Eggo, Nicholas G Davies, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Alex R Cook, and Mark Jit. Projecting contact matrices in 177 geographical regions: an update and comparison with empirical data for the covid-19 era. *PLoS computational biology*, 17(7):e1009098, 2021.
- [407] Xue-Zhi Li, Junyuan Yang, and Maia Martcheva. *Age structured epidemic modeling*, volume 52. Springer Nature, 2020.
- [408] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge university press, 2008.
- [409] Michele Catanzaro, Marián Boguná, and Romualdo Pastor-Satorras. Generation of uncorrelated random scale-free networks. *Physical review e*, 71(2):027103, 2005.
- [410] Ginestra Bianconi. *Multilayer networks: structure and function*. Oxford university press, 2018.
- [411] Petter Holme and Jari Saramäki. Temporal networks. *Physics reports*, 519(3):97–125, 2012.
- [412] Mason A Porter and James P Gleeson. Dynamical systems on networks. *Frontiers in Applied Dynamical Systems: Reviews and Tutorials*, 4, 2016.
- [413] Reuven Cohen, Shlomo Havlin, and Daniel Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical review letters*, 91(24):247901, 2003.
- [414] Manuel Garcia-Herranz, Esteban Moro, Manuel Cebrian, Nicholas A Christakis, and James H Fowler. Using friends as sensors to detect global-scale contagious outbreaks. *PloS one*, 9(4):e92413, 2014.
- [415] Suyu Liu, Nicola Perra, Márton Karsai, and Alessandro Vespignani. Controlling contagion processes in activity driven networks. *Physical review letters*, 112(11):118702, 2014.
- [416] Paulo Cesar Ventura, Alberto Aleta, Francisco Aparecido Rodrigues, and Yamir Moreno. Modeling the effects of social distancing on the large-scale spreading of diseases. *Epidemics*, 38:100544, 2022.
- [417] J Marro and R Dickman. Nonequilibrium statistical mechanics of lattice models, 1999.
- [418] George Kingsley Zipf. The p 1 p 2/d hypothesis: on the intercity movement of persons. *American sociological review*, 11(6):677–686, 1946.
- [419] Sven Erlander and Neil F Stewart. *The gravity model in transportation analysis: theory and extensions*, volume 3. Vsp, 1990.
- [420] Juan de Dios Ortúzar and Luis G Willumsen. *Modelling transport*. John wiley & sons, 2011.
- [421] Xinhai Li, Huidong Tian, Dejian Lai, and Zhibin Zhang. Validation of the gravity model in predicting the global spread of influenza. *International journal of environmental research and public health*, 8(8):3134–3143, 2011.
- [422] Filippo Simini, Marta C González, Amos Maritan, and Albert-László Barabási. A universal model for mobility and migration patterns. *Nature*, 484(7392):96–100, 2012.
- [423] Michele Tizzoni, Paolo Bajardi, Adeline Decuyper, Guillaume Kon Kam King, Christian M Schneider, Vincent Blondel, Zbigniew Smoreda, Marta C González, and Vittoria Colizza. On the use of human mobility proxies for modeling epidemics. *PLoS computational biology*, 10(7):e1003716, 2014.
- [424] Hugo Barbosa, Marc Barthelemy, Gourab Ghoshal, Charlotte R James, Maxime Lenormand, Thomas Louail, Ronaldo Menezes, José J Ramasco, Filippo Simini, and Marcello Tomasini. Human mobility: Models and applications. *Physics Reports*, 734:1–74, 2018.
- [425] T Déirdre Hollingsworth, Neil M Ferguson, and Roy M Anderson. Will travel restrictions control the international spread of pandemic influenza? *Nature medicine*, 12(5):497–499, 2006.

- [426] Wesley Cota and Silvio C Ferreira. Optimized gillespie algorithms for the simulation of markovian epidemic processes on large and heterogeneous networks. *Computer Physics Communications*, 219:303–312, 2017.
- [427] Vito Latora, Vincenzo Nicosia, and Giovanni Russo. *Complex networks: principles, methods and applications*. Cambridge University Press, 2017.
- [428] Chaoming Song, Tal Koren, Pu Wang, and Albert-László Barabási. Modelling the scaling properties of human mobility. *Nature physics*, 6(10):818–823, 2010.
- [429] Luca Pappalardo, Filippo Simini, Salvatore Rinzivillo, Dino Pedreschi, Fosca Giannotti, and Albert-László Barabási. Returners and explorers dichotomy in human mobility. *Nature communications*, 6(1):8166, 2015.
- [430] Luca Pappalardo, Salvatore Rinzivillo, and Filippo Simini. Human mobility modelling: exploration and preferential return meet the gravity model. *Procedia Computer Science*, 83:934–939, 2016.
- [431] Hugo Barbosa, Fernando B de Lima-Neto, Alexandre Evsukoff, and Ronaldo Menezes. The effect of recency to human mobility. *EPJ Data Science*, 4:1–14, 2015.
- [432] Laura Alessandretti, Piotr Sapiezynski, Vedran Sekara, Sune Lehmann, and Andrea Baronchelli. Evidence for a conserved quantity in human mobility. *Nature human behaviour*, 2(7):485–491, 2018.
- [433] Markus Schläpfer, Lei Dong, Kevin O’Keeffe, Paolo Santi, Michael Szell, Hadrien Salat, Samuel Anklesaria, Mohammad Vazifeh, Carlo Ratti, and Geoffrey B West. The universal visitation law of human mobility. *Nature*, 593(7860):522–527, 2021.
- [434] Junyu Lv, Chen Zhao, and An Zeng. Bursty visitation of locations in human mobility. *Physica A: Statistical Mechanics and its Applications*, 567:125674, 2021.
- [435] Giuliano Cornacchia and Luca Pappalardo. Sts-epr: Modelling individual mobility considering the spatial, temporal, and social dimensions together. *Procedia Computer Science*, 184:258–265, 2021.
- [436] Shan Jiang, Yingxiang Yang, Siddharth Gupta, Daniele Veneziano, Shounak Athavale, and Marta C González. The timegeo modeling framework for urban mobility without travel surveys. *Proceedings of the National Academy of Sciences*, 113(37):E5370–E5378, 2016.
- [437] Luca Pappalardo and Filippo Simini. Data-driven generation of spatio-temporal routines in human mobility. *Data Mining and Knowledge Discovery*, 32(3):787–829, 2018.
- [438] Dirk Brockmann, Lars Hufnagel, and Theo Geisel. The scaling laws of human travel. *Nature*, 439(7075):462–465, 2006.
- [439] Anastasios Noulas, Salvatore Scellato, Renaud Lambiotte, Massimiliano Pontil, and Cecilia Mascolo. A tale of many cities: universal patterns in human urban mobility. *PloS one*, 7(5):e37027, 2012.
- [440] Laura Alessandretti, Ulf Aslak, and Sune Lehmann. The scales of human mobility. *Nature*, 587(7834):402–407, 2020.
- [441] Valentina Costantino, David J Heslop, and C Raina MacIntyre. The effectiveness of full and partial travel bans against covid-19 spread in australia for travellers from china during and after the epidemic peak in china. *Journal of travel medicine*, 27(5):taaa081, 2020.
- [442] Aldo Carranza, Marcel Goic, Eduardo Lara, Marcelo Olivares, Gabriel Y Weintraub, Julio Covarrubia, Cristian Escobedo, Natalia Jara, Leonardo J Basso, et al. The social divide of social distancing: Lockdowns in santiago during the covid-19 pandemic. Technical report, 2020.
- [443] Shengjie Lai, Nick W Ruktanonchai, Liangcai Zhou, Olivia Prosper, Wei Luo, Jessica R Floyd, Amy Wesolowski, Mauricio Santillana, Chi Zhang, Xiangjun Du, et al. Effect of non-pharmaceutical interventions to contain covid-19 in china. *nature*, 585(7825):410–413, 2020.
- [444] Meirui Qian and Jianli Jiang. Covid-19 and social distancing. *Journal of Public Health*, pages 1–3, 2020.

