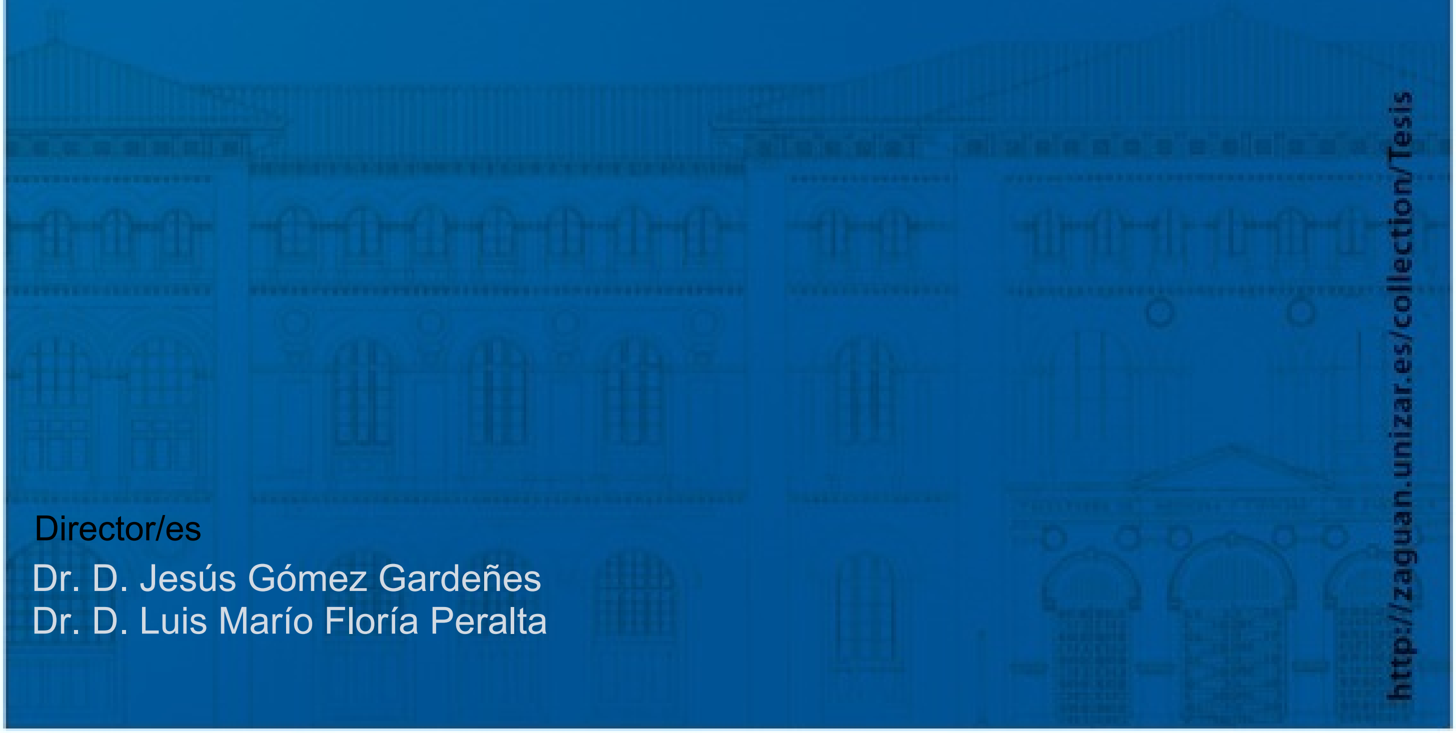
TESIS DE LA UNIVERSIDAD
DE ZARAGOZA
2021
351

David Soriano Paños

Physics of interdependent dynamical processes.



Director/es
Dr. D. Jesús Gómez Gardeñes
Dr. D. Luis Marío Floría Peralta

ISSN 2254-7606

Prensas de la Universidad
Universidad Zaragoza

Servicio de Publicaciones

ISSN 2254-7606

Tesis Doctoral

PHYSICS OF INTERDEPENDENT DYNAMICAL PROCESSES.

Autor

David Soriano Paños

Director/es

Dr. D. Jesús Gómez Gardeñes
Dr. D. Luis Marío Floría Peralta

UNIVERSIDAD DE ZARAGOZA
Escuela de Doctorado

2021

Physics of interdependent dynamical processes

David Soriano Paños

September 3, 2021

A mi madre y a mi hermana.

Abstract

The observation in real systems of emergent macroscopic properties which are absent at the microscopic scale calls reductionist theories into question. Instead, systemic approaches are needed to account for the nontrivial interaction patterns among the constituents of social, biological or economic systems, which gave rise to the birth of the scientific field called complex systems science. An usual avenue to characterize complex systems has been to connect the structure of connections existing in complex systems with the observed collective behaviors by focusing on single isolated dynamics. Nonetheless, complex systems are not immutable but instead reorganize themselves according to both external and internal inputs. This thesis revisits typical models studying dynamics in complex systems to accommodate this flow of information, either between both microscopic and macroscopic scales or encoded in the interdependencies between simultaneous dynamical processes.

The first part of the thesis tackles the study of intertwined dynamics on static contact networks. By adapting the compartmental models introduced in the XX century to the nature of each dynamics, we address four different problems: the spread of interacting pathogens, whose coexistence might be beneficial or detrimental for their evolution, the control of ongoing epidemic outbreaks with the use of digital contact tracing, the emergence of social movements triggered by small initially coordinated groups and the competition between honesty and corruption in modern societies. We find that the flow of information changes the critical properties of the dynamics and forces us to rethink some of the conclusions drawn on the role of the contact structure when studying each dynamics individually.

The second part of the thesis focuses on the impact of recurrent mobility on epidemic spreading in urban environments. We derive a simple model which enables to readily incorporate the distribution of the population in real cities and their usual commuting patterns without any loss of information. We prove that the effects of containment policies based on mobility reductions are not universal but strongly depend on the interplay between the structural features of cities and the epidemiological parameters of the virus propagating across the population. In particular, we discover and characterize a new phenomenon, the epidemic detriment, by which mobility can be beneficial to contain an ongoing outbreak. We finally explore three different real case studies, showing that our model provides insights onto the mechanisms turning urban centers into

important contagion hotspots in recent outbreaks and that further tailored frameworks built upon our simple model capture the evolution of real diseases such as COVID-19 or Dengue.

Resumen

La emergencia de fenómenos colectivos a escalas macroscópicas no observados en escalas microscópicas cuestiona la validez de las teorías reduccionistas. Para explicar estos fenómenos se necesitan enfoques sistémicos que den cuenta de los patrones de interacción no triviales existentes entre los constituyentes de los sistemas sociales, biológicos o económicos, lo que ha dado lugar al nacimiento de la disciplina conocida como ciencia de los sistemas complejos. Una vía habitual para caracterizar los sistemas complejos ha sido la búsqueda de la conexión entre la estructura de interacciones y el comportamiento colectivo observado en sistemas reales mediante el estudio individual de dinámicas aisladas. No obstante, los sistemas complejos no son inmutables y se encuentran constantemente intercambiando información mediante estímulos internos y externos. Esta tesis se centra en la adaptación de modelos sobre diferentes dinámicas en el campo de los sistemas complejos para caracterizar el impacto de este flujo de información, ya sea entre escalas microscópicas y macroscópicas de un mismo sistema o mediante la existencia de interdependencias entre procesos dinámicos que se propagan de forma simultánea.

La primera parte de la tesis aborda el estudio dinámicas acopladas en redes de contacto estáticas. Adaptando los modelos compartimentales introducidos en el siglo XX a la naturaleza de cada dinámica, caracterizamos cuatro problemas diferentes: la propagación de patógenos que interactúan, cuya coexistencia puede ser beneficiosa o perjudicial para su evolución, el control de brotes epidémicos con el uso del rastreo de contactos digital, la aparición de movimientos sociales desencadenados por pequeñas minorías sociales bien coordinadas y la competencia entre honestidad y la corrupción en las sociedades modernas. En todas estas dinámicas, encontramos que el flujo de información cambia las propiedades críticas del sistema así como algunas de las conclusiones extraídas sobre el papel de la estructura de contactos al estudiar cada dinámica de forma individual.

La segunda parte de la tesis se centra en el impacto de la movilidad recurrente en la propagación de epidemias en entornos urbanos. Derivamos un modelo sencillo que permite incorporar fácilmente la distribución de la población en las ciudades reales y sus patrones habituales de desplazamiento sin ninguna pérdida de información. Demostramos que los efectos de las políticas de contención basadas en la reducción de la movilidad no son universales y dependen en gran medida de las características estructurales de las ciudades y los parámetros epidemiológicos del virus circulante en la población. En particular, descubrimos y caracterizamos un nuevo fenómeno, el detrimento epidémico, que refleja el efecto beneficioso de la movilidad en algunos escenarios para contener un brote epidémico. Por último, exploramos tres casos de estudio reales, mostrando que nuestro modelo permite capturar algunos de los mecanismos que han convertido a los núcleos urbanos en importantes focos de contagio en recientes epidemias y que el modelo desarrollado puede servir como base para desarrollar marcos teóricos más realistas que reproducen la evolución de distintas enfermedades como la COVID-19 o el dengue.

Contents

List of Figures

List of Tables

List of publications

14. S. Hazarie, D. Soriano-Paños, A. Arenas, J. Gómez-Gardeñes and G. Ghoshal. Communications Physics **4** (1), 1-10 (2021).
13. W. Cota, D. Soriano-Paños, A. Arenas, S. Ferreira and J. Gómez-Gardeñes. New Journal of Physics **23**, 073019 (2021).
12. A. Reyna-Lara, D. Soriano-Paños, S. Gómez, C. Granell, J. T. Matamalas, B. Steinegger, A. Arenas and J. Gómez-Gardeñes. Phys. Rev. Research **3**, 013163 (2021).
11. A. Arenas, W. Cota, J. Gómez-Gardeñes, S. Gómez, C. Granell, J. T. Matamalas, D. Soriano-Paños and B. Steinegger. Phys. Rev. X **10**, 041055 (2020).
10. A. Arenas, J. Gómez-Gardeñes, D. Soriano-Paños, C. Granell. Europhysics News **51** (5), 38-40 (2020).
9. F. Bauza, D. Soriano-Paños, J. Gómez-Gardeñes and L. M. Floría. Chaos **30**, 063107 (2020).
8. D. Soriano-Paños, J. H. Arias-Castro, A. Reyna-Lara, H. J. Martínez, S. Meloni and J. Gómez-Gardeñes. Phys. Rev. Research **2**, 013312 (2020).
7. D. Soriano-Paños, G. Ghoshal, A. Arenas and J. Gómez-Gardeñes. Journ. Stat. Mechanics **2**, 024006 (2020).
6. D. Lu, F. Bauza, D. Soriano-Paños, J. Gómez-Gardeñes and L. M. Floría. Phys. Rev. E **101**, 022306 (2019).
5. D. Soriano-Paños, F. Ghanbarnejad, S. Meloni and J. Gómez-Gardeñes. Phys. Rev. E **100**, 062308 (2019).
4. D. Soriano-Paños, Q. Guo, V. Latora and J. Gómez-Gardeñes. Phys. Rev. E **99**, 062311 (2019).
3. D. Soriano-Paños, J. H. Arias-Castro, F. Naranjo-Mayorga and J. Gómez-Gardeñes. Eur. Phys. J. Spec. Top. **227**, 661-672 (2018).
2. D. Soriano-Paños, L. Lotero, A. Arenas and J. Gómez-Gardeñes. Phys. Rev. X **8** , 031039 (2018).

1. J. Gómez-Gardeñes, D. Soriano-Paños and A. Arenas. Nat. Phys. **14**, 391-395 (2018).

Chapter 1

Introduction

> *The ability to reduce everything to simple fundamental laws does not imply the ability to start from those laws and reconstruct the universe.*
>
> P. W. Anderson [1].

Physics is the scientific discipline devoted to the characterization of the laws governing the phenomena observed in nature. Since the publication of the Newton's laws in 1687, different generations of scientists have formulated enlightening and accurate theories which have allowed us to acquire a deep but not totally complete understanding of the principles determining the systems' behavior across different spatial and energy scales, ranging from the physics of the universe to the study of elementary particles. In light of the broad application of physics, a natural question arises: can we derive from first principles universal laws which remain valid to characterize any physical system?

The answer to the previous question has confronted two school of thoughts, *Reductionism* and *Emergentism* and transcends the scope of physics, involving the rest of scientific disciplines. Reductionists argue that these universal laws should exist and, consequently, that the behavior of every system could be explained as a sum of the contributions of each of its components (Constructionist hypothesis). Consequently, an extreme reductionist seeks for a *Theory of everything* and defends the existence of a natural hierarchy by which particle physics should be able to describe not only the physics of the universe but eventually also chemistry, biology etc. In contrast, emergentism breaks this hierarchical structure and defends that different *emergent* properties appear as a function of the level of description at which the system is studied. The existence of emergent properties inherent to the complexity of the system renders an exclusively reductionist approach as incomplete in order to derive and understand the different laws governing the evolution of processes across different scales.

The controversy between the two school of thoughts has persisted over centuries and covers different areas of knowledge [2–4]. From a physicist's perspective, the spontaneous symmetry breaking observed in many-body physical systems questions the validity of the constructionist hypothesis. This mechanism implies that some symmetries of the equations governing the interactions among the constituents of a system are not present when studying the system as a whole. For instance, the translational symmetry of the Hamiltonian governing the interaction among atoms is broken when the latter are arranged conforming crystals or the rotational symmetry of the Hamiltonian governing spin alignment is lost in the ferromagnetic transition. Beyond these examples, gauge symmetry breaking has proven to be a fundamental mechanism involved in the phase transitions giving rise to diverse relevant phenomena observed macroscopically such as superconductivity or superfluidity.

Far from being something derived from the symmetry breaking in materials, the emergence of unexpected collective phenomena also occurs in living systems. Indeed, routinary aspects such as motion, cognition or ultimately life cannot be generated by a single cell but instead arise from the nontrivial combination of the different hierarchical levels existing inside organisms. Climbing into the natural biological hierarchy, collections of entities also display complex emergent phenomena such as the formation of fish banks as a defense mechanism from possible predators or the flocking behavior giving rise to mesmerizing coherent structures of starling's swarms. From a social perspective, language shifts, the irruption of coordinated social movements or cultural evolution are a few examples of the dynamical richness resulting from human interactions.

The ubiquity of emergent phenomena across different levels of description in nature calls for a rethinking of reductionist theories. Indeed, the aforementioned examples constitute irreducible problems where the collective behavior cannot be understood by summing the contributions of the individual components of the system. Therefore, characterizing these systems require systemic approaches modeling the relevance of the interactions between those components as entities are aggregated across different scales. From the seminal work published by Anderson in 1972 [1], the seek for theories connecting the local interactions with the macroscopic behavior of real systems has given rise to the so-called complex systems science.

Complex systems

The ambiguity of the word *complexity* and the great diversity of systems with emergent properties makes delimiting an accurate definition of a complex system difficult. Nonetheless, there is an extended consensus on some of the properties which are always present in complex systems:

- **Collective behavior:** As anticipated when describing the examples above, complex systems display emergent collective phenomena which cannot be characterized when studying individually their components. Interestingly, complex systems usually exhibit long-range coherent behavior emerging from local interactions, even in spite of the absence of a central driver in the system.

- **Nonlinearity:** The nonlinearity of the equations governing the evolution of complex systems is a natural consequence of their irreducibility and is also needed to account for the complex response of real systems to external perturbations.

- **Adaptability:** Along this line, complex systems are not immutable but instead evolve over time. In this sense, complex systems have the ability of restructuring themselves to increase their resilience and improve their adaptation to changing environmental conditions.

The holistic approach through the lens of complex systems has been really useful to shed light on economic, social, biological and political dynamics. Notwithstanding, one of the most important challenges still faced by complex systems science is that of determining the features driving the spontaneous emergence of complex collective phenomena from the assembling of a set of locally interacting components in these different disciplines. The parallelism between such goal and the connection between the microscopic and macroscopic scales in physical systems has turn statistical physics into an indispensable tool to tackle the study of the roots of complexity.

One of the most important advances coming from statistical physics in this endeavor corresponds to the discovery of the self-organized criticality. The existence of long-range correlations arising from local interactions in materials constitutes the fingerprint of the divergence of the correlation length occurring in phase transitions. Therefore, the macroscopic collective behaviors locates complex systems in the surroundings of critical points where the order parameters behave according to power-law functions and, accordingly, spatial scale invariance should be observed. Self-organized criticality entails that this critical behavior arises naturally from the internal dynamics of the system, i.e. critical points are the equilibria of the underlying dynamics. The Bak-Tang-Weisenfeld model [5] proposed in 1987 to mimic the dynamics of sand piles was the first example of dynamics displaying self-organized criticality. In this model, simple local interaction rules give rise to power-law distributions for the size of avalanches triggered by the sequential addition of grains. Following this seminal work, other models have been proposed inspired by different dynamics such as forest fires [6] or earthquakes [7]. Despite the relevance of the latter results, there are still important concerns on whether self-organized criticality can represent the main mechanism driving the emergence of complexity [8].

In the previous examples, the system is considered to be mostly homogeneous and the scale invariance shown by the different order parameters emerges as a direct consequence of the imposed dynamical rules. Nonetheless, the structure of connections in most of the biological and social systems are far from being homogeneous and usually correspond to fat-tailed distributions where highly heterogeneous nodes coexist. Remarkably, there has been a growing debate on whether these fat-tailed distributions corresponds to power-law functions [9–11]. This controversy was ignited by the seminal work published by Barabasi and Albert in 1999 [12], who proved that a power-law distribution for the number of connections of the nodes in a network emerges naturally from including a preferential attachment mechanism governing its growth. This result has profound implications, for it entails that scale invariance can be conceived as something inherent to the structure of the interactions rather than to the nature of the dynamics analyzed.

The characterization of the complexity inherent to the networks of interactions constituted a shift of the paradigm which encouraged complex systems scientists to study the connection between structure and function in real systems, giving rise to the so-called network science. On the one hand, from an empirical point of view, the study of the microscopic and mesoscopic structural features of real networks has provided insights into different social and biological problems ranging from the polarization of societies [13] or the roots of success in different disciplines [14, 15] to the structural causes of brain disorders [16]. On the other hand, from a theoretical point of view, many efforts have been devoted to determine how the heterogeneous nature of networks alters the phase diagrams or the critical properties of the mean-field theories capturing diverse phenomena such as epidemics, consensus or synchronization.

Our contribution to complex systems science

Despite the advances made in complex systems' modeling, the vast majority of the latter models cannot capture their natural adaptability. This limitation comes from the fact that they usually study each dynamics individually, thus neglecting the impact of possible interdependencies between different processes occurring simultaneously. Moreover, even when addressing single dynamics, models usually consider a bottom-up scheme exploring the casual relation between the structure of the underlying network and the observed behavior, but do not include the possible feedback loop shaping the microscopic rules as a function of the macroscopic dynamical state of the system. The main goal of this thesis is to overcome the limitations of former models and explore the physics and new phenomena arising from accounting for the environmental adaptability of complex systems.

The document is divided into two well-separated parts. In the first part of the thesis, we will study the effects of accounting for the time-varying nature of the rates governing the evolution of dynamical processes on contact networks. These rates vary according to the information obtained by the system under study, either from the advance of other simultaneous processes or from its own dynamical state. According to how the information flow is defined, we have *cooperative* processes in which both dynamics are accelerated and benefit from each other or *competitive* processes, characterized by a negative feedback from the system to the dynamics. Specifically, we will explore four different problems: the spread of interacting pathogens, which can be either cooperative or competitive according to the nature of the diseases caused by them, the positive feedback between information spread and consensus, the control of the spread of a virus with contact tracing policies and the competition between norm violating behaviors and punishment in social dynamics. In all these cases, we will prove that interactions among coexisting dynamics shape the nature of the phase transitions undergone by the system and alter some of the important conclusions regarding the role of network which were previously drawn by studying each dynamics individually.

The second part of the thesis instead focuses on studying how the underlying structure of connections governing the advance of one dynamical process is shaped by a simultaneous process. Within the possible topics, given its relevance to understand the fast evolution of recent outbreaks, we will characterize the interplay between human mobility and disease spreading. In this sense, from a microscopic point of view, agents constantly change their acquaintances driven by their movements, thus making the representation of populations as static contact networks unrealistic. In this part of the thesis, we will revisit the alternatives to contact networks proposed in the literature for such task and propose a new dynamical model which allows us to address the impact of commuting patterns on the spread of diseases. At a theoretical level, we will show that the effect of mobility is not universal but depends on the structural features of the population and on the characteristic of the pathogen as well. In more applied terms, we will prove that the model allows getting insights into the outcome of outbreaks in real cities. Specifically, we will derive the features making specific areas vulnerable to the spread of diseases and show how models based on complex systems can be leveraged to anticipate the advance of epidemics, forecast its evolution and design policies to keep them under control.

Part I

Intertwined dynamics on contact networks

Presentation of Part I

The advent of network science from the beginning of the XXI century boosted by the increasing availability of real and complete datasets has fueled the advance in our knowledge of complex systems. In particular, the incorporation of complex networks to theoretical frameworks addressing different spreading processes has allowed for disentangling the main mechanisms governing their evolution. One paradigmatic example of network science success in such endeavor was the correct characterization of the spread of computer viruses, which remained elusive for the mean-field theories originally proposed to deal with viruses diffusion across virtual spaces. Specifically, the frameworks incorporating mean-field assumptions were not able to explain how the low prevalence of the viruses, associated with critical or near-critical scenarios, and the large lifetimes of the endemic states, typical from supercritical regimes, were observed simultaneously. This conflict was solved by Pastor Satorras et al. in their seminal work [17] where they stated that this dichotomy is caused by the heterogeneous nature of the underlying structure, the Internet network, which makes the existence of sustained endemic states affecting the most vulnerable (connected) nodes compatible with a negligible global impact of the virus [18–20].

The previous finding still constitutes one of the most important milestones in network science and paved the way to the application of networks to address different problems from a wide variety of disciplines ranging from natural sciences such as biology [21–23], ecology [24–26] or epidemiology [27–29] to human sciences such as sociology [30–32] or economy [33, 34]. Typically, the most common way of incorporating complex networks into these disciplines has been to focus specifically on one spreading process and analyze the information that the structural analysis of the underlying complex network can provide about its evolution. In this sense, the interplay between dynamics and structure in individual processes has shedded light onto very diverse problems such that the estimation of arrival times of imported cases into a given area from epidemic hotspots [35–37], the emergence of polarized and echo-chamber states in social media [13, 38, 39] or the functionality of proteins assemblies [40–42].

Despite the aforementioned successes, neglecting the possible interdependencies existing among simultaneous spreading processes prevents these individual frameworks from capturing many other observed phenomena. For instance, the application of the Susceptible-Infected-Susceptible (SIS) model to the spread of Pneunomia cannot reproduce the sudden surge in deaths caused by this disease around 1918 [43] or the seasonal patterns of Influenza viruses

[44]. Along the same line, the use of rumor models for the diffusion of ideas does not explain the important irruption of recent social movements starting from a small set of initial spreaders in social platforms such as the Arab Spring [45], the 15-M movement [46] or, more recently, the Black Lives Matter movement [47].

The latter examples highlight the need for more general approaches which, apart from accounting properly for the inherent nature of each spreading process, incorporate how the possible mutual feedback existing among them can shape the rules governing their evolution. In this part of the thesis we tackle this challenge and characterize the physics arising from the coexistence of different spreading processes over the same population. Our focus is therefore to disentangle the coupling mechanisms between these processes giving rise to those phenomena which are unexpected when studying both dynamics individually.

In Chapter 2, we will derive a theoretical framework to analyze the spread of interacting diseases across a networked population. This interaction either entails a cooperation between both diseases, meaning that one pathogen makes hosts more susceptible to contract the other one, or a competition, which usually takes place in multi-strain diseases, for which contracting one strain confers partial or total immunity with respect to the others. We will check that the coexistence of both diseases can drastically alter the phase diagram of the usual SIS model giving rise to new phases emerging from the interaction among them.

Furthermore, in this chapter we will also address a very timely topic, the effectiveness of control policies based on digital proximity tracing apps to keep an ongoing outbreak under control. In the language of interacting processes, this problem can be seen as a competition between two spreading processes, the contagions driven by human-human interactions in the physical space and a cascade of detections triggered in the virtual network constructed by the application's users. We will derive a theoretical framework to incorporate both processes and use real networks constructed from face-to-face interactions recorded via proximity sensors to test the framework in realistic setups. Interestingly, we will see that the competition of both processes has strong implications for the topology of the effective network of contacts across which contagions occur. This fact leads to an outcome of control policies based on contact tracing totally different from others aimed at reducing the transmissibility of infectious individuals such as the increase of the testing capacity.

Chapter 3 is devoted to the study of social dynamics. In particular, we will modify a SIS model which captures the activity of users in social platforms to incorporate the higher predisposition to share one's opinion when this is widely accepted within your close contacts. In more technical words, this assumption corresponds to a positive feedback among consensus dynamics and the diffusion of ideas. After formulating the framework, we will determine under which conditions the cooperation among both dynamics can explain the social movements mentioned above. As a byproduct, we will also prove how the information horizon, controlled by the topology of the social network, crucially shapes the diffusion of ideas when consensus is required.

Finally, to round off this chapter on social dynamics, we will also study the diffusion of corruptive behaviors over a given population. To do so, we will propose a theoretical framework adapting the compartmental models used in epidemics to accommodate different features associated with the corruption. In particular, characterizing the extent of corruption involves the study of two competing dynamics, the diffusion of corruptive behaviors driven by temptation induced by the benefits of such conduct and the delation of the corrupts by the honest population to restore honesty in the society. The outcome of such competition is governed by the coupling among both mechanisms along with the underlying network structure, which produce a very rich phase diagram.

Chapter 2

Disease dynamics

> *It is somewhat surprising that so little mathematical work should have been done on the subject of epidemics, and, indeed, on the distribution of diseases in general. Not only is the theme of immediate importance to humanity, but it is one which is fundamentally connected with numbers, while vast masses of statistics have long been awaiting proper examination*
>
> R. Ross and H. Hudson [48].

The complexity of the spatio-temporal patterns observed during different epidemic waves across the history and the threats posed by contagious diseases on the society have encouraged mathematicians and physicists to work actively on epidemic modeling since the XVIII century. These models have been continuously gaining relevance until becoming indispensable tools to understand the mechanisms responsible for epidemic spreading, provide short-term forecasts about ongoing epidemic outbreaks and design control policies to keep them under control. Despite the recent advances on epidemic modeling, most of these sophisticated frameworks are tailored for one specific disease and neglect the existence of other circulating pathogens which, as motivated in the introduction to this part, may be crucial to shape their long-term evolution.

In this chapter, we propose a theoretical framework to accommodate different coexisting diseases which simultaneously spread across a population and evaluate how the interaction among them can substantially alter their propagation. In Section 2.1 we first lay the foundations for such framework by explaining the basic concepts involved in epidemic modeling and revisiting how the epidemic models have been evolved from their beginnings to the incorporation of complex networks. In Section 2.2 we propose the equations conforming our systemic approach to the spread of simultaneous diseases and study the impact of the existence of positive (cooperation) or negative (competition) interactions among them. Finally, in Section 2.3 we propose a model to accommodate contact tracing policies which can be conceived as a practical example of the competition between spreading phenomena propagating over both virtual and physical spaces.

2.1 Compartmental models: From mean-field theories to complex networks

In 1766, Daniel Bernoulli published a seminal work [49] containing a mathematical analysis to quantify the importance of the *variolation*, i.e., the inoculation of virus extracted from patients with mild smallpox in healthy individuals, for increasing the life expectancy of the population. This work supposed the birth of a new discipline, the mathematical epidemiology, aimed at using mathematics to guide decision-making to reduce the impact of epidemics. Following Bernoulli's seminal work, other important models were published by Willian Hammer [50] or Ronald Ross [51] at the beginning of the XX century. Both authors were the pioneers in stating that the transmission of a virus should be mediated by the close interaction among infected and susceptible hosts, as dictated by the law of mass action typically applied to chemical reactions, which is nowadays a basic pillar to construct epidemiological models.

In particular, Ross was the first to introduce the concept of compartmental models to represent the evolution of malaria. Compartmental models assume that the population is divided into a finite set of compartments which represent the possible states across which an individual can transit during the course of a disease. These transitions are governed by a system of differential equations which are constructed according to the microscopic processes driving the spread of diseases. Despite the eventual relevance of the model proposed by Ross, especially for the modeling of vector-borne diseases [52], compartmental models were mostly unnoticed until 1927, when A. G. McKendrick and W. O. Kermack formulated the standard Susceptible-Infected-Recovered (SIR) model [53–55].

2.1.1 SIR model

The SIR model divides a set of individuals into three different compartments which are Susceptible, corresponding to healthy individuals who have not contracted the disease yet, Infected, containing those individuals who have been already infected and are infectious and Recovered/Removed which gather those infectious individuals who have either overcome the disease or eventually died. In this sense, the SIR model assumes that the disease confers immunity to its hosts so recovered individuals cannot be reinfected again. Regarding the transition rules, each individual is assumed to make k contacts per time unit. If those contacts involve susceptible and infectious individuals, the disease is transmitted with a probability λ. Finally, infected individuals enters into the R compartment at a rate μ related to the typical infectious period.

In the original SIR model, the population is assumed to be statistically equivalent, so the state of a system of N agents is completely described by the differential equations governing the fraction of the population belonging to each of the compartments, denoted by the corresponding lowercase letters. In particular,

$$\dot{s} = -\lambda k s i \ , \tag{2.1}$$

$$\dot{i} = \lambda k s i - \mu i \ , \tag{2.2}$$

$$\dot{r} = \mu i \ . \tag{2.3}$$

From Eq. (2.3), it becomes clear that $\dot{r} = 0 \Rightarrow i = 0$; therefore, the steady state of a SIR dynamics is always characterized by the extinction of the disease. The epidemic trajectory to arrive to such steady configuration is crucially determined by the trade-off between the contagion and recovery rates. This trade-off is usually captured in a pivotal indicator in mathematical epidemiology, the basic reproduction number $\mathcal{R}_0$ [56, 57], which is defined as the expected number of contagions made by an index case in a fully susceptible population. In the SIR model,

$$\mathcal{R}_0 = \frac{\lambda k}{\mu} \ . \tag{2.4}$$

From Eq. (2.2), it can be seen that when $\mathcal{R}_0 > 1$, introducing an index case in a fully susceptible population triggers a cascade of contagions translated into an exponential growth in the number of infected individuals. In contrast, if $\mathcal{R}_0 < 1$, the infection force is not strong enough to get a sustained epidemic state and the outbreak dies out without reaching a macroscopic fraction of the population. In a SIR dynamics, the extent of an outbreak is characterized by the attack rate $r(\infty)$, which reflects the total number of individuals who have become infected over the course of the disease. This quantity can be derived from Eq. (2.3) and Eq. (2.1) yielding the following self-consistent equation:

$$r(\infty) = 1 - (1 - i_0) e^{-\mathcal{R}_0 r(\infty)} \ , \tag{2.5}$$

where i_0 is the fraction of the population initially infected acting as a seed of the outbreak. To get a more illustrative picture on the relevance of the basic reproduction number, we represent in Fig. 2.3 the dependence of the attack rate on $\mathcal{R}_0$ as obtained by solving the latter equation along with the time evolution of an outbreak in both subcritical ($\mathcal{R}_0 < 1$) and supercritical scenarios ($\mathcal{R}_0 < 1$).

2.1.2 SIS model

The Susceptible-Infected-Susceptible (SIS) model was proposed as a variant of the SIR model for those diseases not conferring long-lasting immunity to their hosts such as sexually-transmitted diseases or Influenza. In this scenario, the infected individuals are susceptible again of contracting the pathogen after losing their infectiousness and, as a consequence, enter into the S compartment. Following similar rules to the ones used in the SIR model, the equations governing the evolution of a SIS disease are:

$$\dot{s} = -\lambda k s i + \mu i \ , \tag{2.6}$$

$$\dot{i} = \lambda k s i - \mu i \ . \tag{2.7}$$

Given that the population is closed, it is useful to write $s = 1 - i$. The study of the fixed points of the dynamics gives us two possible steady configurations:

$$(s^*, i^*)_1 = (1, 0) \ ,$$
$$(s^*, i^*)_2 = (1/\mathcal{R}_0, 1 - 1/\mathcal{R}_0) \ ,$$

whose stability changes via a transcritical bifurcation taking place at $\mathcal{R}_0 = 1$. In particular, when $\mathcal{R}_0 < 1$, the epidemic eventually dies out, for the disease-free state is stable and indeed constitutes the only physically meaningful fixed point. In contrast, above the bifurcation, the disease evolves towards an endemic state characterized by a sustained fraction of infected individuals. Therefore, the impact of a SIS disease is totally characterized by the fraction of infected individuals in the steady state, $i(\infty)$. For the sake of visualization, in Fig. 2.1b, we illustrate the evolution of the latter quantity as a function of $\mathcal{R}_0$ along with examples of epidemic trajectories in both regimes.

Finally, from Eq. (2.4), we can obtain an expression for the epidemic threshold which is defined as the minimum infectivity per contact required to observe a macroscopic outbreak and denoted in the following by λ_c. In terms of the basic reproduction number, the epidemic threshold fulfills that $\mathcal{R}_0(\lambda_c) = 1$. This quantity will appear recurrently throughout this thesis to quantify the interplay between structure and dynamics in the different frameworks here studied. In particular, in the case of the SIS model under the mean-field approach, the epidemic threshold is defined as

$$\lambda_c = \frac{\mu}{k} \ . \tag{2.8}$$

In what follows, we focus on SIS dynamics because of the higher analytical tractability of their equations and the existence of a more pronounced phase transition separating the disease-free and the endemic regimes.

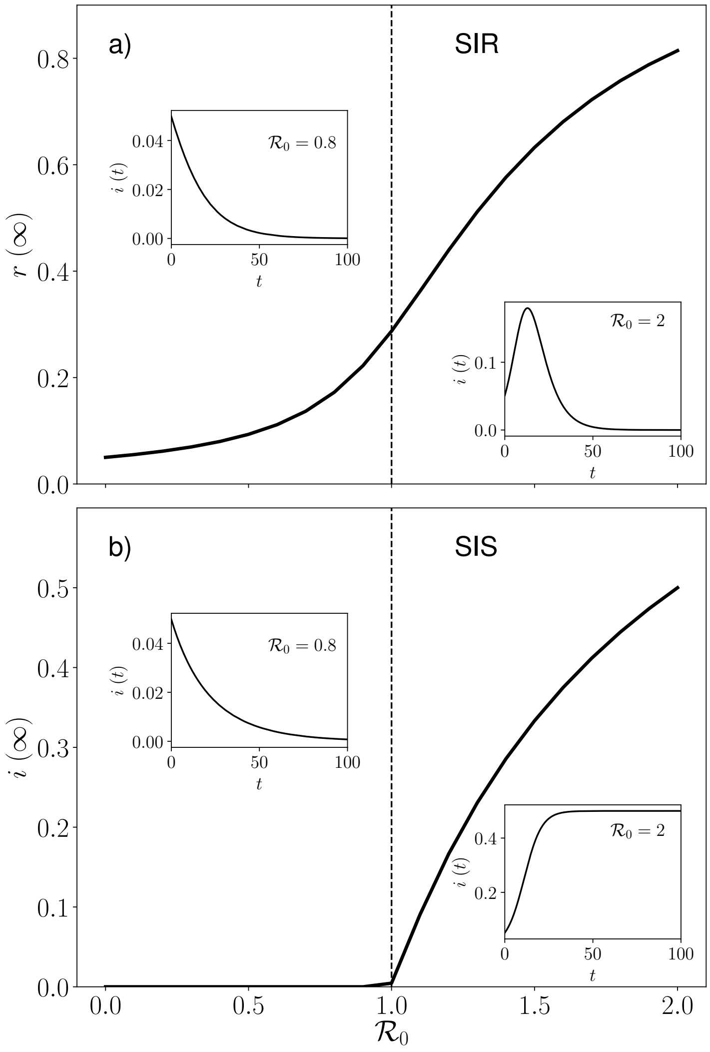


Figure 2.1: Evolution of the impact of an outbreak as a function of the control parameter $\mathcal{R}_0$ for both SIR and SIS dynamics. For the SIR dynamics (Panel a), this impact is quantified by the attack rate $r(\infty)$ whereas, in the SIS model (Panel b), the extent of a disease corresponds to the fraction of population in the infected compartment in the steady state $i(\infty)$. In both panels, the insets represent the time evolution of the fraction of infected population in a subcritical scenario $\mathcal{R}_0 < 1$ and a supercritical scenario $\mathcal{R}_0 > 1$. In both panels, we fix $k = 10$, $\mu = 0.2$, $i_0 = 0.05$ and vary the probability of contagion per contact λ.

2.1.3 Heterogeneous mean field (HMF)

The mean field theories originally proposed for the SIS and SIR models provide important qualitative insights into the nature of epidemics transmitted through direct interactions among healthy and infectious peers, pinpointing the existence of a second-order transition triggered by an increase in the infectiousness of the pathogen. Nonetheless, neglecting the population diversity in terms of their social connectivity, biological features or interaction patterns makes mean field theories fail in giving accurate quantitative predictions on the position of the epidemic threshold or the evolution of the epidemic trajectories. The latest example of this failure resides in the COVID-19 pandemic, whose course has been crucially shaped by the existence of super-spreaders [58–60]. This fact by no means can be explained with the homogeneous assumptions included in mean-field theories.

The first important attempt to account for the heterogeneity of human interactions was the heterogeneous mean field (HMF) theory proposed by Pastor Satorras and Vespignani in 2001 [17]. In this model, the authors abandon the statistical equivalence of the entire population and include heterogeneous populations where the number of interactions of each individual, referred to as the degree or connectivity k, follows a probability distribution $P(k)$. In addition, HMF theory gets its name because all the individuals within the same degree class are equivalent. Under these assumptions, the evolution of the fraction of infected agents within the degree class k, ρ_k, reads as

$$\dot{\rho}_k = -\mu\rho_k + (1-\rho_k)k\sum_{k'} P(k|k')\rho_{k'} \ , \tag{2.9}$$

where $P(k|k')$ denotes the probability that one of the k connections of the agents within the corresponding degree class points to an individual with degree k'. Assuming an absence of degree-degree correlations, so that $P(k'|k) = k'P(k')/\langle k\rangle$, the linearization of the equations yields the following analytical expression for the epidemic threshold:

$$\lambda_c = \mu\frac{\langle k\rangle}{\langle k^2\rangle} \ , \tag{2.10}$$

where

$$\langle k^m\rangle = \int_0^\infty P(k)k^m dk \ . \tag{2.11}$$

Eq. (2.10) tells us that the value of the epidemic threshold is governed by the trade-off existing between the first and the second moments of the degree distribution. In particular, note that the mean field prediction is retrieved by setting $P(k) = \delta_{k\langle k\rangle}$.

2.1.4 Models for synthetic networks

In absence of real data, different methods to construct synthetic networks have been proposed in the literature. Within the very diverse realm of synthetic networks, throughout this thesis we mostly focus on two models because of their historical relevance and the sharp differences existing among them. These models are the Erdös-Rényi (ER) and the Barábasi-Albert (BA) networks.

Paul Erdös and Alfréd Rényi proposed in 1960 a model [61] to generate random networks by assuming that each possible pair of nodes in a network are connected with a given probability p. As a result, the configuration obtained is pretty homogeneous and the probability that a node has k connections follows a binomial distribution so that

$$P(k) = \binom{N-1}{k} p^k (1-p)^{1-k} \ , \tag{2.12}$$

with $\langle k \rangle = p(N-1)$ and $\langle k^2 \rangle = \langle k \rangle (1-p) + \langle k \rangle^2$. Plugging both values into Eq.(2.10), we obtain that

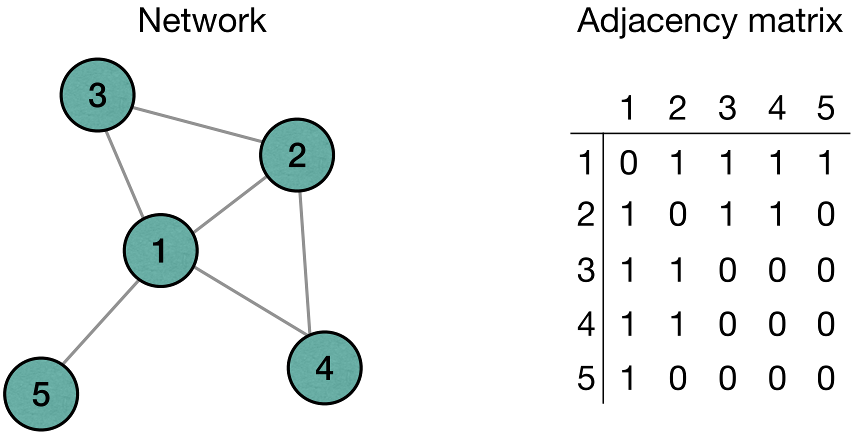
$$\lambda_c^{ER} = \frac{\mu}{1 - \langle k \rangle / N + \langle k \rangle} \ , \tag{2.13}$$

which is slightly lower than the one predicted by the MF equations due to the existence of practically negligible heterogeneities arising from the randomness of the connections.

Nonetheless, when analyzing the degree distributions of real networks [11], one realizes that most of them are highly heterogeneous networks with fat-tailed degree distributions characterized by the existence of a few set of individuals, usually denoted as *hubs*, whose connections are orders of magnitude larger than the average connectivity in the network. To capture this fact, Laszlo Barábasi and Reka Albert published a seminal work in 1999 [12] in which they state that the formation of a link in a network is not a random event but instead is driven by a preferential attachment mechanism. This preferential attachment consists in assuming that the probability for a new node setting a link with a node already existing in the network is proportional to the connectivity of the latter. As a result, the authors obtained the following degree distribution:

$$P(k) \sim k^{-3} \ , \tag{2.14}$$

for which $\langle k^2 \rangle \gg \langle k \rangle^2$. As a consequence, $\lambda_c^{BA} << \lambda_c^{MF}$, which clearly suggests that contact heterogeneity is an important driver for spreading phenomena. In particular, subsequent works have proved that, despite having low infectiousness, a given pathogen or computer virus is able to persist endemically in localized states composed of the most connected nodes and their neighborhood in this kind of networks [18–20, 62].



$$\begin{array}{c|ccccc} & 1 & 2 & 3 & 4 & 5 \\ \hline 1 & 0 & 1 & 1 & 1 & 1 \\ 2 & 1 & 0 & 1 & 1 & 0 \\ 3 & 1 & 1 & 0 & 0 & 0 \\ 4 & 1 & 1 & 0 & 0 & 0 \\ 5 & 1 & 0 & 0 & 0 & 0 \end{array}$$

Figure 2.2: Example of a complex network with $N = 5$ nodes and its associated adjacency matrix.