- [445] Daniel Gros et al. The great lockdown: was it worth it. *CEPS Policy Insights*, 11, 2020.
- [446] Alessandro Spelta, Andrea Flori, Francesco Pierri, Giovanni Bonaccorsi, and Fabio Pammolli. After the lockdown: simulating mobility, public health and economic recovery scenarios. *Scientific reports*, 10(1):1–13, 2020.
- [447] Luis Pedauga, Francisco Sáez, and Blanca L Delgado-Márquez. Macroeconomic lockdown and smes: the impact of the covid-19 pandemic in spain. *Small Business Economics*, pages 1–24, 2021.
- [448] Hiroyasu Inoue and Yasuyuki Todo. The propagation of economic impacts through supply chains: The case of a mega-city lockdown to prevent the spread of covid-19. *PloS one*, 15(9):e0239251, 2020.
- [449] Serpil Aday and Mehmet Seckin Aday. Impact of covid-19 on the food supply chain. *Food Quality and Safety*, 4(4):167–180, 2020.
- [450] Dabo Guan, Daoping Wang, Stephane Hallegatte, Steven J Davis, Jingwen Huo, Shuping Li, Yangchun Bai, Tianyang Lei, Qianyu Xue, D'Maris Coffman, et al. Global supply-chain effects of covid-19 control measures. *Nature human behaviour*, 4(6):577–587, 2020.
- [451] Ulrike M Vieten. The “new normal” and “pandemic populism”: the covid-19 crisis and anti-hygienic mobilisation of the far-right. *Social Sciences*, 9(9):165, 2020.
- [452] Isaac Yen-Hao Chu, Prima Alam, Heidi J Larson, and Leesa Lin. Social consequences of mass quarantine during epidemics: a systematic review with implications for the covid-19 response. *Journal of travel medicine*, 27(7):taaa192, 2020.
- [453] Stephen Reicher and Clifford Stott. On order and disorder during the covid-19 pandemic. *British Journal of Social Psychology*, 59(3):694–702, 2020.
- [454] Francisco Javier Candel, Jesús San-Román, Pablo Barreiro, Jesús Canora, Antonio Zapatero, Mar Carretero, Antonio Lastra, and Francisco Javier Martínez-Peromingo. Integral management of covid-19 in madrid: Turning things around during the second wave. *The Lancet Regional Health–Europe*, 2, 2021.
- [455] Mario Fontán-Vela, Pedro Gullón, and Javier Padilla-Bernáldez. Reply to “integral management of covid-19 in madrid: Turning things around during the second wave”. *The Lancet Regional Health–Europe*, 3, 2021.
- [456] Felix Lobo. Reply to “integral management of covid-19 in madrid: Turning things around during the second wave”. *The Lancet Regional Health–Europe*, 3, 2021.
- [457] María-Victoria Zunzunegui, Fernando J García López, Miguel Ángel Royo-Bordonada, and Isabel Cienfuegos. Reply to “integral management of covid-19 in madrid: Turning things around during the second wave”. *The Lancet Regional Health–Europe*, 3, 2021.
- [458] Ángel López-Encuentra, Pablo Lázaro y de Mercado, Víctor Abraira, Sergio García Rull, and Francisco Pozo Rodríguez. Reply to:“integral management of covid-19 in madrid: Turning things around during the second wave”. *The Lancet Regional Health–Europe*, 3, 2021.
- [459] Mario Fontán-Vela, Pedro Gullón, and Javier Padilla-Bernáldez. Selective perimeter lockdowns in madrid: a way to bend the covid-19 curve? *European Journal of Public Health*, 31(5):1102–1104, 2021.
- [460] García-García David, Herranz-Hernandez Rafael, Rojas-Benedicto Ayelén, León-Gomez Inmaculada, Larrauri Amparo, Peñuelas Marina, Guerrero-Vadillo Maria, Ramis Rebeca, and Gomez-Barroso Diana. Perimeter confinements of basic health zones and covid-19 incidence in madrid, spain. *BMC Public Health*, 22:216, 2021.
- [461] Yige Li, Eduardo A Undurraga, and José R Zubizarreta. Effectiveness of localized lockdowns in the covid-19 pandemic. *American journal of epidemiology*, 2022.
- [462] Covid 19 -tia zonas básicas de salud. [https://datos.comunidad.madrid/catalogo/dataset/covid19\\_tia\\_zonas\\_basicas\\_salud](https://datos.comunidad.madrid/catalogo/dataset/covid19_tia_zonas_basicas_salud) [Online; accessed: 3. May. 2022 (in Spanish)].

- [463] Richard C Dicker, Fatima Coronado, Denise Koo, and R Gibson Parrish. *Principles of epidemiology in public health practice; an introduction to applied epidemiology and biostatistics*. 2006.
- [464] Repositorio histórico de las medidas adoptadas durante la crisis sanitaria por la covid-19. <https://www.comunidad.madrid/servicios/salud/comunicados-covid-19-normativa-notas-prensa> [Online; accessed: 3. May. 2022 (in Spanish)].
- [465] Alex Arenas, Wesley Cota, Jesús Gómez-Gardeñes, Sergio Gómez, Clara Granell, Joan T Matamalas, David Soriano-Paños, and Benjamin Steinegger. Modeling the spatiotemporal epidemic spreading of covid-19 and the impact of mobility and social distancing interventions. *Physical Review X*, 10(4):041055, 2020.
- [466] Francesco Parino, Lorenzo Zino, Maurizio Porfiri, and Alessandro Rizzo. Modelling and predicting the effect of social distancing and travel restrictions on covid-19 spreading. *Journal of the Royal Society Interface*, 18(175):20200875, 2021.
- [467] Pietro Coletti, Pieter Libin, Oana Petrof, Lander Willem, Steven Abrams, Sereina A Herzog, Christel Faes, Elise Kuylen, James Wambua, Philippe Beutels, et al. A data-driven metapopulation model for the belgian covid-19 epidemic: assessing the impact of lockdown and exit strategies. *BMC infectious diseases*, 21(1):1–12, 2021.
- [468] Alberto Aleta and Yamir Moreno. Evaluation of the potential incidence of covid-19 and effectiveness of containment measures in spain: a data-driven approach. *BMC medicine*, 18(1):1–12, 2020.
- [469] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, Marco Ajelli, and Hongjie Yu. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*, 368(6498):1481–1486, June 2020.
- [470] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Wen Zheng, Huilin Shi, Alessandro Vespignani, Cecile Viboud, Marco Ajelli, and Hongjie Yu. The impact of relaxing interventions on human contact patterns and SARS-CoV-2 transmission in China. *Sci. Adv.*, 7(19):eabe2584, May 2021.
- [471] SAMUR - Protección Civil. COVID-19 Actualización Científica 7 - Noviembre - 2020 (COVID-19 Scientific Update 7 - November - 2020). [https://www.madrid.es/UnidadesDescentralizadas/Emergencias/Samur-PCivil/Samur/ApartadosSecciones/COVID-19/data/COVID19\\_ActualizacionCientificaSAMUR\\_20201107.pdf](https://www.madrid.es/UnidadesDescentralizadas/Emergencias/Samur-PCivil/Samur/ApartadosSecciones/COVID-19/data/COVID19_ActualizacionCientificaSAMUR_20201107.pdf), 2020.
- [472] Plan de medidas para responder al impacto del covid 19 en el sector transporte y movilidad. <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data> [Online; accessed: 3. May. 2022 (in Spanish)].
- [473] Bryan Grenfell and John Harwood. (meta) population dynamics of infectious diseases. *Trends in ecology & evolution*, 12(10):395–399, 1997.
- [474] Alberto Aleta, Andreia NS Hisi, Sandro Meloni, Chiara Poletto, Vittoria Colizza, and Yamir Moreno. Human mobility networks and persistence of rapidly mutating pathogens. *Royal Society open science*, 4(3):160914, 2017.
- [475] Jianying Wang, Lei Dong, Ximeng Cheng, Weijun Yang, and Yu Liu. An extended exploration and preferential return model for human mobility simulation at individual and collective levels. *Physica A: Statistical Mechanics and Its Applications*, 534:121921, 2019.
- [476] Suttikiat Changruenngam, Dominique J Bicout, and Charin Modchang. How the individual human mobility spatio-temporally shapes the disease transmission dynamics. *Scientific Reports*, 10(1):11325, 2020.
- [477] Alexandru Topîrceanu and Radu-Emil Precup. A novel geo-hierarchical population mobility model for spatial spreading of resurgent epidemics. *Scientific Reports*, 11(1):14341, 2021.
- [478] Valeria Saladino, Davide Algeri, and Vincenzo Auriemma. The psychological and social impact of covid-19: new perspectives of well-being. *Frontiers in psychology*, 11:2550, 2020.

- [479] Benjamin YQ Tan, Nicholas WS Chew, Grace KH Lee, Mingxue Jing, Yihui Goh, Leonard LL Yeo, Ka Zhang, Howe-Keat Chin, Aftab Ahmad, Faheem Ahmed Khan, et al. Psychological impact of the covid-19 pandemic on health care workers in singapore. *Annals of internal medicine*, 173(4):317–320, 2020.
- [480] Carolina S Romero, Carlos Delgado, Juan Catalá, Carolina Ferrer, Carlos Errando, Adina Iftimi, Ana Benito, Jose De Andres, Maria Otero, et al. Covid-19 psychological impact in 3109 healthcare workers in spain: The psimcov group. *Psychological medicine*, pages 1–7, 2020.
- [481] Frederik Verelst, Elise Kuylen, and Philippe Beutels. Indications for healthcare surge capacity in european countries facing an exponential increase in coronavirus disease (covid-19) cases, march 2020. *Eurosurveillance*, 25(13):2000323, 2020.
- [482] Benedetta Armocida, Beatrice Formenti, Silvia Ussai, Francesca Palestra, and Eduardo Missoni. The italian health system and the covid-19 challenge. *The Lancet Public Health*, 5(5):e253, 2020.
- [483] Jill E Hobbs. Food supply chains during the covid-19 pandemic. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroéconomie*, 68(2):171–176, 2020.
- [484] Warwick McKibbin and Roshen Fernando. The economic impact of covid-19. *Economics in the Time of COVID-19*, 45, 2020.
- [485] Amory Martin, Maryia Markhvida, Stéphane Hallegatte, and Brian Walsh. Socio-economic impacts of covid-19 on household consumption and poverty. *Economics of disasters and climate change*, 4(3):453–479, 2020.
- [486] Hien Lau, Veria Khosrawipour, Piotr Kocbach, Agata Mikolajczyk, Justyna Schubert, Jacek Bania, and Tanja Khosrawipour. The positive impact of lockdown in wuhan on containing the covid-19 outbreak in china. *Journal of travel medicine*, 27(3):taaa037, 2020.
- [487] David Koh. Covid-19 lockdowns throughout the world. *Occupational Medicine*, 70(5):322–322, 2020.
- [488] Xuefei Ren. Pandemic and lockdown: a territorial approach to covid-19 in china, italy and the united states. *Eurasian Geography and Economics*, 61(4-5):423–434, 2020.
- [489] The Lancet. India under covid-19 lockdown. *Lancet (London, England)*, 395(10233):1315, 2020.
- [490] Thomas AJ Meunier. Full lockdown policies in western europe countries have no evident impacts on the covid-19 epidemic. *MedRxiv*, 2020.
- [491] Moritz UG Kraemer, Chia-Hung Yang, Bernardo Gutierrez, Chieh-Hsi Wu, Brennan Klein, David M Pigott, Open COVID-19 Data Working Group†, Louis Du Plessis, Nuno R Faria, Ruoran Li, et al. The effect of human mobility and control measures on the covid-19 epidemic in china. *Science*, 368(6490):493–497, 2020.
- [492] Alberto Aleta, David Martin-Corral, Ana Pastore y Piontti, Marco Ajelli, Maria Litvinova, Matteo Chinazzi, Natalie E Dean, M Elizabeth Halloran, Ira M Longini Jr, Stefano Merler, et al. Modelling the impact of testing, contact tracing and household quarantine on second waves of covid-19. *Nature Human Behaviour*, 4(9):964–971, 2020.
- [493] Giovanni Bonaccorsi, Francesco Pierri, Matteo Cinelli, Andrea Flori, Alessandro Galeazzi, Francesco Porcelli, Ana Lucia Schmidt, Carlo Michele Valensise, Antonio Scala, Walter Quattrociocchi, et al. Economic and social consequences of human mobility restrictions under covid-19. *Proceedings of the National Academy of Sciences*, 117(27):15530–15535, 2020.
- [494] Guglielmo Briscese, Nicola Lacetera, Mario Macis, and Mirco Tonin. Compliance with covid-19 social-distancing measures in italy: the role of expectations and duration. Technical report, National Bureau of Economic Research, 2020.
- [495] Linda Thunström, Stephen C Newbold, David Finnoff, Madison Ashworth, and Jason F Shogren. The benefits and costs of using social distancing to flatten the curve for covid-19. *Journal of Benefit-Cost Analysis*, 11(2):179–195, 2020.
- [496] Joseph A Lewnard and Nathan C Lo. Scientific and ethical basis for social-distancing interventions against covid-19. *The Lancet Infectious Diseases*, 20(6):631–633, 2020.

- [497] World Health Organization et al. Rational use of personal protective equipment for coronavirus disease (covid-19): interim guidance, 27 february 2020. Technical report, World Health Organization, 2020.
- [498] Holger J Schünemann, Elie A Akl, Roger Chou, Derek K Chu, Mark Loeb, Tamara Lotfi, Reem A Mustafa, Ignacio Neumann, Lynora Saxinger, Shahnaz Sultan, et al. Use of facemasks during the covid-19 pandemic. *The Lancet Respiratory Medicine*, 8(10):954–955, 2020.
- [499] Shanay Rab, Mohd Javaid, Abid Haleem, and Raju Vaishya. Face masks are new normal after covid-19 pandemic. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews*, 14(6):1617–1619, 2020.
- [500] Thomas H Lee. Creating the new normal: the clinician response to covid-19. *NEJM Catalyst Innovations in Care Delivery*, 1(2), 2020.
- [501] Jens O Zinn. A monstrous threat': how a state of exception turns into a 'new normal. *Journal of Risk Research*, 23(7-8):1083–1091, 2020.
- [502] T Thanh Le, Zacharias Andreadakis, Arun Kumar, R Gómez Román, Stig Tollefsen, Melanie Saville, Stephen Mayhew, et al. The covid-19 vaccine development landscape. *Nat Rev Drug Discov*, 19(5):305–306, 2020.
- [503] Jieliang Wang, Ying Peng, Haiyue Xu, Zhengrong Cui, and Robert O Williams. The covid-19 vaccine race: challenges and opportunities in vaccine formulation. *AAPS PharmSciTech*, 21(6):1–12, 2020.
- [504] Melanie Schuster, Juhani Eskola, Philippe Duclos, et al. Review of vaccine hesitancy: Rationale, remit and methods. *Vaccine*, 33(34):4157–4160, 2015.
- [505] Amiel A Dror, Netanel Eisenbach, Shahar Taiber, Nicole G Morozov, Matti Mizrahi, Asaf Zigrin, Samer Srouji, and Eyal Sela. Vaccine hesitancy: the next challenge in the fight against covid-19. *European journal of epidemiology*, 35(8):775–779, 2020.
- [506] Jagdish Khubchandani, Sushil Sharma, James H Price, Michael J Wiblishauser, Manoj Sharma, and Fern J Webb. Covid-19 vaccination hesitancy in the united states: a rapid national assessment. *Journal of Community Health*, 46(2):270–277, 2021.
- [507] Daniel Romer and Kathleen Hall Jamieson. Conspiracy theories as barriers to controlling the spread of covid-19 in the us. *Social Science & Medicine*, 263:113356, 2020.
- [508] Heidi J Larson, Caitlin Jarrett, Elisabeth Eckersberger, David MD Smith, and Pauline Paterson. Understanding vaccine hesitancy around vaccines and vaccination from a global perspective: a systematic review of published literature, 2007–2012. *Vaccine*, 32(19):2150–2159, 2014.
- [509] LP Wong, PF Wong, and S AbuBakar. Vaccine hesitancy and the resurgence of vaccine preventable diseases: the way forward for malaysia, a southeast asian country. *Human vaccines & immunotherapeutics*, 16(7):1511–1520, 2020.
- [510] Grete H Porteous, Neil A Hanson, Lila Ann A Sueda, Carli D Hoaglan, Aaron B Dahl, Brooks B Ohlson, Brian E Schmidt, Chia C Wang, and R Eliot Fagley. Resurgence of vaccine-preventable diseases in the united states: anesthetic and critical care implications. *Anesthesia & Analgesia*, 122(5):1450–1473, 2016.
- [511] Gregory A Poland, Robert M Jacobson, and Inna G Ovsyannikova. Trends affecting the future of vaccine development and delivery: the role of demographics, regulatory science, the anti-vaccine movement, and vaccinomics. *Vaccine*, 27(25-26):3240–3244, 2009.
- [512] Measles in England - Public health matters. <https://publichealthmatters.blog.gov.uk/2019/08/19/measles-in-england>, Sep 2021. [Online; accessed 11. Sep. 2021].
- [513] Kristina L Bajema, Ryan E Wiegand, Kendra Cuffe, Sadhna V Patel, Ronaldo Iachan, Travis Lim, Adam Lee, Davia Moyse, Fiona P Havers, Lee Harding, et al. Estimated sars-cov-2 seroprevalence in the us as of september 2020. *JAMA internal medicine*, 2020.
- [514] David Lazer, Katherine Ognyanova, Matthew Baum, James Druckman, Jon Green, Adina Gitomer, Matthew Simonson, Roy H Perlis, Mauricio Santillana, Alexi Quintana, et al. The covid states project# 43: Covid-19 vaccine rates and attitudes among americans. 2021.

- [515] Mélodie Monod, Alexandra Blenkinsop, Xiaoyue Xi, Daniel Hebert, Sivan Bershan, Simon Tietze, Marc Baguelin, Valerie C Bradley, Yu Chen, Helen Coupland, et al. Age groups that sustain resurging covid-19 epidemics in the united states. *Science*, 371(6536):eabe8372, 2021.
- [516] Andrew T Levin, William P Hanage, Nana Owusu-Boaitey, Kensington B Cochran, Seamus P Walsh, and Gideon Meyerowitz-Katz. Assessing the age specificity of infection fatality rates for covid-19: systematic review, meta-analysis, and public policy implications. *European journal of epidemiology*, 35(12):1123–1138, 2020.
- [517] Us census 2021. <https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html>.
- [518] Jonas Dehning, Johannes Zierenberg, F. Paul Spitzner, Michael Wibral, Joao Pinheiro Neto, Michael Wilczek, and Viola Priesemann. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science*, Jul 2020.
- [519] PhD Russell M. Viner. Susceptibility to SARS-CoV-2 Infection Among Children and Adolescents Compared With Adults: A Systematic. *JAMA Pediatr.*, 175(2):143–156, Feb 2021.
- [520] Ying Liu, Albert A. Gayle, Annelies Wilder-Smith, and Joacim Rocklöv. The reproductive number of COVID-19 is higher compared to SARS coronavirus. *J. Travel Med.*, 27(2):taaa021, March 2020.
- [521] Ying Liu and Joacim Rocklöv. The reproductive number of the Delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *J. Travel Med.*, Aug 2021.
- [522] Lixin Lin, Yanji Zhao, Boqiang Chen, and Daihai He. Multiple covid-19 waves and vaccination effectiveness in the united states. *International journal of environmental research and public health*, 19(4):2282, 2022.
- [523] Anastasia S Lambrou, Philip Shirk, Molly K Steele, Prabasaj Paul, Clinton R Paden, Betsy Cadwell, Heather E Reese, Yutaka Aoki, Norman Hassell, Xiao-yu Zheng, et al. Genomic surveillance for sars-cov-2 variants: predominance of the delta (b. 1.617. 2) and omicron (b. 1.1. 529) variants—united states, june 2021–january 2022. *Morbidity and Mortality Weekly Report*, 71(6):206, 2022.
- [524] Rigel Suzuki, Daichi Yamasoba, Izumi Kimura, Lei Wang, Mai Kishimoto, Jumpei Ito, Yuhei Morioka, Naganori Nao, Hesham Nasser, Keiya Uriu, et al. Attenuated fusogenicity and pathogenicity of sars-cov-2 omicron variant. *Nature*, pages 1–6, 2022.
- [525] Tegan K Boehmer, Jourdan DeVies, Elise Caruso, Katharina L van Santen, Shichao Tang, Carla L Black, Kathleen P Hartnett, Aaron Kite-Powell, Stephanie Dietz, Matthew Lozier, et al. Changing age distribution of the covid-19 pandemic—united states, may–august 2020. *Morbidity and Mortality Weekly Report*, 69(39):1404, 2020.
- [526] Jeffrey E Harris. Data from the covid-19 epidemic in florida suggest that younger cohorts have been transmitting their infections to less socially mobile older adults. *Review of Economics of the Household*, 18(4):1019–1037, 2020.
- [527] Robert Verity, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, Hayley Thompson, Patrick GT Walker, Han Fu, et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *The Lancet infectious diseases*, 20(6):669–677, 2020.
- [528] Selene Ghisolfi, Ingvild Almås, Justin C Sandefur, Tillman von Carnap, Jesse Heitner, and Tessa Bold. Predicted covid-19 fatality rates based on age, sex, comorbidities and health system capacity. *BMJ global health*, 5(9):e003094, 2020.
- [529] Seung-Ji Kang and Sook In Jung. Age-related morbidity and mortality among patients with covid-19. *Infection & chemotherapy*, 52(2):154, 2020.
- [530] Clara Bonanad, Sergio García-Blas, Francisco Tarazona-Santabalbina, Juan Sanchis, Vicente Bertomeu-González, Lorenzo Fáfila, Albert Ariza, Julio Nunez, and Alberto Cordero. The effect of age on mortality in patients with covid-19: a meta-analysis with 611,583 subjects. *Journal of the American Medical Directors Association*, 21(7):915–918, 2020.