2.1.5 Microscopic Markov Chain Approach (MMCA)

The HMF equations reveal the relevance of contact heterogeneity for disease spreading; however, they cannot account for the impact of the specific complex structure of connections governing the interactions among the elements of a system and determining the evolution of spreading phenomena running on top of them. These complex patterns are usually embedded in the so-called complex networks, which represent the core of network science [63–65]. Mathematically, a complex network $\mathcal{G}$ is defined as a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where $\mathcal{V} = \{1, \ldots, N\}$ represents the vertices of the network and $\mathcal{E} = \{e_1, \ldots, e_L\}$ the set of links connecting them. Within the different classes of networks, here we focus on unweighted undirected networks, in which all the interactions are reciprocal and their strength is equal regardless of the agents involved. An useful object to represent such networks is the adjacency matrix $\mathbf{A}$, whose entries $A_{ij} = 1$ if two agents i and j interact and $A_{ij} = 0$ otherwise. Fig. 2.2 depicts a toy complex network with $N = 5$ along with its adjacency matrix.

Multiple theoretical frameworks have been proposed over the last years to accommodate complex networks in epidemic modeling [66–72]. All these options differ from one another in the amount of information regarding the structure of the complex network introduced in the equations. Theoretically, the more information is included, the better the formalism reproduces the results obtained from explicit agent-based simulations based on Monte Carlo routines. Nonetheless, increasing the complexity of the formalism can be counterproductive, for it is usually made at the expense of enlarging the set of equations of the framework. This clearly increases the computational cost of their numerical resolution and can also preclude getting any analytical insights into the specific role of the network in the process. Among all the available options, throughout this thesis we will adopt the Microscopic Markov Chain Approach (MMCA) [73] which optimizes the compromise existing between having a framework as simple as possible while reproducing fairly enough the results from the simulations.

Generally speaking, the MMCA is a time-discrete formalism which assumes that the evolution of a disease is well-characterized by the probability that each individual i is in each of the m available compartments at each time step t, hereinafter denoted by $\rho_i^m(t)$. Therefore, the number of equations in the formalism scales as $N(\Omega_C - 1)$ where Ω_C represents the number of compartments of the chosen compartmental model. In particular, for the SIS model, the evolution of the former probabilities reads as

$$\rho_i^I(t+1) = (1-\mu)\rho_i^I(t) + (1-\rho_i^I(t))\Pi_i(t) \ . \tag{2.15}$$

Eq. (2.15) constitutes the fundamental basis on which the rest of formalisms developed in this thesis are constructed. It basically tells us that the probability for an agent i being infected at $t+1$ can be expressed as the sum of the probability for i being already infected and not recovering and that of being susceptible at t and contracting the disease $\Pi_i(t)$. In this sense, assuming that an agent interacts with all his/her neighbors in each time step, $\Pi_i(t)$ is given by

$$\Pi_i(t) = 1 - \prod_{j=1}^{N} \left(1 - \lambda A_{ij} p_j(t)\right) \ , \tag{2.16}$$

which corresponds to the complementary probability of not contracting the disease in any of these contacts. The numerical iteration of the system of $N(\Omega_C - 1)$ coupled discrete equations informs about the spatio-temporal evolution of SIS diseases on complex networks.

At the microscopic scale, the MMCA equations constitutes a probabilistic description of the evolution of the state of each agent in the system under study. Nonetheless, for a single outbreak, these states are well-defined due to the compartmentalization of the epidemic model and live in a discrete space. Therefore, the information provided by the MMCA equations should be understood as the average probability that an individual i is in compartment m at time t computed over an ensemble of stochastic epidemic trajectories generated with the same parameters and initial conditions. To obtain an order parameter informing us about the global epidemic state of the network, we must compute the epidemic size by averaging the probabilities that each individual remains infected in the steady state. Therefore, the epidemic size obtained from the

MMCA equations reads as

$$i(\infty) = \frac{1}{N}\sum_{i}^{N} \rho_i^I(\infty) \ . \tag{2.17}$$

Beyond understanding numerically the interplay between structure and dynamics on the evolution of the epidemic size, the MMCA approach allows for getting an analytical computation of the epidemic threshold λ_c as well. To do so, let us assume that the disease has arrived to a steady configuration for infectiousness values $\lambda \simeq \lambda_c$. Mathematically, this implies that

$$\rho_i^I(t+1) = \rho_i^I(t) = \epsilon_i^I \ll 1 \ , \tag{2.18}$$

where we assume that the epidemic size close to the epidemic threshold is negligible. The latter condition allows us to linearize Eq. (2.15) obtaining

$$\mu\epsilon_i = \lambda \sum_{j}^{N} A_{ij}\epsilon_j \ . \tag{2.19}$$

From the last expression, it becomes clear that computing the epidemic threshold involves solving the spectrum of the adjacency matrix $\mathbf{A}$. In particular, as we are interested in the minimum infectiousness required to observe a macroscopic outbreak, the epidemic threshold is given by

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{A})} \ , \tag{2.20}$$

where $\Lambda_{\max}(\mathbf{A})$ represents the largest eigenvalue of matrix $\mathbf{A}$.

2.2 Interacting diseases

The theoretical formalism described thus far focuses on characterizing the spread of a single virus and the role that the contact network plays in their diffusion across the population. In this sense, the contagion/recovery rates are assumed to be constant and not to depend on any environmental factor which could interact with the pathogen under study. This individualistic approach, though useful to get some insights into the complex nature of epidemic spreading, is not always valid to characterize the evolution of some diseases, whose course is crucially shaped by the presence of other pathogens and possible interactions existing among them.

One empirical example of the crucial role of these interactions was the simultaneous propagation of the Spanish Flu and Pneumonia in 1918 [43], for a sudden increase in the death rate caused by Pneumonia was reported coinciding with the onset of the Spanish Flu. Another typical example of the impact of reinforcement between coeval pathogens is the higher risk of contracting TB for HIV patients, which is estimated to be around between 16-27 times according to the WHO [74]. In general, we will denote by cooperation all the scenarios in which the transmission of several diseases is favored by their coexistence.

In contrast, there are other scenarios in which there is a competition between coexisting diseases, for the presence of one pathogen is detrimental to the propagation of other infections. These examples usually correspond to multi-strain diseases in which the different variants/serotypes of the virus compete. This competition is a consequence of the incompatibility of observing different variants inside the same host and the total or partial immunity that one serotype of a disease confers to the other circulating serotypes. Some paradigmatic examples are the different Dengue [75] or Influenza [76] strains propagating across the population or, much more recently, the complex and variable landscape depicted by the different variants of the SARS-CoV-2 virus [77]. Nonetheless, these detrimental interactions can also occur between pathogens corresponding to different diseases, such as the reported negative interaction between the Zika virus and Dengue [78].

In the light of the ubiquity of interaction processes among coexisting diseases, several works [79–90] have tackled the extension of usual epidemic models to accommodate the mechanisms driving the competition or cooperation among them. Interestingly, different theoretical results have been already obtained on the physics behind the rapid surge of cases in presence of cooperation or the extinction of evolutionary impaired variants in supercritical scenarios ($\mathcal{R}_0 > 1$) as a result of the competition with other diseases. Nonetheless, most of the mechanisms have been found at the expense of relying on mean-field assumptions that neglect the structure of the underlying network of connections. Likewise, the frameworks found in the literature are usually specifically tailored for addressing either strong cooperation or mutually exclusive competition between diseases. Therefore, a general mathematical framework, able to characterize the interaction between two pathogens in between these two extreme scenarios when incorporating the complex nature of human contacts is still missing in the literature. In this section, we tackle this challenge by proposing a new model built upon the MMCA equations for the SIS model, which are generalized to account for the coexistence of two pathogens and their mutual influence.

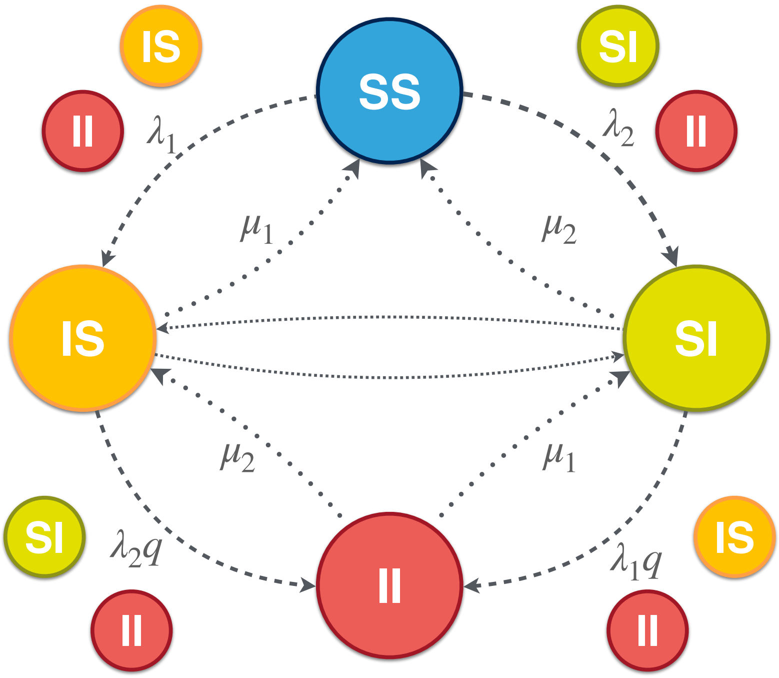


Figure 2.3: Contagion and recovery microscopical processes considered in our model. Note that the contagion processes involving already infected individuals are influenced by the parameter q which encodes the interaction between the two coexisting diseases.

.

2.2.1 MMCA equations for interacting diseases

Let us first assume that our system is composed of N agents whose interactions are governed by a complex network represented by its adjacency matrix $\mathbf{A}$. For the sake of simplicity, we also assume that we deal with two SIS diseases which are individually characterized by their infectiousnesses λ_s and their recovery rates μ_s with $s = 1, 2$. Because of the presence of two pathogens, there are four available states: totally susceptible individuals (SS), agents infected by the first pathogen but susceptible to the second one (IS), its counterpart (SI), and those who have contracted both diseases (II). Regarding the interaction between both pathogens, we assume that having contracted one disease modulates the probability of contracting the other disease by a given factor q. To be able to capture the mutually exclusive competing diseases, we must impose a rule to forbid double simultaneous contagions of totally healthy individuals. In our case, we have decided that, if a susceptible individual receives potential contagions from both pathogens when interacting with her neighborhood, she is equally likely to contract each disease. A schematic representation of the different compartments and the flows among them can be found in Fig. 2.3.

In what follows, we denote the probability that an individual, say i, is in each of the γ available states ($\gamma = \{II, IS, SI, SS\}$) at time t by $[\rho^{\gamma}]_i^t$. Under the microscopic processes described above, the time evolution of these probabilities read as

$$\begin{aligned}[\rho^{II}]_i^{t+1} &= [\rho^{SI}]_i^t(1-\mu_2)[\Pi^{SI}]_i^t + [\rho^{IS}]_i^t(1-\mu_1)[\Pi^{IS}]_i^t \\ &\quad + [\rho^{II}]_i^t\,(1-\mu_1)\,(1-\mu_2)\ ,\end{aligned} \tag{2.21}$$

$$\begin{aligned}\left[\rho^{IS}\right]_i^{t+1} &= [\rho^{SI}]_i^t\mu_2[\Pi^{SI}]_i^t + [\rho^{IS}]_i^t\,(1-\mu_1)\,(1-[\Pi^{IS}]_i^t) + [\rho^{II}]_i^t\mu_2(1-\mu_1) \\ &\quad + \left[\rho^{SS}\right]_i^t[\Pi^{SS}]_i^t[f^{IS}]_i^t\ ,\end{aligned} \tag{2.22}$$

$$\begin{aligned}\left[\rho^{SI}\right]_i^{t+1} &= [\rho^{IS}]_i^t\mu_1[\Pi^{IS}]_i^t + [\rho^{SI}]_i^t\,(1-\mu_2)\,(1-[\Pi^{SI}]_i^t) + [\rho^{II}]_i^t\mu_1(1-\mu_2) \\ &\quad + \left[\rho^{SS}\right]_i^t[\Pi^{SS}]_i^t[f^{SI}]_i^t\ ,\end{aligned} \tag{2.23}$$

$$\left[\rho^{SS}\right]_i^{t+1} = 1 - [\rho^{II}]_i^{t+1} - \left[\rho^{IS}\right]_i^{t+1} - \left[\rho^{SI}\right]_i^{t+1}\ . \tag{2.24}$$

The rationale for each equation is similar to the one used in the original MMCA for individual SIS diseases. In presence of two pathogens, the probability that a totally susceptible agent i contracts the disease, $[\Pi^{SS}]_i^t$, is given by

$$[\Pi^{SS}]_i^t = 1 - \prod_j^N \left[1 - \lambda_1 A_{ij}[\rho^{IS}]_j^t - \lambda_2 A_{ij}[\rho^{SI}]_j^t - (\lambda_1 + \lambda_2 - \lambda_1\lambda_2)A_{ij}[\rho^{II}]_j^t\right]. \tag{2.25}$$

In its turn, f_{IS} and f_{SI} encode the probabilities for a susceptible individual contracting the first and second disease respectively. Based on the aforementioned selection rule, both quantities read as

$$[f^{IS}]_i^t = \frac{[g^{IS}]_i^t(1-[g^{SI}]_i^t) + 0.5[g^{IS}]_i^t[g^{SI}]_i^t}{[g^{SI}]_i^t + [g^{IS}]_i^t - [g^{IS}]_i^t[g^{SI}]_i^t}\ , \tag{2.26}$$

$$\left[f^{IS}\right]_i^t = \frac{[g^{SI}]_i^t(1-[g^{IS}]_i^t) + 0.5[g^{IS}]_i^t[g^{SI}]_i^t}{[g^{SI}]_i^t + [g^{IS}]_i^t - [g^{IS}]_i^t[g^{SI}]_i^t}\ . \tag{2.27}$$

In the former expressions, $[g^{IS}]_i^t$ ($[g^{SI}]_i^t$) represents the probability of having a potential contagion from infected individuals by the first (second) disease. Therefore:

$$[g^{IS}]_i^t = 1 - \prod_j^N \left(1 - \lambda_1 A_{ij}([\rho^{IS}]_j^t + [\rho^{II}]_j^t)\right)\ , \tag{2.28}$$

$$\left[g^{SI}\right]_i^t = 1 - \prod_j^N \left(1 - \lambda_2 A_{ij}([\rho^{SI}]_j^t + [\rho^{II}]_j^t)\right)\ . \tag{2.29}$$

On the other hand, the probabilities of contracting one disease at time t, providing an agent i is already infected by one pathogen, $[\Pi^{IS}]_i^t$ or $[\Pi^{SI}]_i^t$, must incorporate the factor q encoding the interaction between pathongens. In particular,

$$[\Pi^{IS}]_i^t = 1 - \prod_j^N \left[1 - A_{ij}\lambda_2 q([\rho^{SI}]_j^t + [\rho^{II}]_j^t)\right] \ , \tag{2.30}$$

$$\left[\Pi^{SI}\right]_i^t = 1 - \prod_j^N \left[1 - A_{ij}\lambda_1 q([\rho^{IS}]_j^t + [\rho^{II}]_j^t)\right] \ . \tag{2.31}$$

Eqs. (2.21-2.31) allow us to monitor the time evolution of interacting SIS diseases running on complex networks. To validate these equations in both cooperative and competitive scenarios, we compare their predictions with the results from Monte Carlo simulations. Each of these simulations corresponds to a single epidemic trajectory in which the state of each agent is perfectly defined at each time step. For each simulation, we start by infecting a set of individuals, acting as the infectious seed of the disease. Once initialized, we let the system evolve driven by stochastic probabilities defined according to the microscopic contagion and recovery processes assumed in the model until the stationary regime is reached. Note that in the Monte Carlo simulations, this regime does not resemble a frozen configuration in which the state of each agent is immutable. Instead, it corresponds to a dynamic equilibrium in which the state of each agent is allowed to change but its average over a given time window remains constant. To reduce the number of parameters, in what follows, we focus on identical diseases so that $\lambda_1 = \lambda_2 = \lambda$ and $\mu_1 = \mu_2 = \mu$.

2.2.2 Cooperative dynamics

Recall that, under cooperation, being infected by one disease increases the vulnerability to the other one. Mathematically, this is reflected in the formalism by exploring the region of the parameters' space for which $q > 1$. To analyze the impact of cooperation, we first consider that the structure of contacts corresponds to an ER network of $N = 10^3$ agents with average degree $\langle k \rangle = 8$.

To validate the MMCA equations, we focus on the evolution of the total epidemic size ρ as a function of the infectiousness of the pathogens λ and their coupling strength q. According to the equations, as there are three possible infectious states, the epidemic size should be computed as

$$\rho = \frac{1}{N}\sum_{i=1}^N [\rho^{IS}]_i^\infty + [\rho^{SI}]_i^\infty + [\rho^{II}]_i^\infty \ . \tag{2.32}$$

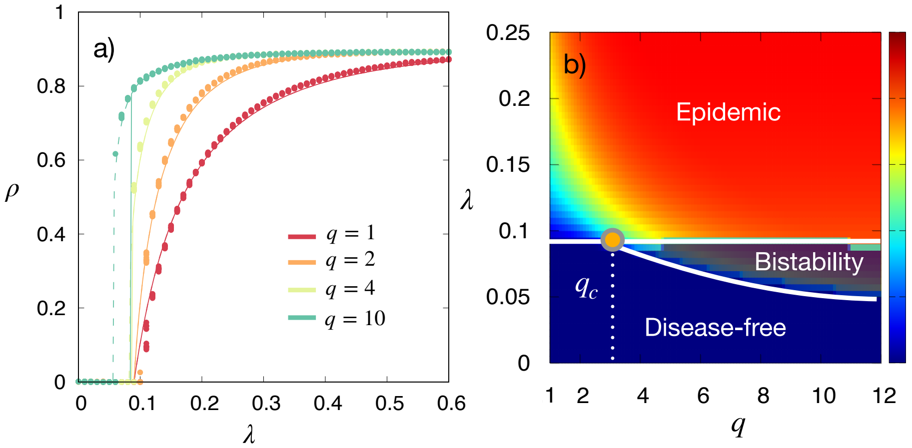


Figure 2.4: Panel (a) shows the epidemic size ρ as a function of the infectivity λ for several values of the interaction parameter q. The curves (solid and dashed) correspond to the results obtained by iterating Eqs.(2.21-2.23) when performing forward (solid) and backward (dashed) continuation schemes (see text for details). In its turn, points represent the results from 50 realizations of Monte Carlo simulations. In panel (b) we plot the phase diagram $\rho(q, \lambda)$ for the cooperative case obtained by the numerical iteration of the equations. The solid curves denote the points where epidemic onsets take place. The critical point at q_c pinpoints the value above which bistable solutions are observed. In both panels we use an ER network of $N = 1000$ nodes with $\langle k \rangle = 8$, while $\mu = 0.75$.

In the case of the simulations, the epidemic size is obtained by averaging the fraction of the population in the infectious compartments over a window of 50 time steps when the system has arrived to an stationary state so that no important fluctuations are observed.

Fig. 2.4a contains the comparison between both approaches, showing that the MMCA equations capture very accurately the steady state of the system. Regarding the physics of the cooperation, we observe that, when the pathogens are weakly cooperating ($q \simeq 1$), the epidemic curve resembles an ordinary SIS model with a smooth second order transition taking place at the epidemic threshold λ_c whose expression is given by Eq. (2.20). Nonetheless, as we increase the coupling strength, this transition becomes more abrupt until a first order transition emerges at a given value q_c and hysteresis cycles appear.

To characterize the hysteresis cycles, we iterate adiabatically the equations. In particular, we follow both forward (backward) continuation schemes in which we set the initial conditions for $\lambda + \delta\lambda$ ($\lambda - \delta\lambda$) as the steady state for λ with a small noise introduced as a perturbation to avoid unstable absorbing states. The existence of bistable solutions is more evident in the phase diagram represented in Fig. 2.4b. Interestingly, we can observe that the bistable area is progressively enlarged and invades the subcritical area $\lambda < \lambda_c$ as cooperation increases, which suggests that cooperation confers higher stability to endemic configurations even in *a priori* controlled scenarios.

The analysis of the phase diagram has revealed the existence of a critical point q_c. Thus far, we have obtained this point numerically by comparing the forward and backward continuation schemes. In what follows, we seek to derive an analytical approximation for its value by analyzing a simple case study. In particular, let us consider a mean field scenario in which every agent is equivalent and fix $\mu = 1$. Automatically, the last constraint implies that the II compartment is unaccessible. If we sum Eqs. (2.22-2.23) and introduce a mean field assumption in the contagion probabilities, implying that all individuals have degree k and neglecting the network structure, we arrive to the following expression for the epidemic size in the steady state:

$$\rho = (1-\rho)\left(1-(1-\lambda\rho)^k\right) + \rho\left(1-(1-q\lambda\rho/2)^k\right) \tag{2.33}$$

The value q_c corresponds to the coupling strength for which a saddle node bifurcation occurs at $\lambda = \lambda_c$ and, as a consequence, a new physical solution emerges. To find the saddle node bifurcation, we compute the Taylor expansion of the nonlinear terms and neglect those terms with $\mathcal{O}(\rho^4)$. Taking into account that $\lambda_c = 1/k$ under a mean field approach, Eq. (2.33) turns into

$$\rho^2\left(\frac{q}{2} - \frac{3k-1}{2k}\right) + \rho^3\frac{k-1}{2k}\left(1 - \frac{q^2}{4} + \frac{k-2}{3k}\right) = 0\ . \tag{2.34}$$

We find two fixed points, $\rho_1 = 0$ and another different solution ρ_2 given by

$$\rho_2 = \frac{3k-1-kq}{(k-1)\left(1-q^2/4+\frac{k-2}{3k}\right)}\ . \tag{2.35}$$

The critical point q_c corresponds to the q value at which ρ_2 becomes physically meaningful. In particular,

$$q_c = \frac{3k-1}{k}\ . \tag{2.36}$$

This expression reveals that the critical coupling strength depends on the connectivity of the underlying population and is bounded so that $2 \leq q < 3$. In particular, assuming $k = 8$ yields $q_c = 2.87$, constituting a good approximation for the value computed numerically in Fig. 2.4.

Finally, we want to assess the interplay among the cooperation between pathogens and the heterogeneity of the underlying contact structure. To do so, we rely on a model introduced by Gómez-Gardeñes and Moreno [91] which allows for interpolating between ER and BA networks by tuning a parameter

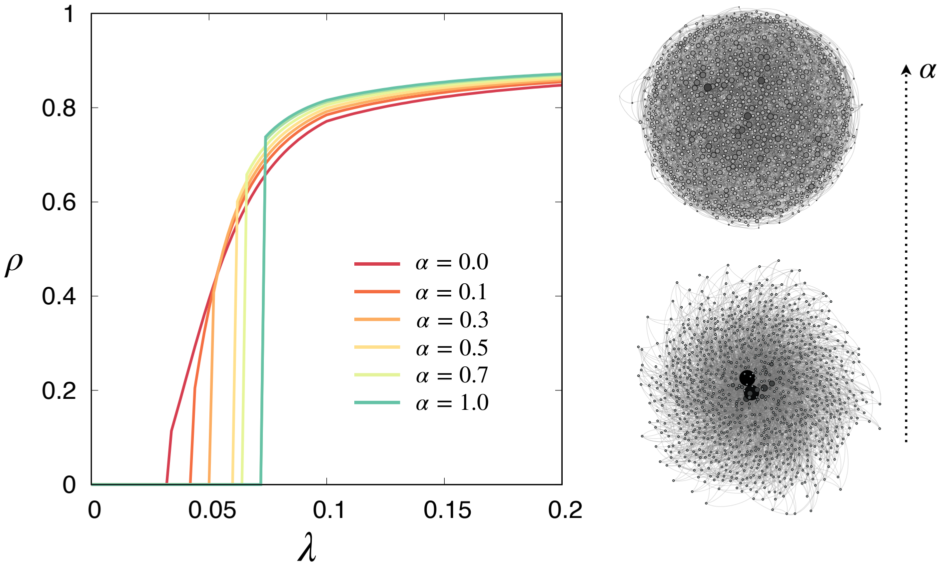


Figure 2.5: Epidemic curves $\rho(\lambda)$ computed following the forward continuation scheme for different networks characterize by the parameter α. The interaction strength is fixed to $q = 10$ and the rest of the parameters are the same as in Fig. 2.4. The pictures on the right illustrate two of the networks with $\alpha = 0$ (bottom) and $\alpha = 1$ (top) respectively. In both of them, the size of the nodes is proportional to its degree.

α. This parameter determines the fraction of nodes of the final network which form the initial bulk from which a BA network is constructed and, therefore, drawing their links randomly rather than according to the preferential attachment mechanism. Indeed, $\alpha = 0$ recovers the BA model whereas $\alpha = 1$ generates an ER network.

Fig. 2.5 shows the epidemic curves $\rho(\lambda)$ for different networks in the strong cooperation regime, for $q = 10$. As expected by the analysis made in Sec. 2.1, increasing the heterogeneity leads to a decrease of the epidemic threshold. Despite the higher vulnerability of heterogeneous configurations, the jump in the order parameter is much less pronounced because of the localization of the endemic states in the most connected nodes. In contrast, in homogeneous configurations, all nodes are very similar in terms of connectivity which causes that, once the transition takes place, a macroscopic set of them contributes to the endemic state, thus explaining the more abrupt behavior observed.

2.2.3 Competitive dynamics

Now we shift our attention to the competitive case in which the spread of one disease is hampered by the presence of another one, so $q < 1$. A particular case

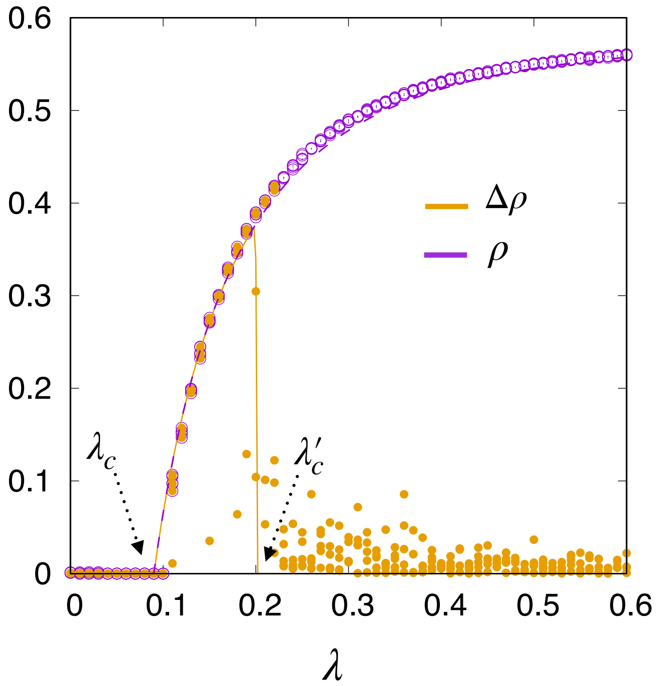


Figure 2.6: Evolution of the total epidemic size ρ (purple) and the difference in pathogen abundance $\Delta\rho$ (orange) as a function of the infectiousness for mutually exclusive diseases, $q = 0$. The underlying topology is an ER network with $N = 1000$ nodes and $\langle k \rangle = 8$. Dots shows the results from 30 realizations of the Monte Carlo simulations for each λ value whereas solid lines represent the results obtained from the numerical iteration of Eqs. (2.21-2.31).

of interest given its frequency in nature is that of mutually exclusive diseases for which contracting one disease automatically protects the host from getting infected by the other ones. Previous works on mutually exclusive diseases have revealed the existence of two different outcomes for competing diseases. Namely, depending on the model and the topology of contacts, the competition may end up with the prevalence of one pathogen over the other one which is eventually extinguished or instead with the coexistence of both pathogens forming disjoint clusters in the population. To capture the spread of mutually exclusive diseases, we fix $q = 0$ and consider the emergence of a new strain in an ongoing epidemic outbreak. To do so, we introduce an asymmetry in the number of individuals initially infected by each disease. In particular, we assume that initially 2% of the population carries one pathogen whereas 1% of the population is infected by the other one. Apart from the total epidemic size, another useful indicator to quantify the outcome of the competition is the difference in the individual epidemic sizes $\Delta\rho$, which is computed as

$$\Delta\rho = |\rho^{IS}(\infty) - \rho^{SI}(\infty)| \ , \tag{2.37}$$

with

$$\rho^{IS}(\infty) = \frac{1}{N}\sum_{i=1}^{N}[\rho^{IS}]_i^{\infty} \ , \tag{2.38}$$

$$\rho^{SI}(\infty) = \frac{1}{N}\sum_{i=1}^{N}[\rho^{SI}]_i^{\infty} \ . \tag{2.39}$$

Fig. 2.6 shows the evolution of both order parameters as a function of the infectiousness λ, having a fair agreement again between Monte Carlo simulations and the numerical results obtained from the iteration of the equations. In addition to the usual transition separating the disease-free state from the endemic regime, another transition occurs at $\lambda = \lambda_c'$. Note that this transition does not affect the total prevalence of the disease but instead shapes the outcome of the competition between both diseases. Namely, for $\lambda < \lambda_c'$, the dominance regime is the stable solution and the most abundant pathogen prevails over the other one. This happen because, for low infectiousness values, the dominant disease is able to block the spreading paths of the weak one, thus precluding its fostering across the population. In contrast, when infectiousness is high enough, the weak pathogen is able to spread despite the existence of the other pathogen, thus leading to the solution of coexistence in which the infected population is equally distributed among both diseases.

Unlike the cooperative scenario, unfortunately, it is not possible to find an analytical closed expression for the second threshold λ_c', for its computation involves solving a self-consistent equation whose terms depend on the fixed points of the individual SIS dynamics of the dominant pathogen (see Appendix A). To numerically study the value of the second threshold as a function of the underlying network structure, we make the following definition:

$$\lambda_c' \coloneqq \{\min(\lambda) \mid \rho(\lambda) \geq 2/N, \Delta(\lambda) < 0.01\rho(\lambda)\} \ .$$

We are now interested in determining the influence of the network of contacts on the different regimes arising from the cooperation. For this purpose, we represent in Fig. 2.7 the values for λ_c and λ_c' as a function of the parameter α controlling network's heterogeneity. Interestingly, we notice that, the more heterogeneous the network is, the lower the ratio λ_c'/λ_c is, implying that the dominance regime loses stability in favor of the coexistence solution. Thinking of spreading pathways, this occurs because the presence of highly connected nodes allows the weak pathogen to find potential hosts despite the existence of the dominant strain.

Finally, we analyze in Fig.2.7 how the phase diagram is altered for non-mutually exclusive diseases so that $q \neq 0$. In particular, we observe that for

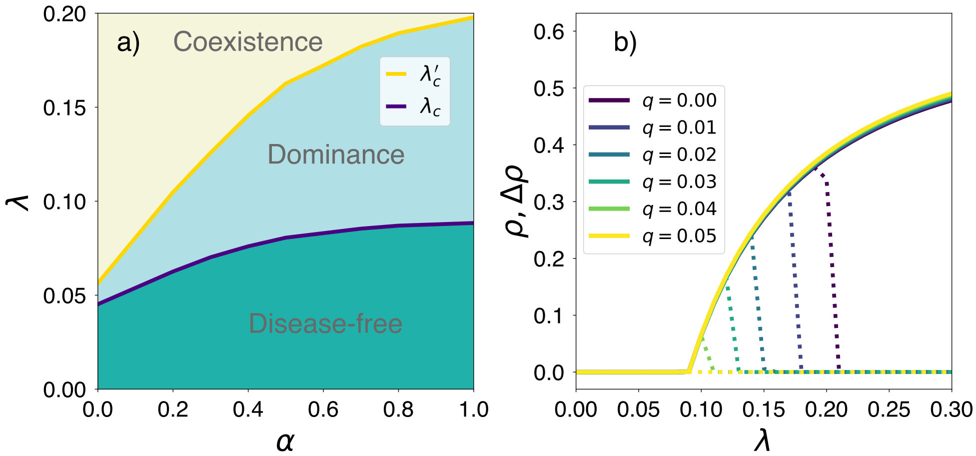


Figure 2.7: Panel a): Phase diagram of the spread of two mutually exclusive diseases with $q = 0$. The control parameters are the heterogeneity of the underlying network governed by α and the infectiousness λ. Three different regimes appear: Disease-free state characterized by $\rho = \Delta\rho = 0$, the dominance regime for which $\rho \neq 0$ and $\Delta\rho \neq 0$ and the coexistence regime in which a non-finite fraction of infected individuals ($\rho \neq 0$) is shared between both pathogens ($\Delta\rho = 0$). The solid lines represent the boundaries between the different regimes. Panel b) Evolution of ρ (solid lines) and $\Delta\rho$ (dashed lines) as the competition between both pathogens is weakened (color code). In both panels, we set the same network and recovery parameter as for the previous figures.

very low coupling strength q, the competition between both pathogens still leads to the extinction of the weak one, albeit contracting both diseases is no longer forbidden. Nonetheless, the extent of the dominant regime starts shrinking losing stability for $q = 0.05$, from which the coexistence between both pathogens is the unique stable configuration in supercritical scenarios when $\lambda > \lambda_c$.

2.3 Contact tracing on complex networks

So far, we have characterized uncontrolled epidemic scenarios where the trajectory of ongoing outbreaks is univocally determined by the structure of the network of contacts, the biological features of the pathogens or the interactions existing among them. Albeit theoretically enlightening, these scenarios hardly happen in reality, for the exponential growth in the number of infected individuals in supercritical epidemic scenarios usually force the authorities to promote interventions aimed at keeping them under control. The latest example of the relevance of control policies to shape disease evolution is the spread of the SARS-CoV-2 virus, responsible for the COVID-19 disease. From the first case officially reported in December 2019 in Wuhan, COVID-19 has represented the most important pandemic humankind has faced since the Spanish Flu in 1918, taking a huge humanitarian, social and economic toll worldwide. In particular, as of May 2021, more than 152 million of people have been diagnosed with COVID-19, of whom 4 million have died because of the disease.

The impact of COVID-19 pandemic has not been uniform, with important differences between countries as a function of the strategies [92–94] chosen to cope with the spread of the virus. Regardless of their nature, control policies seek to lower the effective reproduction number $\mathcal{R}_e$ defined as the average number of contagions that an infectious individual makes during her infectious period. Note that this number differs from $\mathcal{R}_0$ in the sense that it quantifies the stage of an ongoing outbreak rather than the potential infections made by an index case in a fully susceptible population. Qualitatively, the effective reproduction number reads as

$$\mathcal{R}_e(t) \sim \lambda k \tau \rho^{\mathrm{S}}(t) \ , \tag{2.40}$$

where λ represents the probability of infection per contact, k the average number of contacts, τ the infectious period and $\rho^{\mathrm{S}}(t)$ the fraction of population susceptible at time t. Therefore, the reduction of the effective reproduction number can be achieved with strategies targeting each of these four components. For instance, the promotion of prophylactic measures such as hand hygiene or mask wearing is translated into a lower λ value, the vaccination of the population leads to a depletion of the pool of susceptible individuals ρ^{S} or lockdown policies entail a strong reduction in our sociability k.

The explosion of COVID-19 cases took many countries worldwide by surprise, forcing their authorities to deploy aggressive lockdown policies as an emergency tool to drastically reduce the number of contacts of the population. Despite the effectiveness of the latter interventions, they have had important economic side effects as a consequence of stopping the work activity for long periods of time. In contrast, the previous experience acquired by other countries, mainly located in South East Asia, during past epidemiological crises such as SARS in 2003 or H1N1 in 2009 has allowed them to anticipate the spread of the current pandemic and avoid the emergence of macroscopic outbreaks. To do so, they have relied on strict border closure policies restricting the possible landing of infectious individuals from other countries and strict contact tracing and case isolation policies preventing community transmission in case of the emergence of small local outbreaks.

The success of the aforementioned contact tracing policies rely on the identification of the close contacts of a diagnosed individual, mainly thanks to digital proximity applications, and their immediate isolation, therefore reducing

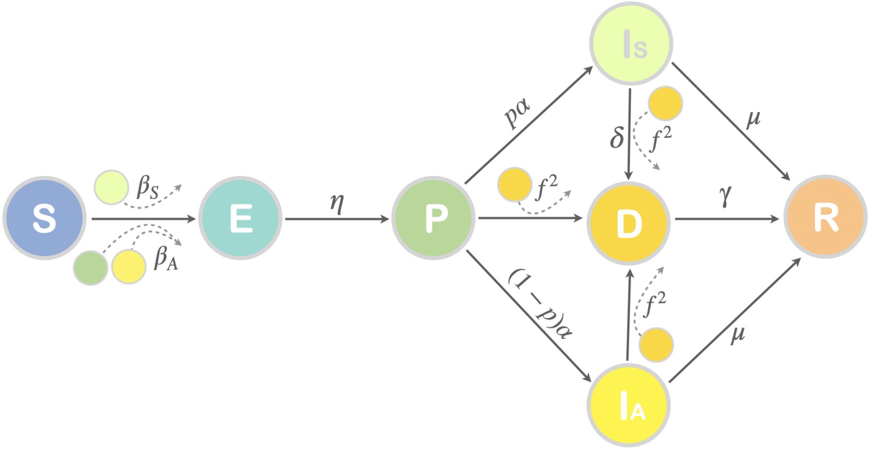


Figure 2.8: Compartmental scheme of the model proposed to accommodate contact tracing in epidemic modeling. Note that both contagion and detection events are driven by pairwise interactions whereas the rest of transitions are governed by constant rates encoding the typical epidemiological times.

the number of secondary contagions caused by the index case of an outbreak. In terms of the effective reproduction number, contact tracing policies entail a reduction of the infectious period τ of the close contact depending on the moment in which she is identified and an immediate reduction in her sociability k. Therefore, in presence of contact tracing, epidemic spreading acts as a self-inhibited dynamics, for activating contact tracting requires having detected infectious individuals, whose spreading capability is highly precluded when the control mechanism acts.

Interestingly, the pairwise nature of the events triggering contact identification resembles the interactions by which a virus spread across a population. Therefore, we can identify contact tracing as a spreading phenomenon taking place across the virtual network constructed from the users of digital applications. Regarding their interaction with the actual pathogen, we can identify contact tracing as a competing dynamics with disease spreading, for it leads to a dismantling of the contact network sustaining the spread of the pathogen across a population. In what follows, we propose a compartmental model tailored for the spread of COVID-19 to capture the interplay between network structure, disease spreading and the implementation of contact tracing policies.

2.3.1 The model

The specific features associated with the transmission of SARS-CoV-2 virus such as the important fraction of asymptomatic infections or the crucial role played by the contagions made before developing symptoms demands to revisit the compartmental models explained in this chapter. In particular, to accommodate these different epidemiological states, we assume that the population is divided into the following compartments: Susceptible (S), Exposed (E), Pre-symptomatic (P), Infectious Asymptomatic (I_A), Infectious symptomatic (I_S) and Recovered (R). Finally, to include the contact tracing mechanism and the possible detection by symptoms of the patients, we also include an extra compartment gathering the individuals Detected (D) and considered as active cases.

A schematic representation of the model can be found in Fig. 2.8. The flow of the compartmental model reads as follows: For each contact of susceptible agents with infectious peers in a given state m ($m = \{\mathrm{P}, \mathrm{I}_A, \mathrm{I}_S\}$), they contract the disease with a probability λ_m, entering into the E compartment where they incubate the virus for a typical time given by η^{-1}. Once they leave the exposed compartment, they enter into the Pre-symptomatic compartment where they have enough viral load to propagate the pathogen, staying there for α^{-1} time steps. Afterwards, a fraction $1-p$ of the patients are Asymptomatic individuals who do not show symptoms before the eventual end of their infectiousness after a typical time μ^{-1} and a fraction p are Symptomatic individuals. Symptomatic patients can either be tested and detected after a typical time δ^{-1} or end their infectiousness after a period of μ^{-1} days. Finally, the contact tracing mechanism adds an extra detection channel during a typical time γ^{-1} for those contacts of detected individuals who are equipped with the application and have incubated enough viral load to be infectious. Specifically, we assume that a fraction f of the population is registered in the application so that the probability of observing a successful contact tracing event is f^2.

Let us consider that our system is composed of a set of N agents. Under the microscopic processes described above, the model is characterized by the probabilities for each individual i being in each of the m available compartments, denoted by ρ_i^m. The time evolution of these equations read as

$$\begin{aligned}
\rho_i^{\mathrm{S}}(t+1) &= \rho_i^{\mathrm{S}}(t)\left(1-\Pi_i^{\mathrm{S}\to\mathrm{E}}(t)\right) \ , \\
\rho_i^{\mathrm{E}}(t+1) &= (1-\eta)\rho_i^{\mathrm{E}}(t) + \rho_i^{\mathrm{S}}(t)\Pi_i^{\mathrm{S}\to\mathrm{E}}(t) \ , \\
\rho_i^{\mathrm{P}}(t+1) &= (1-\Pi_i^{\mathrm{P}\to\mathrm{D}}(t))(1-\alpha)\rho_i^{\mathrm{P}}(t) + \eta\rho_i^{\mathrm{E}}(t) \ , \\
\rho_i^{\mathrm{I}_A}(t+1) &= (1-\Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t))(1-\mu)\rho_i^{\mathrm{I}_A}(t) + (1-\Pi_i^{\mathrm{P}\to\mathrm{D}}(t))(1-p)\alpha\rho_i^{\mathrm{P}}(t) \ , \\
\rho_i^{\mathrm{I}_S}(t+1) &= (1-\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t))(1-\mu)\rho_i^{\mathrm{I}_S}(t) + (1-\Pi_i^{\mathrm{P}\to\mathrm{D}}(t))p\alpha\rho_i^{\mathrm{P}}(t) \ , \\
\rho_i^{\mathrm{D}}(t+1) &= (1-\gamma)\rho_i^{\mathrm{D}}(t) + \Pi_i^{\mathrm{I}_S\to\mathrm{D}}\rho_i^{\mathrm{I}_S}(t) + \Pi_i^{\mathrm{I}_A\to\mathrm{D}}\rho_i^{\mathrm{I}_A}(t) + \Pi_i^{\mathrm{P}\to\mathrm{D}}(t)\rho_i^{\mathrm{P}}(t) \ , \\
\rho_i^{\mathrm{R}}(t+1) &= \rho_i^{\mathrm{R}}(t) + \gamma\rho_i^{\mathrm{D}}(t) + \mu(1-\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t))\rho_i^{\mathrm{I}_S}(t) \\
&\quad + \mu(1-\Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t))\rho_i^{\mathrm{I}_A}(t) \ .
\end{aligned} \tag{2.41}$$

Note that the flows involving the detection of infectious individuals have been prioritized over the transitions between the different epidemiological compartments to reflect the fact that, once individuals test positive, they are considered as active cases regardless of their epidemiological state. The contagion term of the susceptible individuals should take into account the different infectiousness of the peers with whom they contact. Therefore,

$$\Pi_i^{\mathrm{S}\to\mathrm{E}}(t) = 1 - \prod_{j=1}^{N}\left\{1 - A_{ij}\left[\lambda_{\mathrm{P}}\rho_j^{\mathrm{P}}(t) + \lambda_{\mathrm{I}_A}\rho_j^{\mathrm{I}_A}(t) + \lambda_{\mathrm{I}_S}\rho_j^{\mathrm{I}_S}(t)\right]\right\} . \tag{2.42}$$

Finally, the detection terms driven by pairwise interventions can be expressed as

$$\Pi_i^{\mathrm{P}\to\mathrm{D}}(t) = \Pi_i^{\mathrm{I}_A\to\mathrm{D}}(t) = 1 - \prod_{j=1}^{N}\left(1 - A_{ij}f^2\rho_j^{\mathrm{D}}(t)\right) \ , \tag{2.43}$$

$$\Pi_i^{\mathrm{I}_S\to\mathrm{D}}(t) = 1 - (1-\delta)\prod_{j=1}^{N}\left(1 - A_{ij}f^2\rho_j^{\mathrm{D}}(t)\right) \ . \tag{2.44}$$

Both probabilities differ in that symptomatic agents are also detected with direct testing without the need of being identified as a close contact of a reported case.