- [531] Albert Esteve, Iñaki Permanyer, Diederik Boertien, and James W Vaupel. National age and coresidence patterns shape covid-19 vulnerability. *Proceedings of the National Academy of Sciences*, 117(28):16118–16120, 2020.
- [532] Our World in Data. <https://ourworldindata.org/us-states-vaccinations>. [Online; accessed: 2020-04-08].
- [533] The new york times github covid-19 repository. <https://github.com/nytimes/covid-19-data>. [Online; accessed: 2020-04-08].
- [534] Arno Karlen. *Plague's progress: a social history of man and disease*. Gollancz, 1995.
- [535] Jo N Hays. *Epidemics and pandemics: their impacts on human history*. Abc-clio, 2005.
- [536] Zhen Wang, Chris T Bauch, Samit Bhattacharyya, Alberto d’Onofrio, Piero Manfredi, Matjaž Perc, Nicola Perra, Marcel Salathé, and Dawei Zhao. Statistical physics of vaccination. *Physics Reports*, 664:1–113, 2016.
- [537] Calum NL Macpherson. Human behaviour and the epidemiology of parasitic zoonoses. *International journal for parasitology*, 35(11-12):1319–1331, 2005.
- [538] Bruno B Chomel, Albino Belotto, and François-Xavier Meslin. Wildlife, exotic pets, and emerging zoonoses. *Emerging infectious diseases*, 13(1):6, 2007.
- [539] Julien Fosse, Henri Seegers, and Catherine Magras. Foodborne zoonoses due to meat: a quantitative approach for a comparative risk assessment applied to pig slaughtering in europe. *Veterinary research*, 39(1):1–16, 2008.
- [540] Emilie Alirol, Laurent Getaz, Beat Stoll, François Chappuis, and Louis Loutan. Urbanisation and infectious diseases in a globalised world. *The Lancet infectious diseases*, 11(2):131–141, 2011.
- [541] Marcia C Inhorn and Peter J Brown. The anthropology of infectious disease. *The Anthropology of Infectious Disease*, pages 31–67, 2013.
- [542] Carl-Johan Neiderud. How urbanization affects the epidemiology of emerging infectious diseases. *Infection ecology & epidemiology*, 5(1):27060, 2015.
- [543] Noel G Brizuela, Néstor García-Chan, Humberto Gutierrez Pulido, and Gerardo Chowell. Understanding the role of urban design in disease spreading. *Proceedings of the Royal Society A*, 477(2245):20200524, 2021.
- [544] Mary E Wilson. Travel and the emergence of infectious diseases. *Emerging infectious diseases*, 1(2):39, 1995.
- [545] Nils Daulaire. Globalization and health. *Development*, 42(4):22–24, 1999.
- [546] Stacey Knobler, Adel Mahmoud, Stanley Lemon, and Leslie Pray. The impact of globalization on infectious disease emergence and control: Exploring the consequences and opportunities. 2006.
- [547] Julie Fournet and Alain Barrat. Contact patterns among high school students. *PloS one*, 9(9):e107878, 2014.
- [548] Bing Wang, Lang Cao, Hideyuki Suzuki, and Kazuyuki Aihara. Safety-information-driven human mobility patterns with metapopulation epidemic dynamics. *Scientific reports*, 2(1):887, 2012.
- [549] Joshua M Epstein, Jon Parker, Derek Cummings, and Ross A Hammond. Coupled contagion dynamics of fear and disease: mathematical and computational explorations. *PloS one*, 3(12):e3955, 2008.
- [550] Shalini Quadros, Shalini Garg, Rupesh Ranjan, Guruprasad Vijayasarathi, and Mohammed A Mamun. Fear of covid 19 infection across different cohorts: a scoping review. *Frontiers in Psychiatry*, 12:708430, 2021.

- [551] Jizhou Huang, Haifeng Wang, Miao Fan, An Zhuo, Yibo Sun, and Ying Li. Understanding the impact of the covid-19 pandemic on transportation-related behaviors with human mobility data. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3443–3450, 2020.
- [552] Juanjuan Zhang, Maria Litvinova, Yuxia Liang, Yan Wang, Wei Wang, Shanlu Zhao, Qianhui Wu, Stefano Merler, Cécile Viboud, Alessandro Vespignani, et al. Changes in contact patterns shape the dynamics of the covid-19 outbreak in china. *Science*, 368(6498):1481–1486, 2020.
- [553] Alberto d’Onofrio and Piero Manfredi. Information-related changes in contact patterns may trigger oscillations in the endemic prevalence of infectious diseases. *Journal of Theoretical Biology*, 256(3):473–478, 2009.
- [554] Jiangzhuo Chen, Achla Marathe, and Madhav Marathe. Coevolution of epidemics, social networks, and individual behavior: A case study. In *Advances in Social Computing: Third International Conference on Social Computing, Behavioral Modeling, and Prediction, SBP 2010, Bethesda, MD, USA, March 30-31, 2010. Proceedings 3*, pages 218–227. Springer, 2010.
- [555] Eve Dubé, Caroline Laberge, Maryse Guay, Paul Bramadat, Réal Roy, and Julie A Bettinger. Vaccine hesitancy: an overview. *Human vaccines & immunotherapeutics*, 9(8):1763–1773, 2013.
- [556] Pavan V Ganapathiraju, Christiaan B Morssink, and James Plumb. Endgame for polio eradication? options for overcoming social and political factors in the progress to eradicating polio. *Global public health*, 10(4):463–473, 2015.
- [557] Malik Sallam. Covid-19 vaccine hesitancy worldwide: a concise systematic review of vaccine acceptance rates. *Vaccines*, 9(2):160, 2021.
- [558] Leonardo Carlucci, Ines D’ambrosio, and Michela Balsamo. Demographic and attitudinal factors of adherence to quarantine guidelines during covid-19: the italian model. *Frontiers in psychology*, 11:559288, 2020.
- [559] W Song, FJ Sawafta, BM Ebrahem, and MA Jebril. Public attitude towards quarantine during the covid-19 outbreak. *Epidemiology & Infection*, 148:e220, 2020.
- [560] Hamzeh Al Zabadi, Noor Yaseen, Thair Alhroub, and Maryam Haj-Yahya. Assessment of quarantine understanding and adherence to lockdown measures during the covid-19 pandemic in palestine: community experience and evidence for action. *Frontiers in Public Health*, 9:570242, 2021.
- [561] Patricio Goldstein, Eduardo Levy Yeyati, and Luca Sartorio. Lockdown fatigue: The diminishing effects of quarantines on the spread of covid-19. *CID Working Paper Series*, 2021.
- [562] Arielle Kaim, Maya Siman-Tov, Eli Jaffe, and Bruria Adini. Factors that enhance or impede compliance of the public with governmental regulation of lockdown during covid-19 in israel. *International Journal of Disaster Risk Reduction*, 66:102596, 2021.
- [563] Neil Ferguson. Capturing human behaviour. *Nature*, 446(7137):733–733, 2007.
- [564] Herbert W Hethcote. An immunization model for a heterogeneous population. *Theoretical population biology*, 14(3):338–349, 1978.
- [565] Christopher M Kribs-Zaleta and Jorge X Velasco-Hernández. A simple vaccination model with multiple endemic states. *Mathematical biosciences*, 164(2):183–201, 2000.
- [566] Xianning Liu, Yasuhiro Takeuchi, and Shingo Iwami. Svir epidemic models with vaccination strategies. *Journal of Theoretical biology*, 253(1):1–11, 2008.
- [567] Boris Shulgin, Lewi Stone, and Zvia Agur. Pulse vaccination strategy in the sir epidemic model. *Bulletin of mathematical biology*, 60(6):1123–1148, 1998.
- [568] Almut Scherer and Angela McLean. Mathematical models of vaccination. *British medical bulletin*, 62(1):187–199, 2002.
- [569] E Shim, Z Feng, M Martcheva, and Carlos Castillo-Chavez. An age-structured epidemic model of rotavirus with vaccination. *Journal of mathematical biology*, 53:719–746, 2006.

- [570] Ying-Hen Hsieh. Age groups and spread of influenza: implications for vaccination strategy. *BMC infectious diseases*, 10:1–12, 2010.
- [571] Meagan C Fitzpatrick and Alison P Galvani. Optimizing age-specific vaccination. *Science*, 371(6532):890–891, 2021.
- [572] Liang Mao and Yan Yang. Coupling infectious diseases, human preventive behavior, and networks—a conceptual framework for epidemic modeling. *Social science & medicine*, 74(2):167–175, 2012.
- [573] Zhen Wang, Michael A Andrews, Zhi-Xi Wu, Lin Wang, and Chris T Bauch. Coupled disease–behavior dynamics on complex networks: A review. *Physics of life reviews*, 15:1–29, 2015.
- [574] Alessandro Rizzo, Mattia Frasca, and Maurizio Porfiri. Effect of individual behavior on epidemic spreading in activity-driven networks. *Physical Review E*, 90(4):042801, 2014.
- [575] Qian Yin, Zhishuang Wang, Chengyi Xia, and Chris T Bauch. Impact of co-evolution of negative vaccine-related information, vaccination behavior and epidemic spreading in multilayer networks. *Communications in Nonlinear Science and Numerical Simulation*, 109:106312, 2022.
- [576] Yoshiro Iwamura and Jun Tanimoto. Realistic decision-making processes in a vaccination game. *Physica A: Statistical Mechanics and its Applications*, 494:236–241, 2018.
- [577] Shunjiang Ni, Wenguo Weng, and Hui Zhang. Modeling the effects of social impact on epidemic spreading in complex networks. *Physica A: Statistical Mechanics and its Applications*, 390(23–24):4528–4534, 2011.
- [578] Lucila G Alvarez-Zuzek, Cristian E La Rocca, José R Iglesias, and Lidia A Braunstein. Epidemic spreading in multiplex networks influenced by opinion exchanges on vaccination. *PloS one*, 12(11):e0186492, 2017.
- [579] Marzena Fügenschuh and Feng Fu. Overcoming vaccine hesitancy by multiplex social network targeting. In *International Conference on Complex Networks and Their Applications*, pages 576–587. Springer, 2022.
- [580] Alexandra Teslya, Hendrik Nunner, Vincent Buskens, and Mirjam E Kretzschmar. The effect of competition between health opinions on epidemic dynamics. *PNAS nexus*, 1(5):pgac260, 2022.
- [581] Mark Granovetter. Threshold models of collective behavior. *American journal of sociology*, 83(6):1420–1443, 1978.
- [582] Duncan J Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, 2002.
- [583] Duncan J Watts and Peter Sheridan Dodds. Influentials, networks, and public opinion formation. *Journal of consumer research*, 34(4):441–458, 2007.
- [584] Damon Centola and Michael Macy. Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3):702–734, 2007.
- [585] Byungjoon Min and Maxi San Miguel. Competing contagion processes: Complex contagion triggered by simple contagion. *Scientific reports*, 8(1):10422, 2018.
- [586] Giulia Cencetti, Diego Andrés Contreras, Marco Mancastroppe, and Alain Barrat. Distinguishing simple and complex contagion processes on networks. *Physical Review Letters*, 130(24):247401, 2023.
- [587] Marcel Salathé and Sebastian Bonhoeffer. The effect of opinion clustering on disease outbreaks. *Journal of The Royal Society Interface*, 5(29):1505–1508, 2008.
- [588] Ellsworth Campbell and Marcel Salathé. Complex social contagion makes networks more vulnerable to disease outbreaks. *Scientific reports*, 3(1):1905, 2013.
- [589] Feng Fu, Nicholas A Christakis, and James H Fowler. Dueling biological and social contagions. *Scientific reports*, 7(1):43634, 2017.
- [590] Damon Centola. The complex contagion of doubt in the anti-vaccine movement. In *Meeting the challenge of vaccination hesitancy*, pages 126–136. The Aspen Institute Washington DC, 2020.

- [591] Laurent Hébert-Dufresne, Dina Mistry, and Benjamin M Althouse. Spread of infectious disease and social awareness as parasitic contagions on clustered networks. *Physical Review Research*, 2(3):033306, 2020.
- [592] Zirou Qiu, Baltazar Espinoza, Vitor V Vasconcelos, Chen Chen, Sara M Constantino, Stefani A Crabtree, Luojun Yang, Anil Vullikanti, Jiangzhuo Chen, Jörgen Weibull, et al. Understanding the coevolution of mask wearing and epidemics: A network perspective. *Proceedings of the National Academy of Sciences*, 119(26):e2123355119, 2022.
- [593] Our World in Data. Daily covid-19 vaccine doses administered. <https://ourworldindata.org/grapher/daily-covid-19-vaccination-doses?tab=chart&country=~USA>, 2023. Accessed on 2023-10-26.
- [594] USA Facts. What's the nation's progress on vaccinations? <https://usafacts.org/visualizations/covid-vaccine-tracker-states/>, 2023. Accessed on 2023-10-26.
- [595] Our World in Data. Daily share of the population receiving a covid-19 vaccine dose. <https://ourworldindata.org/grapher/daily-covid-19-vaccination-doses?tab=chart&country=~USA>, 2023. Accessed on 2023-11-2.
- [596] Eve Dubé and Noni E MacDonald. Vaccine acceptance: barriers, perceived risks, benefits, and irrational beliefs. In *The vaccine book*, pages 507–528. Elsevier, 2016.
- [597] Matteo Bizzarri, Fabrizio Panebianco, and Paolo Pin. Epidemic dynamics with homophily, vaccination choices, and pseudoscience attitudes. *arXiv preprint arXiv:2007.08523*, 2020.
- [598] Rohan S Mehta and Noah A Rosenberg. Modelling anti-vaccine sentiment as a cultural pathogen. *Evolutionary Human Sciences*, 2:e21, 2020.
- [599] Johannes Müller, Aurélien Tellier, and Michael Kurschilgen. Echo chambers and opinion dynamics explain the occurrence of vaccination hesitancy. *Royal Society Open Science*, 9(10):220367, 2022.
- [600] Steve Rathje, James K He, Jon Roozenbeek, Jay J Van Bavel, and Sander van der Linden. Social media behavior is associated with vaccine hesitancy. *PNAS nexus*, 1(4):pgac207, 2022.
- [601] Alberto Aleta, Guilherme Ferraz de Arruda, and Yamir Moreno. Data-driven contact structures: from homogeneous mixing to multilayer networks. *PLoS computational biology*, 16(7):e1008035, 2020.
- [602] Xiaoyi Yuan and Andrew Crooks. From cyber space opinion leaders and the diffusion of anti-vaccine extremism to physical space disease outbreaks. In *Social, Cultural, and Behavioral Modeling: 10th International Conference, SBP-BRIMS 2017, Washington, DC, USA, July 5-8, 2017, Proceedings 10*, pages 114–119. Springer, 2017.
- [603] Steven Lloyd Wilson and Charles Wiysonge. Social media and vaccine hesitancy. *BMJ global health*, 5(10):e004206, 2020.
- [604] Gaurav Jain, Avinash Bapu Sreenivas, Samrat Gupta, and Amit Anand Tiwari. The dynamics of online opinion formation: polarization around the vaccine development for covid-19. In *Causes and symptoms of socio-cultural polarization: role of information and communication technologies*, pages 51–72. Springer, 2022.
- [605] Ana Lucía Schmidt, Fabiana Zollo, Antonio Scala, Cornelia Betsch, and Walter Quattrociocchi. Polarization of the vaccination debate on facebook. *Vaccine*, 36(25):3606–3612, 2018.
- [606] Will Jennings, Gerry Stoker, Hannah Willis, Viktor Valgardsson, Jen Gaskell, Daniel Devine, Lawrence McKay, and Melinda C Mills. Lack of trust and social media echo chambers predict covid-19 vaccine hesitancy. *MedRxiv*, pages 2021–01, 2021.

# Figure List

1.1	Bills of mortality . . . . .	9
1.2	Farr's curve depiction . . . . .	10
1.3	Black Death and COVID-19 spatial spread . . . . .	20
1.4	Different scales structure used in epidemic modeling . . . . .	32
2.1	SIR model . . . . .	46
2.2	Age-structured contact matrices . . . . .	58
2.3	ER and BA degree distributions and networks . . . . .	64
2.4	Metapopulation scheme . . . . .	69
2.5	EPR model scheme . . . . .	83
3.1	Madrid's epidemiological trajectory . . . . .	102
3.2	Madrid's mobility patterns . . . . .	103
3.3	Epidemic impact curves for varying control parameter values . . . . .	104
3.4	Epidemic impact when varying simultaneously $\kappa$ and $\chi$ for selected $\Theta$ scenarios . . . . .	106
4.1	Greater Boston Area attractiveness field and distribution . . . . .	117
4.2	Invasion event share and invasion time profiles per $\rho$ group . . . . .	119
4.3	Invasion spatial patterns . . . . .	120
4.4	Disease's prevalence and infection time profiles per $\rho$ group . . . . .	121
4.5	Contagion spatial patterns . . . . .	123
4.6	Infections at home locations and outside . . . . .	124
4.7	Contribution to infections per $\rho$ group . . . . .	126
4.8	Event and trip decomposition by origin and destination of infections . . . . .	127
4.9	Contagion events and locations attractiveness . . . . .	128
5.1	Proposed baseline scenario . . . . .	139
5.2	Comparison of spreading dynamics . . . . .	141
5.3	Attack rates scatter plots . . . . .	142
5.4	US map of attack rates and vaccine hesitancy . . . . .	143
5.5	Death scatter plots . . . . .	145
5.6	Attack rate scatter plots by age . . . . .	146
5.7	Data and model comparison . . . . .	149
6.1	Homogeneous thresholds in Erdős-Rényi networks . . . . .	162
6.2	Homogeneous thresholds in Barabási-Albert networks . . . . .	165
6.3	Comparison of impact at $n_Z = 0$ and $n_Z = 0.5$ on Erdős-Rényi networks . . . . .	167
6.4	Varying anti-vaccine attitude ( $n_Z$ ) curves on Erdős-Rényi networks . . . . .	168
6.5	Comparison of impact at $n_Z = 0$ and $n_Z = 0.5$ on Barabási-Albert networks . . . . .	170
6.6	Varying anti-vaccine attitude ( $n_Z$ ) curves on Barabási-Albert networks . . . . .	171
6.7	US states' population shares by attitude towards vaccination. . . . .	172
6.8	Epidemiological impact across US states for varying vaccination rates $\alpha$ . . . . .	173
6.9	Pairwise scatter plots of prevalence and vaccine attitudes . . . . .	175



# Table List

3.1	Madrid's population data (2020) . . . . .	101
3.2	Epidemic impact with the threshold $\Theta = 20$ . . . . .	107
3.3	Epidemic impact with the threshold $\Theta = 500$ . . . . .	107
5.1	Averted deaths . . . . .	147
6.1	Vaccination attitude categories and activation threshold attribution . . . . .	159



# **Appendix**



# Appendix A

## Codes

A lot of a scientist's time today, aside from paperwork, is spent coding, debugging, curating data for models, analyzing results, and creating visual representations. Computers have long become the primary laboratory for a growing segment of the scientific community. As a result, the ideal scientist should not only be an expert in their field but also possess advanced skills in information and computational management. Quite a task!