2.3.2 Results

To quantify the relevance of contact tracing in realistic setups, we apply our model to static networks extracted from face-to-face interactions measured in three different contexts: a hospital [95], a school [96] and a gallery of science [97]. To transform the temporal nature of face-to-face networks to a static configuration, we define a minimum duration of 5 minutes for the interaction

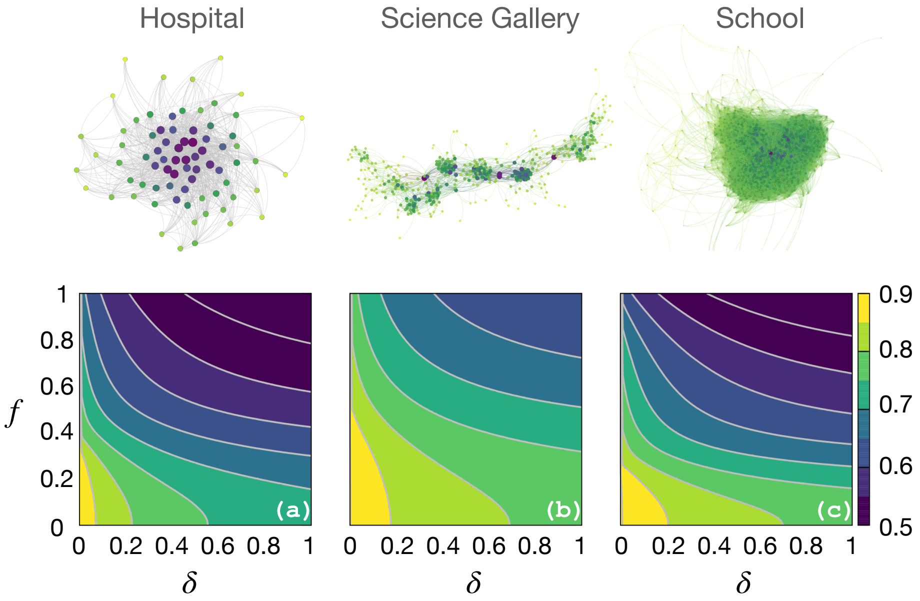


Figure 2.9: Attack rate $r(\infty)$ as a function of the symptomatic detection rate δ and the penetration of the digital contact tracing app f. The underlying contact networks are the static networks obtained from face-to-face interactions recorded in a hospital (Panel a), a science gallery (Panel b) and a school (Panel c) (see text for further details). In all the panels, the grey lines denotes different contour lines of the surfaces represented.

between two agents to be recorded by the application as a close contact between two individuals. The main structural properties of the resulting static contact networks are summarized in Table 2.1. Regarding the epidemiological parameters, most of them are extracted from the literature as reflected in Table 2.2 but we leave $\lambda_{\mathrm{P}} = \lambda_{\mathrm{I}_S} = \lambda_{\mathrm{I}_A}$ as a free parameter which is tuned in the three different networks to obtain an attack rate in absence of detection ($\delta = f = 0$) of $r(\infty) \simeq 0.9$.

Figs. 2.9a-c show the dependence of the attack rate $r(\infty)$ on the symptomatic detection governed by δ and the fraction of the population equipped with the application f. In all these panels, we check that improving the detection mechanisms by decreasing the time from the onset of symptoms to detection (increasing δ) or achieving a higher penetration of the application (increasing f) yields a substantial reduction in the extent of an outbreak. Interestingly, the impact of promoting contact tracing is much more pronounced in all the cases when symptomatic detection is efficient, for the latter is required as the spark needed to trigger the cascade of detections on the virtual space. Regarding the role of the contact network, some quantitative differences are observed between the different diagrams but qualitatively all of them are

Network	N	$< k >$	r
Hospital	75	30	-0.18076
Science Gallery	410	13	0.22575
School	784	60	0.22814

Table 2.1: **Characteristics of the three proximity networks**. For each network we show the number of nodes N, the average degree of the nodes $\langle k \rangle$, and the assortatitivity measured as the Pearson correlation between the degrees of adjacent nodes r. We also report the reference were these networks where presented and analyzed.

Parameter	Value	Description	Reference
η	1/2.5 day^{-1}	Probabilty E $\rightarrow$ P	[98]
α	1/2.5 day^{-1}	Probabilty P $\rightarrow$ $\mathrm{I_A}$, $\mathrm{I_S}$	[98]
p	0.65	Fraction of Symptomatic	[99]
μ	1/7 day^{-1}	Probability $\mathrm{I_A}$, $\mathrm{I_S}$ $\rightarrow$ R	[100, 101]
γ	1 day ^{1}	Detection window length	Assumed

Table 2.2: Epidemiological parameters of the compartmental model.

pretty similar. For this reason, in what follows, we focus on illustrating the physics of both detection mechanisms with a specific configuration, the school proximity network.

To obtain a more clear picture on the role of each detection mechanism, we keep one fixed and study the effect of enhancing the other one on the evolution of the infection curves $C(t)$. Here we define $C(t)$ as the daily incidence, which according to our formalism reads as

$$C(t) = \sum_{i=1}^{N} C_i(t) = \sum_{i=1}^{N} \rho_i^{\mathrm{S}}(t)\Pi_i^{\mathrm{S}\rightarrow\mathrm{E}}(t) \ . \tag{2.45}$$

Fig. 2.10 reveals that, despite their common overall positive effect on the attack rate, the impact of symptomatic detection and contact tracing on the dynamics is remarkably different. Namely, symptomatic detection *flattens* the epidemic curve by reducing and delaying the epidemic peak whereas contact tracing *bends* the epidemic curve by reducing but also anticipating the peak of contagions. In terms of the standard SIR model, the former behavior corresponds to a decrease of the reproduction number $\mathcal{R}$ to obtain a new harmless supercritical scenario whereas the latter phenomenon implies that contact tracing is able to bring the system to a subcritical regime. This difference resides in the fact that contact tracing is able to suppress the transmission from agents in both the presymptomatic and asymptomatic epidemic states. Finally, it is worth noting that, despite the high effectiveness of contact tracing, it does not allow for entirely mitigating the epidemic outbreak but instead keep the system

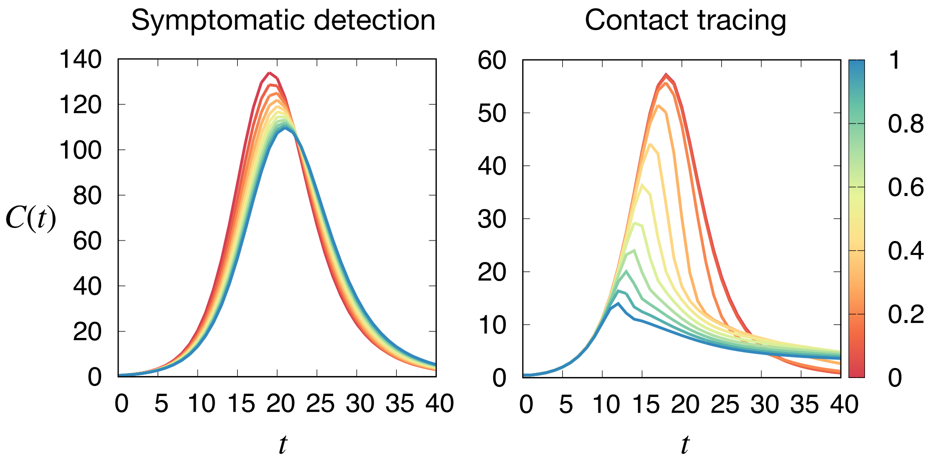


Figure 2.10: Evolution of the daily cases $C(t)$ as detection mechanisms are enforced. Panel a): The color of the lines denote the value of the symptomatic detection rate δ whereas the penetration of the application is fixed to $f = 0$. Panel b): The color of the lines denotes the penetration of the application f whereas the symptomatic detection rate is fixed to $\delta = 0.05$. Both panels have been obtained by studying the school proximity network.

with a low basal incidence which is needed to keep this mechanism working.

To further characterize the differences between both control mechanisms, let us compute numerically the expression of the effective reproduction number $\mathcal{R}(t)$, hereby defined as the number of contagions made by an average infectious individual becoming exposed at time t during the infectious period [102]. Note that, at variance with the problems tackled throughout this chapter, the contact tracing mechanism cannot be incorporated in the analytical expression of the epidemic threshold λ_c or the basic reproduction number $\mathcal{R}_0$, for contact tracing events involve terms $\mathcal{O}(\epsilon^2)$ which vanish when linearizing the equations.

To estimate the evolution of $\mathcal{R}(t)$, one must account for all the possible combinations for the course of the disease in infected patients. These combinations are encoded in joint probabilities and determine $\mathcal{I}_i(t)$, defined as the expected number of contagions made by agent i after becoming exposed at time t. The derivation of the expression for this quantity is a bit cumbersome so we refer the reader to Appendix B for further details. With this information at hand, the effective reproduction number at time t reads as

$$\mathcal{R}(t) = \frac{\sum_{i=1}^{N} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)\mathcal{I}_i(t)}{\sum_{i=1}^{N} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)} \,. \tag{2.46}$$

We represent the time dependence of the effective reproduction number

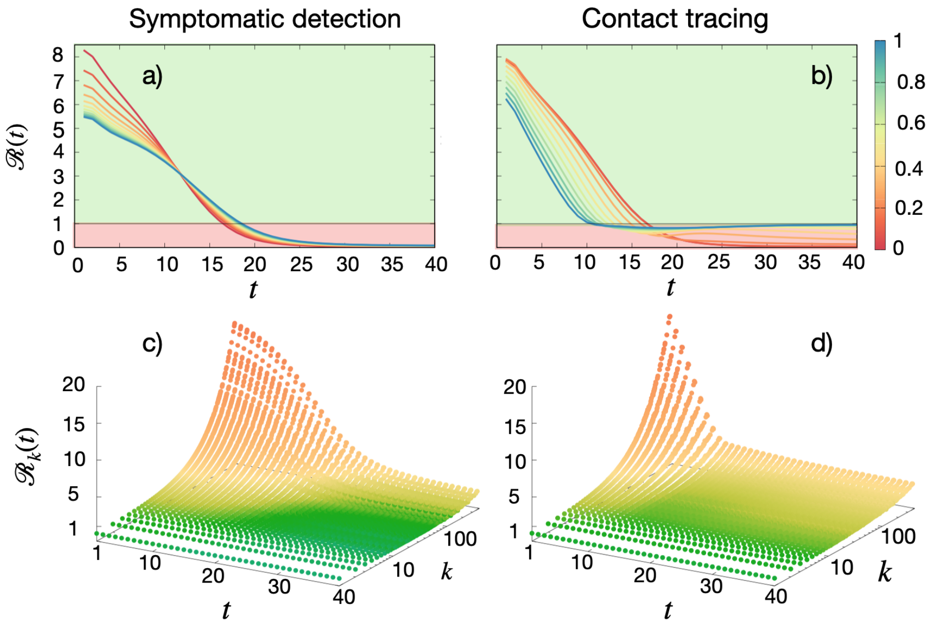


Figure 2.11: Panels a-b: Evolution of the global effective reproduction number $\mathcal{R}(t)$ defined according to Eq. (2.46) as symptomatic detection is enhanced (Panel a) or contact tracing is promoted (Panel b). The values of δ and f are the same as the one explored in Fig. 2.9. Panels c-d: Evolution of the effective reproduction number for each degree class k, $\mathcal{R}_k(t)$ computed by Eq. (2.47). The parameters governing detection processes are set to $(\delta, f) = (1, 0)$ in Panel c) and $(\delta, f) = (0.05, 1)$ in Panel d).

in Fig. 2.11 when symptomatic detection (Panel a) or contact tracing (Panel b) are the dominant detection mechanisms respectively. Firstly, we observe that at very early stages of the outbreak, symptomatic detection works better because of the time lag needed for contact tracing to be activated. Once contact tracing is working, it leads to a more abrupt decrease of the reproduction number, forcing the system to enter in a subcritical regime as anticipated before. Note that the basal incidence aforementioned is reflected by a sustained value of the reproduction number around $\mathcal{R} = 1$, once contact tracing controls the first epidemic peak.

Finally, to get some microscopic/mesoscopic insights into the impact of both control mechanism, we compute the time evolution of the effective reproduction number associated with each degree class k, denoted by $\mathcal{R}_k(t)$. This indicator tell us the average number of contagions made by individuals becoming exposed at time t with k neighbors in the contact network. Therefore,

$$\mathcal{R}_k(t) = \frac{\sum_{(i\,|k_i=k)} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)\mathcal{I}_i(t)}{\sum_{(i\,|k_i=k)} \rho_i^S(t-1)\Pi_i^{S\to E}(t-1)}\ . \tag{2.47}$$

We represent the time evolution of these queantities for two scenarios characterized by $(\delta, f) = (1, 0)$ (Panel c) and $(\delta, f) = (0.05, 1)$ (Panel d). There it becomes clear that contact tracing is so effective because it dismantles the contact network and hinders the contagions made by the most connected nodes. This result is aligned with the different papers studying the impact of contact tracing under different theoretical approaches [103, 104] on contact networks.

2.4 Conclusions

In this chapter we have proposed a general framework to accommodate the interaction existing between two pathogens which are simultaneously propagating over the same population. This interaction may lead to a cooperative scenario in which contracting one pathogen improves the diffusion of the other one or to a competitive scenario where one disease confers total or partial immunity with the respect to its counterpart, thus hindering its spread. In both cases, we have checked how new physics emerges when dealing with interacting diseases which cannot be explained with the usual compartmental models addressing each one individually.

In the case of cooperation, we have revealed the existence of a critical coupling from which the smooth second order transition in the epidemic size of the SIS model, separating the disease-free and endemic states, turns into an abrupt first order one. As a consequence, we have found that the cooperation between diseases leads to bistable solutions in which the emergence of macroscopic epidemic states in subcritical outbreaks is possible. Finally, we have found that the first order transition is much more abrupt in homogeneous networks than in heterogeneous networks. This occurs because epidemic states are mainly localized on the most connected nodes in heterogeneous configurations whereas in homogeneous networks population-wide outbreaks emerge abruptly as a result of the similar vulnerability of all the nodes.

For competitive diseases, we have revealed the existence of a second threshold in the supercritical regime separating the solution of dominance of one pathogen over the other one from the equilibrium corresponding to their coexistence. Interestingly, the dominance solution leading to the extinction of one pathogen emerges despite being in a supercritical scenario, which should theoretically ensure its survival over the population in an isolated scenario. We have found that this phenomenon mainly occurs for strongly competing diseases because the presence of the most abundant pathogen blocks the spreading paths of the weak pathogen. As a consequence of this mechanism, the effective degree for the weak pathogen is reduced turning the hypothetically supercritical scenario into a subcritical one, thus explaining its extinction. For this reason, the dominance of one pathogen over the other one is less stable in heterogeneous configurations where the highly connected nodes offer the weak pathogen more chances to keep reproducing and survive.

In both cases, we have focused on identical interacting diseases characterized by the same epidemiological parameters. Nonetheless, the formalism could be applied to a more general scenario in which the latter assumption is broken. For instance, in the case of nonidentical competing diseases, the prevalent disease in the dominance solution is not the most abundant one but instead corresponds to the one with the higher infectiousness, as observed in the subsequents COVID-19 variants with higher infectiousness becoming dominant worldwide. In the case of cooperative diseases, the existence of a disease with much higher infectiousness than the other one would lead to asymmetric scenarios where one disease is highly extended on the population whereas the other one remains silent. As a consequence, the hysteresis cycles caused by the cooperation between both of them would be enlarged due to the stability of the epidemic states caused by the most contagious disease. Finally, this formalism paves the way to the elaboration of other frameworks addressing the spread of simultaneous interacting spreading units different from diseases. For example, a more socially inspired model constructed from this framework could be useful to study the intertwined spread of different opinions across social platforms.

Beyond competing spreading phenomena in the physical space, we have also addressed the impact of digital contact tracing on epidemic spreading. Contact tracing can be conceived as a spreading process, running on top a virtual network constructed from the users of these applications, which is competing or hindering the spread of contagious diseases across the physical network. We have revealed that control tracing outperforms other control mechanisms such as improving testing capacity, due to its ability to dismantle the network of contacts and detect infectious individuals at the very beginning of their infectious period. Finally, we have also observed that, despite optimizing the trade-off between the functionality of a society and the control of an ongoing outbreak, contact tracing policies cannot lead to the total extinction of outbreaks because they need the emergence of symptomatic individuals to be functional. This extinction arises from the combination of these interventions with lockdowns of the population, which are widely studied in Chapter 6.

Chapter 3

Social dynamics

> *A person whose life is intertwined with the group [through friendship, kinship, organizational membership, informal support networks, or shared relations with outsiders]. . . has a big stake in the group's fate. When collective action is urgent, the person is likely to contribute his or her share even if the impact of that share is not noticeable.*
>
> B. Fireman and W.A. Gamson [105].

Nowadays, the physical barriers previously separating people from faraway regions have been removed by the existence of a vast interconnected virtual network composed of different social platforms. As a consequence, the time elapsed between the emergence of a given cultural trait or social trend in localized minorities, its progressive transmission across the population and its eventual rooting in the society has been tremendously accelerated, giving rise to a complex and fast-shifting cultural landscape. In light of this complex scenario, disentangling the microscopic rules driving the exchange of information across the network connecting people worldwide and its further penetration in the society has become essential to understand the nature of different social phenomena.

The direct interaction required between the spreader of an idea and its possible receptor in the virtual network reminds of the physical contact needed to transmit a disease from infectious to susceptible individuals. For this reason, different approaches resembling the compartmental models used in mathematical epidemiology have been proposed to deal with the spread of ideas. In Sec. 3.1 we explain these models and study the main similarities and differences existing between such frameworks and the models described in the previous chapter. Most of these differences come from the social flavor of human interactions in the case of ideas diffusion, whose nature is different from the biological processes governing the spread of pathogens between hosts.

Beyond the processes governing their spread across the population, the ideas of one individual or her opinions about a given topic are not immutable and usually change over time. In particular, these changes are far from being random processes and are usually highly influenced by the opinion of the rest of the population with which one individual is in close contact. In Sec. 3.2, we give a brief overview about the models used to accommodate the rules driving the alignment of different opinions.

Although consensus and ideas diffusion have been already well-characterized independently, less work has been done on the physics arising from the interdependencies between both processes. To fill this gap, we propose in Sec. 3.3 an unifying theoretical framework characterizing the simultaneous evolution of both dynamics and the feedback existing between them. This framework allows us to propose a mechanism responsible for the sudden irruption of social movements starting from strongly active minorities. We check that this phenomenon is the fingerprint of the cooperation between consensus dynamics, i.e. the extent to which an idea penetrates the society, and spreading dynamics, i.e. how fast ideas are propagated.

Finally, in Sec. 3.4, we exploit the parallelisms between social dynamics and epidemic spreading to characterize the diffusion of corrupt behaviors across a given set of agents. To do so, we show how the standard SIR introduced in Chapter 2 model can be modified to account for the factors driving the onset of corruption in a given society such as the greed, the social punishment or the coercion. In terms of interacting spreading processes, the model represents a competition between two behaviors, honesty and corruption.

3.1 Compartmental models for social dynamics

In the context of epidemics, the propagation of a virus is usually mediated by pairwise interactions between infected hosts and susceptible individuals whose success probabilities depend on the infectiousness of the pathogen. Likewise, the infectious period of the disease's hosts is tightly linked with the trade-off between the speed at which the antibody response is deployed in our organisms and the reproduction rate of the virus. Therefore, in absence of any pharmaceutical interventions, mathematical models assume both the infectiousness and the infectious period as constant variables which are characteristic of each disease.

In social dynamics, the contagion/recovery events are not univocally determined by constant rates. Instead, these success probabilities for the microscopic processes are usually shaped by the interplay between the type of information spreading across the network and the dynamical state of the system. For instance, in rumor dynamics, those spreading a rumor do not have an established typical time during which they actively propagate the information to other peers. Instead, it seems more logical to assume that a person spreads actively a rumor until realizing that it is already well-known in the population.

Motivated by the latter fact, Maki-Thomson [106], based on the seminal work published by Kendall and Daley in 1964 [107], proposed a variant of the SIR model to account for the existence of recovery events mediated by pairwise interactions of the population. In this model, the Susceptible, Infected and Recovered compartments should be interpreted as ignorants (people not knowing the rumor), spreaders (people aware of the rumor and actively sharing it) and stiflers (people who have lost interest in the rumor) respectively. The transition from ignorant to spreader is equivalent to the contagions occurring in epidemic spreading. In contrast, spreaders become stiflers because of pairwise interactions with other spreaders or stiflers which underline the lack of novelty of the rumor. The mean field evolution of the fraction of individuals in each compartment for this model reads as

$$\dot{s} = -\lambda k s i \ , \tag{3.1}$$

$$\dot{i} = \lambda k s i - \alpha k i (i + r) \ , \tag{3.2}$$

$$\dot{r} = \alpha k i (i + r) \ . \tag{3.3}$$

It is useful to rewrite the former system of equations considering the closed population constraint so that $s + i + r = 1$. After doing so and rearranging terms, we obtain

$$\dot{s} = -\lambda k s i \ , \tag{3.4}$$

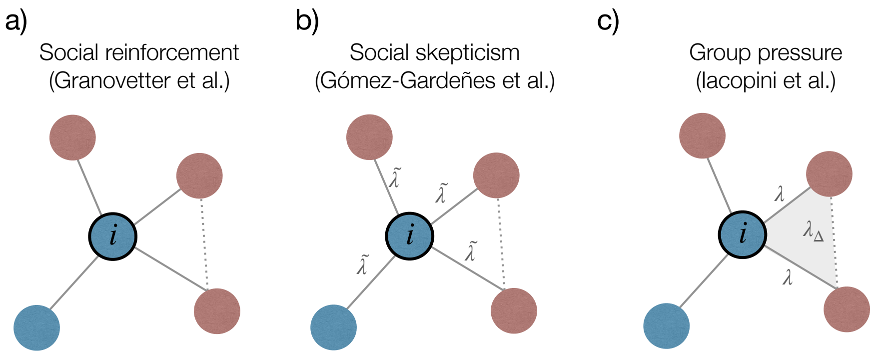
$$\dot{i} = (\lambda + \alpha) k s i - \alpha k i \ . \tag{3.5}$$

From Eq. 3.5, it becomes clear that the basic reproduction number $\mathcal{R}_0$ value is given by

$$\mathcal{R}_0 = \frac{\lambda + \alpha}{\alpha} > 1 \ . \tag{3.6}$$

Therefore, regardless of its virality λ, the rumor is able to reach a supercritical regime and spread across the network until the number of stiflers and spreaders is high enough to prevent other spreaders from propagating it. This is in sharp contrast with the SIR model, in which there exists a clear epidemic threshold separating the subcritical and supercritical behaviors of the system. Phenomenologically, the lack of a threshold can be associated with the emergence of rumors or fake news across social media independently of their reliability.

Apart from the different recovery mechanisms, the diffusion of ideas, opinions or cultural traits is usually intertwined with different environmental fea-



$$P_i(S \to I) = \Theta\left(\sum_k A_{ij}\sigma_j^I - T\right)$$

$$P_i(S \to I) = \tilde{\lambda}_i \sum_j A_{ij}\sigma_j^I$$
$$\tilde{\lambda}_i = \lambda e^{-\beta \sum A_{ij}\sigma_j^S}$$
$$\beta > 0$$

$$P_i(S \to I) = \lambda \sum_j A_{ij}\sigma_j^I + \lambda_\Delta N_\Delta$$
$$N_\Delta = \sum_{j \neq k} (1 - \delta_{ik})(1 - \delta_{ij}) A_{ij}A_{jk}A_{ki}\sigma_j^I\sigma_k^I$$
$$\lambda_\Delta > \lambda$$

Figure 3.1: Different models proposed to include complex contagion mechanisms in social dynamics. Panel a): Contagion events take place only if the number of spreaders surrounding a given agent i exceeds a critical mass. Panel b) The probability of adopting an idea λ is hindered when many of the neighbors surrounding an individual ignore that idea. Panel c) To include the effects of group pressure in the adoption of ideas, the transmission of the idea via collective interactions dictated by λ_Δ is higher than the contagion probability via pairwise contacts λ.

tures such as the credibility they have progressively acquired within the population or their legitimacy [108]. The simple contagion processes included in the usual epidemic models, driven by pairwise interactions and occurring at constant rates, cannot capture the impact of these features. For example, the standard SIR model cannot justify the behavioral patterns observed for the adoption of technological [109] or medical [110] innovations or the emergence of social movements [111, 112]. To overcome this problem, compartmental models applied to social dynamics should include more complex contagion mechanisms. These complex contagion mechanisms usually rely on collective or higher-order interaction terms which account for the presence of social reinforcement or group effect shaping the aforementioned patterns. In what follows, we briefly describe some compartmental models proposed to capture these phenomena.

The most paradigmatic framework along this line is the thresholding model proposed by Granovetter in 1978 [113] and used in subsequent works to characterize the spread of information [114, 115]. In this kind of models, it is assumed that there exists a critical mass of spreaders needed for one ignorant to adopt an idea. Therefore, threshold models represent the natural framework to capture the social reinforcement mechanism. A similar approach was followed by Gómez-Gardeñes et al. in [116], where they include the social skepticism of individuals to start spreading an idea which is not shared by their acquaintances. As a consequence, the diffusion of the idea is noticeably hindered until it is widely accepted across the population, leading to first order transitions in the size of spreaders. Finally, Iacopini et al. in [117] recently proposed a model using simplicial complexes which allow them to include the effects of group pressure on the diffusion of ideas. These models include higher-order links or hyperlinks, involving more than two individuals, where diffusion of ideas is accelerated due to the group effect. Once again, this assumption gives rise to first order transitions as the virality of the idea is high enough to start activating the so-called hyperlinks. In Figure 3.1 we sketch the three different alternatives previously described to illustrate the differences existing among them in terms of the contagion processes.

3.2 Consensus dynamics

As with the propagation of interacting diseases, the evolution of an opinion on a given topic and their penetration in the society cannot be understood as an isolated process but is instead highly influenced by its coexistence and interactions with the rest of opinions/attitudes about the same topic. In the literature, there is a vast number of publications addressing opinion dynamics (see [30] for an extensive review on the topic). These works can be mainly separated into two categories depending on how opinions are defined:

- *Discrete models*: These models assume that the different opinions about a given topic are characterized by a discrete variable $\sigma_i \in \mathbb{N}$. Most of these models assume the existence of two opposite opinions and are usually inspired in the Ising model with $s = 1/2$. In 1982, S. Galam [118] applied the Ising model, by providing its usual parameters with a socially inspired meaning, to understand the emergence of a strike within the workers of a plant. Other examples of this type of models are the voter model [119] in which a randomly selected agent adopts the opinion of one of her neighbors or the majority rule model [120] where all the individuals in a given group are forced to adopt the most popular opinion among its members.

- *Continuous models*: There, agents' opinion lies in a continuous space and therefore are represented by a variable $x \in \Re$. Some important models along this line are the so-called bounded confidence frameworks such as the Deffuant [121] or the Hegselmann-Krause [122] models which measure how two individuals influence mutually as a function of the distance between their opinions. Following this rationale, Starnini et al. [38] included a rewiring mechanism of the contact network towards those with more aligned opinions, giving rise to a rich phase diagram composed of different behaviors observed in social networks such as consensus, radicalization or polarization.

3.2.1 Kuramoto model in complex networks

Within the very diverse realm of models applied to opinion dynamics, we have chosen the Kuramoto model following [123]. The Kuramoto model represents one of the most used theoretical frameworks to study the collective synchronization of non-identical oscillators emerging as a result of their local interactions. Synchronization phenomena are ubiquitous in nature and are central actors in many different areas such as biology, economy or technology [124]. From a social perspective, synchronization has been applied to the study of social coordination [125] or behavior alignment within groups [126]. In our case, we will assume that each oscillator encodes the opinion of a given agent and the synchronized state represents the consensus eventually reached within the population.

Formally, the Kuramoto model was first conceived as a mean-field theory [127]. Specifically, Kuramoto studied how the phases $\vec{\theta}$ of a system of N non-identical oscillators, endowed with natural oscillating frequencies $\vec{\omega}$, change as the coupling between them is strengthened. Assuming that the interaction between two oscillators shorten the gap between their phases, the evolution of these phases reads as[1]

$$\dot{\theta}_i = \omega_i + K \sum_{j=1}^{N} \sin(\theta_j - \theta_i) \ . \tag{3.7}$$

This simple model allows for capturing the phase transition towards the synchronization as the parameter K varies. Specifically, for low K values, the oscillators' evolution is driven by their individual frequencies which produces a non-coherent collective state. In contrast, for high K values, the coupling term is dominant, forcing all the opinions to align among each other, thus producing a globally synchronized population. To quantify the degree of synchrony, Kuramoto proposed the order parameter r, measuring the global phase coherence

---

[1] For the sake of direct comparison with the model considering the contact network, here we have rescaled the coupling constant so that $K = NK^{\text{Kuramoto}}$.

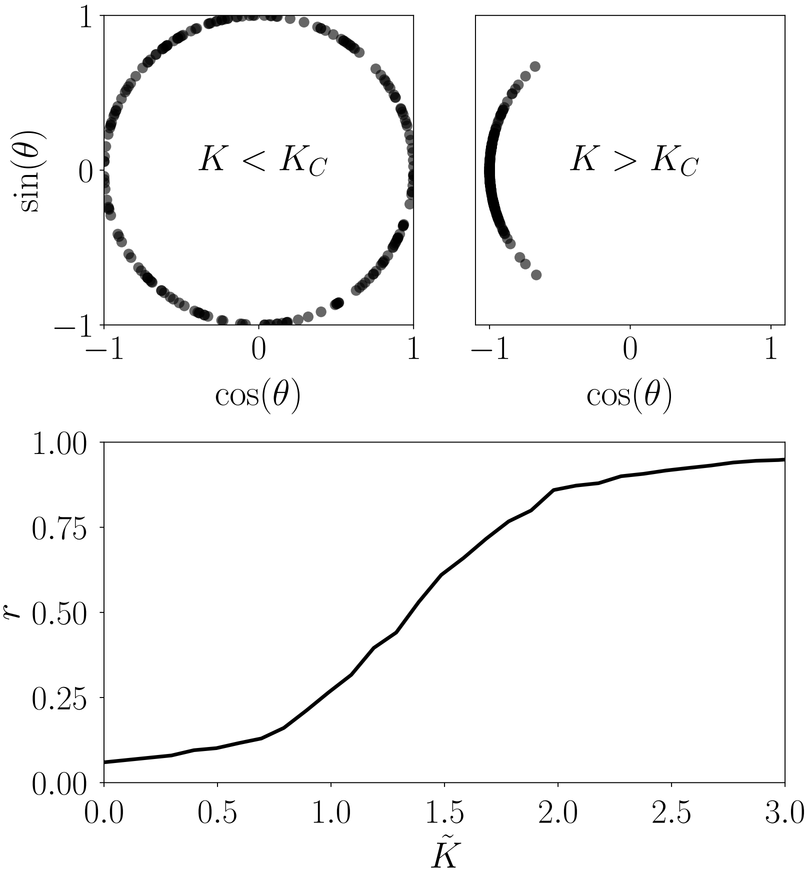


Figure 3.2: Top: Phases configuration in the subcritical regime of the Kuramoto model, corresponding to a desynchronized population (Left), and the supercritical regime characterized by the emergence of synchronization (Right). Bottom: Evolution of the global phase coherence r as a function of the re-scaled coupling parameter $\tilde{K} = K/K_c$, where K_c is computed according to Eq. 3.7. The system is composed of 200 oscillators with natural frequencies normally distributed with $\langle\omega\rangle = 0$ and $\sigma_\omega^2 = 1$.

across the set of oscillators. Mathematically, this parameter is expressed as

$$r(t)e^{i\phi(t)} = \frac{1}{N}\sum_{j=1}^{N} e^{i\theta_j(t)} \ , \tag{3.8}$$

where $\phi(t)$ is the average phase. Interestingly, plugging Eq. 3.8 into Eq. 3.7 makes evident the mean-field character of the Kuramoto model since it turns into

$$\dot{\theta}_i = \omega_i + KNr\sin\left(\phi - \theta_i\right) \ . \tag{3.9}$$

From the latter expression, one can derive the critical value of the coupling K_c above which a synchronized state can emerge. Assuming that the distribution of the natural frequencies $g(\omega)$ is symmetric and unimodal, this value reads as

$$K_c = \frac{2}{N\pi g(0)} \ . \tag{3.10}$$

To get more insights into the physics of the Kuramoto model, Figure 3.2 represents the evolution of the order parameter as a function of the coupling along with two snapshots of the phases' configuration in both subcritical and supercritical regimes.

To better reflect the complex and local nature of interactions, the Kuramoto model was extended to accommodate complex networks [128]. In this case, the evolution of the phase of one oscillator, say i, is given by

$$\dot{\theta}_i = \omega_i + K \sum_{j=1}^{N} A_{ij} \sin(\theta_j - \theta_i) \ , \tag{3.11}$$

where $\mathbf{A}$ corresponds to the adjacency matrix governing the interactions. In its turn, the critical coupling K_c is obtained as

$$K_c = \frac{2}{\pi g(0) \Lambda_{\max}(\mathbf{A})} \ . \tag{3.12}$$

From the extension of the Kuramoto model, the interplay between the structure of complex networks and synchronization has been widely studied in the literature [128, 129]. In particular, a seminal work published in 2007 [130] revealed that the route to global synchronization involves the progressive synchronization of local clusters driven by the most connected nodes. Building upon this phenomenon, the same authors discovered the emergence of the so-called explosive synchronization in complex networks, characterized by a first order transition in the synchronization of the network, when applying the Kuramoto model to highly heterogeneous configurations [131]. Specifically, the explosive synchronization [132, 133] arises from hindering the local synchronization routes described above by imposing higher natural frequencies for the *hubs* of the network. This delays the synchronization transition which now occurs at a higher coupling strength, leading to the abrupt formation of a globally synchronized state.

3.3 Kuramoto-SIS model

The introduction of complex mechanisms in compartmental epidemic models or the design of specific configurations in opinion dynamics governed by the Kuramoto model provoke first-order (explosive) transitions in the system under study. This is a highly desired property in our model aimed at capturing the abrupt irruption of movements in social networks. Notwithstanding, although treated independently so far, information spreading and consensus dynamics are highly intertwined in real systems. In particular, note that, in presence of different ideas, imposing a complex contagion mechanism driven by group pressure is equivalent to assume that individuals adopt an idea in virtue of the consensus found among their acquaintances. Likewise, it has been reported that the active diffusion of an idea boosts reaching the degree of homophily needed for the creation of social consensus [134, 135]. Therefore, in terms of interacting spreading processes, there exists a cooperation between both consensus and information spreading, for each dynamics enhances each other.

As illustrated in the Chapter 2, the existence of interdependencies between dynamical processes [136, 137] can be captured by unifying frameworks in which the rates governing each individual dynamics depends on the evolution of the other one. Let us assume that our system is composed of N nodes which form a multiplex network of $L = 2$ layers. In multiplex networks, there exists a one-to-one correspondence between nodes in each of the layers. However, the topologies of the two layers can in general be different and are described by the adjacency matrices $A^{[1]} = \{a_{ij}^{[1]}\}$ and $A^{[2]} = \{a_{ij}^{[2]}\}$ respectively. Therefore, multiplex networks accommodate the different connection patterns existing between the same set of individuals as a function of the nature of their interactions. We denote the degree of node i in the first (second) layer by $k_i^{[1]} = \sum_{j=1}^{N} a_{ij}^{[1]}$ $(k_i^{[2]} = \sum_{j=1}^{N} a_{ij}^{[2]})$.

In general grounds, if we assume that the state of each dynamics is described by a couple of vectors, $\mathbf{x(t)} = \{x_1(t), x_2(t), \ldots, x_N(t)\} \in \Re^N$ and $\mathbf{y(t)} = \{y_1(t), y_2(t), \ldots, y_N(t)\} \in \Re^N$, the evolution of the system can be written as

$$\begin{cases} \dot{x}_i = F_{\xi_i}(\mathbf{x}, A^{[1]}) \ , \\ \dot{y}_i = G_{\eta_i}(\mathbf{y}, A^{[2]}) \ . \end{cases} \qquad i = 1, 2, \ldots N \tag{3.13}$$

Note that F_{ξ_i} and G_{η_i} are in principle arbitrary functions so the previous expression holds for every pair of dynamics running on top of a multiplex network of $L = 2$ layers. To tailor the model for intertwined dynamics, as stated above, the parameters ξ_i and η_i should depend on the dynamical state of the other process. To highlight the role of local interactions for the emergence of collective states, we assume that the parameter ξ_i (η_i) depends on the dynamical states in layer 2 (1), $y_j(t)$ $(x_j(t))$, of the j neighboring nodes of i in the first (second) layer such that $a_{ij}^{[1]} = 1$ $(a_{ij}^{[2]} = 1)$. Formally, Eq. (3.13) is therefore completed with

$$\begin{cases} \xi_i(t) = f(\{y_j(t) | a_{ij}^{[1]} = 1\}) \\ \eta_i(t) = g(\{x_j(t) | a_{ij}^{[2]} = 1\}) \end{cases} \qquad i = 1, 2, \ldots N \tag{3.14}$$

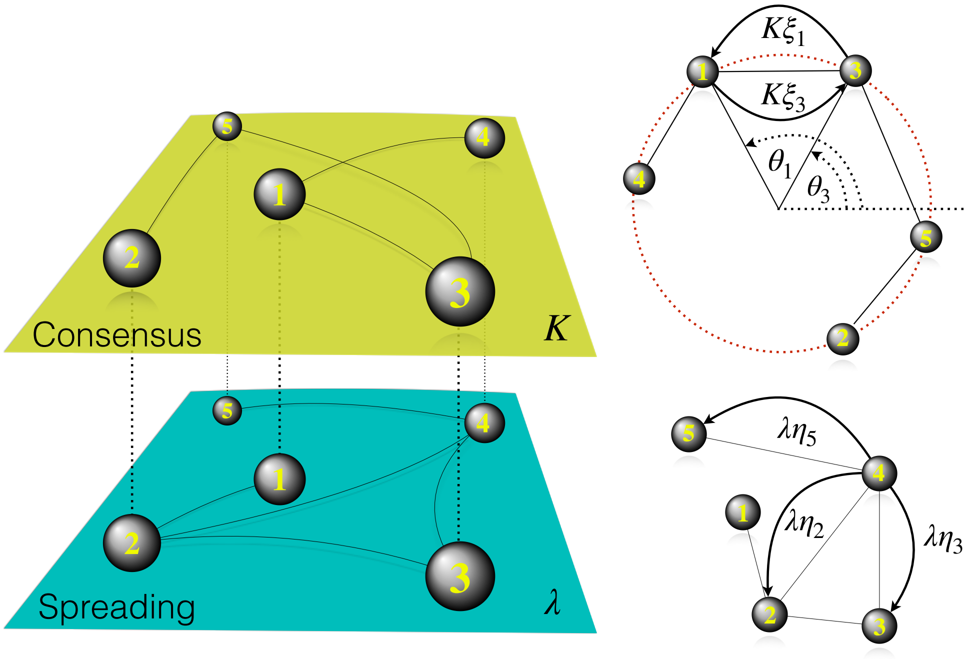


Figure 3.3: Left: Schematic representation of our model on a multiplex network with $L = 2$ layers and $N = 5$ nodes. The first (top) layer accounts for the consensus dynamics, which is modeled by a Kuramoto model as in Eq. (3.15), whereas the second (bottom) layer describes the spreading of information according to the SIS model as in Eq. (3.16). Right: The coupling strength K between opinions (top) as well as the contagion rate λ (bottom) have been modified as in Eq. (3.17) and Eq. (3.18) respectively in order to mutually couple the synchronization process to the spreading of information.

where f and g are two assigned functions.

3.3.1 Model equations

In our case, layer 1 contains the consensus dynamics governed by the Kuramoto model. Therefore, the dynamical state of the agent i at the first layer $x_i(t)$ represents the phase of an oscillator, so that $x_i = \theta_i \in (-\pi, \pi]$, whose time evolution reads as

$$\dot{\theta}_i(t) = F_{\xi_i}(\theta, A^{[1]}) = \omega_i + K\xi_i(t) \sum_{j=1}^{N} a_{ij}^{[1]} \sin\left[\theta_j(t) - \theta_i(t)\right] \ . \tag{3.15}$$

In its turn, layer 2 governs information diffusion according to a SIS model. Therefore, the state of each agent i in this layer corresponds with the probability of being a spreader so that $y_i(t) = \rho_i^I(t)$, which evolves according to

$$\dot{p}_i(t) = -\mu\rho_i^I(t) + (1-\rho_i^I(t))\left[1 - \prod_{j=1}^{N}\left(1 - a_{ij}^{[2]}\lambda\eta_i(t)\rho_j^I(t)\right)\right] \,. \tag{3.16}$$

For both individual dynamics, the control parameters K and λ are modulated by the factors ξ_i and η_i encoding their mutual influence. To reflect the social reinforcement mechanism, we assume that the willingness of individuals to change their opinions increases with the number of spreaders surrounding them. A possible choice for ξ_i capturing this fact is

$$\xi_i(t) = \frac{\sum_{j=1}^{N} a_{ij}^{[1]}\rho_j^I(t)}{k_i^{[1]}} \,. \tag{3.17}$$

Note that, for each agent i, this factor depends on the dynamical state in layer 2 of her neighbors in layer 1. In its turn, we consider that social pressure prevents individuals from spreading their ideas unless an extended consensus about them exists within their acquaintances. To reflect this fact, we choose

$$\eta_i(t) = \frac{1}{1+\exp[-\alpha(r_i(t)-r^*)]} \,, \tag{3.18}$$

where $r_i(t)$ represents the degree of local consensus (phase coherence) among the different opinions, residing in layer 1, of agent i neighbors in layer 2. Therefore,

$$r_i(t)e^{i\psi_i(t)} = \frac{1}{k_i^{[2]}}\sum_{j=1}^{N} a_{ij}^{[2]}e^{i\theta_j(t)} \tag{3.19}$$

The use of the Fermi function and the introduction of another control parameter α allows us to tune the relevance of the group pressure for information spreading. Specifically, for low α values, the local consensus does not play any role and $\eta_i(t) = 0.5$, which yields to a standard SIS model where the infectiousness is rescaled by a constant factor. In contrast, when α is high enough, a minimum consensus r^* among the neighbors of one individual is required for her to start spreading her opinion. For the sake of simplicity, hereinafter we fix $r^* = 0.5$. Figure 3.3 schematizes the model proposed for a toy multiplex configuration with $L = 2$ layers and $N = 5$ nodes.