There's still so much for me to learn and master in order to produce high-quality, well-organized code effortlessly (yes, I admit to some laziness). Honestly, some of my code is a bit embarrassing, but I've committed to transparency and decided to share this usually hidden work. So, here are the repositories containing the code developed for the projects in this thesis. As of now, some parts need improvement for clarity and reproducibility. However, at the very least, I've provided the basic code for others to see how the simulations were created. This should offer the fundamental components needed to replicate the results presented in my thesis.

- Chapter 3: <https://github.com/phononautomata/madrid>
- Chapter 4: <https://github.com/phononautomata/expidemics>
- Chapter 5: <https://github.com/phononautomata/hesitancy>
- Chapter 6: <https://github.com/phononautomata/threshold>

I will try to keep these links updated in the future in order to offer a more polished code, pipeline, and documentation.



# Resumen

Esta tesis, profundamente influenciada por el advenimiento de la pandemia de COVID-19, explora la modelización de propagación de epidemias en poblaciones humanas, así como sus consecuentes impactos. Aparte del tema principal, dos hilos subyacentes guían los trabajos presentados: i) la utilización de datos reales con el fin de ofrecer una capa de mayor realismo a los modelos y sus respectivas conclusiones, y ii) la aplicación de modelos mecanicistas computacionales para describir la propagación de epidemias con el fin de ofrecer una explicación basada en principios fundamentales. Los resultados aquí proporcionados buscan contribuir humildemente al conocimiento acumulado de la literatura del campo y aportar perspectivas para enriquecer el debate y la formulación de políticas de salud pública.

Los Capítulos 1 y 2 conforman *Part I: Overture*, y respectivamente ofrecen, con el inevitable sesgo de un físico de formación, una revisión extensa del campo y la base teórica para enmarcar y comprender la colección subsiguiente de trabajos presentados.

*Part II: Metapopulation models* alberga los capítulos 3 y 4 y está dedicada a trabajos relacionados con la propagación espacial de enfermedades infecciosas y el papel de la movilidad, principalmente en entornos urbanos.

En el Capítulo 3, se examina la efectividad de los confinamientos perimetrales a escala urbana, implementados en la ciudad de Madrid durante la pandemia de COVID-19. Usamos este caso real como inspiración para estudiar, en un contexto idealizado, cuán efectiva puede ser esta estrategia para contener una epidemia. Usando un modelo metapoblacional informado por datos reales de movilidad, encontramos que en sistemas urbanos altamente interconectados, las epidemias se propagan con rapidez y estos confinamientos son ineficaces. Las aproximaciones realizadas, incluida la vigilancia perfecta y el cumplimiento total de las medidas, resaltan aún más las limitaciones de tal estrategia. Esta investigación se suma a la evidencia que indica que las restricciones de movilidad son inefectivas a menos que sean lo suficientemente estrictas. Así, los esfuerzos deberían dirigirse a entender y modificar los patrones de contacto poblacionales, que impulsan la transmisión de enfermedades infecciosas.

En el Capítulo 4, se profundiza en el papel de la movilidad de exploración y retorno preferencial en la propagación espacial de epidemias y se trata de avanzar en los modelos metapoblacionales estándar que generalmente ignoran características de movilidad realistas.

Utilizamos el modelo d-EPR, informado por datos de movilidad de alta resolución del área metropolitana de Boston. Se analizaron diferentes configuraciones y escenarios, revelando que solo el modelo d-EPR con una distribución de movilidad heterogénea mostró un impacto distinto entre los grupos de movilidad. Los hallazgos indican que los exploradores, aquellos con una elevada movilidad, contribuyen significativamente en la propagación de la epidemia.

En *Part III: Single-population behavioral structured models*, se presentan modelos considerando estructura de edad (capítulo 5) y estructura de red de contactos (6). Aquí, se evalúa el papel del comportamiento humano en epidemias, particularmente en relación con la reticencia hacia las vacunas.

El Capítulo 5 evalúa el impacto de la reticencia hacia las vacunas, exacerbada por preocupaciones y desinformación durante la pandemia de COVID-19. Aprovechando encuestas realizadas en EE. UU. (a nivel nacional y de estado) sobre predisposición a la vacunación, y matrices de contacto de alta resolución estructuradas por grupos de edad, se buscó evaluar el impacto epidemiológico de la reticencia a la vacunación en brotes secundarios de COVID-19, especialmente a la luz de la rápida propagación de la variante Delta. Los hallazgos indican una correlación directa entre el tamaño de los brotes secundarios y la reticencia frente a la vacuna. La edad juega un papel crucial: mientras que los grupos más jóvenes influyen significativamente en las tasas de ataque (prevención), los ancianos afectan notablemente las tasas de mortalidad. Es crucial destacar que el modelo del estudio se alinea bien con los datos reales de vacunación y mortalidad durante la ola Delta.

En el Capítulo 6, analizamos la interacción entre dinámicas de enfermedades infecciosas, vacunación y dinámica de opinión sobre vacunación con un modelo de umbrales en redes homogéneas y heterogéneas. En general, a medida que aumentan las tasas de vacunación, emerge un comportamiento pro-vacunación en cascada, ayudando a controlar la propagación y favoreciendo la aparición de una creciente fase libre de enfermedad en una parte significativa del espacio de parámetros. Sin embargo, para redes heterogéneas, las tasas de vacunación deben ser excepcionalmente altas para prevenir brotes epidémicos grandes. El impacto de las actitudes antivacunas en las tasas de vacunación es mínimo en niveles más bajos pero se vuelve significativo a medida que estas tasas aumentan. La introducción de individuos antivacunas eleva efectivamente el umbral pro-vacunación del sistema, lo que puede conducir a una mayor prevalencia de enfermedades. El estudio concluye caracterizando el impacto en escenarios más realistas, aprovechando datos de encuestas sobre vacunación en todos los estados de EE. UU.

Finalmente, en *Parte IV: Closure*, su capítulo homónimo 7 resume el trabajo presentado en este texto y ofrece unas reflexiones finales.

# Clausura

## Resumen y principales conclusiones

Esta disertación comprende varios trabajos sobre la modelización de la propagación de epidemias en poblaciones humanas y el análisis de su impacto. En *Part I: Overture*, presentamos una extensa introducción al campo en el Capítulo 1 y cubrimos los fundamentos—y algunos aspectos avanzados—del marco teórico utilizado a lo largo de esta disertación en el Capítulo 2. Las contribuciones originales están estructuradas en dos partes. *Part II: Metapopulation models*, que contiene los Capítulos 3 y 4, se centra en la propagación espacial de enfermedades infecciosas y el papel de la movilidad. *Part III: Single-population behavioral structured models* incluye los Capítulos 5 y 6, y se centra en la integración de aspectos del comportamiento humano en modelos con estructura de edad o de contacto. Aunque los trabajos fueron discutidos y se ofrecieron comentarios finales en cada capítulo correspondiente, los resumiremos nuevamente y ofreceremos algunas reflexiones finales de carácter general.

Comenzamos *Part II* con el Capítulo 3, donde examinamos una situación real motivada por la pandemia de COVID-19. Como ya se ha explicado, la pandemia de COVID-19, junto con la consiguiente crisis sanitaria y socioeconómica, tomó por sorpresa a todas las autoridades. La ausencia de planes preventivos competentes contra brotes de enfermedades infecciosas emergentes, como fue el caso en España, resultó en que las autoridades tuvieran que improvisar, viéndose obligadas a implementar medidas de control altamente agresivas para doblegar la curva y evitar un impacto catastrófico en la población. La aplicación de confinamientos perimetrales en la Comunidad de Madrid, así como en la ciudad de Madrid a nivel de Zonas Básicas de Salud, puede verse como una de esas medidas que fueron tanto disruptivas como de limitada efectividad para contener la propagación. Así, en este trabajo, desarrollamos un modelo de metapoblaciones de la ciudad de Madrid a nivel de distrito, informado por flujos de movilidad de datos reales proporcionados por el Ministerio de Transporte, Movilidad y Agenda Urbana de España, evaluamos el potencial éxito de la estrategia de confinamiento perimetral a la hora de contener la propagación de la enfermedad y mitigar su impacto en la población. Aunque el modelo está simplificado y no

está diseñado para predecir escenarios futuros o retrodecir la experiencia real de Madrid, demuestra con suficiencia que en sistemas urbanos altamente interconectados, la enfermedad se propaga fácil y rápidamente a través de todo el sistema. Los brotes locales tienden a sincronizarse, haciendo prácticamente imposible contener la enfermedad en áreas localizadas mientras se espera que el resto del sistema funcione con normalidad. Debe señalarse que nuestro modelo asume una información perfecta del sistema de vigilancia, un cumplimiento absoluto de las reglas y ninguna permeabilidad excepcional. Incluso con estas simplificaciones significativas, que teóricamente deberían favorecer esta estrategia, los resultados indican la ineficacia de los confinamientos perimetrales en un contexto urbano.

Se ha puesto considerable énfasis en las restricciones de movilidad a diferentes escalas espaciales antes y durante esta pandemia. Sin embargo, nuestro estudio contribuye a la evidencia establecida de que las reducciones de movilidad, en el mejor de los casos, solo permiten un retraso modesto en la progresión de la epidemia. Esta reducción debe ser absurdamente grande en magnitud, interrumpiendo significativamente los patrones de movilidad normales, y debería implementarse casi inmediatamente desde los primeros casos detectados. En consecuencia, argumentamos que el enfoque debería ponerse en cambio en los patrones de contacto y mezcla de la población, que en última instancia impulsan la transmisión de enfermedades similares a la influenza.

Pasando al Capítulo 4, este trabajo resuena estrechamente con el anterior, ya que también trata sobre la propagación espacial de una epidemia y el papel de la movilidad. Impulsados por avances en la modelización de la movilidad humana y conocimientos como los mecanismos de exploración y retorno preferencial, y reconociendo que la literatura sobre metapoblación epidémica aún tiene que explorar la implementación de modelos de movilidad microscópica más realistas, identificamos una oportunidad para contribuir con algo novedoso. Dentro de la familia de modelos EPR, seleccionamos el modelo d-EPR, buscando un equilibrio entre realismo y facilidad de implementación. Como la etapa de exploración del modelo sigue el modelo de gravedad de movilidad humana, a las localizaciones donde se mezcla la población se les asignó un valor que refleja su atractivo o relevancia. Esta asignación está informada por datos reales del Área Metropolitana de Boston, gracias a la abundante información de movilidad de alta resolución de registros de teléfonos móviles. Además, para evaluar adecuadamente las fortalezas y debilidades del modelo d-EPR, investigamos la propagación espacial de la enfermedad bajo diferentes configuraciones y escenarios. Estos se refieren tanto a la distribución de homogeneidad/heterogeneidad de actitudes de movilidad en la población, como a variaciones del modelo d-EPR original, donde se eliminan algunas

características fundamentales. Encontramos que sólo el modelo d-EPR bajo una distribución heterogénea del parámetro que caracteriza la movilidad de los agentes es capaz de mostrar algún impacto diferencial entre grupos de movilidad en algunos de los observables epidemiológicos evaluados. Específicamente, las características no markovianas en los mecanismos de exploración y retorno preferencial se antojan fundamentales para observar que este tipo de movilidad juega un papel importante en el desarrollo de la epidemia. Los exploradores impulsan la invasión de la enfermedad y, por lo tanto, ayudan a difundir la enfermedad en todo el sistema, contribuyen proporcionalmente más que los perfiles de movilidad retornadores, y significativamente más rápido. En cuanto a la prevalencia de la enfermedad, los retornadores tienden a verse claramente menos afectados que los exploradores. Sin embargo, los tiempos medios de infección por grupo de movilidad muestran poca diferencia. Al investigar de manera más detallada del proceso de infección, se descubre que los agentes tienden a infectarse con más frecuencia fuera que en sus lugares de origen. Esto se debe simplemente al hecho de que incluso para valores de bajos y medios del perfil de movilidad, los agentes tienden a pasar más tiempo fuera de sus localizaciones de origen y a visitar lugares que, de media, tienden a ser más atractivos. Como las localizaciones más atractivas tienden a experimentar un mayor número de sucesos de contagio, claramente, la exploración facilita la difusión de la enfermedad. El hecho de que incluso agentes con baja-media movilidad puedan acceder con facilidad a localizaciones muy atractivas hace que la epidemia se extienda fácilmente por todo el sistema en un breve intervalo de tiempo.

En resumen, el acoplamiento del modelo de movilidad humana microscópica d-EPR con un modelo epidemiológico aporta nuevas perspectivas en la interacción entre movilidad y la propagación de una epidemia. Debemos admitir, no obstante, que nuestro modelo e implementación presenta algunas limitaciones que deberán abordarse en trabajos futuros, tales como la comparación del modelo frente a la simulación de una epidemia con movilidad basada en trayectorias reales, o directamente la consideración de modelos de movilidad que incorporen características de movilidad aún más realistas. Mientras tanto, este es un pequeño paso hacia el reconocimiento de la heterogeneidad y complejidad empíricamente observada en el ámbito de la movilidad humana, así como en la evaluación de su papel y consecuencias en la difusión de epidemias.

Iniciando *Part III*, el Capítulo 5 examina los efectos de la reticencia a la vacunación del COVID-19 en poblaciones estructuradas por edad. Esta investigación, impulsada por la pandemia de COVID-19, coincide con el período en el que se iniciaron las campañas de vacunación tras un esfuerzo científico sin precedentes y se logró desarrollar vacunas efectivas en un tiempo récord. Paradójicamente, el rápido desarrollo de estas

vacunas, junto con las *infodemias* generalizadas en las redes sociales e Internet, despertó preocupaciones en ciertos sectores de la población. Esto resultó en la difusión de desinformación y falsedades flagrantes sobre la seguridad y eficacia de las vacunas contra la COVID-19. Utilizando encuestas sobre la aceptación de la vacuna realizadas en los EE. UU., junto con matrices de contacto detalladas por edad para los estados del país, exploramos el impacto potencial de brotes secundarios de COVID-19. La pregunta central era determinar hasta qué punto la reticencia a la vacuna podría provocar brotes secundarios significativos, especialmente en un entorno dominado por una variante del SARS-CoV-2 más transmisible y preocupante. Esta cuestión puede entenderse plenamente al recordar que, después de medio año del inicio de la campaña de vacunación, la ola de la variante Delta altamente transmisible estaba cobrando fuerza en EE. UU. y en el resto del mundo. Nuestro enfoque involucró un modelo SIR segmentado en 85 franjas de edad. Aunque la sencillez del modelo SIR quizás no capture completamente las complejidades de la COVID-19, la dinámica de transmisión del SARS-CoV-2 y el proceso de vacunación, no obstante, se nutre de datos específicos del estado sobre prevalencia, demografía, encuestas de aceptación de la vacuna y patrones de contacto estructurados por edad de alta resolución.

El análisis arrojó una clara conexión entre la magnitud de los brotes secundarios en cada estado y el nivel de reticencia a la vacunación presente. Al profundizar en la influencia de la distribución por edades, descubrimos una fuerte asociación entre la prevalencia de brotes y la proporción de individuos aún susceptibles (tras la ola inicial) en cada grupo de edad, siendo las poblaciones más jóvenes las que más significativamente contribuyen a las tasas de transmisión. No obstante, se aconseja precaución al proyectar estas correlaciones en las cifras de mortalidad esperadas, ya que el impacto de la edad en las tasas de mortalidad tiende a estar inversamente relacionado. Las proyecciones de nuestro modelo sugirieron que reducir la reticencia a la vacunación incluso en un punto porcentual podría prevenir numerosas muertes, subrayando la importancia de considerar la estructura de edad en la dinámica de transmisión. Notablemente, no todos los estados con altos niveles de reticencia encabezaron la lista de muertes prevenibles; en cambio, el porcentaje de individuos reticentes entre las poblaciones mayores emergió como un factor crucial. La comparación de los resultados de nuestro modelo con datos reales reveló altas correlaciones tanto para las tasas de vacunación predichas contra las cifras reales como para las muertes estimadas frente a las fatalidades registradas durante el auge de la variante Delta en EE. UU. En resumen, a pesar de la relativa simplicidad de nuestro modelo, el uso de datos granulares para procesos clave como la demografía, los patrones de contacto y las encuestas de actitud hacia la vacuna permitieron desarrollar una representación razonablemente precisa de

las repercusiones de la reticencia a la vacuna durante la pandemia de COVID-19 en curso. Como parte del discurso científico, este trabajo aporta valiosas perspectivas para apoyar la adopción generalizada de vacunas, respaldadas con datos y una modelización sólida.

Concluimos en el Capítulo 6 con una segunda mirada al papel de la reticencia a la vacunación, esta vez con poblaciones estructuradas en redes de contacto y modelando explícitamente la interacción entre la vacunación y las actitudes hacia ella. Presentamos un modelo que acopla la epidemiología de una enfermedad similar a la gripe (usando el modelo SIR), una dinámica de una campaña de vacunación y una dinámica de opinión respecto a la aceptación de la vacuna. Este último aspecto se informa por el modelo de umbral de Watts-Granovetter, con un enfoque ligeramente diferente: nuestro modelo tiene en cuenta el estado de vacunación de los individuos en lugar de simplemente sus opiniones expresadas al respecto. El uso de este modelo de opinión está motivado a la luz de datos de encuestas de EE. UU. que sugieren que el estado de vacunación del entorno social influye significativamente en las decisiones de vacunación personales. Nuestro análisis se lleva a cabo en dos tipos de redes, Erdős-Rényi (ER) y Barabási-Albert (BA), examinando los resultados en una gama de parámetros que incluyen el umbral de activación de actitud pro-vacuna, la proporción inicial de individuos pro-vacuna y la tasa de vacunación.

Nuestros hallazgos revelan dinámicas distintas dentro de las redes ER y BA. En las redes ER, aumentar la tasa de vacunación puede fomentar un bucle de retroalimentación positiva entre la dinámica de opinión y la adopción real de la vacuna, reduciendo efectivamente la prevalencia de la enfermedad y potencialmente llevando a la erradicación de la enfermedad dentro de un espacio de parámetros realista. Sin embargo, la aparición de un estado libre de enfermedad en las redes BA requiere esfuerzos de vacunación más altos y menos prácticos, con una prevalencia general de la enfermedad más alta en comparación con las redes ER para cualquier tasa de vacunación dada. Adicionalmente, la introducción de individuos antivacunas  $n_Z$  en el modelo demuestra impactos variados en la cobertura de la vacunación y la prevalencia de la enfermedad, dependiendo de la posición del sistema en el espacio de parámetros. La presencia de sentimiento antivacunas es insignificante a tasas de vacunación bajas, pero se vuelve influyente a medida que las tasas de vacunación aumentan. Para un soporte inicial pro-vacunas dado  $n_A(0)$ , los entornos con diferentes umbrales de activación convergen a medida que la fracción de individuos antivacunas aumenta, tornando el umbral irrelevante en  $n_A + n_Z = 1$  debido al agotamiento del grupo de individuos indecisos. Además, emergen no linealidades en la cobertura de vacunación y la prevalencia para valores aumentados de la proporción de vacunación. Finalmente,

el modelo se aplica a los diferentes estados de EE. UU. para caracterizar el impacto epidemiológico resultado de la heterogeneidad de opinión respecto a la vacunación en estos territorios.