3.3.2 Results

Let us first study the evolution of both order parameters of the coupled dynamics, i.e., the fraction of spreaders ρ and the global consensus within the population r, as the group pressure α is increased. To this aim, we start by infecting a small fraction $\rho_0 = 1\%$ of agents and by initially setting the oscillator phases θ_i at random within a range $\theta_i \in (-\pi, \pi]$. The natural frequencies

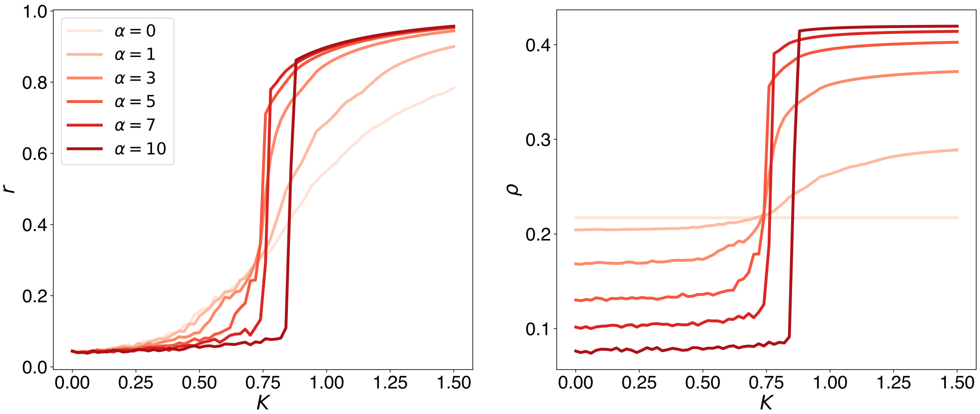


Figure 3.4: Evolution of the global consensus r and the fraction of spreaders ρ as a function of the synchronization control parameter K and the group pressure α. The underlying multiplex network is a BA-ER network with $N = 500$ agents and $\langle k \rangle = 4$ in both layers (see text for more details). The parameters of the SIS model have been fixed to $\lambda = 0.70$ and $\mu = 1$.

of oscillators $\{\omega_i\}$ are also randomly chosen so that $\omega_i \in [-0.5, 0.5]$. Let us first assume that interactions are governed by a multiplex network of $N = 500$ agents whose first layer is a BA network and the second layer contains an ER network, both with $\langle k \rangle = 4$. In what follows, we denote this configuration as BA-ER network. To focus on the interplay between the synchronization parameter and the group pressure, we fix $\mu = 1$ and $\lambda = 0.70$, which corresponds to a supercritical scenario for a hypothetical individual SIS dynamics running on top of the ER layer. The evolution of the dynamics is computed by integrating numerically Eqs.(3.15-3.16) by means of the 4th order Runge-Kutta algorithm with $dt = 0.01$. Once the dynamics is stationary, both order parameters are obtained by averaging them over a time window of $T = 500$ time steps.

Figure 3.4 shows the curves $r(K)$ and $\rho(K)$ for different values of the social pressure α. There, we observe that the cooperation between consensus and spreading phenomena leads to the emergence of first-order transitions in both the agreement on a given topic and the number of spreaders. Namely, in absence of coupling between both dynamics, the fraction of spreaders is roughly constant and the evolution of consensus is smooth as the control parameter of the synchronization dynamics increases. Interestingly, strengthening the social pressure partially hinders the spread of ideas in absence of global consensus, as reflected by the decrease in the number of spreaders for low K values. In contrast, as K increases when social pressure is at play, different clusters synchronized locally appear, favoring the emergence of spreaders as a result of the social pressure, which in turn boosts reaching a global consensus.

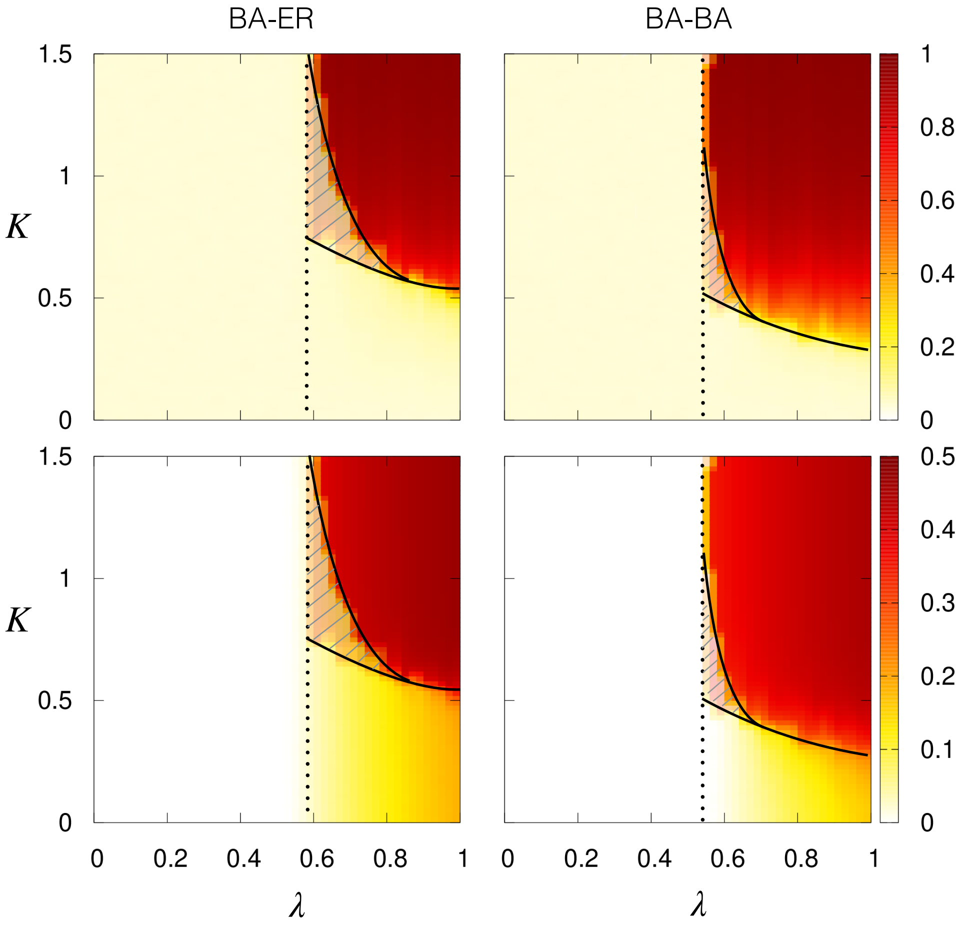


Figure 3.5: Diagrams of the average global consensus r (top) and the fraction of spreaders ρ (bottom) as a function of the virality λ and the coupling parameter K for the BA-ER (left) and BA-BA (right) multiplex networks. The color code encodes the magnitudes of the two order parameters. The striped regions in the panels highlight the parameter region (K, λ) where hysteresis cycles appear due to the coexistence of two stable solutions: total consensus-disagreement in the synchronization layer and active-inactive spreaders at the information spreading layer. The dotted line corresponds to the critical value λ_c ensuring the presence of spreaders. The black solid lines denote the critical values of the coupling K_c separating the two solutions. The rest of the parameters are the same as in Figure 3.4.

To further analyze the interplay between consensus and spreading phenomena, we fix the group pressure to $\alpha = 10$ and represent in Figure 3.5 the diagrams for the global consensus r (top row) and the total fraction of spreaders ρ (bottom row) as a function of (λ, K) for two different multiplex configurations, a BA-ER network (left panels) and a BA-BA network (right panels). To find numerically the hysteresis cycles characteristic of first-order transitions in the phase diagrams, we integrate the equations adiabatically following forward and backward continuation schemes in the parameter K for each λ value. At first sight, we observe that below a given threshold λ_c the only stable state is the absence of both consensus and spreaders within the population. Above this threshold, the ideas are viral enough to ensure the presence of spreaders propagating them regardless of the consensus, whose emergence is governed by K. In particular, there exists another threshold $K_c(\lambda)$ determining the sudden emergence of global consensus. Therefore, three qualitatively different stable configurations appear in the phase diagrams:

1. $\lambda < \lambda_c$: Absence of spreaders and lack of global consensus within the population.
2. $\lambda \geq \lambda_c$, $K < K_c(\lambda)$: Presence of spreaders when consensus has not been reached yet.
3. $\lambda \geq \lambda_c$, $K \geq K_c(\lambda)$: Macroscopic set of spreaders and global consensus within the population.

Interestingly, the transition between the latter two states is highly dependent on the λ value, which encodes the virality of the idea, as illustrated in Figure 3.6 for the BA-ER mutliplex configuration. In particular, for very viral ideas ($\lambda \simeq 1$), the stable solution in the absence of consensus is already composed of a macroscopic set of spreaders and the system undergoes a second-order transition when K is increased in both order parameters. In contrast, for virality values closer to the threshold, the initial small set of spreaders experiences a first order transition triggered by consensus dynamics. Note that the latter behavior captures the sudden irruption of social movements whereas the former one could capture the sustained presence and spread of *fake news* in social media [138].

Regarding the role of the contact networks, there does not seem to be any remarkable differences between the diagrams of both BA-ER and BA-BA networks. This constitutes a surprising result since, in the absence of consensus, the spreading phenomena running on top of the BA network should be favored by the presence of highly connected nodes. In fact, assuming that $\eta_i = \eta \; \forall i$, one arrives to the following estimation of the ratio between both spreading thresholds

$$\frac{\lambda_c^{\mathrm{BA}}}{\lambda_c^{\mathrm{ER}}} = \frac{\Lambda_{\max}(\mathbf{A}^{\mathbf{ER}})}{\Lambda_{\max}(\mathbf{A}^{\mathbf{BA}})} \simeq 0.11 \; , \tag{3.20}$$

which clearly differs from the results obtained. Therefore, the previous assumption does not hold in this system and the observation of similar λ_c values for

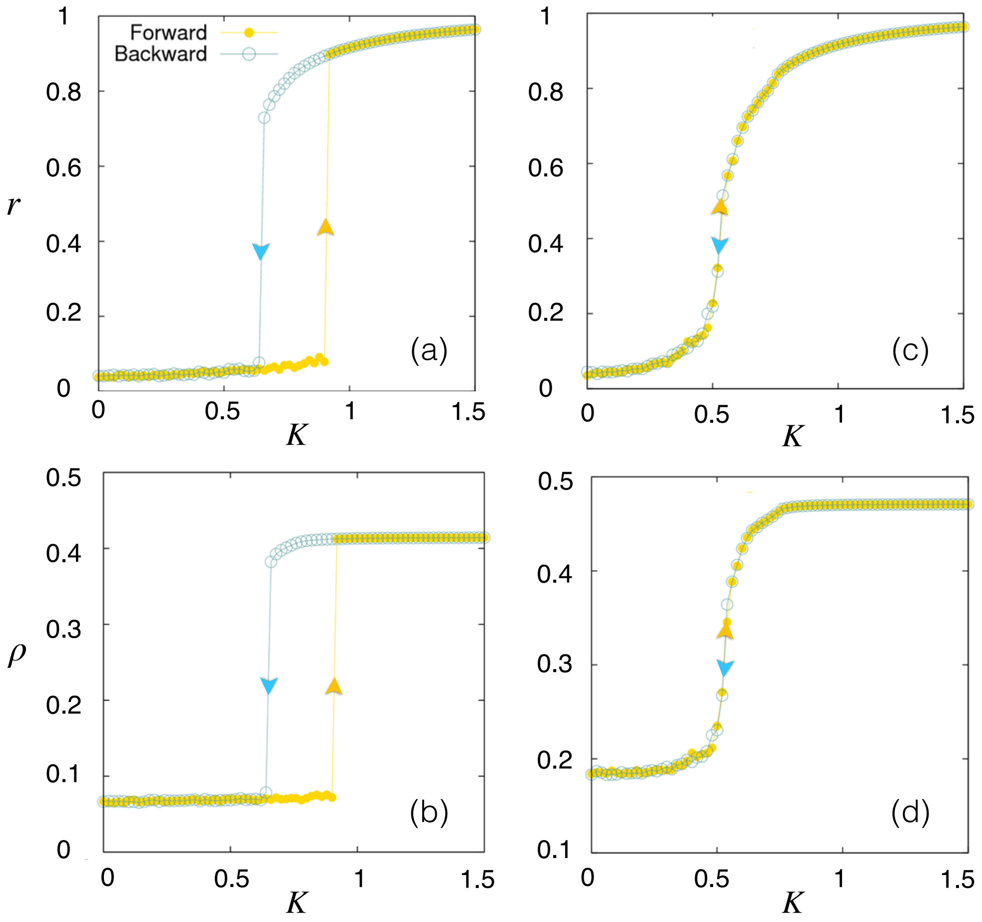


Figure 3.6: Average global consensus r (Top) and fraction of information spreaders ρ (bottom) as a function of the coupling constant K for the BA-ER multiplex configuration. These order parameters have been computed adiabatically by increasing the value of the coupling constant K (Forward) from $K = 0$ or by decreasing it (Backward) from $K = 1.5$. The contagion rate values used are for (a) and (b) $\lambda = 0.70$ and for (c) and (d) $\lambda = 1$. The rest of the model parameters are the same as in previous figures.

both configurations must be attributed to local effects related to the coupling between consensus and information spreading. To confirm this statement, we now represent in Figure 3.7, the evolution of the fraction of spreaders when the consensus mechanism is not working ($K = 0$) as λ increases for different values of the social pressure α. Interestingly, when both dynamics are weakly coupled, the expected behavior is retrieved and the heterogeneity of the BA network favors the emergence of spreaders. Nonetheless, for high α values, the trend is reversed and the ER layer diffuses the ideas more efficiently.

To understand the roots of this phenomenon, we must realize that increas-

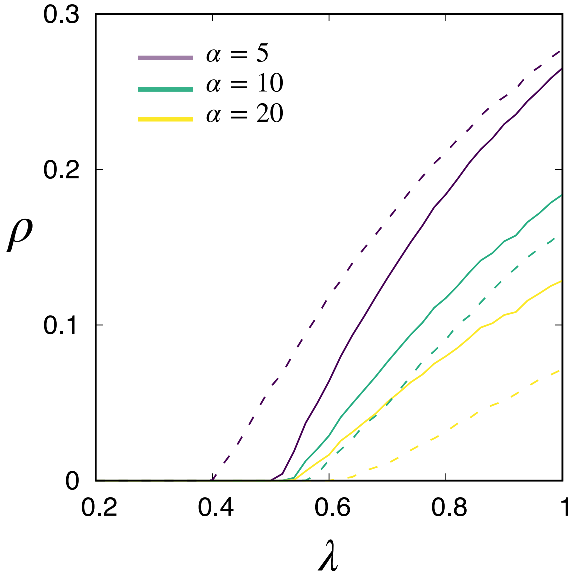


Figure 3.7: Fraction of spreaders in the non-synchronized regime ($K = 0$) as a function of the virality λ for both BA-ER (solid lines) and BA-BA configurations (dashed lines). The line color denotes the value of the social pressure α.

ing the degree of one node has two opposite effects whose trade-off, governed by α, determines the emergence of spreaders across the network. On the one hand, having a larger number of acquaintances favors the faster diffusion of the information by the spreaders. On the other hand, when synchronization is not working, the probability for a highly connected ignorant adopting an idea decreases, for reaching a significant local consensus becomes more difficult given the random nature of the values for the initial phases of the oscillators and their natural frequencies. Mathematically, the local phase coherence in absence of consensus fulfills

$$\frac{dr_i}{dk_i} < 0 \ , \tag{3.21}$$

which according to Eq. 3.18 implies that

$$\frac{d\eta_i}{dk_i} < 0 \ . \tag{3.22}$$

To get a more systematic picture of the competition between such opposite effects, we now use the BA-ER network and numerically study in Fig. 3.8a the evolution of the threshold λ_c in the non-synchronized state as a function of the social pressure α and the average degree $\langle k \rangle$ of the ER layer. As anticipated above, the onset of an idea crucially depends on the interplay between both factors.

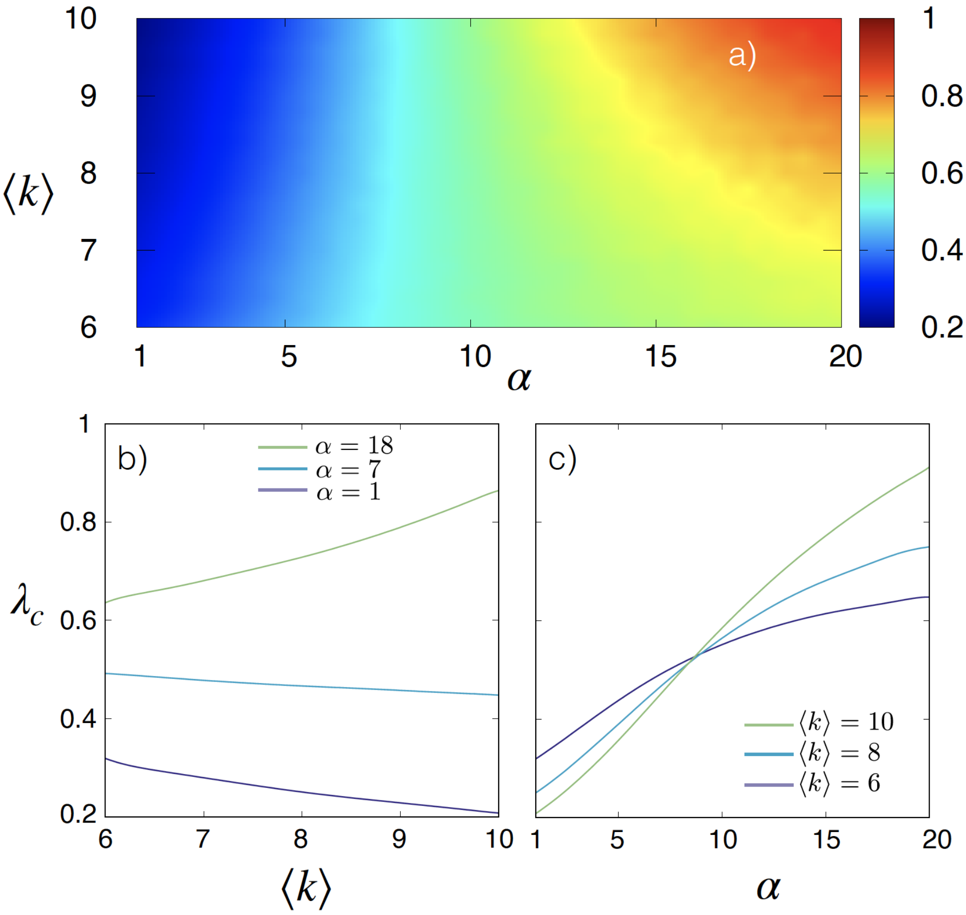


Figure 3.8: (a) Virality threshold λ_c (color code) as a function of the social pressure α and the average degree $\langle k \rangle$ of the diffusion layer for the BA-ER configuration. In panels (b) and (c) we show the virality threshold λ_c as a function of the average degree $\langle k \rangle$ and the social pressure α respectively. In both cases we show these functions for several values of the social pressure α (b) and the average degree $\langle k \rangle$ (c).

Fig. 3.8b confirms the central role of the social pressure when increasing the average degree of the population. Three different behaviors as the average degree increases are observed: a reduction of the threshold for low α values (higher spreading capability of the spreaders wins over the effect of the absence of local consensus), a roughly constant threshold value for intermediate α values (both effects are equally weighted) or an increase of the threshold when α is high (the social pressure dominates over the higher spreading capability). Finally, to round off this analysis, we represent the dependence of λ_c on α for three different configurations of the ER layer in Fig. 3.8c. As expected, the threshold always increases with the social pressure, hindering the spread of ideas in absence of global consensus. Furthermore, this effect is more pronounced as the connectivity of the underlying spreading network increases, for local consensus among the neighbors of the most connected nodes is hardly reachable.

3.4 Modeling the evolution of corruption

Upon this section, we have focused on studying the spread of ideas or opinions across a given set of agents by using compartmental models adapted to capture the social nature of these spreading units. Nonetheless, the study of social dynamics under the lens of complex systems goes beyond the characterization of the diffusion of ideas and tackles a wide diversity of problems ranging from the origins of language [139–141] or cultural traits [142, 143] to the roots of success in different disciplines [14, 15, 144]. In this sense, one of the most active fields in sociology is that of disentangling the mechanisms explaining the different behaviors or conducts observed in the society [145–147].

The characterization of human behavior has been typically done by proposing theoretical frameworks based on evolutionary game theory [148–151]. Under this approach, each conduct is represented with a strategy and a matrix of payoffs determines the cost of the interaction between two agents according to their strategies. Moreover, an update rule is introduced, allowing agents to change their strategy when considering that it does not constitute the optimal one to maximize their benefit. Apart from socially-inspired problems, the characteristics of evolutionary game theory have prompted its use in different disciplines such as biology [152, 153] or economy [154]. Within all these possible applications, the study of corruption [155, 156] has drawn the attention of many scientists over the last decades given its ubiquity in many political, economic and social contexts.

The greediness associated with corrupt behaviors and the trade-off existing between the benefits of this strategy and the punishment risk turns game theory into the natural framework to characterize the evolution of this phenomenon [157–164]. Notwithstanding, following game theoretical approaches usually lead to a huge parameters' space which hinders getting the key aspects determining the viability of corruption in the system. In addition to the large parameters' space, the rule governing the evolution of strategies supposes a hyper-parameter of the model, for its choice may dramatically affect the outcome of the dynamics [165]. To overcome the limitations of game theoretical approaches, we propose a minimal compartmental model which incorporates some of the factors inherent to the evolution of the corruption while not losing the capability of obtaining analytical, general results provided by the basic compartmental models explained in Chapter 2.

The existence of a social norm or convention among a set of agents leads naturally to the division of the population into two disjoint states: the Honest compartment, which comprises the individuals accomplishing the social norm and the Corrupt compartment gathering those violating it. Moreover, the punishment imposed by the society to corrupt behaviors usually represent some sort of freedom privation or isolation, (e.g. expulsion from society, exile, and prison), which differentiates those punished from the corrupt and honest populations. To capture this fact, we must introduce another extra compartment denoted by Ostracism. Once defined the three possible states (Honest, Corrupt and Ostracism), we must propose sensible flows connecting them to reflect the inherent features to the evolution of corruption:

- First, honest agents can be tempted by their corrupt counterparts to mimic their behavior in light of the potential benefits of violating the social norm. The diffusion of corruption via interactions between corrupt and honest agents resembles the nature of the contagions occurring when modeling epidemics, which allows conceptualizing corruption as a spreading unit diffusing across the population. Mathematically, we assume that a honest individual has a probability α of becoming corrupt when interacting with a corrupt peer.
- Second, the corrupt compartment has two escape routes: *i)* the repentance before being punished or *ii)* the punishment. Repentance is the spontaneous decision of a corrupt individual to stop violating the social norm because of the risk of being punished. We assume that this flow, denoted in the following by *warning-to-wrongdoers* (WTW) is quantified by the fraction of the population in the Ostracism. In contrast, corrupt population does not decide to punish themselves, but instead punishment involves the participation of the honest population which wants the social norm to be restored. This way, punishment is driven by pairwise delation events from honest to corrupt individuals. To reflect this fact, for each interaction with a honest agent, we introduce a given probability β for a corrupt individual entering into the Ostracism compartment.
- Finally, we assume that punished agents can be reinserted in the society as honest individuals at a given rate r.

As a result, the HCO model is a 3-state model with four possible flows connecting them, which are schematized in Fig. 3.9. Conceptually, the model resembles the Susceptible-Infected-Recovered-Susceptible (SIRS) where the Susceptible, Infected and Recovered states are identified with the Honest, Corrupt and Ostracism compartments respectively. Furthermore, the spontaneous

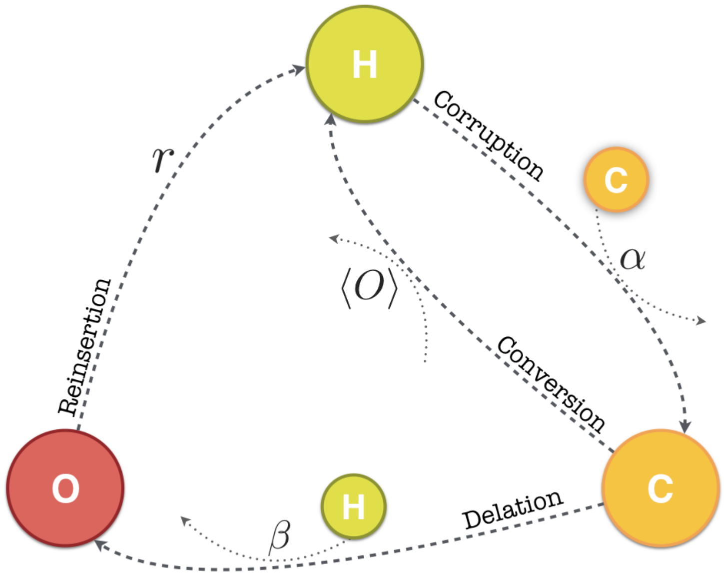


Figure 3.9: *Chart flow* representation of the model. Four flows between population compartments are possible. The flow $O \rightarrow H$ (reinsertion) occurs at a constant rate r. The flow $C \rightarrow H$ (conversion) is fueled by the perception of the delation risk, that we simply quantify by $\langle O\rangle$. However, only pairwise social contacts $C - H$ determine the other two flows, say corruption flow at an infection rate f_α and delation flow at a delation rate f_β.

recovery rate of the SIRS model is changed by the pairwise delation events whereas an additional flow, the WTW, not contemplated in the SIRS model is added in the HCO model to reflect corrupts' risk perception.

3.4.1 Mean-field equations

Let us first characterize the mean-field equations of the model. The evolution of the dynamics is completely characterized by the equations governing the fraction of the population inside the Honest, Corrupt and Ostracism compartments, denoted in the following by ρ^H, ρ^C and ρ^O respectively. From the model assumptions, the time evolution of the latter variables reads as

$$\rho^H(t+1) = (1 - f^\alpha(t))\,\rho^H(t) + \rho^C(t)\rho^O(t) + r\rho^O(t)\ , \tag{3.23}$$

$$\rho^C(t+1) = \left(1 - \rho^O(t)\right)\left(1 - f^\beta(t)\right)\rho^C(t) + f^\alpha(t)\rho^H(t)\ , \tag{3.24}$$

$$\rho^O(t+1) = (1 - r)\,\rho^O(t) + \left(1 - \rho^O(t)\right) f^\beta(t)\rho^C(t)\ , \tag{3.25}$$

where the terms containing the interaction among corrupt and punished population, $\rho_c(t)\rho_o(t)$, encode the WTW effect and $f_\alpha(t)$ and $(f_\beta(t))$ corresponds

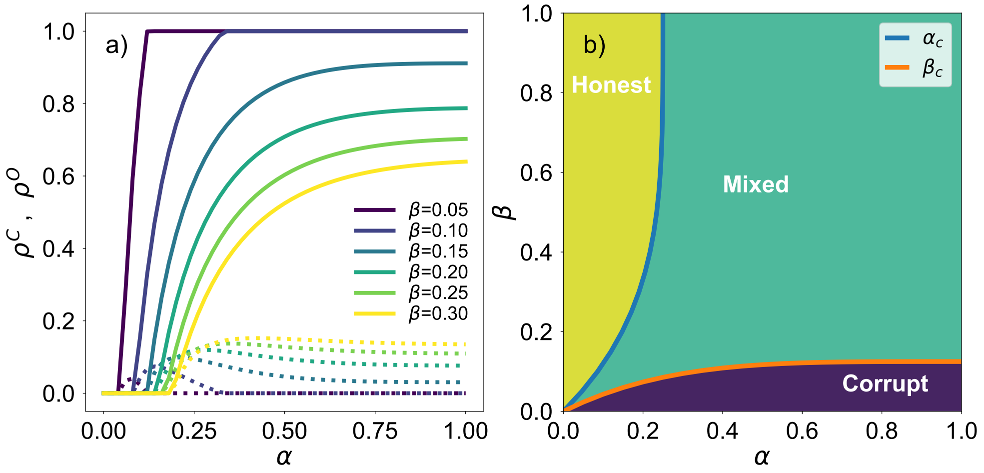


Figure 3.10: Panel a): Evolution of the steady state of the dynamics as a function of the corruption probability α and the delation probability β (color code). The fractions of the population in the Corrupt and the Ostracism comparment, denoted by $\rho^C$and ρ^O, are represented with solid and dotted lines respectively. Panel b) Phase diagram of the system with three different equilibria: the honest and corrupt populations and a mixed state in which agents in the three compartments coexist.

to the probability for a honest (corrupt) individual being corrupted (delated) at time t. Considering that individuals make k contacts per time step, both probabilities are given by

$$f_\alpha(t) = 1 - \left(1 - \alpha\rho^C(t)\right)^k \ , \tag{3.26}$$

$$f_\beta(t) = 1 - \left(1 - \beta\rho^H(t)\right)^k \ . \tag{3.27}$$

Eqs. (3.23)-(3.27) allows tracking the evolution of corruption in the system. To characterize the equilibria of the dynamics, we fix $k = 4$ and represent in Fig 3.10a the occupation of the Corrupt and Ostracism compartments in the steady state as a function of the model parameters. There we observe that, for low α values, the stable configuration corresponds to a honest population. In its turn, for high α values, the equilibrium depends on the delation mechanism: when the delation mechanism works efficiently , i.e. high β values, the stable configuration is a mixed state in which the population is splitted into the three available states. In contrast, when delation events hardly occur, i.e. low β values, corruption is able to invade the entire system; hence, the stable configuration is exclusively composed of corrupt agents. It is worth remarking that an endemic equilibrium characterized by a totally infected population is not feasible in classical epidemiological compartmental models. Therefore, the emergence of this stable configuration obeys the social flavor of the flows introduced in the HCO model to accommodate the diffusion of corruption.

Let us now characterize the transcritical bifurcations changing the stability of the three aforementioned equilibria. First, we assume that the disease has reached the steady state so no fluctuations are observed in the order parameters. Mathematically, this implies that $\rho^m(t+1) = \rho^m(t) = \rho^m$ with $m = \{H, C, O\}$. Introducing this condition in Eqs. (3.24)-(3.25) turns them into

$$0 = \left(\rho^O\left(1 - f^\beta\right) - f^\beta\right)\rho^C + f^\alpha(1 - \rho^O - \rho^C) \ , \tag{3.28}$$

$$0 = -r\rho^O + \left(1 - \rho^O\right) f^\beta \rho^C \ , \tag{3.29}$$

where we have used the normalization condition $\rho^H + \rho^C + \rho^O = 1$.

We first derive the corruption value α_c triggering the emergence of corrupt agents in a population of honest agents; we will call it *corruption threshold.* For this purpose, we assume that $\rho^C = \epsilon^C \ll 1$ and that $\rho^O = \epsilon^O \ll 1$. The latter assumption allows linearizing the equations governing the steady state of the system. In particular, Eqs. (3.26)-(3.27) turn into

$$f_\alpha \simeq k\alpha\epsilon^C \ , \tag{3.30}$$

$$f_\beta \simeq 1 - (1-\beta)^k \ . \tag{3.31}$$

Plugging the former expression into Eq. (3.28) and neglecting $\mathcal{O}(\epsilon^2)$ terms, we obtain

$$\epsilon^C\left(\alpha k - \left(1 - (1-\beta)^k\right)\right) = 0 \ . \tag{3.32}$$

Therefore, the value α_c corresponding to the onset of corruption is given by

$$\alpha_c = \frac{1 - (1-\beta)^k}{k} \ . \tag{3.33}$$

Note that the latter expression qualitatively resembles the epidemic threshold obtained for the mean-field formulation of the SIS model according to Eq. (2.8), considering the numerator as an effective recovery probability μ^{eff}. To characterize the onset of honesty in a corrupt population, we assume $\rho^H = \epsilon^H$ and $\rho^O = \epsilon^O$, which yields $\rho^C = 1 - \epsilon^O - \epsilon^H$. Under these assumptions, we have that

$$f_\alpha \simeq 1 - (1-\alpha)^k \ , \tag{3.34}$$

$$f_\beta \simeq k\beta\epsilon^H \ , \tag{3.35}$$

which turn Eq. (3.29) into

$$\epsilon^O = \frac{k\beta\epsilon^H}{r} \ . \tag{3.36}$$

Substituting this value into Eq. (3.28) and following the same rationale as in the former transition, the value of the delation probability β_c triggering the emergence of honesty, hereinafter referred to as *delation threshold* is given by

$$\beta_c = \frac{r}{r+1}\frac{1-(1-\alpha)^k}{k} \ . \tag{3.37}$$

We observe that both thresholds are qualitatively similar to one another, but the β_c value is modulated by a prefactor depending on the reinsertion rate r which accounts for the impact of the WTW mechanism. In this way, as the punishment times increase (r decreases), the area of the parameters' space corresponding to the corrupt equilibria shrinks, given the higher relevance of the repentance of corrupt individuals. Fig 3.10b shows the phase diagram of the system when fixing $r = 1$ and $k = 4$, highlighting the three stable equilibria and the boundaries separating them.

3.4.2 HCO in complex networks

So far, we have introduced a mean-field theory to explore the qualitative changes in the phase diagrams of epidemic compartmental models induced by the social flavor of the processes involved in the diffusion of corruption. To analyze the effect of network structure in this problem, we now abandon the mean-field assumption and propose a formalism based on the MMCA introduced in Chapter 2.

Recall that the MMCA equations for a given compartmental model describe the probability for each individual i being in each of the available compartments m, denoted in the following by ρ_i^m. Consequently, in the HCO model, the dynamics is completely characterized by a system of $3 \times N$ coupled equations governing the occupation of the Honest, Corrupt and Ostracism states, being N the size (number of agents) of the system. Considering that $\mathbf{A}$ corresponds to the adjacency matrix of the network connecting the population, these equations are given by

$$\rho_i^H(t+1) = \left(1 - f_i^\alpha(t)\right)\rho_i^H(t) + \rho_i^C(t)\langle\rho^O(t)\rangle + r\rho_i^O(t) \ , \tag{3.38}$$

$$\rho_i^C(t+1) = \left(1 - \langle\rho^O(t)\rangle\right)\left(1 - f_i^\beta(t)\right)\rho_i^C(t) + f_i^\alpha(t)\rho_i^H(t) \ , \tag{3.39}$$

$$\rho_i^O(t+1) = (1-r)\,\rho_i^O(t) + \left(1 - \langle\rho^O(t)\rangle\right) f_i^\beta(t)\rho^C(t) \ , \tag{3.40}$$

with

$$f_i^\alpha(t) = 1 - \prod_{j=1}^{N}\left(1 - \alpha A_{ij}\rho_j^C(t)\right) \ , \tag{3.41}$$

$$f_i^\beta(t) = 1 - \prod_{j=1}^{N}\left(1 - \beta A_{ij}\rho_j^H(t)\right) \ . \tag{3.42}$$

It is worth remarking that corruption/delation processes are considered to be local and driven by pairwise interactions whereas the WTW effect depends on global information, as reflected by the terms containing $\langle \rho^O(t) \rangle$. This choice is based on the fact that, in real setups, information on the occupation of the Ostracism compartment, e.g. the number of imprisoned individuals, can be obtained via social media or newspapers.

Let us quantify the impact of the network structure on the evolution of corruption. With this aim, we study how the panels represented in Fig. 3.10 are altered by the network governing the interactions among agents. Specifically, in Fig. 3.11a we represent the fraction of corrupt population in the equilibrium state for ER and BA contact networks, both with $N = 1000$ agents and $\langle k \rangle = 4$. There we can observe that, for high β values, we recover the expected behavior in spreading phenomena where heterogeneous networks (BA) prompt the emergence of infective individuals (corrupt in a honest population) earlier than homogeneous networks (ER) due to the presence of highly connected nodes; this dynamical advantage decreases for low β values. In contrast, beyond the corruption threshold, the fraction of corrupt population is systematically larger in the ER than in the BA network, to the extent that the corrupt equilibrium is destabilized earlier for the latter.

To shed more light on this behavior, we now derive the analytical expressions for both thresholds in a networked configuration. Let us start with the corruption threshold α_c. To derive its value, we repeat the same procedure explained for the mean-field approach but now accounting for the statistical independence of the nodes. Therefore, we assume that

$$\rho_i^C(t+1) = \rho_i^C(t) = \epsilon_i^C \ll 1 \ , \tag{3.43}$$

$$\rho_i^O(t+1) = \rho_i^O(t) = \epsilon_i^O \ll 1 \ , \tag{3.44}$$

$$\tag{3.45}$$

which turn Eq. (3.41)-(3.42) into

$$f_i^\alpha = 1 - \sum_{j=1}^{N} \alpha A_{ij} \epsilon_j^C \ , \tag{3.46}$$

$$f_i^\beta = 1 - (1-\beta)^{k_i} \ , \tag{3.47}$$

being k_i the degree of node i. Introducing both expressions into Eq. 3.39 and keeping only linear terms yields:

$$\epsilon_i^C = \sum_{j=1}^{N} M_{ij} \epsilon_j^C \ , \tag{3.48}$$

with

$$M_{ij} = \frac{A_{ij}}{1-(1-\beta)^{k_i}} \ . \tag{3.49}$$

Eq. (3.48) represents an eigenvalue problem. Accordingly, the corruption threshold α_c is given by

$$\alpha_c = \frac{1}{\Lambda_{\max}(\mathbf{M})} \ . \tag{3.50}$$

Let us analyze the elements of the matrix $\mathbf{M}$ as a function of the delation probability β. First, for high β values, we have that $1-(1-\beta)^{k_i} \simeq 1$. Therefore, the matrix elements read as

$$M_{ij} = A_{ij} \tag{3.51}$$

and the threshold corresponds to

$$\alpha_c \simeq \frac{1}{\Lambda_{\max}(\mathbf{A})} \ , \tag{3.52}$$

which is the epidemic threshold of a SIS model under the MMCA approach with a recovery probability $\mu = 1$. This parallelism explains why the HCO model retrieves the observed behavior in epidemic compartmental models when the delation probability β is large enough.

In contrast, for small β values, one can assume that $1-(1-\beta)^{k_i} \simeq k_i\beta$, yielding

$$M_{ij} = \frac{A_{ij}}{\beta k_i} \ . \tag{3.53}$$

In this scenario, the corruption threshold is straightforwardly computed by recalling the Perron-Frobenius theorem, which states that for every positive matrix $\mathbf{M}$ the largest eigenvalue $\Lambda_{\max}(\mathbf{M})$ fulfills

$$\min \sum_{j=1}^{N} M_{ij} \leq \Lambda_{\max}(\mathbf{M}) \leq \max \sum_{j=1}^{N} M_{ij} \ . \tag{3.54}$$

Given the fact that $\sum\limits_{j=1}^{N} M_{ij} = 1/\beta$ for every row i, the value for the threshold is

$$\alpha_c = \beta \ . \tag{3.55}$$

The previous expression does not depend on the underlying network and makes the antagonistic nature of corruption and delation processes evident. Namely, when β is small enough so no delation events are ensured, the stability of the honest population is directly determined by whether delation or corruption spread faster in the pairwise interactions involving honest and corrupt population.

Let us study the value of the delation threshold β_c. In this case, we must assume:

$$\rho_i^H(t+1) = \rho_i^H(t) = \epsilon_i^H \ll 1 \ , \tag{3.56}$$
$$\rho_i^O(t+1) = \rho_i^O(t) = \epsilon_i^O \ll 1 \ , \tag{3.57}$$
$$\tag{3.58}$$

which leads to

$$f_i^\alpha = 1 - (1-\alpha)^{k_i} \ , \tag{3.59}$$
$$f_i^\beta = \sum_{j=1}^N \beta A_{ij} \epsilon_j^H \ . \tag{3.60}$$

Introducing both values into Eqs. (3.39)-(3.40) and keeping linear terms, we obtain

$$r\epsilon_i^O = \beta \sum_{j=1}^N A_{ij}\epsilon_j^H \ , \tag{3.61}$$
$$\left[1-(1-\alpha)^{k_i}\right]\epsilon_i^H = r\epsilon_i^O + \frac{1}{N}\sum_{j=1}^N \epsilon_j^O \ . \tag{3.62}$$

Merging both expressions, after simple algebraic operations, we obtain:

$$\epsilon_i^H = \beta \sum_{j=1}^N \tilde{M}_{ij}\epsilon_j^H \ , \tag{3.63}$$

with

$$\tilde{M}_{ij} = \frac{\dfrac{k_j}{Nr} + A_{ij}}{1-(1-\alpha)^{k_i}} \ . \tag{3.64}$$

Therefore, the delation threshold β_c is given by:

$$\beta_c = \frac{1}{\Lambda_{\max}(\tilde{\mathbf{M}})} \ . \tag{3.65}$$

To further characterize the impact of the contact network, we represent in Fig. 3.11 the phase diagrams for both the ER and the BA networks here considered. As anticipated above, for small values of both the corruption and delation probabilities, the phase diagrams are independent on the underlying network. This phenomenon occurs because the highly connected nodes in heterogeneous networks are more capable of spreading corruption but are also more exposed to the delation by their honest counterparts. The competition

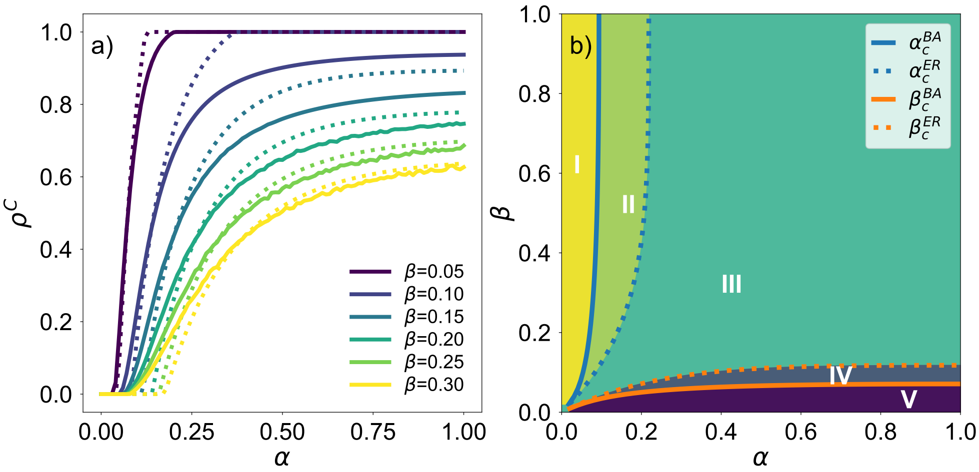


Figure 3.11: Panel a): Evolution of the fraction of corrupt population in the equilibrium ρ^C as a function of the corruption probability α and the delation probability β (color code). The underlying contact structure is an ER network (dotted lines) or a BA network (solid lines), both with $N = 1000$ nodes and $\langle k \rangle = 4$. Panel b): Phase diagram of the system for the two networks with five different regions according to the stable equilibria in each network. Specifically, denoting by $M + N$ the scenario in which the stable equilibrium for the ER network is M and the one for the BA network is N we observe: (I) Honest + Honest, (II) Mixed + Honest, (III) Mixed + Mixed, (IV) Mixed + Corrupt and (V) Corrupt + Corrupt. Finally, the lines show the corruption threshold α_c (blue) and the delation threshold β_c (orange) for both the ER (dotted lines) and the BA (dashed lines) networks.

between both processes loses relevance when either delation ($\beta \simeq 1$) or corruption ($\alpha \simeq 1$) are ensured. In this scenario, the equilibrium of the system, either an entirely honest or corrupt population, is more easily destabilized in heterogeneous networks, provided the higher capacity of the hubs to propagate either corrupt or honest perturbations.

3.4.3 Fear-induced explosive transitions

The interplay between individual awareness and spreading phenomena has been widely studied in the literature, specially in the context of epidemics [166–170]. In most of these works, the individual awareness raised by the access to information on the global advance of the outbreak leads the population to take preemption measures which reduce the probability of contracting the disease and hence mitigate its impact. In contrast, in the HCO model, the mechanism of individual awareness does not affect honest population but instead is introduced via the WTW effect, which reduces the impact of corruption because of the fear of the corrupt population to be punished.

Here we aim at quantifying the effect of introducing individual awareness for the honest population. On the one hand, note that, unlike epidemics, becoming corrupt can be beneficial or tempting for honest agents, given the incentives to follow this strategy. Accordingly, we assume that the awareness mechanism does not alter the corruption probability α. On the other hand, different studies can be found in the literature exploring the relation between corruption dynamics and coercion [171–173]. In this sense, in the HCO model, coercion from corrupt individuals to honest population can be conceptualized as an awareness mechanism preventing the latter from delating the former. To accommodate this effect, we introduce a new parameter, the *social intimidation* γ and define a new delation probability $\tilde{\beta}$ which is given by

$$\tilde{\beta}(t) = \beta\left(1 - \gamma\langle\rho^C(t)\rangle\right) \ , \tag{3.66}$$

which modifies Eq. (3.42), which now reads as

$$f_i^\beta(t) = 1 - \prod_{j=1}^{N}\left(1 - \tilde{\beta}A_{ij}\rho_j^H(t)\right) \ . \tag{3.67}$$

Let us first study how the social intimidation mechanism shapes the evolution of corruption. We focus on the ER network but the results here described also hold for the BA configuration. In Fig. 3.12a we represent the evolution of ρ^C in the equilibrium state by keeping fixed the delation probability β and varying both the social intimidation γ and the corruption probability α. We computed the curves adiabatically, following forward (backward) continuation schemes, where the initial configuration used for $\alpha + \delta\alpha$ ($\alpha - \delta\alpha$) corresponds to the equilibrium state for α with a negligible noise introduced to avoid absorbing states.

The first relevant result is that the corruption threshold α_c remains unaltered by the new mechanism. Mathematically, this result arises because, close to the corruption threshold, the contribution of the social intimidation in Eq. (3.66) is negligible so that

$$\tilde{\beta} \simeq \beta \ . \tag{3.68}$$

In practical terms, with very low corruption levels, the coercion mechanism is not strong enough to inhibit the honest population activity. More strikingly, we find that the smooth second-order transition from the mixed to the corrupt state turns into an abrupt first-order one as social intimidation is strengthened. In particular, we observe how coercion gives rise to the emergence of hysteresis cycles involving two stable states whose composition depends on the social intimidation γ.

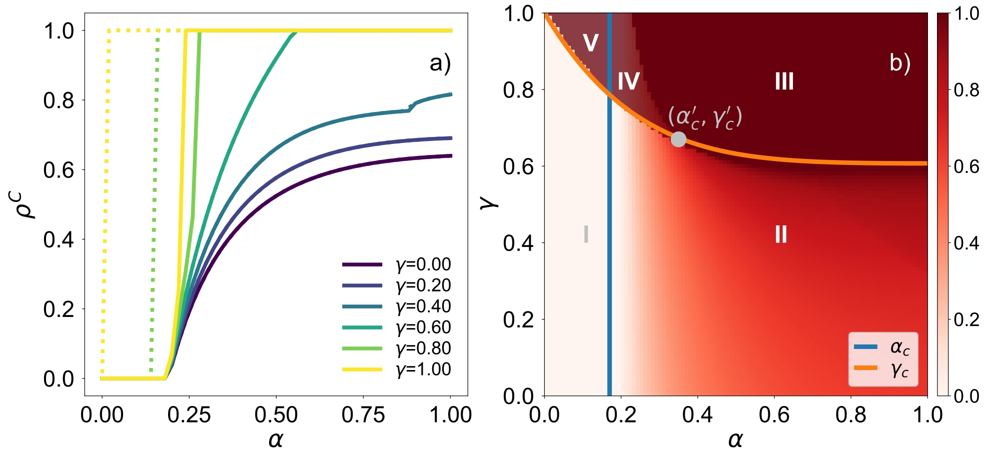


Figure 3.12: Panel a): Evolution of the fraction of corrupt population in the equilibrium ρ^C as a function of the corruption probability α and the social intimidation γ (color code). The underlying contact structure is an ER network with $N = 1000$ nodes and $\langle k \rangle = 4$. The equations of the formalism are iterated adiabatically, following forward (solid lines) and backward (dotted lines) continuation schemes (see text for details). Panel b): Phase diagram of the system in the plane (α, γ). The color code represents the fraction of the population in the corrupt state at the equilibrium. We distinguish five regions as a function of the stable equilibria. Regions I, II, III correspond to the honest, mixed and corrupt states observed in the original HCO model. The shadowed areas highlight the presence of bistable solutions, which consist of the mixed and corrupt equilibria in region IV and the honest and corrupt equilibria in region V. The critical point (α'_c, γ'_c) is obtained numerically whereas the corruption threshold α_c and the intimidation threshold γ are computed by Eq. (3.52) and Eq. (3.70) respectively. In both panels, the rest of the parameters are fixed to $(r, \beta) = (1, 0.3)$.

To have a more complete characterization of the dynamics we represent the phase diagram in Fig. 3.12b. There we distinguish five different equilibria: (I) honest population, (II) mixed state, (III) corrupt population, (IV) bistability between corrupt and mixed states and (V) bistability with corrupt and honest states. Regions III, IV and V are separated from the rest of the phases by the intimidation threshold γ_c, defined as the minimum social intimidation needed to observe a corrupt state. The derivation of this value is straightforward, for close to the corrupt state, we have that $\langle \rho^C \rangle \simeq 1$, implying that

$$\tilde{\beta} = \beta(1 - \gamma) \ . \tag{3.69}$$

We can substitute this value into Eq. (3.65), determining the stability of the corrupt state, and rearrange the terms to obtain

$$\gamma_c = 1 - \frac{1}{\beta \Lambda_{\max}(\tilde{\mathbf{M}})} \ . \tag{3.70}$$

The common point between regions II, III and IV, denoted by (α_c', γ_c') constitutes the critical point of the dynamics, separating the two different types of bifurcations—either smooth or explosive transitions [132, 133, 174]—leading to the corrupt state. Conceptually, we observe that social intimidation, which gains relevance in highly corrupted societies, allows corrupt configurations to persist in the population because of the lack of delation events. Interestingly, this phenomenon also takes places when corruption is not enough to destabilize a honest society, which corresponds to the region V described above.

3.5 Conclusions

The social nature inherent to the fostering of ideas, opinions, innovations or cultural traits demands the sophistication of the compartmental models explained in Chapter 2. Unlike contagious diseases, the propagation of an idea or the adoption of a social conduct does not only depend on intrinsic probabilities, encoding its virality, but is also highly influenced by other environmental features such as its penetration in the society or its reliability among the population. In particular, in this chapter we have focused on unraveling possible explanations for two different social dynamics: the recent explosive irruption of some movements in social networks such as the Arab Spring or the Black Lives Matter movements and the evolution of corruption in the society.

Concerning the irruption of social movements, we have revisited how compartmental models usually applied in mathematical epidemiology have been modified to accommodate the complex contagion mechanisms driving the spread of ideas across social platforms. Likewise, ideas or opinions are not unique or immutable but are usually part of a complex ecosystem in which different points of views about a topic coexist and interact, leading to dynamic changes in the opinions' landscape. In this chapter, we have also explained the most paradigmatic frameworks addressing the evolution of opinion dynamics, putting our focus on the Kuramoto model, provided its relevance and ubiquity in network science to tackle other problems such as synchronization dynamics.

Following the spirit of this thesis, we have proposed the KM-SIS model, which consists in an unifying formalism integrating both consensus and spreading dynamics on top of multiplex networks, and studied the physics arising from the interdependencies existing among both processes. In this sense, our complex contagion mechanism entails that an individual is willing to adopt an idea whenever she observes a high degree of consensus among her acquaintances. In its turn, we assume that agents are more likely to synchronize their ideas with the rest of the population when being surrounded by a large fraction of spreaders, thus mimicking the social reinforcement mechanism.

The natural coupling between both dynamics gives rise to the emergence of first order transitions separating the social movement state, characterized by the existence of a high degree of consensus on a given topic and a large number of people spreading it, from an inactive and poorly aligned population. The existence of bistability between both states can explain why punctual perturbations forcing the population to leave the latter stable configuration can give rise to a solid sustained social movement. It is worth mentioning that different coupling rules, e.g., assuming that the spread of ideas is hindered when high degrees of consensus are observed, lead to interesting physics like the emergence of chaotic trajectories [136].

Regarding the evolution of corruption, we have exposed the shortcomings of following game theoretical approaches to address this phenomenon and proposed an alternative, the HCO model, which modifies the flows of the SIR model to accommodate the features inherent to corruption. Specifically, the main novelties of the formalism are that recovery events are substituted by delation events from the honest population to the corrupt individuals and that two additional flows are introduced: the WTW effect hindering corruption because of the possible eventual punishment and the reinsertation of the punished individuals in the society.

The mechanisms introduced in the HCO model give rise to a rich phase diagram with three different equilibria corresponding to a honest population, a corrupt population and a mixed population characterized by the coexistence between the three different types of individuals. Interestingly, we have checked how another social mechanism, the coercion made by corrupt individuals, alters the nature of the transitions between these three equilibria. Specifically, the smooth second order transitions inherited from epidemic models become abrupt first order ones with bistable equilibria, where both corrupt and honest societies can be stable configurations.

Finally, we have revealed how the social flavor of the interactions introduced in both models partially hinders the spreading advantage of highly connected nodes. Specifically, in the KM-SIS model, we have found that the spread of mistrusted ideas is favored as the average degree of the network decreases, pinpointing that not contrasting what we read or receive in social platforms with enough sources of information can provide an individual with a distorted picture of the consensus existing on it. In addition, we have observed that corrupt societies are more vulnerable in heterogeneous configurations, for highly connected nodes are more capable of spreading corrupt behaviors but are also more exposed to the delation of their honest counterparts.

The formalisms presented here constitute the starting point for the elaboration of more sophisticated frameworks aimed at capturing different social dynamics. In this sense, we think that the KM-SIS could be extended to address the recent and relevant emergence of polarization and echo-chambers in social networks [13, 38]. This phenomenon has been usually characterized by means of the activity-driven (AD) model [175] and a preferential rewiring mechanism to prioritize the links with those with aligned opinions. Introducing the latter mechanism in our model and quantifying its interplay with the spreading dynamics remains as a future work. Furthermore, the HCO model can constitute the core of a more refined model taking into account the presence of different gangs or corrupt organizations in the population and the outcome of the competition among them.

Part II

Intertwined dynamics on metapopulations

Presentation of Part I

The global spread of diseases can be conceived as a reaction-diffusion process where the reaction stage corresponds to the contagions between infectious hosts and susceptible individuals and the diffusion stage is associated with the movements of the hosts, fostering the spatial unfolding of the contagious diseases. The comparison between modern epidemics with important pandemics occurring in the past before the XX century yields a straightforward result: there has been an inversion in the time scales governing the spread of epidemic outbreaks globally. In terms of reaction-diffusion processes, pandemics occurring before the XX century like the Black Death [176] were characterized by diffusion times much slower than the reaction time scales, which gave rise to well-defined subsequent epidemic outbreaks which were happening one after another, being propagated typically to geographically neighboring areas or as dictated by the most usual trading routes. In contrast, recent pandemics such as the Influenza A caused by the virus H1N1 in 2009 or much more recently the COVID-19 disease have been characterized by the existence of important outbreaks occurring simultaneously in faraway regions as a consequence of the shortening of the time scale associated with diffusion processes.

The acceleration of the spatial dissemination of recent epidemic outbreaks is one of the side effects of the vast interconnected society to which we have evolved as a result of the striking growth experienced by the international mobility network [177]. In this sense, different studies have proven that international connections provide key information to predict the global expansion of epidemic outbreaks initially localized in small geographical areas. For instance, the airport connectivity has proved to be a very reliable proxy to estimate the arrival times of SARS epidemic to different countries in 2003 [35, 178], the most likely infectious routes followed by the virus H1N1 in 2009 [179] or recently the risk of importing COVID-19 cases [180, 181].

Unlike the well-defined role of the airport mobility network for the importation and exportation of cases, the complex relationship existing between the distribution of the population across a city, their local recurrent mobility patterns and the evolution of local outbreaks has been less explored in the literature. In this part of the thesis, we aim at shedding light on this problem by proposing a theoretical framework to accommodate recurrent mobility patterns in epidemic modeling. In terms of the common thread of this thesis, the interplay between recurrent mobility and epidemic spreading can be characterized as an intertwined dynamics. On the one hand, mobility influences epidemic spreading, by changing both the contact structures of the population and their spatial distribution. On the other hand, the advance of an epidemic also affects the mobility of the population, because of the usual mobility restriction policies implemented to stop the advance of threatening outbreaks.

Chapter 4 is devoted to the study of the Movement-Interaction-Return (MIR) model, which constitutes the core of this part of the thesis and the formalism on top of which the rest of the results here presented are built. The application of the MIR model to real cities, without making any statistical assumption about the properties of their population or their mobility patterns, reveals unexpected phenomena arising from the interplay between mobility and epidemics which have not been reported in the literature yet. In particular, we found that promoting mobility does not always foster epidemic spreading but sometimes may lead to its extinction, a phenomenon that we have denoted by epidemic detriment. Finally, leveraging the resolution of the available data for Colombian cities, we introduce the socio-economic diversity of the population and study its impact for the evolution of diseases in these cities.

Chapter 5 explores the physics behind the emergence of the so-called epidemic detriment. For this purpose, we introduce a synthetic metapopulation, the star-like network, which allows us to control the population distribution and their movements. This simple configuration allows us to explain the emergence of the epidemic detriment following macroscopic arguments in some regions of the parameters' space but fails in capturing this phenomenon in other scenarios. Afterwards, we propose an extension of the original MIR model including tunable trips duration which reveals the microscopic nature of the epidemic detriment. We conclude the chapter by proposing a model which allows for incorporating heterogeneous degree distributions within the populations and decouple the number of interactions of each individual from mobility. This model reveals that the epidemic detriment is rooted in the dismantling of the contact structures of vulnerable populations driven by mobility.

Finally, Chapter 6 reveals how the theoretical insights gained from the previous chapters can be leveraged to increase the resilience of real cities to epidemic outbreaks by the smart modification of the mobility patterns of their population. Beyond theoretical results, we also show that the framework developed constitutes an useful tool to forecast the evolution of diseases in real epidemic scenarios and assess the impact of those interventions promoted to keep them under control. In particular, starting from the MIR model, we propose two formalisms which allows us to predict the most vulnerable areas in the city of Cali for the spread of Dengue disease and the course of COVID-19 pandemic across Spain.

Chapter 4

Recurrent mobility and epidemics

> *Why listen to me? I can only predict epidemics and plagues.*
>
> L. Kramer [182].

The importance of our mobility patterns, both across global and local scales, for the course of contagious diseases demands to revisit the epidemic models explained in the first part of the thesis. Upon this point, the population under study has been always embedded into contact networks, whose nodes represent individuals and links encode the nontrivial interaction patterns existing among them. Contact networks constitute static configurations which, despite explaining the social factors driving the reaction phase of disease, cannot account for the spatial diffusion of the hosts.

To quantify the impact of human mobility, one must seek for alternative representations of the system. In particular, to include the effect of long-haul displacements, one typical approach has been to construct the airport mobility network [35, 178, 183, 184] where each node corresponds to an airport and links represent the flow of passengers between each pair of them. In this case, the network provides very insightful information about the diffusion stage of the disease at the expense of neglecting the specificities of its reaction phase.

To incorporate simultaneously both reaction and diffusion processes, we use more complex objects to represent the population which are called metapopulations. The basic notions about these structures as well as some previous attempts existing in the literature to account for the interplay between mobility and epidemics are explained in Section 4.1. Section 4.2 presents the Movement-Interaction-Return (MIR) model, which enables to accurately capture the impact of recurrent mobility on disease spreading and represents the theoretical framework of reference for the rest of the models studied in this part of the thesis. Finally, in Section 4.3 we propose an extension of the MIR model to account for the socio-economic diversity of the residents in a city and study the new phenomena arising from the existence of different types of individuals.

4.1 Metapopulations

The concept of metapopulations was initially introduced in the field of ecology to study the impact of the coexistence of different subpopulations within the same ecosystem [185–187]. In general grounds, a metapopulation is defined as a set of spatially separated populations which interact among each other because of the mobility of their constituents. In the language used in network science, the usual representation of a metapopulation is a complex network where each node, usually referred to as patch, corresponds to the geographical area populated by each subpopulation. In its turn, the links connecting the different subpopulations now encode the mobility patterns of the population rather than their interaction patterns. Formally, a metapopulation with N_P patches and L links is defined as a tuple $\mathcal{G} = (\mathcal{V}, \mathcal{N}, \mathcal{E})$, being $\mathcal{V} = \{\sqsubseteq_{\infty}, \ldots, \sqsubseteq_{\mathcal{N}_{\mathcal{P}}}\}$ the set of patches in the metapopulation, $\mathcal{N} = \{\backslash_{\infty}, \ldots, \backslash_{\mathcal{N}_{\mathcal{P}}}\}$ the sizes of the different subpopulations and $\mathcal{E} = \{\rceil_{\infty}, \ldots, \rceil_{\mathcal{L}}\}$ the mobility patterns of the population.

Given their nature, metapopulations are the most natural framework to incorporate reaction-diffusion processes, for reaction processes constitute the internal dynamics of each subpopulation and diffusion processes are captured by the exchange of particles between subpopulations governed by their links. In this sense, metapopulations have allowed for studying very diverse phenomena ranging from the stability of ecosystems [188] and the dynamics determining the conservation or extinction of species [189, 190] to the diffusion of tumor cells [191]. In the context of mathematical epidemiology, the use of metapopulations has constituted an important step forward in the process to turn simple epidemic models into valuable forecasting tools to anticipate the trajectory of outbreaks and devise interventions to mitigate them [192, 193].

For the sake of visualization, Fig. 4.1 illustrates the differences between the representation of a set of agents as elements of a contact network or as subpopulations within a metapopulation. It is worth remarking that metapopulations offer a versatile framework for epidemic modeling, for they enable to characterize the spread of diseases across different spatial scales depending on the definition of patch and subpopulation. In this way, metapopulations can accommodate from a set of neighborhoods within a city to a set of countries within a continent [194–196].

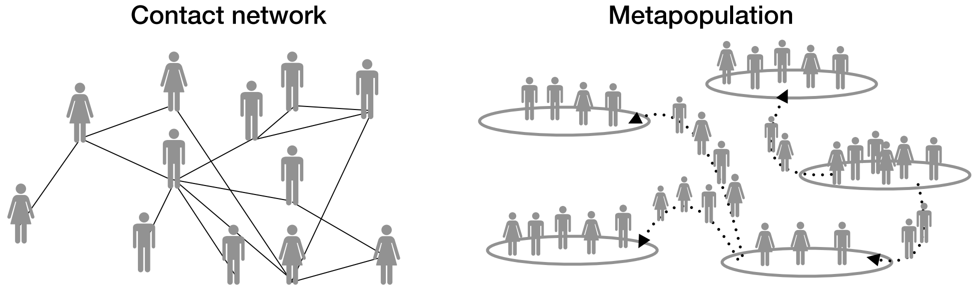


Figure 4.1: Schematic representation of a given population as a contact network (Left) where nodes identify individuals and links their interactions or as a metapopulation (Right) where each node, also known as patch, represents a geographical area and links govern the movements of the different subpopulations across the network.

4.1.1 Mobility frameworks

To run an epidemic model on top of a metapopulation, one has to choose how the mobility patterns and the diffusive behavior of the different subpopulations are introduced. In absence of data, one has also to determine how these links are synthetically constructed given the attributes of the different subpopulations exchanging individuals. Different alternatives have been proposed in the literature [197] but the most paradigmatic ones are:

- Gravity model: The gravity model for human mobility inspired in the Newton's laws for body motion was first introduced by George K. Zipf in 1946 [198]. In this model, it is assumed that the population flow between two locations is directly proportional to the product of their populations and inversely proportional to the distance among them. Mathematically, if we denote the population of a node i as N_i and its distance with another node j as d_{ij}, the weighted link connecting them, T_{ij}, is estimated by

  $$T_{ij} = K \frac{N_i N_j}{d_{ij}^2} \ . \tag{4.1}$$

  The latter expression can be generalized by assuming an arbitrary dependence on the *masses* (populations) and general functions to shape the relevance of geographical distances as well. Therefore, in general, the weights are computed as

  $$T_{ij} = K m_i m_j f(d_{ij}) \ , \tag{4.2}$$

  where m_i corresponds to the mass associated with node i and $f(d_{ij})$ is a decreasing function with the distance which usually ranges from power-law to exponentially decaying functions, thus hindering the movements between faraway regions. The size of the different subpopulations are the typical choices for the mass of each patch, although other variables have been also explored such as the GDP per capita [199, 200].

- Radiation model: The gravity model assumes that the flows between two locations obey only physical variables but does not account for other incentives driving people's movements [201]. To incorporate the attractiveness of each location in terms of, for instance, the number of job opportunities, the authors in [202] develop the so-called radiation model. In this model, it is assumed that an individual chooses the closest location to its residence which maximizes the number of available opportunities. Assuming that the number of job positions is proportional to the population in each area, the expected value for T_{ij} reads

  $$\langle T_{ij} \rangle = T_i \frac{m_i n_j}{(m_i + s_{ij})(m_i + n_j + s_{ij})} \ , \tag{4.3}$$

  where T_i is the total number of flows departing location i, m_i and n_j encode the population at the origin and destination respectively and s_{ij} is the total population competing for resources inside the circle centered in i with radius d_{ij}, excluding the inhabitants of both the origin and destination. Note that this model, unlike the gravity model, introduces a less severe penalization to long-range displacements connecting distant population centers.

4.1.2 Previous epidemic models on metapopulations

Apart from constructing the links, it is necessary to determine the nature of the mobility patterns. In this sense, the seminal model introduced in [203, 204] took advantage of the knowledge acquired in the study of random walkers dynamics on contact networks [205] and assume that individuals move randomly across the different patches according to the links of the metapopulation. To simplify the analysis, the authors also make use of a HMF approach, similar to the one explained in Chapter 2, and consider that all patches within the same degree class in the metapopulation are statistically equivalent. Note that the HMF assumption in metapopulations has further implications, for it entails that the attributes of each subpopulation, such as its size, should be linked to its connectivity in the mobility network.

Under these assumptions, the evolution of the infected population inside patches with degree k, I_k is written as

$$\begin{aligned}\partial_t I_k = &- p_k I_k + (1-p_k)\left[-\mu I_k + \lambda \frac{I_k S_k}{N_k}\right] \\ &+ k\sum_{k'} P\left(k'|k\right) p_{k'} d_{k'k}\left[(1-\mu)I_{k'} + \lambda \frac{I_{k'} S_{k'}}{N_{k'}}\right] \ .\end{aligned} \tag{4.4}$$

In the latter equation, N_k represents the population inside patches with degree k, p_k encodes the probability that one individual inside a patch of degree k moves and $d_{k'k}$ corresponds with the diffusion rate between patches with connectivities k and k'. The diffusion rate is assumed to follow the gravity model, being the masses proportional to the connectivity, so that

$$d_{k'k} \propto (kk')^{\theta}. \tag{4.5}$$

Note that Eq. (4.4) contains both reaction-diffusion processes involved in the interplay between mobility and epidemics. The evolution of the occupation of each compartment is completed with the time evolution of the spatial distribution of the population, which evolves driven by mobility as

$$\partial_t N_k(t) = -p_k N_k(t) + k\sum_{k'} P\left(k'|k\right) p_{k'} d_{k'k} N_{k'}(t) \ . \tag{4.6}$$

Nonetheless, an usual assumption is to consider the diffusion time scales much shorter than the duration of the outbreak, so $\tau^{diff} \ll \tau^{react}$. This allows for neglecting the time evolution of the underlying populations and assume that the number of individuals inside each patch is given by the steady state of the random walker dynamics.

Apart from proposing a set of deterministic equations to track the spatio-temporal spread of diseases, the authors in [204] reveal the existence of a new phase transition between the existence of localized outbreaks inside single subpopulations and the observation of widespread epidemics across the entire metapopulation. This phase transition is characterized by the so-called invasion threshold $\mathcal{R}^*$, which is computed from a branching process characterizing the evolution of the number of invaded subpopulations.

Despite the theoretical relevance or the results described above and others found in subsequent works [179, 192, 206], assuming that the inhabitants of a given city behave as random walkers exploring the different neighborhoods without a fixed residence is far from reflecting the nature of our daily movements. In contrast, there is growing evidence about the prominent recurrent nature of human mobility patterns [207–210] in real datasets. As an example, we represent in Fig. 4.2 the results from surveys carried out in the city of Medellín in Colombia to understand the usual mobility patterns of its residents [211, 212]. Specifically, we represent the volume of people arriving and

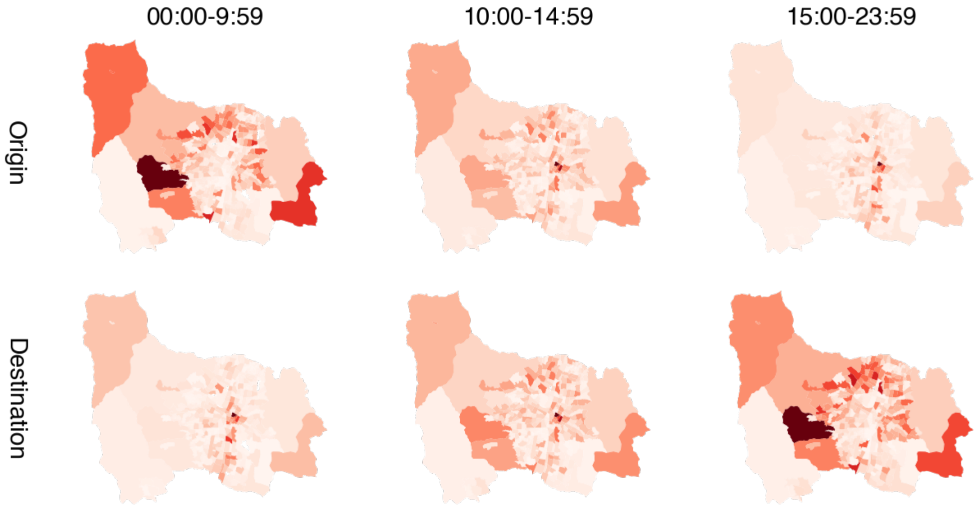


Figure 4.2: Origin and destination of the trips recorded in the surveys made to the population of Medellín in 2017 to characterize their mobility habits. The color of each area is proportional to the number of individuals entering or departing from them. The title show the time range during which trips were aggregated.

departing in each specific neighborhood of the city at the beginning, middle and end of the workday. In addition to the clear distinction between residential areas and workplaces, the figure clearly suggests that urban mobility is governed by the regular trips followed by the commuters.

There are several attempts made in the literature aimed at including commuters in epidemic modeling. First works [213, 214] on the topic adapt the previously explained HMF equations to include a preferential return rate τ_k to the associated patch of each agent, typically identified as the place where his or her residence is located. In this case, agents are not only characterized by their location in the network but also by their associated patch, which enlarges the set of equations needed to characterize the evolution of the population. In particular, Eq. (4.6) turns into

$$\partial_t N_{kk}(t) = -\sigma_k N_{kk}(t) + \tau^{-1} k \sum_{k'} N_{kk'}(t) P\left(k'|k\right) \ , \tag{4.7}$$

$$\partial_t N_{kk'}(t) = \sigma_{kk'} N_{kk}(t) - \tau^{-1} N_{kk'}(t) \ . \tag{4.8}$$

Plugging these equations into the branching process, the authors of [213, 214] derive analytically an expression that relates the invasion threshold $\mathcal{R}^*$ with the mobility rate of the population, the time scale associated with the movements and the features of the mobility network. They find that both promoting mobility and increasing the permanence times at the destinations boost epidemic spreading, for both interventions promote the mixing of individuals from different subpopulations.

The validity of the previous findings when studying real metapopulations remains uncertain due to the strong theoretical assumptions introduced in the HMF equations. Specifically, there does not exist solid basis to couple the size of each subpopulation with its number of connections in the mobility network. For this reason, Belik et al. [215, 216] propose an alternative model which abandons the statistical equivalence of all patches within the same degree class k. In this model, the evolution of the infected and susceptible associated with each patch k and located at each patch n, I_n^k and S_n^k respectively, is driven by

$$\partial_t I_n^k = \frac{\lambda}{N_n} S_n^k \sum_m I_n^m - \mu I_n^k + \sum_m \left(\omega_{mn}^k I_m^k - \omega_{nm}^k I_n^k \right) \ , \tag{4.9}$$

$$\partial_t S_n^k = -\frac{\lambda}{N_n} S_n^k \sum_m I_n^m + \sum_m \left(\omega_{mn}^k S_m^k - \omega_{nm}^k S_n^k \right) \ , \tag{4.10}$$

where ω_{nm}^k encodes the diffusion rate from patch n to patch m for individuals with residence in k. In this general framework, strictly commuting patterns are therefore included by considering $\omega_{nm}^k = \delta_n^k \omega_{nm}$. Armed with these equations, the authors study the propagation of diseases across simple synthetic configurations, finding that outbreaks in structured lattices can be characterized as continuous waves propagating across a media with a well-defined front propagation. In this model, the effect of decreasing the dwelling times at the destination is always detrimental for disease propagation and universal regardless of the size of the network.

The model proposed in [215, 216] constitutes the first step towards understanding the microscopic processes behind the complex interplay between recurrent mobility and epidemic spreading. Nonetheless, it still contains some theoretical limitations. Specifically, the model assumes a constant rate governing contagion events across the entire metapopulation, which implies that the number of interactions made by one agent is constant and independent of the place in which the individual is located. This limits the application of the model to real scenarios; for example, the model cannot capture the recently reported evidence about the positive correlation existing between population density and the number of COVID-19 cases in different countries [217–219].

4.2 MIR model

Here we present our proposal to characterize the impact of urban mobility in epidemic spreading, the Movement-Interaction-Return (MIR) model. Unlike previous formalisms, we follow a discrete-time approach and assume that each

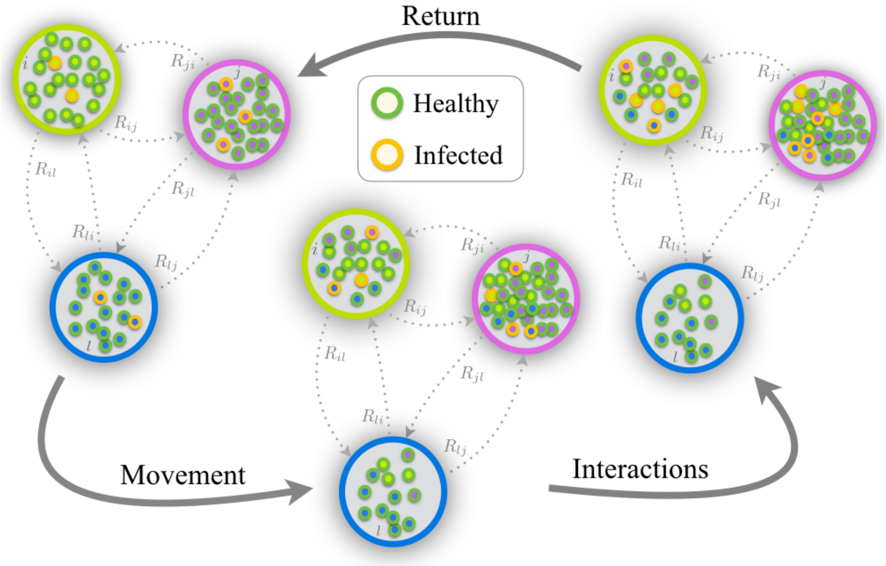


Figure 4.3: Schematic representation of one time step of the movement-interaction-return (MIR) metapopulation model. The metapopulation is composed of $N_P = 3$ patches. At the movement stage, some of the local agents decide to move to the other patches according to the mobility probability p and the mobility patterns encoded in matrix $\mathbf{R}$. After moving, the agents interact in a well-mixed way and change their epidemic state (healthy or infected) according to a SIS model. Finally, they come back to their home patches and a new time step starts.

time step is composed by three stages, which are schematically depicted in Fig. 4.3:

1. Movement: At the beginning of the day, the population decide whether or not to move with probability p. This probability acts as a control parameter which allows us to activate or deactivate mobility at convenience. If an individual decides to move, he or she decides the destination according to the origin-destination (OD) matrix $\mathbf{R}$ whose elements R_{ij} encode the weights of the links connecting patches i and j in the metapopulation.

2. Interaction: Once all the movements have been completed, the reaction phase starts and individuals interact, changing their epidemiological state accordingly. Note that this phase aims at capturing the contacts made at the workplace, school etc. Formally, we introduce a mean-field assumption within each patch, which implies that all the individuals placed at the same location make the same number of contacts. As stated above, this number of contacts is not constant across the metapopulation but is usually a function f of the attributes of the patches.

3. Return: Finally, as our main goal is to reflect the recurrent nature of

commuting mobility, we force all the individuals to return to their associated residential patches. Note that here we get rid of the dwelling times, which will be introduced in Chapter 5.

The iteration of these three stages simulates the daily dynamics of contagious diseases in our metapopulation. Note that our framework also presents some drawbacks. First, we do not include the continuous nature of human mobility and neglect the influence of more complex mobility patterns beyond strictly back-and-forth displacements. This assumption simplifies the equations of the model and is supported by statistics [220] reporting that commuting patterns represent around 30-50% of the daily movements. Second, we neglect the contribution of the interactions made with the household members after the return stage. According to some studies on the contact matrices of the population in different environments [221, 222], this hypothesis seems reasonable[1]. In Appendix C we relax the latter assumption and prove that the qualitative results here obtained are valid in that scenario as well.

4.2.1 Model equations

We now introduce the set of Markovian equations composing the MIR model. To do so, we start by considering a metapopulation with N_P patches and assume that each patch i is populated by n_i residents, whereas the flows between the different patches are captured in the OD matrix $\mathbf{R}$. Regarding the type of compartmental model, we consider that the disease follows a SIR or SIS dynamics, although the formalism here proposed can be easily generalized to account for more complex compartmental schemes, as shown in Chapter 7. Under these conditions, the course of a SIR disease is entirely characterized by the time evolution of the fraction of individuals living in each patch i who are either infected or recovered, denoted in the following as ρ_i and r_i respectively. This time evolution is given by

$$\rho_i(t+1) = (1-\mu)\rho_i(t) + (1-\rho_i(t)-r_i(t))\Pi_i(t) \ , \tag{4.11}$$

$$r_i(t+1) = r_i(t) + \mu\rho_i(t) \ . \tag{4.12}$$

For the SIS dynamics, we can get rid of the equations for the recovered individuals, so the evolution of the disease is characterized by

$$\rho_i(t+1) = (1-\mu)\rho_i(t) + (1-\rho_i(t))\Pi_i(t) \ . \tag{4.13}$$

Unless specified, we will work with SIS dynamics due to the higher simplicity of the equations. The first term in the r.h.s of Eq. (4.13) contains those infected

---

[1] In the case of Spain, $\langle k^H \rangle$=3.2 and $\langle k^O \rangle = 8.6$, being k^H and k^O the contacts inside and outside the household respectively.

individuals associated with i not recovering at time t. In its turn, the second term in the r.h.s of Eq.(4.13) reflects the contagions of susceptible population. In this sense, $\Pi_i(t)$, which represents the probability that a susceptible agent living inside i contracts the disease, is given by

$$\Pi_i(t) = (1-p)P_i(t) + p\sum_j R_{ij}P_j(t) \ , \tag{4.14}$$

where the first term encodes the contagions in the residential patch and the second one corresponds to the contagions occurring when one individual from patch i moves across the metapopulation. Therefore, $P_i(t)$ corresponds to the probability that a susceptible individual, regardless of his or her residential patch, contracts the disease inside patch i. Given the well-mixing hypothesis introduced in the model, this probability is

$$P_i(t) = 1 - \left(1 - \lambda\frac{\tilde{I}_i(t)}{\tilde{n}_i}\right)^{f_i} \ , \tag{4.15}$$

where $\tilde{n}_i$ represents the effective population of patch i, i.e. the number of individuals placed at patch i after the movement stage and $\tilde{I}_i(t)$ is the effective number of infected individuals inside patch i at time t. Taking into account the demographical distribution of the population and the mobility patterns, both quantities are easily computed as

$$\tilde{n}_i = \sum_j n_j \left[(1-p)\delta_{ij} + pR_{ji}\right] \ , \tag{4.16}$$

$$\tilde{I}_i(t) = \sum_j n_j\rho_j(t) \left[(1-p)\delta_{ij} + pR_{ji}\right] \ . \tag{4.17}$$

4.2.2 Validation on synthetic networks

Eqs. (4.13)-(4.17) form a closed set of equations which allows for characterizing the spatio-temporal spread of contagious diseases. In absence of data, we first construct synthetic metapopulations with $N_P = 10^3$ patches which are homogeneously populated by $n = 5000$ agents. For the sake of simplicity, let us assume that interaction processes takes place all-to-all within each patch, implying that $f_i = \tilde{n}_i$. Regarding the mobility patterns, we include both ER and BA networks, which allows us to analyze the role of heterogeneities in terms of spatial connections.

To validate the equations, we compare their predictions about the epidemic size, defined as the total fraction of population remaining infected in the stationary state, with results obtained from computationally expensive Monte Carlo simulations. To compute this order parameter, we start by infecting a 1

% of the population of each node and let the system evolve until no fluctuations are observed in the number of infected individuals. According to the variables of the MIR model, the epidemic size ρ is computed as

$$\rho = \frac{\sum\limits_{i=1}^{N_P} n_i \rho_i(\infty)}{\sum\limits_{i=1}^{N_P} n_i} \ . \tag{4.18}$$

In Fig. 4.4 we represent this indicator as a function of the rescaled infectivity $\tilde{\lambda}$ for several mobility scenarios in the ER (Panel a) and BA (Panel b) metapopulations. The rescaled infectivity is computed by dividing the infectivity λ by the threshold of a well-mixed population of size n, so $\tilde{\lambda} = \lambda n/\mu$. There we observe a fair agreement between the numerical results obtained by iterating our formalism and the ones yielded by the simulations. Apart from the agreement, qualitatively we observe that promoting mobility from the static case ($p = 0$) to the fully active population ($p = 1$), leads to an increase in the epidemic size, pinpointing that mobility accelerates epidemic diffusion as suggested in previous works. This acceleration gains relevance in case of heterogeneous mobility networks, for heterogeneity favors the gathering of a higher number of individuals inside the most connected areas.

4.2.3 Epidemic threshold

The analysis made in synthetic metapopulations has revealed that the impact of mobility is highly influenced by the structural properties of the underlying metapopulation. In addition to the computational cost saved by the Markovian equations, they allow us to get an analytical expression of the epidemic threshold to disentangle the interplay between mobility and epidemics. To do so, let us consider that the disease has reached an endemic equilibrium, so no fluctuations are observed in the fraction of population infected. Therefore, $\rho_i(t+1) = \rho_i(t) = \rho_i^*$, which turns Eq. (4.13) into

$$\mu \rho_i^* = (1 - \rho_i^*)\Pi_i^* \ . \tag{4.19}$$

Recall that the epidemic threshold is defined as the minimum infectivity required to observed a stable outbreak. As a consequence, close to the epidemic threshold, the endemic equilibrium is characterized by a negligible fraction of infected individuals so that $\rho_i = \epsilon_i \ll 1$. This allows for linearizing the equations of the formalism by neglecting the $\mathcal{O}(\epsilon^2)$ terms. After doing so, Eq. (4.15) can be casted into:

$$P_i^* = 1 - \left(1 - \lambda \frac{\tilde{I}_i}{\tilde{n}_i}\right)^{f_i} \simeq \frac{\lambda f_i}{\tilde{n}_i} \sum_{j=1}^{N_P} \left[\delta_{ij}(1-p) + pR_{ji}\right] \epsilon_j n_j \ . \tag{4.20}$$

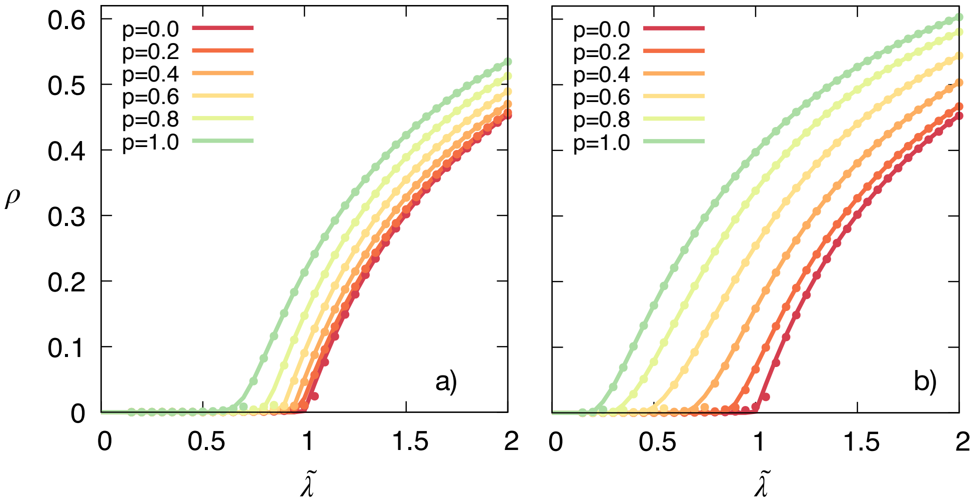


Figure 4.4: Evolution of the epidemic size ρ as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility p (color code). The rescaled infectivity is obtained by dividing the actual infectivity λ by the expected threshold for an isolated metapotulation, i.e. $\tilde{\lambda} = \lambda n/\mu$. In panel a) the metapopulation is a ER network whereas panel b) represents the curves for a BA network. both with $N_P = 10^3$ patches, $\langle k\rangle \sim 4$ and random weights assigned to the links. In both panels, the solid lines represent the numerical results obtained by iterating Eqs. (4.11)-(4.17) whereas dots show the average of the epidemic size values obtained from 50 Monte Carlo simulations. The recovery probability is set to $\mu = 0.2$.