Para reforzar el realismo y el impacto práctico, las futuras direcciones de este trabajo deberían centrarse en integrar la demografía por edad en las redes de contacto, y en una representación más detallada de la toma de decisiones que tenga en cuenta consideraciones de coste-beneficio y los efectos de las redes sociales, como individuos altamente influyentes, desinformación, polarización y la creación de cámaras de eco.

\* \* \*

Esto concluye el resumen y las observaciones clave del trabajo principal desarrollado durante mi formación predoctoral. Aunque los estudios abordan diversas preguntas y problemas dentro de la epidemiología y el modelado de la propagación de epidemias, no se centran estrechamente en ningún problema singular, sino que abarcan una amplia gama de temas. Sin embargo, ciertos principios fundamentales están entrelazados en todo este cuerpo de trabajo. Uno de esos principios es el compromiso con métodos mecanicistas que modelan los procesos elementales que conducen a las dinámicas complejas observadas desde los primeros principios. Dada esta complejidad, los métodos numéricos y computacionales han sido los medios primarios y preferidos para implementar y resolver estos modelos, en lugar de confiar únicamente en enfoques analíticos. Aunque las simulaciones de procesos dinámicos utilizando enfoques mecanicistas pueden ser intensivas en computación y exigir en términos de uso de memoria, ofrecen al modelador la oportunidad de crear microcosmos intrincados que emulan procesos de la vida real con un nivel de detalle razonablemente alto. Este enfoque permite la exploración de diversas configuraciones, adaptaciones y escenarios hipotéticos, proporcionando una comprensión más clara de las dinámicas en juego.

Otro hilo conductor de estos trabajos es la integración de datos del mundo real para informar nuestros modelos siempre que sea posible, ofreciendo así una representación más fundamentada de los fenómenos bajo escrutinio dentro del alcance de las limitaciones de los modelos. Mientras que los modelos teóricos pueden iluminar los mecanismos fundamentales de dinámicas complejas y la exploración de parámetros puede revelar el espectro completo de comportamientos potenciales dentro de un modelo, dichos modelos están atados a la realidad solo por puntos de referencia empíricos. A veces, los modelos pueden volverse tan complejos y cargados de parámetros que lo que debería ser un emocionante viaje de descubrimiento se convierte en un desafío abrumador. En estos casos, los datos del mundo real resultan

de gran valor para parametrizar aspectos clave como los patrones de movilidad, la mezcla social o la progresión de la enfermedad, asegurando que nuestros esfuerzos permanezcan dirigidos y nuestros análisis sean precisos. Además, para lograr relevancia en nuestra investigación y posiblemente influir en la formulación de políticas o contribuir constructivamente al discurso público, los datos robustos y de alta calidad son indispensables.

## Perspectivas

Como el lector habrá notado a través del viaje que comenzó en el Capítulo 1, el campo de la epidemiología matemática y el modelado de epidemias ha evolucionado significativamente desde sus inicios en los siglos XIX y principios del XX. Ha crecido exponencialmente, muy parecido a los fenómenos que pretende modelar, y ha experimentado diversas olas de descubrimiento e innovación, siendo la pandemia de COVID-19 el último catalizador. La pandemia ha estimulado un extenso cuerpo de trabajo, tanto teórico como empírico, en el cual la investigación presentada aquí ocupa humildemente su lugar. Sin embargo, como bien entendemos, el progreso científico suele ser un proceso gradual, con pequeños pasos que se hilan sobre el conocimiento y las percepciones compartidas del pasado. Con el tiempo, estas contribuciones pueden acumularse y llevar a avances significativos, abriendo nuevas vías y desencadenando una explosión de descubrimientos hasta que esa área particular de estudio madura y se estabiliza, esperando nuevas preguntas y paradigmas para reavivar el ciclo de innovación.

En un campo tan vasto, interdisciplinario y maduro en muchos aspectos clave, parece difícil, para alguien que apenas comienza en este campo, prever qué esperar o qué caminos valen la pena seguir en los años venideros. Sin embargo, ofreceré, desde mi corta experiencia, un breve y general comentario sobre lo que siento que necesitamos para impulsar el campo hacia adelante.

Reflexionando sobre la evolución de este campo, o en efecto, cualquier campo de estudio, las hipótesis y modelos a menudo comienzan desde un estado bastante básico y sin refinar. Inicialmente basados en la evidencia y la observación, pueden ser bastante rudimentarios, pero a lo largo de los años e incluso décadas, estos modelos se refinan o experimentan cambios de paradigma que impulsan la progresión y sofisticación del campo. Tomemos, por ejemplo, el modelado de la fuerza de infección, que inherentemente implica modelar el proceso de contacto o mezcla entre individuos. Este concepto comenzó con la suposición de interacción de acción de masas y solo después de un siglo evolucionó con el advenimiento de la epidemiología de redes. Incluso cuando

nos encontramos al borde de una posible revolución en el estudio de las interacciones de orden superior, la suposición de *well-mixing* o mezcla homogénea sigue predominando en gran parte de la literatura y ciertas áreas dentro de la epidemiología matemática. Nuestro propio trabajo presentado aquí no está exento de esta dependencia de la suposición de mezcla homogénea. En defensa de esta suposición y nuestro trabajo, sigue siendo una simplificación justificable bajo circunstancias específicas, lo que explica su persistencia. Sin embargo, no hay duda de que explorar alternativas teóricas y empíricas a esta suposición, aquellas que son menos intensivas en computación y datos que los enfoques de redes, sería intrigante. Tales alternativas podrían proporcionar una representación más precisa del proceso de contagio.

Estrechamente relacionado con el enfoque de mezcla homogénea, la suposición de poblaciones homogéneas se extiende no solo a las interacciones sociales, sino también a otros atributos potencialmente relevantes que afectan a la propagación de epidemias, como el sexo, la edad, los patrones de movilidad o el estatus socioeconómico. Décadas después de los trabajos fundacionales en epidemiología matemática, el esfuerzo por medir e incorporar la heterogeneidad, así como para entender su impacto, comenzó, y continúa hasta el día de hoy. Reconocer la heterogeneidad de la población no es solo un reconocimiento de la diversidad; es un requisito crítico. La heterogeneidad influye significativamente en la dinámica de transmisión de enfermedades y prevalencia, ofreciendo una descripción más matizada y precisa de la propagación de epidemias y sus repercusiones. Este entendimiento permite una asignación de recursos más efectiva y la formulación de estrategias de control o mitigación eficientes.

Consideremos la respuesta internacional a la pandemia de COVID-19 como ejemplo. Acciones amplias e indiscriminadas se implementaron en muchas regiones, previniendo probablemente una pérdida abrumadora de vidas. Sin embargo, la profunda turbulencia psicológica, social y económica causada por estas medidas todavía resuena en las sociedades. Ojalá, un entendimiento más matizado de cómo las sociedades complejas y heterogéneas interactúan con la propagación de epidemias pudiera guiarnos en el desarrollo de estrategias más equilibradas. Dichas estrategias apuntarían a contener o mitigar la propagación mientras se minimiza el daño colateral a otros aspectos de la vida de las personas.

La enfermedad subyacente que afecta todos estos problemas planteados que, de nuevo, también invade otros campos, es la insaciable demanda de datos. Como se señaló anteriormente, el acceso a datos de alta calidad y gran detalle es esencial para desarrollar modelos que puedan pronosticar con precisión el desarrollo y progresión de una epidemia en entornos de la vida real o informar proyecciones y escenarios hipotéticos donde, por ejemplo, se desplieguen restricciones de movilidad o campañas

de vacunación.

En España, al igual que en muchos otros lugares del mundo, investigadores y modeladores a menudo se encuentran con obstáculos sustanciales para acceder a los datos necesarios. Estos desafíos abarcan más que la mera disponibilidad de datos; incluyen inquietudes sobre la calidad de los datos, la necesidad de digitalización, discrepancias en la estandarización a través de las regiones y la granularidad de los datos. Lamentablemente, se ha vuelto demasiado común que los investigadores emprendan búsquedas arduas por estos valiosos datos para alimentar sus modelos. Agravando este asunto está la dependencia de datos en manos de corporaciones privadas, las cuales a menudo (aunque comprensiblemente) limitan el acceso a la comunidad investigadora y al público. Datos relativos a la movilidad y el comportamiento humano, por ejemplo, tienden a estar controlados por entidades privadas, lo que dificulta los esfuerzos de investigación. Además, los datos necesarios para parametrizar adecuadamente los modelos pueden no existir, usualmente debido a la falta de estudios de campo y encuestas exhaustivas—esto es particularmente cierto para la comprensión de los patrones de mezcla de la población.

Es crucial reconocer que estas deficiencias de datos no son solo el resultado de la reticencia de las partes involucradas; también provienen de un problema cultural más amplio: la falta de reconocimiento del papel esencial que la ciencia y la investigación desempeñan en abordar y mitigar los desafíos más significativos que enfrenta la sociedad. Aunque los investigadores por sí solos no pueden resolver este problema, llevamos la responsabilidad de enfatizar la importancia del acceso a los datos tanto a las autoridades como a las empresas privadas. Se necesitan esfuerzos colaborativos para abogar por una mejora en la gestión de datos que beneficie el bien común y nos equipe para enfrentar los desafíos sociales. Un mejor acceso a la información no es simplemente una conveniencia; es un requisito vital para tomar decisiones informadas, implementar medidas de salud pública efectivas y, en última instancia, contribuir a nuestro bienestar.