Introducing this expression into Eq. (4.15) and rearranging terms, we obtain

$$\Pi_i^* = \sum_{j=1}^{N_P} n_j \epsilon_j \left[\delta_{ij}(1-p)^2 \frac{f_j}{\tilde{n}_j} + p(1-p)\left(\frac{R_{ij} f_j}{\tilde{n}_j} + \frac{R_{ji} f_i}{\tilde{n}_i}\right) + p^2 \sum_{l=1}^{N_P} \frac{R_{il} R_{jl} f_l}{\tilde{n}_l} \right] . \tag{4.21}$$

Plugging this value into Eq. (4.19) and neglecting nonlinear terms, it turns into

$$\mu \epsilon_i = \lambda \sum_{j=1}^{N_P} M_{ij} \epsilon_j \ , \tag{4.22}$$

where we have introduced a new matrix $\mathbf{M}$ whose elements are

$$M_{ij} = n_j \left[\delta_{ij}(1-p)^2 \frac{f_j}{\tilde{n}_j} + p(1-p)\left(\frac{R_{ij} f_j}{\tilde{n}_j} + \frac{R_{ji} f_i}{\tilde{n}_i}\right) + p^2 \sum_{l=1}^{N_P} \frac{R_{il} R_{jl} f_l}{\tilde{n}_l} \right] . \tag{4.23}$$

When $f_i = \tilde{n}_i$ the elements of matrix $\mathbf{M}$ are further simplified into

$$M_{ij} = n_j \left[\delta_{ij}(1-p)^2 + p(1-p)\left(R_{ij} + R_{ji}\right) + p^2 (\mathbf{R}\mathbf{R}^{\mathbf{T}})_{ij} \right] . \tag{4.24}$$

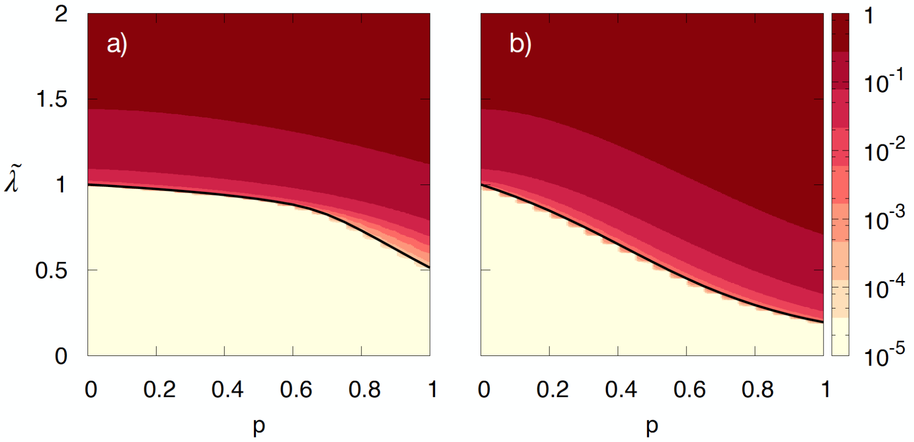


Figure 4.5: Epidemic size ρ obtained by the numerical iteration of Eqs. (4.11)-(4.17) (color code) as a function of the mobility p and the rescaled infectivity $\tilde{\lambda}$. The black solid line shows the analytical estimation for the epidemic threshold computed by Eq. 4.25 for the ER (panel a) and the BA (panel b) metapopulations used in Fig. 4.4.

Eq. (4.22) is an eigenvalue equation; thus, computing the epidemic threshold involves solving the spectrum of the matrix **M**. Specifically, as we are interested in the minimum infectivity needed to trigger outbreaks, the epidemic threshold is given by

$$\lambda_c = \frac{\mu}{\Lambda_{\text{max}}(\mathbf{M})} \ . \tag{4.25}$$

Note that this expression resembles the epidemic threshold derived for SIS dynamics on contact networks but here, rather than the adjacency matrix **A**, a more complex matrix **M** governs the interactions of the population. Specifically, the elements M_{ij} tell us the expected number of interactions made by one individual from patch i with the population residing in patch j. In particular, the first term corresponds to those contacts taking place between individuals from the same patch and not moving. The second (third) term encodes the contacts made where the agent from i (j) moves to patch j (i) and interacts with the population remaining there. Finally, the fourth term contains all the interactions taking place in a third node different from the patches associated with the agents in contact.

Fig. 4.5 shows the dependence of the epidemic size on the mobility p and the infectivity λ along with the estimation of the epidemic computed by Eq. (4.25). There we observe that the analytical expression of the epidemic threshold accurately captures the phase transition between the disease-free state and the emergence of a stable endemic regime in both synthetic metapopulations.

4.2.4 Cali: a real metapopulation

Upon this point, we have validated the equations of the MIR model and proposed an analytical expression which allows us to determine the conditions, related to the spatial distribution and the mobility patterns of the population, triggering local outbreaks. However, we have not fully exploited our model, for one of its main advantages with respect to other alternatives existing in the literature is that no assumptions are made concerning the statistical equivalence of the different patches. This implies that data about the demographical distribution and the mobility of residents in real cities can be readily incorporated into our equations to study the relevance of commuting for disease spreading.

In what follows, we apply the machinery of the formalism to the spread of SIS diseases across Santiago de Cali, in Colombia. Santiago de Cali is the third most populated city in the country with 2.5 million of inhabitants which are distributed across 22 official administrative units. Therefore, our metapopulation is composed of $N_P = 22$ patches. Regarding the population distribution, we collect the data from census information published by the local authorities [223]. For the mobility patterns, we resort to surveys made to a representative sample of the population indicating the usual places among which they usually commute. In particular, the elements of the OD matrix $\mathbf{R}$ are estimated as

$$R_{ij} = \frac{W_{ij}}{\sum\limits_{l=1}^{N_P} W_{il}} \ , \tag{4.26}$$

being W_{ij} the number of trips recorded in the survey between locations i and j. The resulting metapopulation is schematized in Fig 4.6a. Note that different alternatives have been proposed in the literature to capture the mobility patterns of the populations such as census surveys [224, 225] or the location history inferred from call detailed records [226, 227], or the activity on social platforms [228, 229] or mobile-phone devices [230, 231].

Once the metapopulation is constructed, we can repeat the same process performed for the synthetic metapopulations to check whether the equations remain valid in presence of heterogeneously distributed subpopulations and more complex mobility patterns, as the ones characterizing the Cali metapopulation. Note that, in what follows, the rescaled infectivity $\tilde{\lambda}$ is computed as the ratio between the infectivity λ and the threshold for the static case $\lambda_c(p=0)$. Thus, $\tilde{\lambda} = \lambda n_{\max}/\mu$, where $n_{\max}$ represents the maximum number of residents within a patch.

Fig. 4.6b-c proves that our formalism is able to reproduce the results from Monte Carlo simulations even in this case, which highlights the generality of

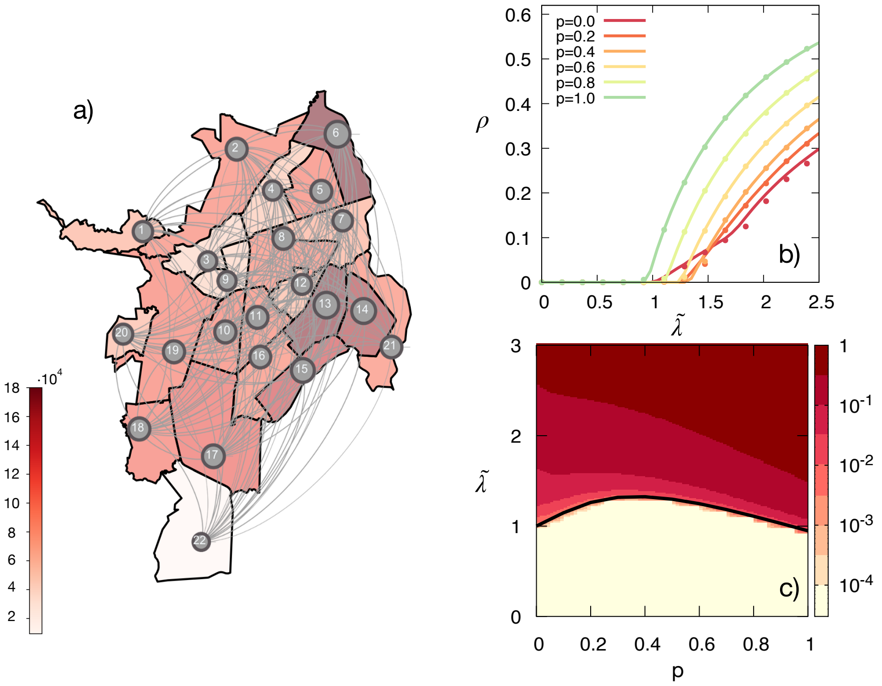


Figure 4.6: Spread of diseases across real metapopulations. Panel a) Schematic representation of the Cali metapopulation with $N_P = 22$ patches. The color of each patch is proportional to its population. Panel b) epidemic size ρ obtained by the numerical iteration of Eqs. (4.11)-(4.17) (solid lines) and by averaging 50 Monte Carlo realizations (dots) as a function of the mobility p (color code) and the rescaled infectivity $\tilde{\lambda}$. Panel c): epidemic size ρ obtained by the numerical iteration of Eqs. (4.11)-(4.17) (color code) as a function of the mobility p and the rescaled infectivity $\tilde{\lambda}$. The black solid line shows the analytical estimation for the epidemic threshold computed by Eq. (4.25). In all panels, the recovery probability is set to $\mu = 0.2$.

its equations. Interestingly, some important quantitative differences appear when comparing the diagrams with the ones previously obtained for synthetic metapopulations, which were homogeneously populated. First, the static curve contains secondary transitions beyond the epidemic threshold, which are characterized by discontinuities in the derivative of the theoretical curve. This is a direct consequence of population heterogeneities, which make some patches more prone to develop early epidemic outbreaks than others. Specifically, when $p = 0$ the matrix $\mathbf{M}$ is diagonal and each individual patch displays its own epidemic threshold $(\lambda_c)_i = \lambda n_i \mu$.

Remarkably, increasing mobility in real heterogeneous metapopulations has a counterintuitive effect on the dynamics. Namely, rather than boosting epidemic spreading as observed in synthetic metapopulations, there exists a non-monotonic dependence of the threshold on the mobility. The initial increase of the threshold implies that mobility prevents the endemic equilibrium from being reached, therefore hindering the spread of diseases. This phenomenon, hereinafter denoted as epidemic detriment, constitutes the most important result obtained in this thesis and has profound implications for the efficiency of control policies restricting population movements. Due to its relevance, we devote the entire Chapter 5 to understand the physical roots behind its emergence.

4.2.5 Perturbative analysis

Mathematically, we can explore the origin of the epidemic detriment by performing a perturbative analysis of matrix **M** [232]. Focusing on the case $f_i = \tilde{n}_i$, after rearranging the terms in Eq. (4.24), the elements M_{ij} read as

$$M_{ij} = M^0_{ij} + pM^1_{ij} + p^2 M^2_{ij} \tag{4.27}$$

with

$$\begin{aligned} M^0_{ij} &= n_j \delta_{ij} \ , && (4.28)\\ M^1_{ij} &= n_j (R_{ij} + R_{ji} - 2\delta_{ij}) \ , && (4.29)\\ M^2_{ij} &= n_j \left[(\mathbf{R} \cdot \mathbf{R^T})_{ij} - R_{ij} - R_{ji} + \delta_{ij} \right] \ . && (4.30) \end{aligned}$$

The perturbative analysis tells us that the leading eigenvalue of matrix **M** can be approximated by

$$\Lambda_{\max}(\mathbf{M}) = \Lambda_{\max}(\mathbf{M^0}) + p\Lambda^1 + p^2\Lambda^2 \ , \tag{4.31}$$

with

$$\begin{aligned} \Lambda^1 &= \langle v^0_{max} | \mathbf{M^1} | v^0_{max} \rangle \ , && (4.32)\\ \Lambda^2 &= \sum_{j \neq max} \frac{\langle v^0_j | \mathbf{M^1} | v^0_{max} \rangle \langle v^0_{max} | \mathbf{M^1} | v^0_j \rangle}{\Lambda_{\max} - \Lambda_{\mathrm{j}}} + \langle | v^0_{max} | \mathbf{M^2} | v^0_{max} \rangle \ . && (4.33) \end{aligned}$$

In both expressions, $\{v^0_j\}$ represents the eigenvector basis of matrix $\mathbf{M^0}$ and, in particular, v^0_{max} represents the eigenvector associated with the leading eigenvalue of matrix. Taking into account that the matrix $\mathbf{M^0}$ is diagonal, the eigenvalues correspond to the population of each patch and the eigenvector basis is the canonical basis, which simplifies the calculus of the perturbative corrections. In particular,

$$\Lambda^0 = n_{max} \ . \tag{4.34}$$

$$\Lambda^1 = 2n_{max}(R_{maxmax} - 1) \ . \tag{4.35}$$

$$\Lambda^2 = n_{max}\Bigg\{(1 - R_{maxmax})^2 + \sum_{j \neq max} \left[\frac{n_j(R_{maxj} + R_{jmax})^2}{n_{max} - n_j} + R^2_{maxj}\right]\Bigg\} \ . \tag{4.36}$$

Note that the linear correction is always negative, ensuring that for low enough p values, the leading eigenvalue of the critical matrix $\mathbf{M}$ decreases, which is reflected in the increase of the epidemic threshold responsible for the epidemic detriment. Interestingly, the second correction is always positive, thus explaining the non-monotonic behavior observed for the epidemic threshold as a function of the trade-off between these two opposite contributions. Finally, let us remark that, if we decrease the heterogeneity of the population distribution, the second term becomes dominant, thus shrinking the epidemic detriment regime, which finally vanishes for homogeneous populations, as observed in the synthetic metapopulations.

4.3 Multiplex metapopulations

The MIR model represents a step forward towards capturing the complex interplay between recurrent mobility and disease spreading. Apart from understanding the central result of this thesis, the detrimental effect of the mobility, in what follows we focus on refining the MIR model to keep shortening the distance between these theoretical frameworks and real epidemic scenarios. One of the key assumptions to construct the equations of the MIR model is to consider all the population within a single patch as statistically equivalent. In plain words, this fact implies that the model assumes that the probability for individuals from the same area being infected or their usual mobility patterns are equivalent and independent of any socio-economic features.

The analysis of epidemic surveillance data calls for a rethinking of the latter assumption. One of the most recent examples about the relevance of social variables to understand the course of a disease is the vastly unequal COVID-19 impact reported for different ethnics, race or socio-economic classes [233–235]. Beyond the unequal impact of diseases, important structural differences appear when analyzing the mobility patterns of the population as a function of different variables like age [236], gender [237, 238] or wealth [239, 240].

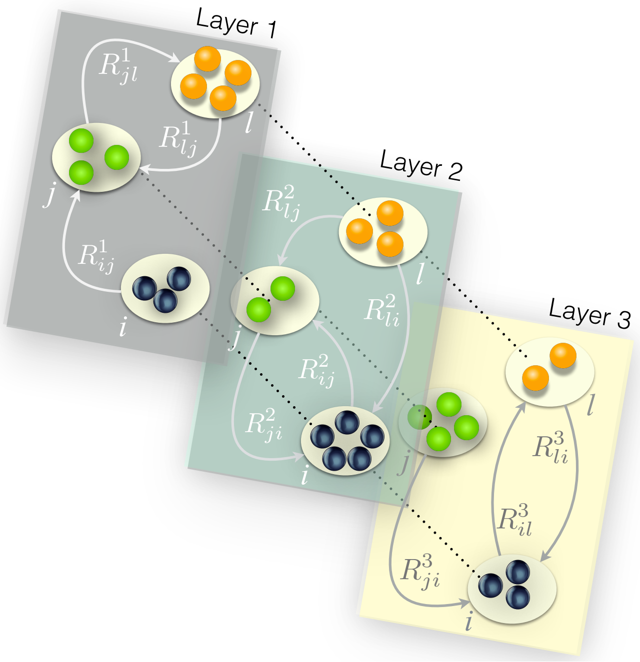


Figure 4.7: Schematic representation of a multiplex network with $N = 3$ patches and $L = 3$ layers. Each layer represents the demographic distribution and the mobility patterns of each type of individuals.

The existence of diverse individuals in each subpopulation, characterized by different spatial distributions and mobility patterns, resembles the problem of accommodating the different types of interactions connecting the same set of individuals in contact networks . Recall that in, the latter case, this problem has been traditionally overcome by using multiplex networks [241, 242] where all the layers contain the same set of agents, but their connection patterns vary from one layer to another as a function of the nature of the interactions.

Here we propose the generalization of multiplex networks applied in the context of metapopulations, which is referred to as multiplex metapopulations. A multiplex metapopulation is a mathematical object composed of L layers containing the same set of geographical areas, but each one encoding the mobility patterns and demographical distribution of the different types of individuals found in each subpopulation. Mathematically, a multiplex metapopulation is characterized by the size of the subpopulations in each layer α, $\vec{n}^{\alpha}$, and the mobility patterns connecting them, usually encoded in OD matrices $\mathbf{R}^{\alpha}$. A schematic representation of a multiplex metapopulation with $L = 3$ layers and $N_P = 3$ patches is shown in Fig 4.7.

4.3.1 Multiplex MIR model

The study of the dependencies between mobility, epidemics and individuals' diversity demands an extension of the original MIR model to include multiplex metapopulations. As in the original MIR model, we propose the equations for SIR and SIS diseases. In the case of a SIR disease, the system is completely described by the time evolution of the fraction of individuals of each type α associated with each patch i inside the I and R compartments, defined in the following as ρ_i^α and r_i^α. Taking into account the assumptions of the MIR model, this evolution reads as

$$\rho_i^\alpha(t+1) = (1-\mu)\rho_i^\alpha(t) + (1-\rho_i^\alpha(t) - r_i^\alpha(t))\Pi_i^\alpha(t)\ , \tag{4.37}$$

$$r_i^\alpha(t+1) = r_i^\alpha(t) + \mu\rho_i^\alpha(t)\ . \tag{4.38}$$

In the case of SIS diseases, we obtain

$$\rho_i^\alpha(t+1) = (1-\mu)\rho_i^\alpha(t) + (1-\rho_i^\alpha(t))\Pi_i^\alpha(t)\ , \tag{4.39}$$

where $\Pi_i^\alpha(t)$ represents the probability that one individual from layer α and patch i contracts the disease. Incorporating the multiplex metapopulation, this probability reads as

$$\Pi_i^\alpha(t) = (1-p)P_i^\alpha(t) + p\sum_j R_{ij}^\alpha P_j^\alpha(t)\ , \tag{4.40}$$

where $P_i^\alpha(t)$ is the probability that an individual from layer α contracts the disease inside i at time t. To compute it, we denote the probability of contagion per contact between individuals from layers α and β by $\lambda^{\alpha\beta}$ and focus on the case in which interactions take place all-to-all within the patch. Therefore,

$$P_i^\alpha(t) = 1 - \prod_{\beta=1}^{L}\prod_{j=1}^{N_P}\left(1-\lambda^{\alpha\beta}\rho_j^\beta\right)^{n_{j\to i}^\beta}\ , \tag{4.41}$$

where $n_{j\to i}^\beta$, encoding the number of individuals from layer β moving from j to i, is given by

$$n_{j\to i}^\beta = n_j^\beta\left[(1-p)\delta_{ij} + pR_{ji}^\beta\right]\ . \tag{4.42}$$

Note that Eq. (4.41) is slightly different from the expression introduced in Eq. (4.15) for the original MIR model. Namely, in the all-to-all contacts scenario, the former, which explicitly accounts for the different infectious states of each layer and patch, is more accurate than the latter, which assumes that an individual makes $\tilde{n}_i$ with an infectious cloud resulting from averaging all of them. Nonetheless, it is worth remarking that the differences in the predictions are negligible and that both approaches yield the same estimation for the epidemic threshold.

To find the analytical expression of the epidemic threshold, we follow the same procedure as when studying the original MIR model. Therefore, we assume that the disease reaches an endemic equilibrium characterized by a negligible number of infected individuals so that $\rho_i^\alpha(t+1) = \rho_i^\alpha(t) = \rho_i^\alpha$. This turns Eq. (4.39) into

$$\mu\rho_i^\alpha = (1-\rho_i^\alpha)\Pi_i^\alpha \ . \tag{4.43}$$

The next step is to assume that we are close to the epidemic threshold so that $\rho_i^\alpha = \epsilon_i^\alpha \ll 1$. This allows for linearizing Eq. (4.41) by getting rid of $\mathcal{O}(\epsilon^2)$ terms. In particular,

$$P_i^\alpha(t) \simeq \sum_{\beta=1}^{L}\sum_{j=1}^{N_P} n_{j\to i}^\beta \lambda^{\alpha\beta}\rho_j^\beta \ . \tag{4.44}$$

After introducing the latter expression into Eq. (4.40) and the result into Eq. (4.43), we obtain

$$\mu\epsilon_j^\alpha = \sum_{\beta=1}^{L}\sum_{j=1}^{N_P} \lambda^{\alpha\beta}\mathcal{M}_{ij}^{\alpha\beta}\epsilon_j^\beta \ , \tag{4.45}$$

with

$$\mathcal{M}_{ij}^{\alpha\beta} = n_j^\beta\left[\delta_{ij}(1-p)^2 + p(1-p)\left(R_{ij}^\alpha + R_{ji}^\beta\right) + p^2(\mathbf{R}^\alpha\mathbf{R}^{\beta T})_{ij}\right] \ . \tag{4.46}$$

Note that the terms of matrix $\mathcal{M}$ encode the same microscopic contact processes as the ones described for matrix $\mathbf{M}$ but also account for the diversity of the residents in the different subpopulations. Let us remark that, in general, it is not possible to derive a closed expression for the epidemic threshold due to the different contagion probabilities as a function of the interacting individuals.

4.3.2 Synthetic metapopulations

We first aim at validating the equations with simple synthetic multiplex metapopulations. For the sake of simplicity, we assume that $\lambda^{\alpha\beta} = \lambda$ and introduce a multiplex network composed of $L = 2$ layers. These layers correspond to an ER and a BA metapopulations respectively, with $N_P = 10^3$ patches which are homogeneously populated with $n = 500$ individuals in each layer. In this scenario, the epidemic threshold is analytically estimated as

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathcal{M})} \ . \tag{4.47}$$

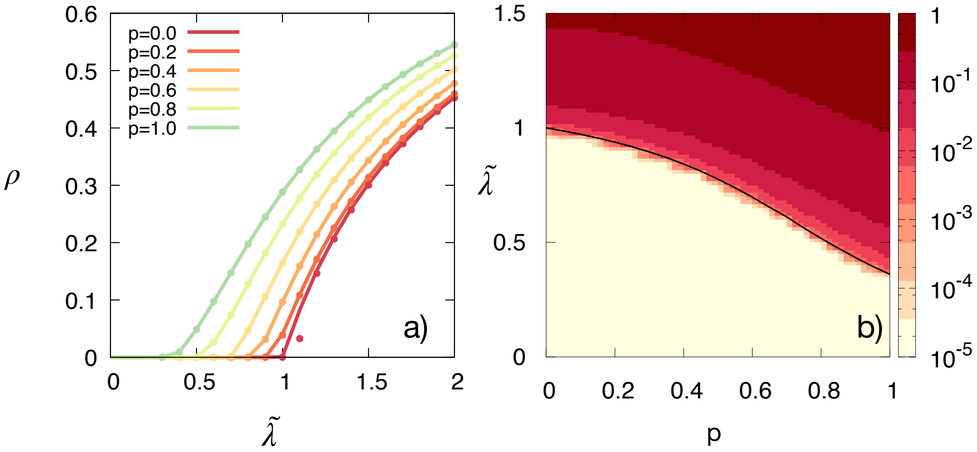


Figure 4.8: Panel a) epidemic size ρ as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility p (color code). Lines show the results obtained by the iteration of the Markovian equations whereas dots correspond with the average of 50 Monte Carlo realizations. Panel b) epidemic size (color code) as a function of $\tilde{\lambda}$ and p. The black line depicts the analytical estimation of the epidemic threshold computed by Eq. (4.47). In both panels, the underlying multiplex network is an ER-SF configuration with $N_P = 1000$ patches, which are homogeneously populated with $n = 500$ individuals in each layer. The recovery rate is set to $\mu = 0.2$.

Let us first study the spread of SIS diseases across multiplex metapopulations. Fig. 4.8a represents the evolution of the epidemic size as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility p, finding a very fair agreement between results from the equations and the ones obtained via Monte Carlo simulations. Note that the epidemic size is now computed as

$$\rho = \frac{\sum\limits_{\alpha=1}^{L}\sum\limits_{i=1}^{N_P} n_i^\alpha \rho_i^\alpha(\infty)}{\sum\limits_{\alpha=1}^{L}\sum\limits_{i=1}^{N_P} n_i^\alpha} . \tag{4.48}$$

In Fig. 4.8b we plot the surface $\rho(\tilde{\lambda}, p)$ along with the analytical estimation for the epidemic threshold provided by Eq. (4.47), which captures the boundary between the endemic and the disease-free equilibria. Furthermore, it is worth remarking that the existence of different mobility networks in the ER-BA multiplex configuration makes the epidemic threshold lie in between the two curves obtained when analyzing monoplex configurations in Section 4.2.

We now abandon the hypothesis of homogeneous infectivities and study the spread of SIR diseases across the ER-BA configuration. In particular, let us study one scenario in which the inter-layer coupling is much smaller than the intra-layer coupling. Mathematically, this implies that $\lambda^{\text{ER-BA}} \ll \lambda^{\text{ER-ER}}$

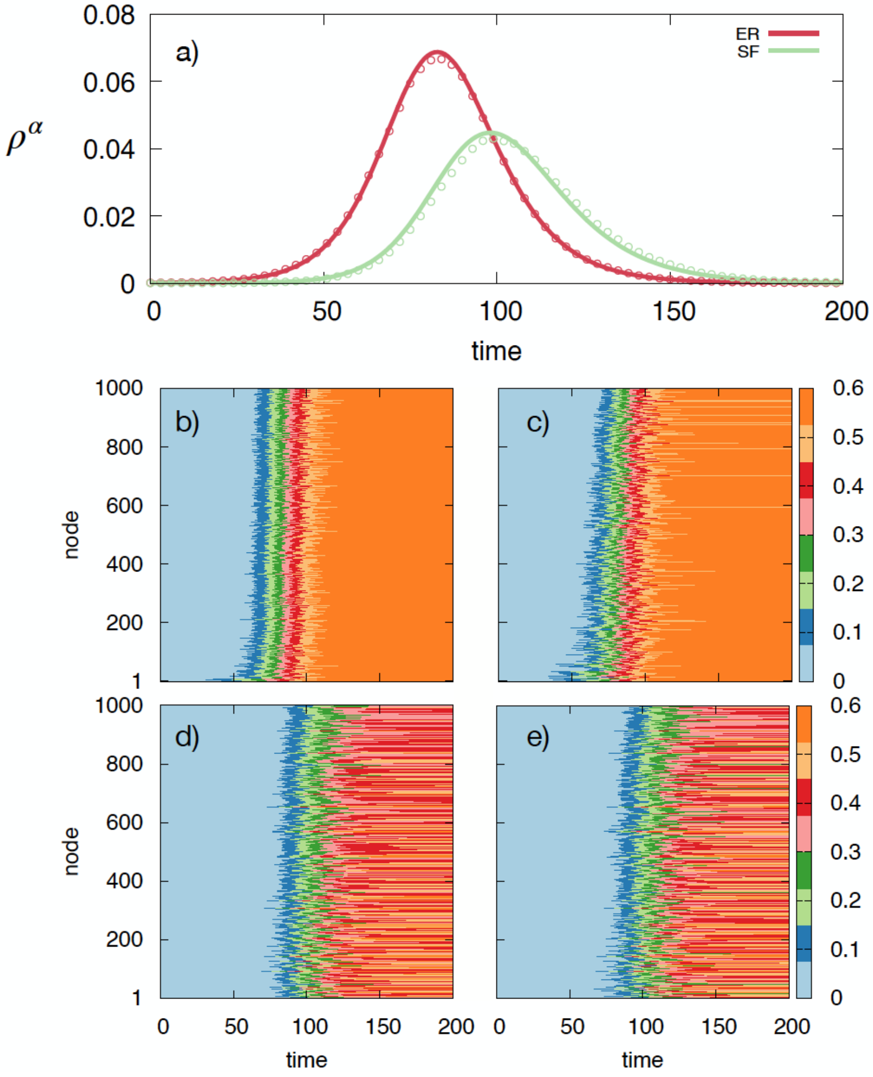


Figure 4.9: Panel a) Time evolution of the epidemic size in each layer ρ^α according to the Markovian equations (lines) and the Monte Carlo simulations (dots). Panels b)-c): Evolution of the fraction of infected agents in the ER layer associated to each patch according to the Markovian equations (Panel b) and the MC simulations (Panel c). Panels d)-e): Evolution of the fraction of infected agents in the BA layer associated to each patch according to the Markovian equations (Panel d) and the MC simulations (Panel e). In all panels, the dynamics starts with one infected individual placed randomly in the ER layer, the mobility is fixed to $p = 0.3$ and the recovery rate to $\mu = 0.2$ (see text for detailed information of the intra-layer and interlayer couplings).

and that $\lambda^{\text{ER-BA}} \ll \lambda^{BA-BA}$. Specifically, we assume that both layers are in a supercritical regime such that $\lambda^{\text{ER-ER}} = 1.5\mu/500$ and $\lambda^{\text{BA-BA}} = 1.1\mu/500$ but weakly coupled among each other so that $\lambda^{\text{ER-BA}} = 0.025\mu/500$.

We are interested in unraveling the spatio-temporal propagation of SIR diseases in this scenario. For this purpose, we start by infecting a single individual in the ER layer and let the system evolve until the disease dies out. Note that the weakened inter-layer coupling can be thought as the result of taking control measures to isolate the affected group from the unaffected layer. We analyze the time evolution of the infected individuals in each layer, denoted as ρ^α and computed as

$$\rho^\alpha(t) = \frac{\sum\limits_{i=1}^{N_P} n_i^\alpha \rho_i^\alpha(t)}{\sum\limits_{i=1}^{N_P} n_i^\alpha} \,. \tag{4.49}$$

Fig. 4.9 shows that the disease is able to invade the BA layer despite the weak inter-layer coupling. Once it has been transferred, the supercritical regime existing in the BA layer triggers a cascade a contagions which ends up with a macroscopic fraction of the population becoming infected. This phenomenon highlights the relevance of anticipating and taking preemption measures in unaffected social groups in light of possible imperfect isolation policies of the affected groups.

Finally, we give a microscopic description of the epidemic trajectory by analyzing the time evolution of the recovered population in each patch and layer. Interestingly, the equations of the formalism shows a fair agreement with the results from stochastic Monte Carlo simulations, which strengthens the validity of the proposed framework. Regarding the spatio-temporal patterns, we observe that in both cases mobility tends to synchronize the different epidemic trajectories, but heterogeneities are observed in the BA layer due to the important role of the most connected subpopulations.

4.3.3 Medellín: a real multiplex metapopulation

Upon this point, we have limited our analysis to synthetic metapopulations. Nonetheless, in the same fashion as monoplex configurations, the formalism allows for readily incorporating real data once the demographical distribution and the mobility patterns of different types of agents are available. In this sense, the division of a given population into disjoint groups can be made attending different qualitative or quantitative indicators such as nationality, age, gender or economic incomes. In our case, we characterize the spread

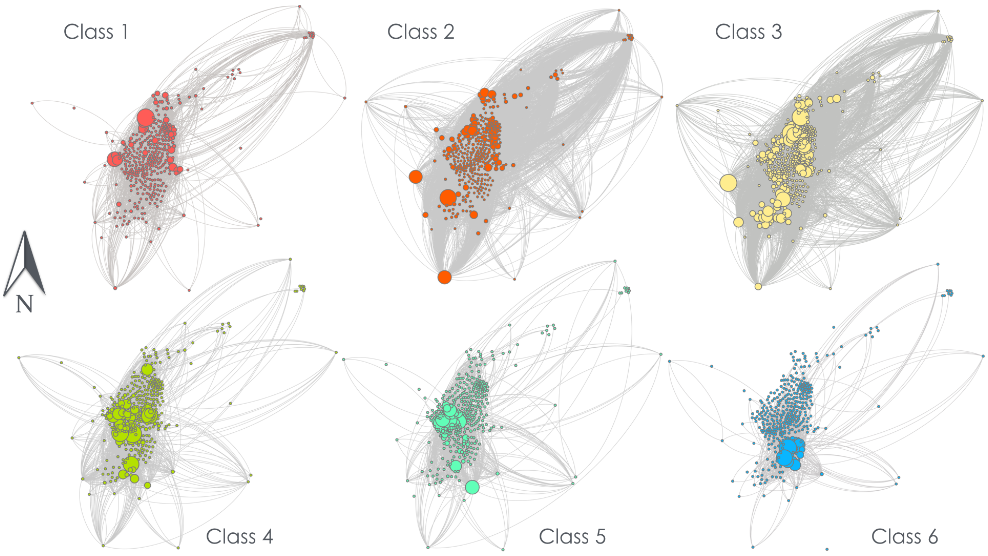


Figure 4.10: Mobility patterns and demographical distribution of each socio-economic stratum in the city of Medellín. The size of the patch is proportional to the number of residents whereas the edge thickness encodes the entries of the origin destination matrices **R**.

of diseases across Colombian cities, for Colombian society is stratified in six socio-economic groups according to the quality of the household of each individual, ranging from stratum 1 typically gathering those with lowest incomes to stratum 6, corresponding to the wealthiest individuals. Important differences concerning the timeline, distance and transportation modes involved in the commuting trips of each stratum have been already reported in the literature [211, 243]; we are interested in addressing the impact of these differences on the spread of contagious diseases.

Let us focus on the city of Medellín, which is the second most populated city in the country with roughly 3.5 million of inhabitants. We use the data published in [211, 243] to construct the multiplex metapopulation of Medellín, with $N_P = 413$ patches and $L = 6$ layers, each one identified with one stratum; the different layers are represented in Fig. 4.10. For the sake of simplicity, we focus on the simple scenario $\lambda^{\alpha\beta} = \lambda$ and represent the evolution of the epidemic size as a function of the mobility and the infectivity in Fig. 4.11. Remarkably, the infection curves allow us to get some insightful information about the underlying mobility network. The first striking result is that the underlying population distribution is pretty heterogeneous, for the values of the epidemic size shown in Fig. 4.11a are well below the prediction of mean field theories for the SIS model, providing a $\tilde{\lambda}$ ($\mathcal{R}_0$) value.

Furthermore, the evolution shown by the epidemic threshold differs not

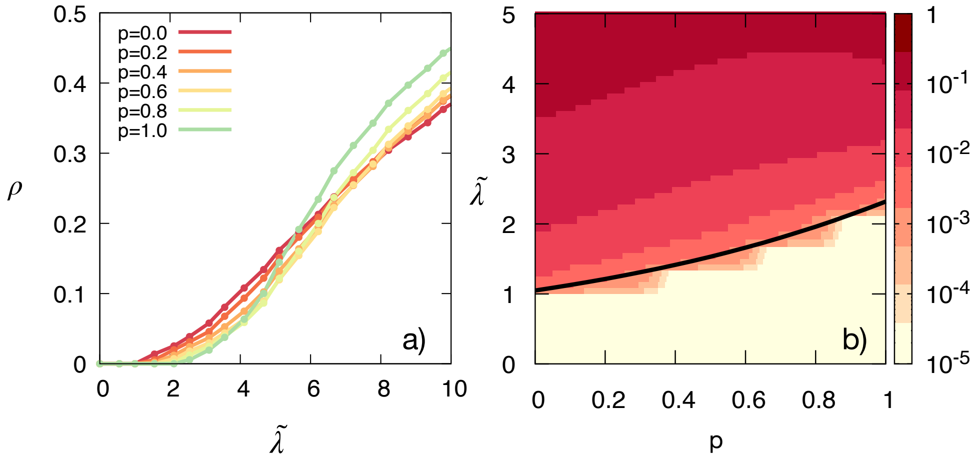


Figure 4.11: Panel a): Epidemic size of a SIS dynamics ρ as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility (color code). Dots show the results obtained by averaging 50 realizations of Monte Carlo simulations whereas solid lines correspond to the endemic equilibrium obtained by the iteration of the formalism. Panel b): Surface $\rho(\tilde{\lambda}, \mu)$ along with the analytical estimation for the epidemic threshold provided by Eq. (4.47).

only quantitatively but also qualitatively from the threshold in Cali. Note that, the threshold displays a monotonic increasing trend, meaning that mobility is always detrimental to the spread of diseases. It is also worth stressing that the crossing points in the supercritical regimes between the curves associated with high mobility p and those with low mobility p take place at high infectivity values. This result suggests that, in the case of Medellín, contact heterogeneities during the reaction phase are more relevant than the mixing of the populations driven by mobility to characterize the course of a disease.

One question arising from the previous figures is whether the impact of diseases is homogeneous across the different socio-economic strata. To answer this question, we represent in Figs. 4.12a-b the epidemic curves of each individual layer for low (Panel a) and high (Panel b) mobility values. There we observe that, when epidemic spreading is totally determined by the demographical distribution (low mobility), important differences emerge between the curves of the poorest and the richest strata, which are considerably shrunk when increasing mobility.

Both results clearly point out the existence of an important demographical segregation in Medellín; in particular, the smaller threshold displayed by low strata suggest that they occupy more crowded residential areas. Despite the differences in terms of the mobility patterns observed among strata, the mixing promoted by mobility is higher than that given by the residential dis-

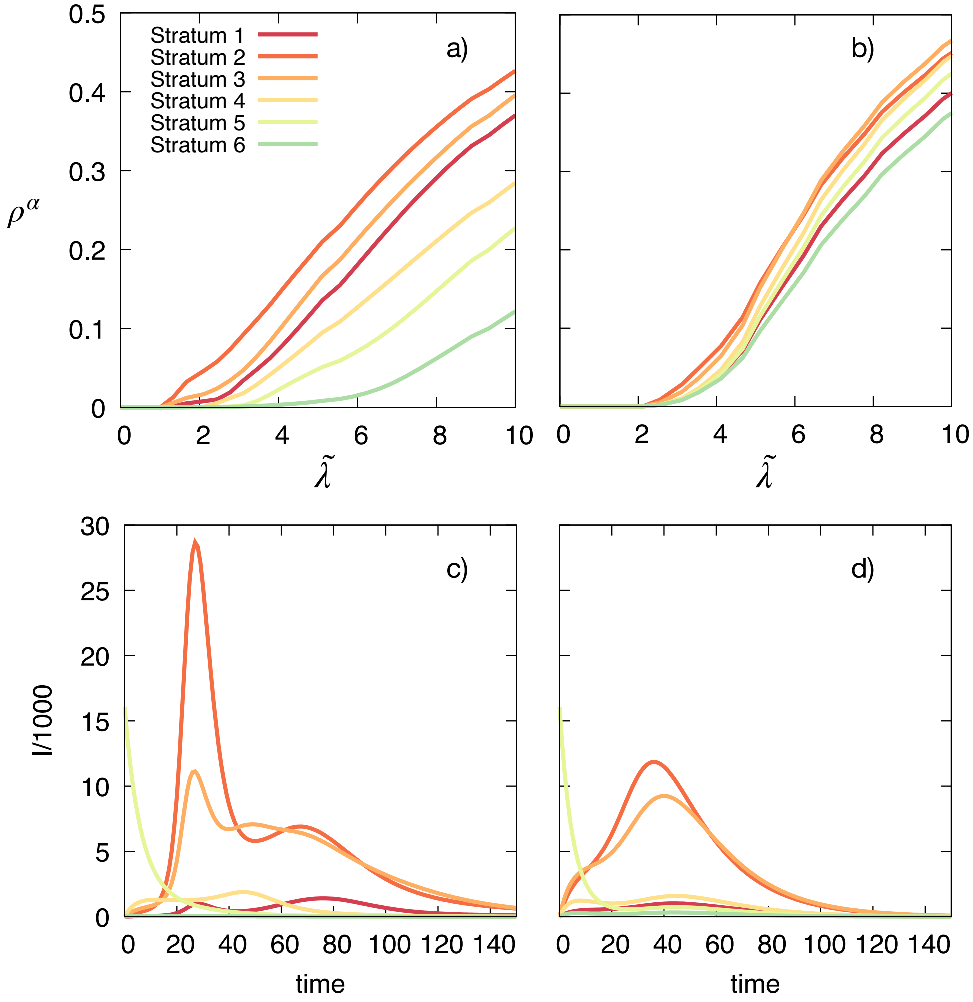


Figure 4.12: Panels a)-b): Epidemic size of a SIS dynamics in each socio-economic stratum α (color code) as a function of the rescaled infectivity $\tilde{\lambda}$ by setting $p = 0.05$ (Panel a) and $p = 1$ (Panel b). Panels c)-d): Time evolution of the number of infected individuals within a SIR dynamics with $p = 0.05$ (Panel c) and $p = 1$ (Panel d). In both scenarios, the epidemic start with a 15% of infected individuals inside stratum 5 and we fix the recovery rate to $\mu = 0.2$ and the rescaled infectivity to $\tilde{\lambda} = 4$.

tribution, which leads to the homogenization of the epidemic curves for high mobility values. More quantitatively, we observe that, despite the emergence of a global epidemic detriment affecting the populations a whole, increasing mobility has different outcomes as a function of the socio-economic stratum. Namely, mobility hinders epidemic spreading inside the lowest strata while promote contagions in the highest ones.

To shed more light on the segregation found for low mobility values, we start an outbreak inside stratum 5 and analyze the transient dynamics, i.e. the time evolution of the number of infected individuals in each stratum, in Fig. 4.10c. There we observe that contagions between different layers do not occur randomly but instead take place hierarchically such that the disease is almost exclusively transmitted between close strata. This hierarchy vanishes when activating mobility, as shown in Fig. 4.10d. Unlike the original SIR model, in both cases, secondary outbreaks are observed within the different strata arising from the interplay between mobility and demographic structures.

Finally, we aim to account for the impact of homophily, which recurrently appears in social networks [244]. Homophily is defined as the higher likelihood of interacting with those sharing similar cultural, social or economic features. In the context of the multilayer formalism here proposed, the homophily is easily introduced by strengthening intra-layer ties while weakening inter-layer interactions. Mathematically,

$$\lambda^{\alpha\beta} = \begin{cases} \lambda & \text{if } \alpha = \beta \\ \gamma\lambda & \text{otherwise} \end{cases} , \tag{4.50}$$

with $\gamma \in [0, 1]$. Therefore, γ is a parameter encoding the strength of the homophily introduced in the population.

Let us first explore how homophily shapes the epidemic threshold. For this purpose, we introduce the new definition of $\lambda^{\alpha\beta}$ into Eq. (4.45) so that the epidemic threshold can be expressed as

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathcal{M}')} , \tag{4.51}$$

where the elements of the new matrix $\mathcal{M}'$ are given by

$$\left(\mathcal{M}'\right)_{ij}^{\alpha\beta} = \Big(1 - (1-\gamma)(1-\delta^{\alpha\beta})\Big)\, \mathcal{M}_{ij}^{\alpha\beta} . \tag{4.52}$$

Fig. 4.13a represents the dependence of the epidemic threshold as a function of the mobility and the homophily parameter. As expected, promoting the homophily by reducing γ leads to a consistent decrease in the epidemic threshold as a result of the lower number of interactions made by the population. In its turn, at first sight, Fig. 4.13a suggests that increasing the mobility always has a detrimental effect regardless of the γ value.

We confirm this intuition in Fig. 4.13b where we explicitly plot the value of the threshold as a function of the mobility for different γ values. Despite the apparent lack of new physics, a closer examination of the curves reveals the existence of abrupt changes in the slope of the leading eigenvalue for low γ values encoding homophilic populations. This fact pinpoints the existence

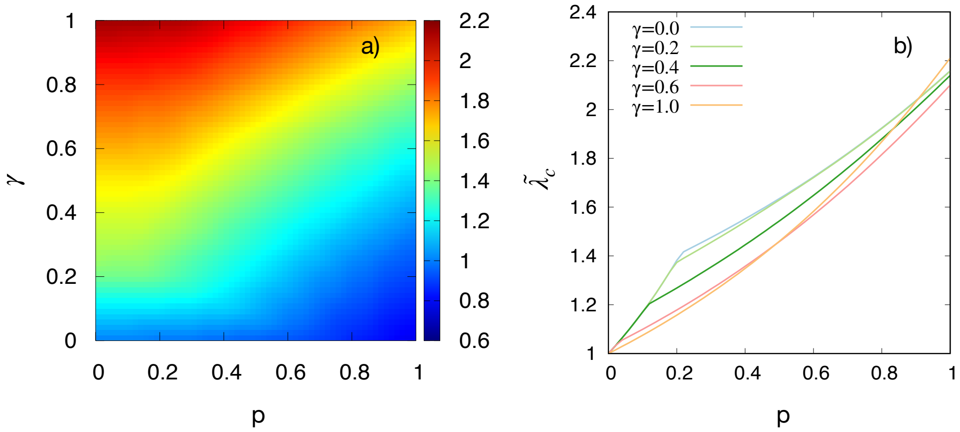


Figure 4.13: Panel a) Evolution of the epidemic threshold $\tilde{\lambda}_c$ as a function of the mobility p and the homophily γ. Note that the epidemic threshold has been rescaled by the value in the static case without homophily so that $\tilde{\lambda}_c(p,\gamma) = \lambda_c(p,\gamma)/\lambda_c(0,1)$. Panel b) Dependence of the epidemic threshold on the mobility p for different homophily values γ. To focus on the qualitative dependence, the rescaled threshold is computed as the ratio between the actual threshold and the one associated with the static case for each γ value. Therefore $\tilde{\lambda}_c(p,\gamma) = \lambda_c(p,\gamma)/\lambda_c(0,\gamma)$.

of a collision between the first and the second eigenvalues of matrix $\mathcal{M}'$, a phenomenon which has been reported for other dynamics running on top of multiplex networks and which is responsible for interesting physics such as superdiffusion processes [245].

In our case, the collision of the two leading eigenvalues has profound implications for the implications of targeted control policies acting on specific areas. Let us recall that the components of the leading eigenvector $\mathbf{v_{max}}$ reflect the fraction of infected individuals belonging to each patch and each layer close to the epidemic threshold. To focus on the spatial distribution of cases, we can define a new vector $\mathbf{V_{max}}$ obtained by coarse-graining the leading eigenvector by summing all the contributions of each layer to the same area i. Therefore

$$(\mathbf{V}_{\max})_i = \frac{\sum_{\alpha=1}^{L} x_i^\alpha \left(\mathbf{v}_{\max}\right)_i^\alpha}{\sqrt{\sum_{j=1}^{N}\left[\sum_{\alpha=1}^{L} x_j^\alpha \left(\mathbf{v}_{\max}\right)_j^\alpha\right]^2}}\ , \tag{4.53}$$

where x_i^α accounts for the fraction of residents in i belonging to the layer α and the denominator is introduced to normalize $\mathbf{V_{max}}$.

Fig. 4.14a illustrates how mobility shapes the components of the coarse-grained eigenvector $\mathbf{V_{max}}$. We observe an abrupt change in the localization of the eigenvector taking place around $p = 0.21$. As anticipated above, this fact tells us that important changes in the spatial distribution of cases occur for

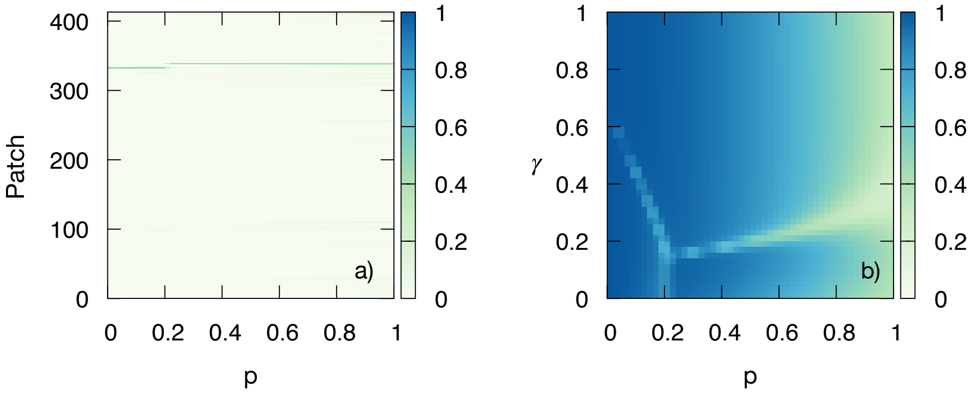


Figure 4.14: Panel a) Component decomposition of the leading eigenvector $\mathbf{V_{max}}$ as a function of the mobility p in the strong homophily regime $\gamma = 0$. Panel b) Inverse participation ratio Y of the leading eigenvector as a function of the mobility p and the homophily γ.

small variations in the mobility parameter, which manifest the need of accurately knowing in advance the demographical and epidemiological conditions of the system before implementing any targeted control policy.

We round off this analysis by analyzing the dependence on both the homophily and the mobility of the inverse participation ratio Y, which quantifies the degree of localization of a normalized eigenvector. In particular, the inverse participation ratio of $\mathbf{V_{max}}$ is given by

$$Y = \sum_{i=1}^{N_P} (V_{max})_i^4 \; . \tag{4.54}$$

From the latter expression, it is straightforward to realize that the inverse participation ratio is bounded so that $Y \in [1/N_P, 1]$. The lower bound corresponds to a completely delocalized eigenvector with all the components adopting the same value ($(V_{max})_i = 1/\sqrt{N_P}$). On the contrary, the upper bound reflects a localized state where the eigenvector is one of the elements of the canonical basis.

Fig. 4.14 represents the inverse participation ratio as a function of the homophily and the mobility. For low mobility values, we observe $Y \simeq 1$, which means that outbreaks close to the threshold affect a few patches. Nonetheless, for low γ values (strong homophily) the localization of the outbreaks changes, as shown by the abrupt variation of the inverse participation ratio reflecting the transitions between different patches. Note that, in contrast, no changes occur for high γ values, denoting the relevance of the complex interplay between socio-economic stratification, mobility and demographics for the emergence of this phenomenon.

4.4 Conclusions

The great development of the international mobility network over the last fifty years has dramatically changed the course of pandemics. The removal of the physical gaps separating faraway regions driven by this phenomenon has lead to an inversion of the time scales responsible for the local and global diffusion of outbreaks. In this sense, while network science has already shed light on the role of mobility in the global diffusion of epidemics, determining exactly how our daily mobility patterns shapes the local contagions remains a more open question, given the limitations of the theoretical frameworks existing in the literature.

In this chapter, we have tackled this problem and proposed a theoretical framework, the MIR model, which allows for readily incorporating data about the demographical distribution and the mobility patterns of the inhabitants of real cities. To do so, we have abandoned the use of contact networks and introduced the concept of metapopulations, which are composed of different subpopulations and constitute the basic skeleton needed to account simultaneously for mobility and epidemic spreading. Afterwards, we have proposed the MMCA equations of the formalism, which enable to quantify the spatio-temporal evolution of the occupation of the different epidemiological states in each subpopulation. These equations represent an important computational advantage with respect to the agent-based simulations in which each microscopic contagion and mobility process should be simulated to obtain the evolution of the disease.

From a theoretical point of view, the linearization of the equations leads to an analytical expression of the epidemic threshold which allows for capturing the impact of the mobility patterns of the population and the underlying population distribution on the onset of diseases. Interestingly, this impact in real metapopulations like Santiago de Cali in Colombia substantially differs from the results for synthetic metapopulations. Specifically, the complexity of the formalism allows us to discover a hitherto unreported phenomenon, the so-called epidemic detriment, for which an increase in the mobility can be beneficial to contain an ongoing epidemic outbreak. The emergence of such phenomenon is characterized mathematically following a perturbative approach and is rooted in the existence of heterogeneous subpopulations. The physical roots and some heuristic rules to control the epidemic detriment are explored in depth in Chapter 5.

To refine the MIR model, we have relaxed the assumption about the statistical equivalence of the individuals in each subpopulation and accounted for the individuals diversity reflected in several factors such as age, race or socio-economic status. To do so, we have introduced a new object, the multiplex metapopulation, which corresponds to a multiplex network where each layer encode the mobility patterns and the spatial distribution of one type of individuals across the same set of patches. Following a similar rationale as in the original MIR model, we have proposed a set of Markovian equations to track the epidemiological evolution of the different type of individuals and derive the behavior of the epidemic threshold as mobility varies, showing different quantitative trends to the ones observed when neglecting the underlying socio-economic diversity.

Finally, we have characterized the spread of diseases across the city of Medellín in Colombia, leveraging the fact that Colombian society is stratified into six different socio-economic classes which are embedded into the different layers of the multiplex metapopulation. In this more realistic setup, the epidemic detriment still emerges and is even stronger than the observed for the city of Cali, as shown by the continuous increasing trend of the epidemic threshold with the mobility. Interestingly, despite the global epidemic detriment, the effect of mobility on the individual threshold associated to each socio-economic class is not universal but remarkable qualitative and quantitative differences are observed across the different strata.

Furthermore, the analysis of the epidemic trajectories have allowed us to get some insights into the features characterizing the underlying demographical distribution. Therefore, we have observed that the residential areas of the different classes are highly spatially segregated, as reflected in the different individual thresholds and the existence of a hierarchical contagion structure between close socio-economic classes. To round off our analysis, we have introduced the mechanism of social homophily, which strengthens the interactions with those sharing similar social features. Remarkably, we have observed that small variations in the mobility in strongly homophilic populations can lead to abrupt changes in the most affected areas and therefore, can compromise the efficiency of targeted control policies.

The MIR model represents the minimal framework needed to capture the evolution of reaction-diffusion process while not making any assumptions about the structure of the mobility network. Apart from the spread of diseases, in upcoming works we seek to generalize the MIR model to accommodate other reaction processes, such as the adoption of cultural traits in hunter-gatherer societies [246, 247] or the exchange of knowledge across different countries [248].

Chapter 5

Epidemic detriment with recurrent mobility

> *If your new theorem can be stated with great simplicity, then there will exist a pathological exception.*
>
> A. Mathesis [249].

The analysis of the impact of recurrent mobility patterns on the onset of epidemics made in the previous chapter reveals that the epidemic threshold exhibits a complex dependence on human mobility which varies as a function of the system under study. In particular, when studying real metapopulations, we report a counterintuitive phenomenon, the so-called epidemic detriment, which consists in the beneficial effect of letting the population move across a given city to contain an ongoing epidemic outbreak.

Far from being an artifact from the equations, we have shown that the epidemic detriment also emerges in the Monte Carlo simulations for both SIS and SIR dynamics and proven mathematically its existence via a perturbative analysis performed on the critical matrix determining the value of the epidemic threshold. Nonetheless, we have little intuition yet about how the features of the metapopulation, i.e. the interplay between the spatial distribution of the population and their mobility patterns, shape this phenomenon.

In this chapter we aim at shedding light on the roots of the aforementioned epidemic detriment. To do so, we first propose in Section 5.1 a toy-model, the star-like metapopulation, which allows us to systematically tune different aspects such as population heterogeneities or flows distribution and quantify their impact on the epidemic threshold. In this section, we check that macroscopic arguments based, for example, on the redistribution of the population driven by mobility cannot justify the emergence of the epidemic detriment in all epidemic scenarios.

Motivated by this fact and to overcome some limitations of the original MIR model, we break the symmetry between the time scales associated with agents' movements and epidemiological processes and introduce in Section 5.2 a new parameter, the return time, encoding how much time individuals spent at the destination before returning to the residence patch. We check that the new time scale has a deep impact on the dynamics by changing quantitatively the behavior of the system even in situations in which the population interacting inside each patch is independent of this parameter. The latter finding suggests that the epidemic detriment is not only determined by the number of interactions that each individual makes but also depends on whom these contacts are made with.

Finally, a typical shortcoming of the metapopulation models is the mean field assumption introduced within each subpopulation. This automatically precludes accounting for the existence of heterogeneous contact structures inside each patch. In Section 5.3 we introduce a theoretical extension of the MIR model which solves this problem by introducing different degree classes in each subpopulation. Unlike the previous formalism, in this case the number of contacts made by each individual remains constant and independent of the location where he/she is placed. The emergence of the epidemic detriment in this scenario allows us to understand the microscopic roots of this phenomenon and its relation with other findings existing in the literature when addressing epidemic spreading on contact networks.

5.1 Epidemic detriment and MIR model

5.1.1 Types of cities

The results shown in Chapter 4 prove that the interplay between mobility and epidemics is not universal but strongly depends on the features of the underlying metapopulation. In particular, we have characterized three different types of metapopulations according to the dependence of the epidemic threshold on the mobility of the population p. In what follows, we denote these categories as Type I,II and III respectively.

- Type I: In these metapopulations, increasing mobility always favors epidemic spreading, which is reflected in a monotonic reduction of the epidemic threshold $\tilde{\lambda}_c$ when increasing the mobility p. Synthetic metapopulations whose nodes are uniformly populated fall into this category.
- Type II: In this case, the epidemic threshold shows a nontrivial behavior with the mobility. For small p values, the epidemic threshold increases with the mobility, giving rise to the so-called epidemic detriment. This trend is maintained until a given value of the mobility, p^*, for which the

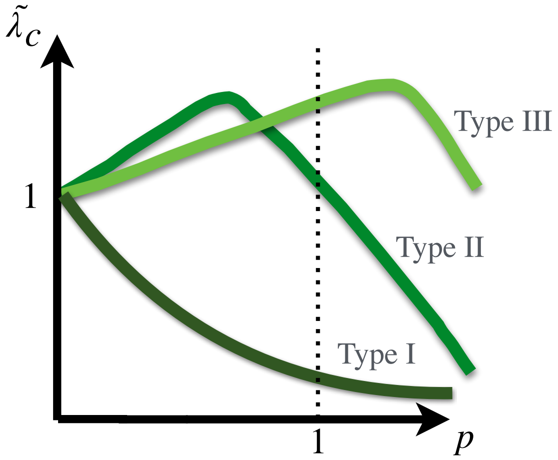


Figure 5.1: Evolution of the rescaled epidemic threshold $\tilde{\lambda}$ as a function of the mobility p for the three different qualitative behaviors observed so far. Type I cities show a continuous reduction of the threshold while increasing p, so mobility boosts epidemic spreading. The epidemic detriment, for which increasing mobility hinders the spread of diseases, is non-monotonic for Type II cities, since it appears for small p values and vanishes after the peak located at p^*, and monotonic for Type III cities.

epidemic threshold reaches its peak. Afterwards, the impact of mobility is reversed and promoting movements boosts epidemic spreading and hence decreases the epidemic threshold. The city of Cali in Colombia constitutes one example of type II metapopulation.

- Type III: Finally, there are other cases, like the city of Medellín in Colombia, where the epidemic detriment appears consistently from the static case $p = 0$ to the case in which the entire population moves according to the mobility network $p = 1$.

The three different qualitative behaviors of the epidemic threshold are captured in Fig. 5.1. The perturbative analysis made in Section 4.2 reveals that the epidemic detriment always emerges, partially or totally, in metapopulations with heterogeneous subpopulations, regardless of the mobility network, thus restricting type I cities to uniformly populated cities. Despite the mathematical characterization of the epidemic detriment, we lack a clear connection between the features of the metapopulation and the qualitative impact of mobility on the onset of epidemics.

5.1.2 Star-like metapopulation

To explain the mechanisms responsible for the emergence of the epidemic detriment, we introduce a new theoretical benchmark, the star-like city. Star networks are simple graphs whose topology is composed of a central node, usually referred to as *hub*, which is linked to a set of peripheral nodes, the so-called *leaves*. The main advantage of these simple configurations is the statistical equivalence of all the leaves, which greatly simplifies the theoretical analysis of different dynamics. For example, the use of star networks has been crucial to analytically characterize the nature of different phenomena such as the onset of SIS diseases [18, 19] and the microscopic origin of the explosive synchronization [131] or the oscillations in chemical reaction-diffusion processes [250].

Fig. 5.2a schematically depicts the structure of the star-like city. In what follows, we denote in the mathematical expressions the hub and the leaves by the subscripts h and l respectively. Note that the metapopulation is fully characterized by four parameters:

1. n_{max}: Number of residents inside the hub so that $n_h = n_{max}$.
2. κ: Number of leaves surrounding the hub. Note that the flows from the central node to the periphery are uniformly distributed so that $R_{hl} = 1/\kappa$.
3. α: Population asymmetry between the hub and the leaves. In this sense, each leaf is populated by $n_l = \alpha n_{max}$ agents with $\alpha \in [0, 1]$.
4. δ: Fraction of the flows departing from the leaves which end in the hub, i.e. $R_{lh} = \delta$. The remaining fraction of the moving population goes to the next leaf in the counterclockwise direction so that $R_{l,l+1} = 1 - \delta$.

At this point, let us recall the expression determining the fraction of individuals in the infected state associated with each patch close to the epidemic threshold. According to Eq. (4.22), we have that

$$\vec{\epsilon} = \mathbf{M}\vec{\epsilon}\,, \tag{5.1}$$

where the elements of the matrix $\mathbf{M}$ in the limit of all-to-all interactions are given by

$$M_{ij} = n_j \left[\delta_{ij}(1-p)^2 + p(1-p)\left(R_{ij} + R_{ji}\right) + p^2(\mathbf{R}\mathbf{R}^{\mathbf{T}})_{ij}\right] \,. \tag{5.2}$$

Note that, in a general scenario, the vector $\vec{\epsilon}$ has as many components as the number of patches N_P in the metapopulation. Nonetheless, the statistical equivalence of the leaves allows us to reduce the dimensionality of the system from N_P to 2 equations, governing the evolution of the disease inside the hub and a generic leaf respectively. Imposing this condition in Eq. (5.1) and unfolding the components of $\vec{\epsilon}$, we obtain

$$\epsilon_h = M_{hh}\epsilon_h + M_{hl}\epsilon_l \ , \tag{5.3}$$

$$\epsilon_l = M_{lh}\epsilon_h + M_{ll}\epsilon_l \ . \tag{5.4}$$

The simplicity of the metapopulation enables to get a straightforward expression for the elements of the matrix. Namely,

$$M_{hh} = n_{max}\left[(1-p)^2 + p^2/\kappa\right] \ , \tag{5.5}$$

$$M_{hl} = \kappa\alpha n_{max}\left[p(1-p)(1/\kappa+\delta) + p^2(1-\delta)/\kappa\right] \ , \tag{5.6}$$

$$M_{lh} = n_{max}\left[p(1-p)(1/\kappa+\delta) + p^2(1-\delta)/\kappa\right] \ , \tag{5.7}$$

$$M_{ll} = \alpha n_{max}\left[(1-p)^2 + 2p(1-p)(1-\delta) + p^2\left(\kappa\delta^2 + (1-\delta)^2\right)\right] \ . \tag{5.8}$$

Note that the matrix $\mathbf{M}$ is a 2×2 matrix, whose leading eigenvalue can be computed analytically. This allows us for getting a simple expression for the epidemic threshold, which now reads as

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})} \ , \tag{5.9}$$

with

$$\Lambda_{\max}(\mathbf{M}) = \frac{\mathrm{tr}(\mathbf{M}) + \sqrt{\mathrm{tr}(\mathbf{M})^2 - 4\det\mathbf{M}}}{2} \ . \tag{5.10}$$

The expression for the epidemic threshold depends on five parameters, the four characterizing the properties of the underlying metapopulation and the recovery rate μ. To focus on the effect of the mobility, we study the rescaled epidemic threshold

$$\tilde{\lambda}_c = \frac{\lambda_c}{\lambda_c(p=0)} = \frac{n_{max}}{\Lambda_{\max}(\mathbf{M})} \ . \tag{5.11}$$

Note that all the elements of the matrix $\mathbf{M}$ depend linearly on the population of the hub n_{max}, which makes the rescaled epidemic threshold independent of this parameter. In plain words, the qualitative impact of the mobility is not a variable of the overall number of agents in the metapopulation but is only influenced by the population asymmetries governed by α and the features of the mobility network determined by (κ, δ). To check whether we can reproduce the three qualitative types of cities observed thus far with the star-like configuration, we explore the mobility value p^* at which the peak of the epidemic threshold is located. Note that $p^* \to 0$ corresponds to type I cities which display a monotonic decrease of the threshold, $p^* \geq 1$ identifies type III cities with a monotonic increase of the threshold and any p^* value in between these extremes is associated with the non-monotonic trend found for type II cities.

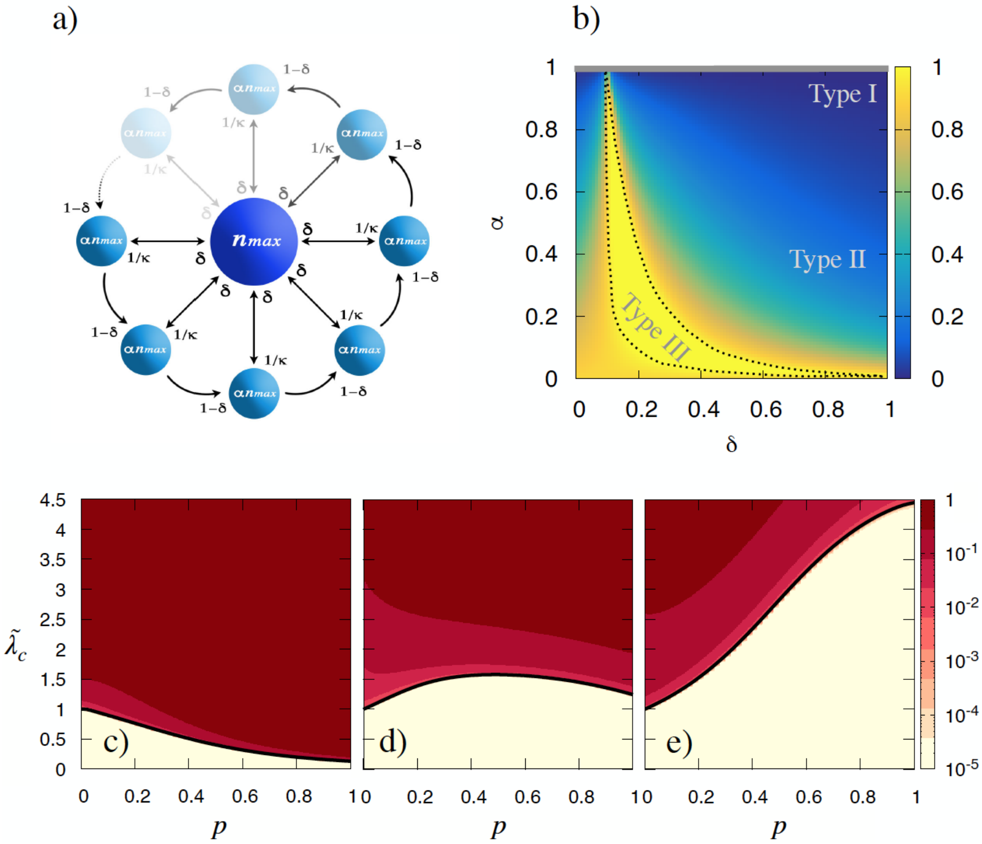


Figure 5.2: Panel a) Schematic representation of the star-like metapopulation. The relevant parameters are the population of the hub n_{max}, the population asymmetry α, the number of leaves κ and the flow from leaves to the hub δ (see text for details). Panel b) p^* (color code) as a function of (α, δ) for a star-like metapopulation with $\kappa = 10$ leaves. Panels c)-e): Epidemic size ρ (color code) as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility p. The solid black line depicts the analytical estimation for the rescaled epidemic threshold provided by Eq. (5.11).

Without loss of generality, let us fix $\kappa = 10$ leaves and represent the value of p^* as a function of (α, δ) in Fig. 5.2b. Note that type I cities are restricted to the areas of the parameters' space with $\alpha \simeq 1$ and therefore to metapopulations with negligible population asymmetries, as suggested by the perturbative analysis included in Chapter 4. The rest of the phase diagram shows a non-trivial boundary between type II and type III cities. In particular, most of the possible configurations correspond to type II cities whereas type III cities are associated with a very specific region of the parameters' space. In Fig. 5.2c-e we illustrate the epidemic diagrams for three different combinations of (α, δ) identifying the three possible types of cities.

Once we have checked that the star-like metapopulation enables to retrieve the three different types of cities observed thus far, we now try to connect the dots between the structural features of the metapopulation and the p^* value. We start by recalling that in the MIR model the number of contacts each individual makes is an extensive variable which depends on the attributes of the place where he/she is located. In particular, under the all-to-all interactions scheme, this number of contacts matches the effective population of the different patches.

Intuitively, one plausible explanation for the epidemic detriment would be that mobility tends to homogenize the underlying population distribution, thus reducing the number of individuals interacting inside the most vulnerable node and, therefore, requiring a higher infectivity to observe a macroscopic outbreak. In the star-like metapopulation, the effective populations of both the hub and the generic leaf read as

$$\tilde{n}_h = n_{max}\left[(1-p) + \alpha\kappa p\delta\right] \ , \tag{5.12}$$

$$\tilde{n}_l = n_{max}\left[\alpha(1-p\delta) + p/\kappa\right] \ . \tag{5.13}$$

Equating both equations, we can estimate the value at which the effective population inside the leaf becomes higher than the one crowded in the hub, denoted in what follows as p^*_{heur}, yielding:

$$p^*_{heur} = \frac{\kappa(1-\alpha)}{(\kappa+1)(1-\delta\alpha k)} \ . \tag{5.14}$$

Fig 5.3a shows the dependence of the latter indicator as a function of the population asymmetry α and the flows from leaves to the hub, governed by δ. Interestingly, two clearly differentiated regions appear in this figure:

- $k\delta\alpha \leq 1$: In this region of the parameters' space, mobility tends to reduce the population gathered inside the hub and increase the population in the leaves, as shown in Fig 5.3b. Remarkably, the value predicted relying on the homogenization of the population is able to capture very accurately the actual value of p^* represented in Fig. 5.2.
- $k\delta\alpha > 1$: In this case, the value derived for p^*_{heur} is always negative, for the distribution of population is no longer homogenizing. Instead mobility tends to accumulate individuals inside the hub at the expense of emptying the leaves, as captured in Fig 5.3c. Therefore, the epidemic detriment observed in Fig. 5.2 and universally predicted by the perturbative analysis cannot be justified based on the redistribution of the population.

The failure of macroscopic arguments based on the sizes of the interacting populations suggests that the epidemic detriment, and more in general the

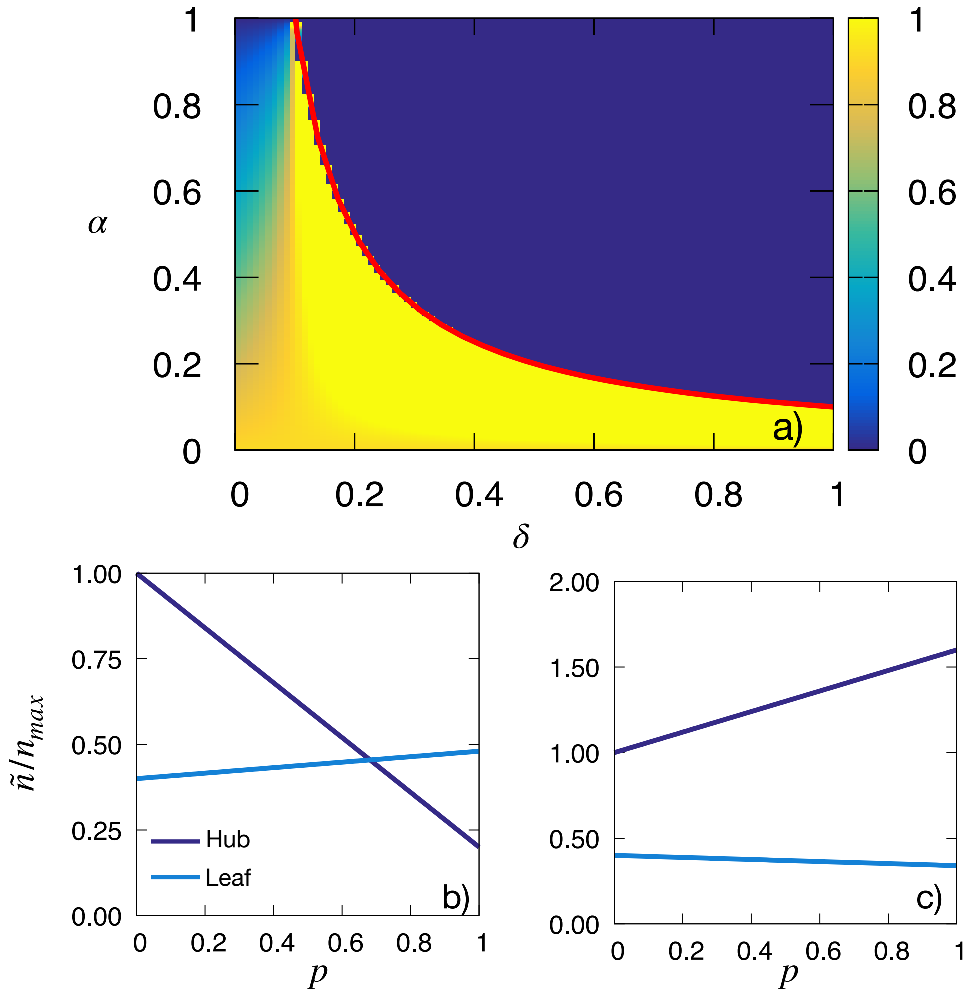


Figure 5.3: Panel a) Value of the heuristic estimation for the position of the peak of the epidemic threshold p^*_{heur} (color code) as a function of the population asymmetry α and the parameter δ governing the flow from the leaves to the hub. The solid red line depicts the function $\alpha = 1/(k\delta)$, which delimit the region where macroscopic arguments work to explain the epidemic detriment (see text for details). Panels b)-c): Dependence of the effective population interacting inside the hub $\tilde{n}_h$ and the leaves $\tilde{n}_l$ on the mobility p for two star-like metapopulations with $(\alpha, \delta) = (0.4, 0.05)$ (Panel b) and $(\alpha, \delta) = (0.4, 0.4)$ (Panel c) respectively. In all the cases, the number of leaves is $\kappa = 10$.

interplay between recurrent mobility and epidemics, are driven by processes taking place across microscopic/mesoscopic scales. In the next sections, we propose two theoretical extensions of the MIR model and provide solid evidence about the crucial role that the modification of the agents' contact structures plays in the emergence of the epidemic detriment.

5.2 MIR model with tunable return time

Upon this point, we have reflected the recurrent nature of human movements in our formalism by forcing all the individuals to come back to their associated patch at the end of each time step. While this hypothesis seems coherent and resembles our daily habits, theoretically it makes us to restrict to a very specific scenario where the time scales governing both contagion/recovery processes and mobility are assumed to be equal.

In this section, we abandon this assumption and generalize our formalism by introducing a new time scale, denoted in what follows as τ, which encodes the number of epidemic time steps that agents spend outside their residence before returning to their associated patch. As explain in Chapter 4, this time scale has been already introduced in the previous HMF approaches existing in the literature [196, 213, 214, 251, 252] and leads to important variations of the epidemic curves. In our case, we are interested in checking whether this time scale can shed light on the origin of the epidemic detriment.

We start by describing the extended version of the MIR model, which is schematized in Fig. 5.4. The symmetry breaking between the time scales involved in the model entails that the initial population of each patch at each time step is not only composed of residents but also contains visitors who have not come back to their respective residences yet. Therefore, we must redefine the rules governing the three stages of each time step.

1. Movements: At the beginning of each time step, the residents in one patch decide whether to leave it with a probability p or remain there with probability $1-p$. To respect the commuting nature of human movements and the new time scale, visitors in one patch are not allowed to move. Mathematically, we introduce a third-order tensor $\mathcal{R}$, whose elements $\mathcal{R}^i_{jk}$ denote the probability that a resident inside i moves from j to k. Note that, in general, the tensor $\mathcal{R}$ allows for accommodating a wide variety of mobility schemes ranging from higher-order paths [253] to random walker dynamics. In our case of interest, the assumptions described above lead to

   $$\mathcal{R}^i_{jk} = \delta^i_j R_{jk} + (1-\delta^i_j)\delta_{jk} \ , \tag{5.15}$$

   where $\mathbf{R}$ is the OD matrix used in the original MIR model.

2. Interaction: This stage remains as in the original MIR model. In this case, both residents and visitors are indistinguishable and make the same number of contacts depending on the attributes of the place where they are located after the movement stage. In each contact between susceptible and infected individuals, the first one contracts the disease with

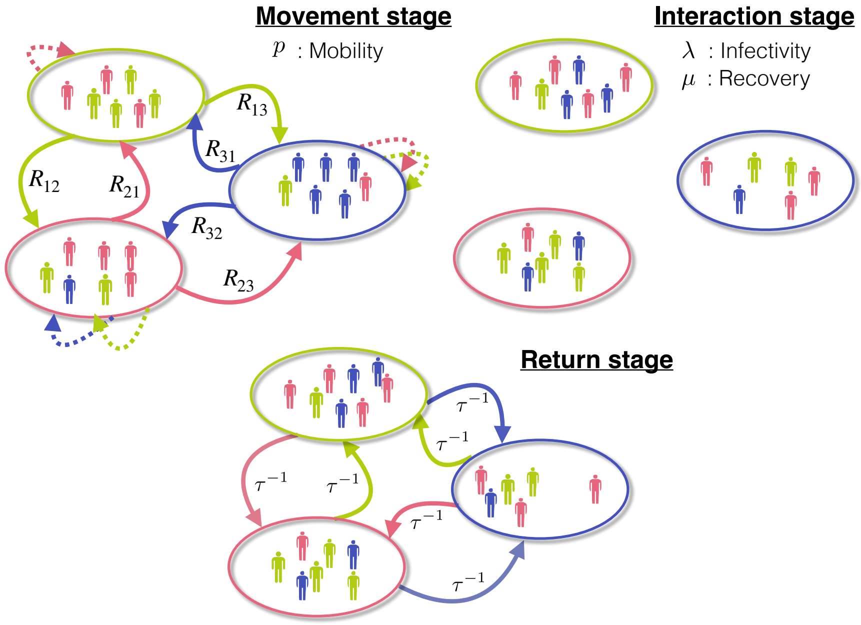


Figure 5.4: Schematic representation of the generalized MIR model. The agents color encodes their residence. Each time step involves three stages: *i*) Movement stage governed by the 3-dimensional tensor $\mathcal{R}$ computed according to Eq. 5.15. At this point, visitors outside their residence are forced to stay at their destination (dashed lines) whereas agents located at their residence are allowed to move according to the degree of mobility p and the OD matrix $\mathbf{R}$. *ii*) Interaction stage in which agents update their dynamical state according to the contagious rate λ and the recovery one μ. *iii*) Return stage where agents outside their residence can come back there with a probability τ^{-1}.

probability λ. Likewise, infected individuals can overcome the disease with a given probability μ.

3. Return: Finally, visitors decide whether returning home with probability τ^{-1} or staying at their destination with probability $1 - \tau^{-1}$.

5.2.1 Model equations

We start by assuming that our metapopulation has N_P patches, each with N^i residents. The introduction of the new time scale makes the population associated with a given patch behave differently depending on whether or not they are located inside their residence. This fact automatically breaks the statistical equivalence of all the residents in a given patch and forces us to enlarge the number of equations needed to fully characterize the evolution of SIS diseases across the metapopulation. In particular, rather than the N_P equations composing the MIR model, we need L equations which correspond with each of the non-zero entries recorded in the OD matrix $\mathbf{R}$.

On top of the equations governing the spread of the pathogen, we need to account simultaneously for the time evolution of the occupation of each patch. For this purpose, let us denote by $n^i_j(t)$ the number of individuals with residence in i who are placed inside patch j at time t. Following the assumptions of our formalism, the time evolution of these quantities is given by

$$\begin{aligned} n^i_j(t+1) = & \left(1-\delta^i_j\right)\left(1-\tau^{-1}\right)\left(pR_{ij}n^i_i(t)+n^i_j(t)\right) \\ & + \delta^i_j\left[\left(1-p\left(1-R_{ii}\right)\right)n^i_i(t)+\tau^{-1}\sum_{k\neq i}\left(pR_{ik}n^i_i(t)+n^i_k(t)\right)\right] . \end{aligned} \tag{5.16}$$

The first line identifies the individuals staying at a destination different from their residential node. This can happen if these individuals move towards the destination or in case they were already there; in both cases, they decide not to come back home later on. In its turn, the second line encodes the number of individuals inside their residential patch. These agents either were already there and decide not to move or return home after visiting a different destination.

Theoretically, we should iterate concomitantly both equations governing the distribution of the population and the course of the disease. Notwithstanding, we must recall that the stationary state of a SIS dynamics is an endemic equilibrium independent of the transient dynamics. To simplify our analysis, we then assume that the epidemic time scale is much longer that the thermalization time for the redistribution of the population. Therefore, we can fix the different populations as their stationary values.

To find them, we must assume that $n^i_j(t+1)=n^i_j(t)=n^i_j\ \forall(i,j)$. Taking into account that the sum of these variables should be equal to the number of residents in each patch i, N^i, we obtain the steady values for the population distribution:

$$n^i_j = \left(1-\delta^i_j\right)\frac{N^iR_{ij}}{1-R_{ii}+\frac{\tau^{-1}}{p(1-\tau^{-1})}} + \delta^i_j\frac{N^i}{1+\frac{p(1-R_{ii})(1-\tau^{-1})}{\tau^{-1}}} . \tag{5.17}$$

Note that $\tau^{-1}=1$ recovers the original MIR model, for $n^i_j=\delta^i_jN^i$. In contrast, $\tau^{-1}=0$ identifies a scenario where agents forget their associated patch and are distributed among the neighboring patches according to the OD matrix $\mathbf{R}$ so that $N^i_j=(1-\delta^i_j)N^iR_{ij}/(1-R_{ii})$.

Under the microscopic rules of this new model, the evolution of the number of infected individuals with residence inside i and placed in j, denoted as ρ^i_j, is given by

$$\begin{aligned}\rho^i_j(t+1) =& \delta^i_j[1-\mu]\left[\left(1-p\left(1-R_{ii}\right)\right)\rho^i_i(t)+\tau^{-1}\sum_{l\neq j}\left(pR_{il}\rho^i_i(t)+\rho^i_l(t)\right)\right]\\ &+\delta^i_j\left(1-p\left(1-R_{ii}\right)\right)\left(n^i_i-\rho^i_i(t)\right)P_i(t)\\ &+\delta^i_j\tau^{-1}\sum_{l\neq j}P_l(t)\left(pR_{il}\left(n^i_i-\rho^i_i(t)\right)+\left(n^i_l-\rho^i_l(t)\right)\right)\\ &+\left(1-\delta^i_j\right)\left(1-\mu\right)\left(1-\tau^{-1}\right)\left(pR_{ij}\rho^i_i(t)+\rho^i_j(t)\right)\\ &+\left(1-\delta^i_j\right)\left(1-\tau^{-1}\right)P_j(t)\left(pR_{ij}\left(n^i_i-\rho^i_i(t)\right)+\left(n^i_j-\rho^i_j(t)\right)\right)\ .\end{aligned} \tag{5.18}$$

The first three lines contain the evolution of the infected individuals located in the same patch where they reside. Specifically, the first line contains those individuals already infected who do not recover at time t and end up inside their residential patch. The second and third lines govern the contagions of susceptible individuals which can happen if the individual does not move and contract the disease inside i (second line) or if he/she moves, contracts the disease in the destination and decide to come back home (third line).

Finally, the last two lines correspond to the evolution of infected individuals who are located inside a patch different from their residence. The fourth line corresponds to those who were already infected and neither recover from the disease nor return to their node and the fifth one encodes those susceptible individuals contracting the disease inside the destination and not returning home.

Assuming a well-mixed scenario and an all-to-all interaction scheme within each patch, the probability for a susceptible individual contracting the disease inside patch i at time t reads as

$$P_i(t)=1-\left(1-\lambda\frac{\tilde{\rho}_i(t)}{\tilde{n}_i}\right)^{\tilde{n}_i}\ , \tag{5.19}$$

where $\tilde{\rho}_i(t)$ and $\tilde{n}_i$ are the number of infected individuals and the population placed inside i at time t after the movement stage respectively. In particular,

$$\tilde{\rho}_i(t)=\sum_{j=1}^{N_P}\delta^i_j\left(1-p\left(1-R_{ii}\right)\right)\rho^i_i(t)+\left(1-\delta^i_j\right)\left(\rho^j_i+pR_{ji}\rho^j_j(t)\right)\ , \tag{5.20}$$

$$\tilde{n}_i=\sum_{j=1}^{N_P}\delta^i_j\left(1-p\left(1-R_{ii}\right)\right)n^i_i+\left(1-\delta^i_j\right)\left(n^j_i+pR_{ji}n^j_j\right)\ . \tag{5.21}$$

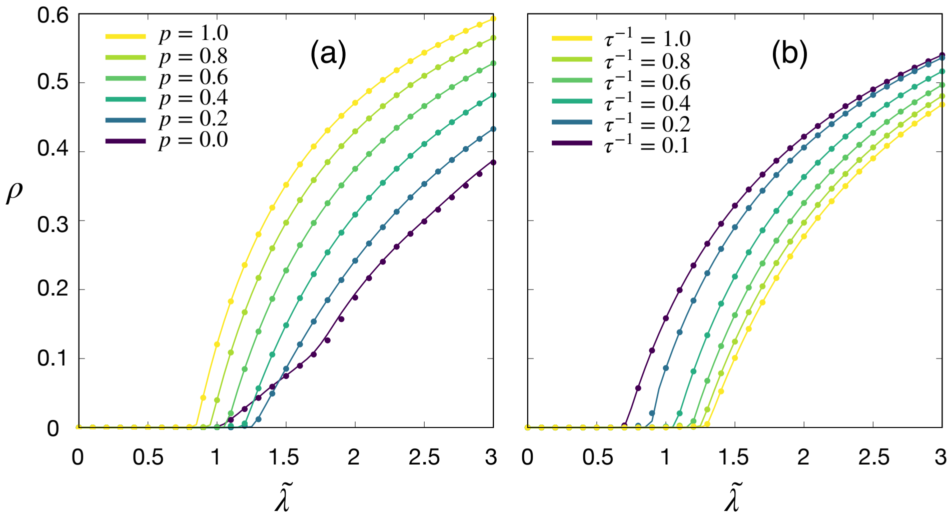


Figure 5.5: Epidemic size ρ as a function of the rescaled infectivity $\tilde{\lambda}$ when varying the mobility p (Panel a) and the permanence times τ (Panel b). The permanence time has been fixed to $\tau^{-1} = 0.5$ in Panel a) whereas the mobility is set to $p = 0.5$ in Panel b). In both cases, solid lines represent the results obtained by iterating Eqs. (5.18)-(5.21) whereas dots show the average of the results obtained from 100 Monte Carlo simulations for each $(p,\tilde{\lambda}))$ value. In both panels, the recovery probability is fixed to $\mu = 0.2$. The underlying metapopulation is a BA network with $N_P = 50$ patches and $\langle k \rangle = 4$ and populations randomly assigned so that $N^i \in [300, 1700]$.

5.2.2 Model validation

First, we aim at validating the equations by comparing their predictions with results from Monte Carlo simulations. To do so, let us consider a BA network with $N_P = 50$ patches and $\langle k \rangle = 4$. Nodes' populations are drawn from an uniform distribution so that $N^i \in [300, 1700]$. We first study the evolution of the epidemic size as a function of the infectivity and the mobility when agents are not forced to come back home after each time step. Based on the variables of the model, the epidemic size is now computed as

$$\rho = \frac{\sum\limits_{i=1}^{N_P} \sum\limits_{j \in k(i)} n_j^i \rho_j^i(\infty)}{\sum\limits_{i=1}^{N_P} N^i} , \tag{5.22}$$

where $k(i)$ represents the neighborhood of patch i. Fig. 5.5 shows that the equations of the extended model fairly reproduce the results from the simulations. Qualitatively, in Fig. 5.5a we observe that the epidemic detriment phenomenon still appears when introducing slightly longer trips. Notwithstanding, Fig. 5.5b shows that increasing the permanence times (decreasing

τ^{-1}) always lead to a reduction in the epidemic threshold, which promotes epidemic spreading. Therefore, there seems to be a competition between two opposite mechanisms, the one driven by mobility tending to increase the epidemic threshold and its decrease caused by increasing the permanence times, whose trade-off governs the onset of diseases.

5.2.3 Epidemic threshold

We now seek to derive an analytical expression for the epidemic threshold to shed more light on the aforementioned antagonistic relation between mobility and permanence times. For this purpose, we try to follow a similar approach as in the original model and assume that the disease has reached an endemic equilibrium characterized by a negligible number of infected individuals. Mathematically, this implies that

$$\rho_j^i(t+1) = \rho_j^i(t) = \rho_j^i \ll n_j^i \ . \tag{5.23}$$

The latter assumption allows us to linearize the equations of the formalism. Specifically, Eq. (5.19) turns into

$$P_i^* \simeq \lambda \tilde{\rho}_i \ . \tag{5.24}$$

Plugging this value into Eq. (5.18) and neglecting $\mathcal{O}(\epsilon^2)$, we characterize the infected individuals close to the threshold for each possible residence-destination (i, j) pair. Let us focus on the case $i = j$, for which we obtain

$$\mu \rho_i^i = p\,(1 - R_{ii})\,(1-\mu)\left(1 - \tau^{-1}\right)\rho_i^i + \sum_{l \neq i}(1-\mu)\tau^{-1}\rho_l^i + \lambda \sum_{l=1}^{N_P}\sum_{m=1}^{N_P} F_{im}^{il}\rho_m^l \ . \tag{5.25}$$

Note that, unlike the rest of examples studied throughout this thesis, the latter equations do not allow us to isolate the infectivity λ from the elements of the critical matrix. This is caused by the terms involving a redistribution of the infected individuals across the different patches without any change in their epidemiological state.

Therefore, we cannot get a closed estimation for the epidemic threshold by using the typical linearization technique. Instead, to estimated this quantity, we here resort to the next-generation matrix method (NGM) [56, 254, 255]. In the context of mathematical epidemiology, the NGM method allows getting an analytical estimation of the basic reproduction number $\mathcal{R}_0$ from the ordinary differential equations (ODEs) governing the evolution of a given disease.

To apply the NGM, we need to turn our discrete-time equations into continuous ones and explore the proximities of the disease-free state. To do so, we define $\dot{\rho}_i^j = \rho_i^j(t+1) - \rho_i^j(t)$ and assume $\rho_i^j \ll n_i^j$, which turn Eq. (5.18) into

$$\dot{\rho}_j^i = \sum_{l=1}^{N_P} \sum_{m=1}^{N_P} \left(\lambda F_{jm}^{il} - V_{jm}^{il}\right) \rho_m^l \ , \tag{5.26}$$

with

$$\begin{aligned} F_{jm}^{il} = \delta_j^i \Bigg\{ & n_i^i \left[1 - p_i'\right] \left[\delta^{il} \delta_m^i \left(1 - p_l'\right) + \left(1 - \delta^{il}\right) \left[pR_{li} \delta_m^l + \delta_{mi}\right]\right] \\ & + \tau^{-1} \left(1 - \delta_m^i\right) \delta_m^l \left(R_{im} + n_m^i\right) \left(1 - p(1 - R_{lm})\right) \\ & + \tau^{-1} \delta_m^l \sum_{o=1}^{N_P} \left(1 - \delta_o^i\right) \left(pR_{io} n_i^i + n_o^i\right) pR_{lo} \\ & + \tau^{-1} \left(1 - \delta_m^i\right) \left(1 - \delta_m^l\right) \left(pR_{im} n_i^i + n_m^i\right) \Bigg\} \\ + (1 - \delta_j^i)(1 - \tau^{-1})(pR_{ij} n_i^i + n_j^i) & \left[\delta_j^l \delta_m^l (1 - p_k') + (1 - \delta_m^l)(pR_{lj} \delta_m^l + \delta_{mj})\right] \end{aligned} \tag{5.27}$$

and

$$\begin{aligned} V_{jm}^{il} = & \ \delta_j^i \delta^{li} \left[\delta_m^i \left(\mu + p_i'(1 - \mu)(1 - \tau^{-1})\right) - (1 - \delta_m^i)(1 - \mu)\tau^{-1}\right] \\ & + (1 - \delta_j^i) \left[\delta^{il} \delta_{jm} (\mu + \tau^{-1}(1 - \mu) - \delta_m^l (1 - \mu)(1 - \tau^{-1}) pR_{lm}\right] \ . \end{aligned} \tag{5.28}$$

In both expressions, we have denoted $p_i' = p(1 - R_{ii})$ for the sake of readability. Note that the tensor **F** gathers all the contagions of susceptible population leading to an increase of the epidemic size whereas tensor **V** contains the transitions between the infectious states. The NGM method states that the basic reproduction number is given by

$$\mathcal{R}_0 = \Lambda_{\max}(\lambda \mathbf{F} \mathbf{V}^{-\mathbf{1}}) = \lambda \Lambda_{\max}(\mathbf{F} \mathbf{V}^{-\mathbf{1}}) \ . \tag{5.29}$$

Recalling that $\mathcal{R}_0$ encodes the number of contagions made by one infectious individual in a fully-susceptible population, the value of the epidemic threshold corresponds to the λ_c value such that $\mathcal{R}_0(\lambda_c) = 1$. Therefore,

$$\lambda_c = \frac{1}{\Lambda_{\max}(\mathbf{F} \mathbf{V}^{-\mathbf{1}})} \tag{5.30}$$

Fig. 5.6 shows that the latter expression accurately captures the value of the epidemic threshold in the scenarios explored in Fig. 5.5 and reproduces the qualitative description previously made on the impact of both the mobility rate and the trip duration.

Beyond generalizing the original MIR model by breaking the time symmetry between contagions and movements or getting an analytical expression

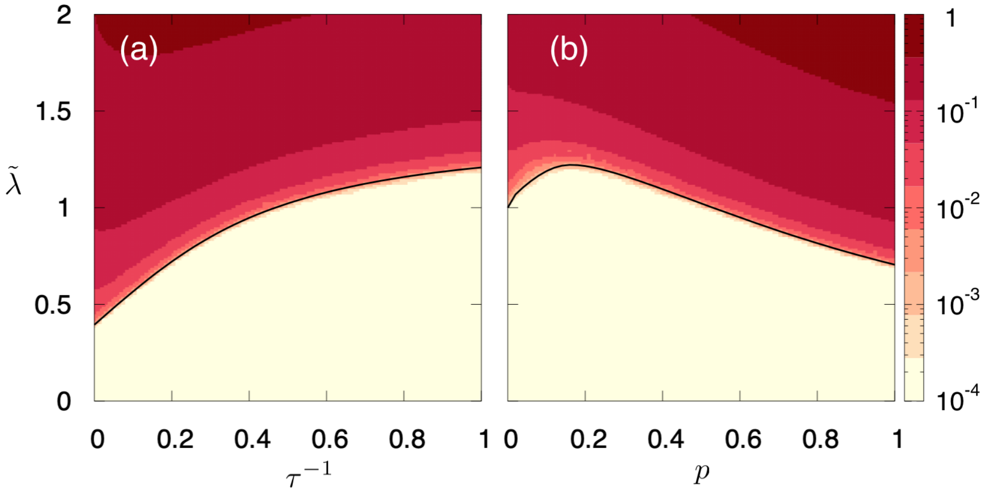


Figure 5.6: Epidemic size ρ (color code) for the BA metapopulation as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility p for $\tau^{-1} = 0.5$ (Panel a) and the trip duration τ for $p = 0.5$ (Panel b). The black solid lines represent the analytical estimation for the epidemic threshold computed from Eq. 5.30. In both panels, the recovery rate is set to $\mu = 0.2$.

for the threshold, here we aim at understanding the origins of the epidemic detriment phenomenon. To do so, we recover the star-like metapopulation introduced in the previous section and set $\alpha = \delta = 0.4$ and $\kappa = 10$ leaves. It is worth remembering that macroscopic arguments based on the redistribution of the population cannot justify the emergence of the detriment in such configuration, for mobility tends to accumulate agents inside the hub.

We represent the evolution of the epidemic threshold $\tilde{\lambda}_c$ as a function of the mobility p and the trip duration τ in Fig. 5.7a. There we observe a striking result: the epidemic detriment vanishes when increasing the permanence times at the destination. Specifically, in Fig. 5.7b we explicitly observe that increasing τ (reducing τ^{-1}) always decreases the value of p^* from the one observed in the original MIR model to $p^* = 0$, at which the epidemic detriment is no longer observed. Apart from the anticipation of the peak p^*, we observe that the effect of mobility is boosted as we increase the new time scale introduced, which is reflected in the higher slope of the threshold for small p values.

Macroscopically, both phenomena suggest that increasing the permanence times comprises the curves of the original MIR model and introduces an effective mobility higher than the actual mobility p. Nonetheless, this argument fails in capturing the behavior of the system for $p = 1$. Note that the effective

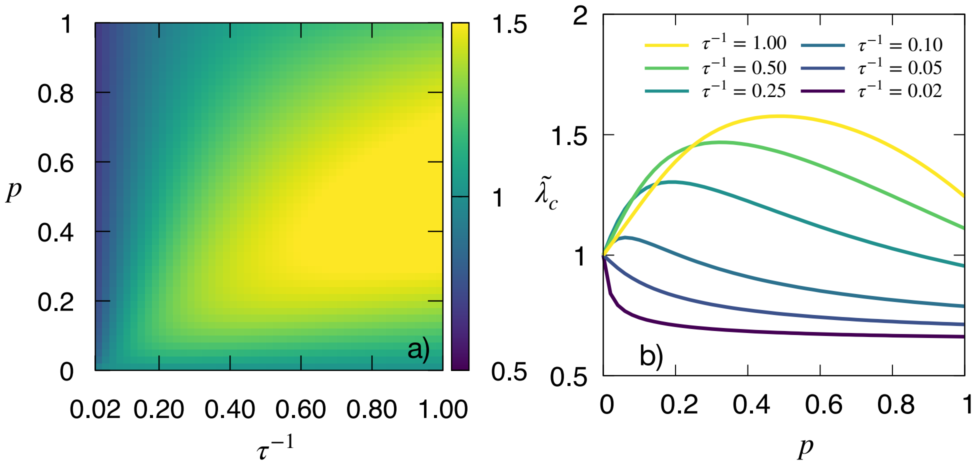


Figure 5.7: Panel a): Rescaled epidemic threshold $\tilde{\lambda}_c$ (color code) as a function of the mobility parameter p and the return probability τ^{-1}. Panel b): Rescaled epidemic threshold as a function of the mobility for different return probabilities (color code). In both panels the recovery rate is fixed to $\mu = 0.2$ and the underlying metapopulation is a star like configuration with $\kappa = 10$ leaves and $(\alpha, \delta) = (0.4, 0.4)$.

population of the hub in this case reads

$$\tilde{n}_h = \alpha k \delta \ , \tag{5.31}$$

which is independent of the permanence time. However, the epidemic threshold is significantly reduced when extending the trip duration; this is the most clear example showing that the interplay between recurrent mobility and epidemics goes beyond the redistribution of the population.

Let us address the role of the architecture of the mobility network. For this purpose, we represent in Fig. 5.8 the same curves for two metapopulations which differ one from another in the parameter δ governing the share of flows going from the leaves to the hub. Interestingly, when $\delta = 1$, so the individuals of the leaf always move to the hub, the value of the rescaled epidemic threshold at $p = 1$ does not depend on the permanence times. In contrast, as stated before, this dependence does appear if $\delta = 0.4$, corresponding to a scenario in which agents from the leaf can either visit the hub or another leaf.

Note that the variability of the possible destinations visited by one individual is crucial to understand the impact of mobility on epidemic spreading. Although the interacting populations remain constant and independent of the permanence time for $p = 1$, the recurrent nature of mobility allow each single individual to escape from the hub, where most contagions occur, and visit a leaf where they rarely get infected. This mechanism is boosted with high return rates to the residence patch, which explains the consistent increase of the

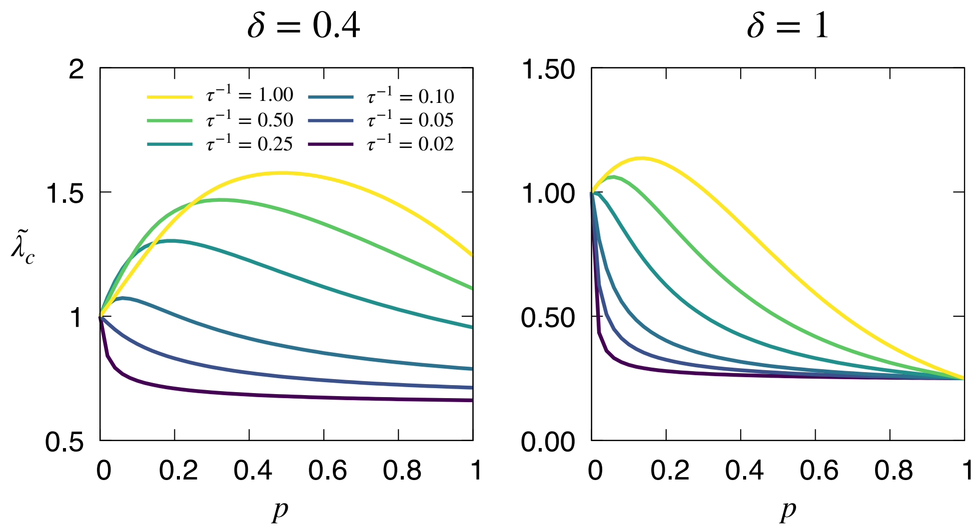


Figure 5.8: Rescaled epidemic threshold $\tilde{\lambda}_c$ as a function of the mobility parameter p and the return probability τ^{-1} (color code).The recovery rate is fixed to $\mu = 0.2$ and the underlying metapopulation is a star like configuration with $\kappa = 10$ leaves and $(\alpha, \delta) = (0.4, 0.4)$ and $(\alpha, \delta) = (0.4, 1)$.

epidemic threshold when decreasing the permanence times τ.

In light of the prominent role of the time scales, we explore the impact of modifying the infectious window of infected individuals, typically encoded by the inverse of the recovery rate μ^{-1}. For this purpose, let us fix the return rate to $\tau^{-1} = 0.2$ and study the curves $\tilde{\lambda}_c(p)$ as a function of the recovery rate μ in Fig. 5.9a. Unlike the original MIR model, we observe that varying the infectious period alters the impact of mobility on epidemic spreading. In particular, when $\mu^{-1} \ll \tau$, the infectious individual typically stays at one destination during the infectious period, which hinders the epidemic detriment and potentiate the negative effect provided by agents' accumulation inside the hub. In contrast, long infectious windows allow them to move and visit less vulnerable areas, which reduces the number of contagions and give rise to the epidemic detriment.

Finally, to further explore the interplay between the two time scales, we fix the mobility rate to $p = 1$ and represent in Fig. 5.9b the evolution of the rescaled epidemic threshold as a function of the permanence time for different infectious windows. The steeper slope of the epidemic threshold for small infectious periods confirms that, rather than the absolute value of the permanence time τ, its relative value compared with the typical infectious window is key to understand how the trip duration hinders the epidemic detriment.

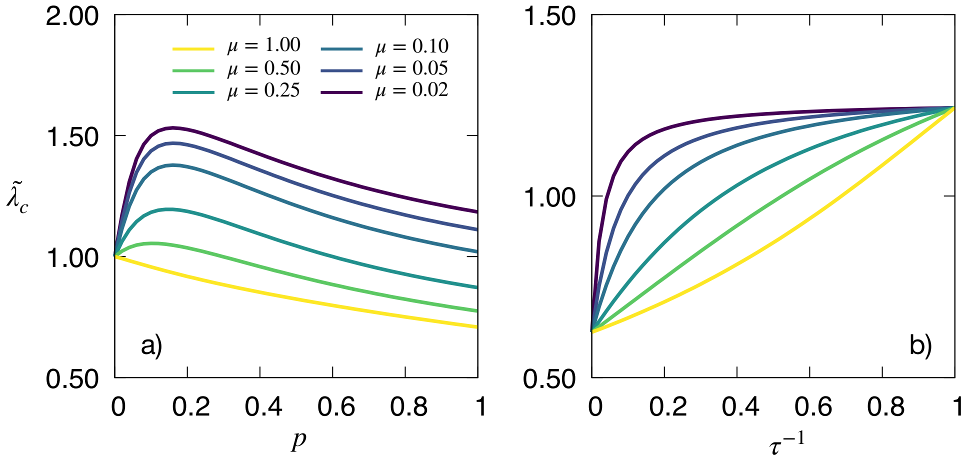


Figure 5.9: Panel a): Rescaled epidemic threshold $\tilde{\lambda}_c$ as a function of the mobility parameter p and the recovery rate μ (color code). The return probability is fixed to $\tau^{-1} = 0.2$. Panel b): Rescaled epidemic threshold $\tilde{\lambda}_c$ as a function of the return probability τ^{-1} and the recovery rate μ (color code). The mobility rate is fixed to $p = 1$. In both panels, the underlying metapopulation is a star like configuration with $\kappa = 10$ leaves and $(\alpha, \delta) = (0.4, 0.4)$.

5.3 MIR model with heterogeneous contact patterns

The introduction of the permanence times in the MIR model reveals that the variation of the contact structures of the population plays a crucial role in the emergence of the epidemic detriment. However, note that, so far we have considered the number of contacts of each agent to be an extensive variable depending on the population gathered in each geographical area. This way, the effect of changing the set of individuals with whom an agent interacts is intermingled with the variation of the overall number of contacts. In general, this fact makes it difficult to attribute the main responsible mechanism for the epidemic detriment.

Furthermore, one of the most important limitations of the frameworks based on metapopulations is that they usually introduce mean field assumptions consisting in considering all the individuals of a given subpopulation as statistically equivalent agents. A few examples can be found in the literature relaxing the latter assumption to accommodate different types of individuals within the same subpopulation [256, 257]. Nonetheless, in all of them, it is assumed that individuals make the same number of interactions within each group, thus hindering the crucial role of heterogeneous contact distributions in epidemic spreading, as found in the models on contact networks explained in Chapter 2.

To overcome both limitations, in this section we propose a new extension of the MIR model in which we assume that each individual carries an attribute, the degree or number of contacts k, which is not altered when visiting different locations from her residential node. Moreover, we abandon the usual mean-field assumption and instead follow a heterogenous mean field (HMF) approach where the individuals from the same residential patch are divided into different degree classes, whose members are assumed to be equivalent.

5.3.1 Model equations

Let us assume that our population is composed of N_P patches, each one populated by n_i $(i = 1, \ldots, N_P)$ residents, giving rise to an overall population of N individuals across the metapopulation. As stated before, we assume that each individual has a given degree k, which encodes his/her number of daily interactions[1], and that individuals with different degree can coexist within the same subpopulation. In particular, we denote by Ω_i the number of different degree classes inside the patch i and define $P_i(k)$ as the probability that a resident inside i has degree k.

Regarding the mobility model, we assume that each individual follows the Movement-Interaction-Return scheme, where the interaction stage is modified to account for the heterogenous nature of the contacts made by the population inside each location. For the sake of simplicity, we fix $\tau = 1$ and restrict the analysis to SIS diseases, although the formalism here proposed is easily generalizable to accommodate more complex compartmental models or account for the impact of trips duration.

Given the assumptions of the model, we need Ω equations to fully describe the evolution of the system, where Ω corresponds with the sum of the different degree classes coexisting within the same patch, so that $\Omega = \sum\limits_{i=1}^{N_P} \Omega_i$. Let us denote the fraction of infected individuals associated with patch i within degree class k at time t by $\rho_i^{[k]}(t)$. Under the dynamical rules previously defined, the time evolution of the latter quantities reads as

$$\rho_i^{[k]}(t+1) = (1-\mu)\rho_i^{[k]}(t) + \left[1 - \rho_i^{[k]}\right] \Pi_i^{[k]}(t) \ , \tag{5.32}$$

where $\Pi_i^{[k]}(t)$ encodes the probability that a susceptible agent with residence inside i and degree k becomes infected at time t. Assuming that individuals with degree k move with probability p_k according to the OD matrix $\mathbf{R}$, we

---

[1] In what follows, the degree refers to the number of contacts of each individual rather than to the number of connections of each subpopulation in the mobility network.

obtain

$$\Pi_i^{[k]}(t) = (1-p_k)\pi_i^{[k]}(t) + p_k \sum_{j=1}^{N} R_{ij}\pi_j^{[k]}(t) \ . \tag{5.33}$$

Note that, unlike the original MIR model, the probability of contracting the disease inside each patch is not the same for all the agents placed there. In particular, assuming an annealed regime where we do not fix any specific contact structure, the probability that a susceptible individual of degree k becomes infectious inside patch i at time t reads as

$$\pi_i^{[k]}(t) = 1 - \left(1 - \lambda \sum_{k'} \tilde{P}_i(k'|k)\tilde{\rho}_i^{[k']}(t)\right)^k \ . \tag{5.34}$$

In the latter expression, $\tilde{\rho}_i^{[k']}(t)$ denotes the fraction of individuals with degree k' inside i who are infectious at time t. Therefore,

$$\tilde{\rho}_i^{[k]}(t) = \frac{\tilde{I}_i^{[k]}(t)}{\tilde{n}_i^{[k]}} \ , \tag{5.35}$$

being $\tilde{I}_i^{[k]}(t)$ and $\tilde{n}_i^{[k]}$ the infected and total population with degree k inside patch i at time t respectively. Taking into account the mobility matrix and the original population distribution, we obtain

$$\tilde{I}_i^{[k]}(t) = \sum_{j=1}^{N_P} n_j P_j(k)\rho_j^{[k]}(t)\left(\delta_{ij}(1-p_k) + (1-\delta_{ij})p_k R_{ji}\right) \ , \tag{5.36}$$

$$\tilde{n}_i^{[k]} = \sum_{j=1}^{N_P} n_j P_j(k)\left(\delta_{ij}(1-p_k) + (1-\delta_{ij})p_k R_{ji}\right) \ . \tag{5.37}$$

In its turn, $\tilde{P}_i(k'|k)$ represents the probability that an individual of degree k sets a link with another individual with degree k' inside patch i. Neglecting degree-degree correlations, this probability reads as

$$\tilde{P}_i(k'|k) = \frac{k'\tilde{P}_i(k')}{\widetilde{\langle k\rangle}_i} = \frac{k'\tilde{n}_i^{[k']}}{\sum_{k''} k''\tilde{n}_i^{[k'']}} \ , \tag{5.38}$$

where we have defined the effective degree distribution of patch i, $\tilde{P}_i(k')$.

5.3.2 Heterogeneous star-like metapopulation

Here we present the generalization of the star-like metapopulation introduced in Section 5.1 to accommodate heterogeneous contact patterns. Specifically,

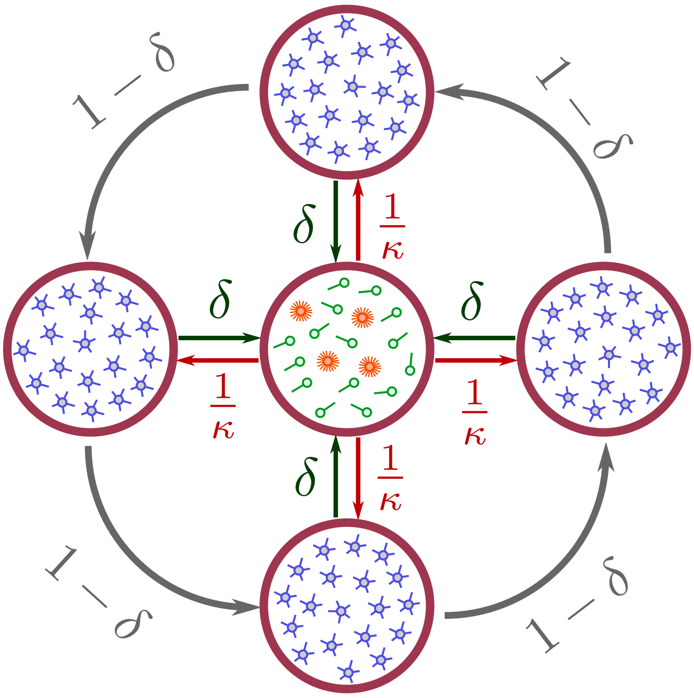


Figure 5.10: Example of a heterogeneous star-like metapopulation network with $N_P = 5$ patches. The bimodal distribution inside the hub is characterized by $k_{max} = 20$ and $\langle k \rangle_h = 5$. In addition, the average degree inside the leaf is fixed to $\langle k \rangle_l = \beta \langle k \rangle_h$, with $\beta = 1$. The population asymmetry is set to $\alpha = 1$.

we keep the population distribution which is governed by the asymmetry parameter α and the mobility network which is fully-determined by the number of leaves κ and their flows governed by δ. Regarding the degree distributions, we assume that all the individuals from the leaves make $\langle k \rangle_l$ contacts, so that the degree distribution is given by

$$P_l(k) = \delta_{k \langle k \rangle_l} \; . \tag{5.39}$$

In its turn, to introduce contact heterogeneities, we consider that the degrees of individuals from the hub follow a bimodal distribution, splitting the population into agents with a single contact and with k_{max} interactions respectively. Note that the share of population belonging to each group is governed by a new parameter denoted as η. Therefore,

$$P_h(k) = \eta + (1-\eta) k_{max} \; . \tag{5.40}$$

Note that the n-th moment of the hub's connectivity distribution is

$$\langle k^n \rangle_h = \sum_k k^n P_h(k) = \eta + (1-\eta) k_{\max}^n \; . \tag{5.41}$$

In particular, fixing the average degree as $\langle k\rangle_h$ automatically determines η, yielding

$$\eta = \frac{k_{\max} - \langle k\rangle_h}{k_{\max} - 1}. \tag{5.42}$$

It is worth remarking that the latter configuration introduces two flavors of contact heterogeneities. First, the bimodal distribution of the contacts made by hub' residents induces local contact heterogeneities there. Likewise, the possible different average degree between individuals from the hub and the leaves represent some sort of global heterogeneity, quantifying how different individuals from distinct locations are. To control this global heterogeneity, we assume that $\langle k\rangle_h = \beta\langle k\rangle_h$ with $\beta \in (0, 1]$. A schematic representation of the proposed toy-model configuration is shown in Fig. 5.10 with $k_{max} = 20$, $\langle k\rangle_h = 5$ and $\alpha = \beta = 1$.

5.3.3 Validation of the equations

Analogously to the rest of the models, first we aim at validating the proposed equations with Monte Carlo simulations. Regarding the different stages involved in one Monte Carlo step, the movement and return stages remain unaltered with respect to the original MIR model but the interaction stage is modified to account for the different contacts made by the population gathered in one patch.

Therefore, in the movement stage, agents with degree k move with a probability p_k and choose their destination according to the OD matrix $\mathbf{R}$. In the interacting stage, we assume that each susceptible individual with degree k makes k interactions. Assuming that the susceptible individual chooses randomly with whom she interacts, the probability of getting infected in each attempt is given by

$$\lambda \frac{\sum\limits_k k\tilde{I}_i^{[k]}}{\sum\limits_k k\tilde{n}_i^{[k]}} \ . \tag{5.43}$$

Note that we do not construct any specific networked configuration within each patch, for we work in the annealed regime [258]. Finally, at the return stage, each individual comes back home and another time step starts.

To focus on the effects on contact heterogeneities, unless specified we assume that both hubs and leaves are equally populated with $n_h = n_l = 10^4$ residents, i.e. $\alpha = 1$. Concerning the mobility, let us assume that the mobility rate is independent of the degree, i.e. $p_k = p$. Fig. 5.11 shows the evolution of the epidemic size ρ, i.e. the fraction of the population infected in the endemic

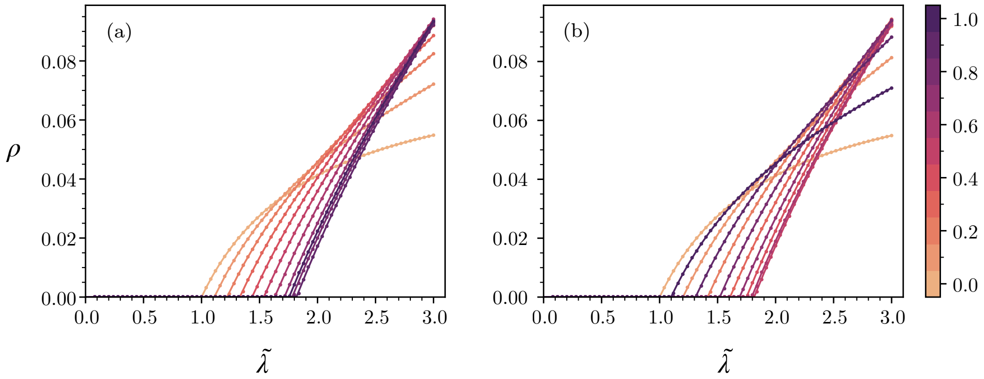


Figure 5.11: Epidemic size ρ as a function of the rescaled infectivity $\tilde{\lambda}$ and the mobility p (color code). Equilibrium regimes of the Markovian equations (lines) and MC simulations (dots) for a star-like metapopulation with $n_h = n_l = 10^4$ ($\alpha = 1$) and $\kappa = 10$. The hub contains individuals with connectivity $\langle k \rangle_h = 100$ ($\eta = 0$ and $k_{\max} = 100$), and the leaves $\langle k \rangle_l = 10$ ($\beta = 0.1$). The mobility patterns are given by (a) $\delta = 0.1$ and (b) 0.9. In both panels, solid lines represent the predictions yielded by the Markovian equations whereas dots correspond with the average epidemic size from 10 realizations of the Monte Carlo simulations. In both panels, the recovery rate is set to $\mu = 0.2$.

equilibrium, as a function of the mobility p and the rescaled infectivity $\tilde{\lambda}$ for two star-like metapopulations, one with $\delta = 0.1$ (Panel a) and another one with $\delta = 0.9$ (Panel b). Note that in this case, we compute the rescaled infectivity by dividing the actual infectivity λ by the expected threshold following the HMF approach in case of a set of spatially isolated patches. This way,

$$\tilde{\lambda} = \frac{\lambda}{\lambda_0} \tag{5.44}$$

with

$$\lambda_0 = \mu \min\{\langle k \rangle_h / \langle k^2 \rangle_h, \langle k \rangle_l / \langle k^2 \rangle_l\} \ . \tag{5.45}$$

In both cases, we neglect the local contact heterogeneities by setting $\eta = 0$ and explore the effect of global heterogeneities governed by $\beta = 0.1$. In this formalism, the epidemic size is computed as

$$\rho = \frac{\sum_{i=1}^{N_P} \sum_{k \in \{k_i\}} n_i P_i(k) \rho_i^{[k]}(\infty)}{N} \ . \tag{5.46}$$

The agreement between the Monte Carlo simulations and the predictions of the formalism is remarkable except some small discordances around the epidemic threshold caused by the difficulties posed by the combination of stochastic fluctuations and the existence of an absorbing state.

Qualitatively, the epidemic detriment emerges despite the uniform population distribution across the metapopulation. This is in sharp contrast with the phenomenology observed in the original MIR model when setting $\alpha = 1$, for in that scenario mobility always promotes epidemic spreading. More importantly, it is worth stressing that in this framework, at variance with the models previously presented, agents preserve their number of interactions regardless of the place to which they move. This way, mobility does not shape the number of interactions of the population and therefore the epidemic detriment is by no means explained by macroscopic arguments based on the redistribution of the population.

5.3.4 Epidemic threshold

To shed light on the interplay between contact heterogeneities and recurrent mobility, we now derive the expression of the epidemic threshold. Let us assume that the disease has reached the endemic equilibrium so that $\rho_i^{[k]}(t+1) = \rho_i^{[k]}(t) = \rho_i^{*[k]}$. In this case, Eq. (5.32) turns into

$$\mu \rho_i^{*[k]} = \left[1 - \rho_i^{*[k]}\right] \Pi_i^{*[k]} \tag{5.47}$$

with

$$\Pi_i^{*[k]} = \left[(1-p_k)\pi_i^{*[k]} + p_k \sum_{j=1}^{N} R_{ij} \pi_j^{*[k]}\right] . \tag{5.48}$$

Now we assume that, close to the epidemic threshold, the endemic equilibrium is characterized by a negligible fraction of infected individuals so that $\rho_i^{*[k]} = \bar{\epsilon}_{ik} \ll 1 \, \forall (i,k)$. The latter assumption allows linearizing the equations of the formalism. In particular, the probability for an individual with degree k contracting the disease inside i reads as

$$\pi_i^{*[k]} = \simeq \lambda k \sum_{k'} \tilde{P}_i(k'|k) \tilde{\rho}_i^{*[k']} . \tag{5.49}$$

Incorporating Eqs. (5.35) and (5.38), the latter expression reads

$$\pi_i^{*[k]} = \frac{\lambda k}{Q_i} \sum_{k'} k' \sum_{j} \left[(1-p_{k'})\, \delta_{ij} + p_{k'} R_{ji}\right] n_j P_j(k') \bar{\epsilon}_{jk'}, \tag{5.50}$$

where

$$Q_i \equiv \sum_{k} k \sum_{j} \left[(1-p_k)\delta_{ij} + p_k R_{ji}\right] n_j P_j(k) \tag{5.51}$$

is the effective number of edges in patch i. Finally, plugging the latter expression into Eq. (5.47) and neglecting $\mathcal{O}(\epsilon^2)$ terms, after some algebra we arrive to the following expression:

$$\mu\bar{\epsilon}_{ik} = \lambda \sum_j \sum_{k'} \bar{M}_{ik}^{jk'} \bar{\epsilon}_{jk'}, \tag{5.52}$$

that defines an eigenvalue problem with

$$\begin{aligned}\bar{M}_{ik}^{jk'} &= kk'P_j(k') \Bigg[(1-p_k)(1-p_{k'})\frac{\delta_{ij}}{Q_i} + (1-p_k)p_{k'}\frac{R_{ji}}{Q_i} + \\ & \qquad p_k(1-p_{k'})\frac{R_{ij}}{Q_j} + p_k p_{k'} \sum_l \frac{R_{il}R_{jl}}{Q_l}\Bigg] n_j.\end{aligned} \tag{5.53}$$

According to its definition, the epidemic threshold is computed as

$$\lambda = \frac{\mu}{\Lambda_{\max}(\bar{\mathbf{M}})}. \tag{5.54}$$

Note that Eq. 5.54 constitutes the exact value of the epidemic threshold in presence of subpopulations with heterogeneous degree distributions. Nonetheless, its computation involves solving the spectrum of a matrix whose dimension corresponds to the sum of all the degree classes coexisting inside each subpopulation. In this sense, for fine-grained spatial datasets, such task can be computationally expensive.

In what follows, we prove that the estimation of the epidemic threshold can be greatly simplified. Let us first perform a similarity transformation $\bar{\epsilon}_{ik} \mapsto k\epsilon_{ik}$, which by definition does not alter the matrix spectrum. After such transformation, Eq. (5.52) turns into

$$\mu\epsilon_{ik} = \bar{\lambda} \sum_j \sum_{k'} M_{ik}^{jk'} \epsilon_{jk'}, \tag{5.55}$$

where the elements of the new matrix $\mathbf{M}$ read as

$$\begin{aligned}M_{ik}^{jk'} &= k'^2 P_j(k') \Bigg[(1-p_k)(1-p_{k'})\frac{\delta_{ij}}{Q_i} + (1-p_k)p_{k'}\frac{R_{ji}}{Q_i} + \\ & \qquad p_k(1-p_{k'})\frac{R_{ij}}{Q_j} + p_k p_{k'} \sum_l \frac{R_{il}R_{jl}}{Q_l}\Bigg] n_j.\end{aligned} \tag{5.56}$$

If we decouple the mobility rate from the degree, i.e $p_k = p$, the matrix' elements become independent of k. This way, we can assume that $\epsilon_{ik} = \epsilon_i$ which allows a dimensionality reduction of the matrix from the total number of degree classes Ω to the number of patches in the metapopulation N_P. Specifically, Eq. (5.55) transforms into

$$\mu\epsilon_i = \lambda \sum_j M_{ij}\epsilon_j, \tag{5.57}$$

and the elements of the reduced matrix $\mathbf{M}$ are given by:

$$M_{ij} = \langle k^2 \rangle_j \left[(1-p)^2 \frac{\delta_{ij}}{Q_i} + p(1-p) \left(\frac{R_{ji}}{Q_i} + \frac{R_{ij}}{Q_j} \right) + p^2 \sum_l \frac{R_{il} R_{jl}}{Q_l} \right] n_j \ . \tag{5.58}$$

Finally, in this scenario, the epidemic threshold is computed as

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})}. \tag{5.59}$$

From the elements of the reduced matrix $\mathbf{M}$, we immediately realize that, when $p = 0$, the epidemic threshold matches the expression for λ_0 provided by Eq. (5.45); thus, our formalism allows for retrieving the HMF equations on contact networks. The connection between this model and the HMF approach on contact networks reveals that the epidemic detriment could emerge as a result of dismantling the contact structures of individuals located inside the most vulnerable node because of the recurrent mobility.

It is also worth remarking that the matrix $\mathbf{M}$ qualitatively resembles the critical matrix obtained in the MIR model. Indeed, the four different contributions to the matrix elements encode the same microscopic processes but are differently weighed according to the modified interaction rules. This reveals the generality of the mobility scheme introduced in this thesis and how this basic model can be easily extended to cover more complex dynamics. Fig. 5.12a shows that the estimation for the epidemic threshold accurately captures the transition from the disease-free to the endemic state and the emergence of the epidemic detriment.

5.3.5 Microscopic roots of the epidemic detriment

One crucial feature of this formalism is that the number of contacts of each individual is independent of the mobility network. As a consequence, we can accurately discriminate the role of both contact heterogeneities and recurrent mobility in shaping the epidemic threshold. To do so, we use a star-like metapopulation where we neglect global heterogeneities $\beta = 1$ and assume the population to be uniformly distributed so that $\alpha = 1$. In this setup, we vary either the local heterogeneities governed by k_{max} or the architecture of the mobility network governed by δ.

Qualitatively, Fig. 5.12b proves that the position of the peak p^* is determined by the mobility flows of the network but is independent on the local contact heterogeneities. Specifically, the peak is anticipated as δ increases, for the scarce overlapping between the mobility patterns of the individuals from the hub and the leaves reduces the beneficial effect of their mixing. In contrast,

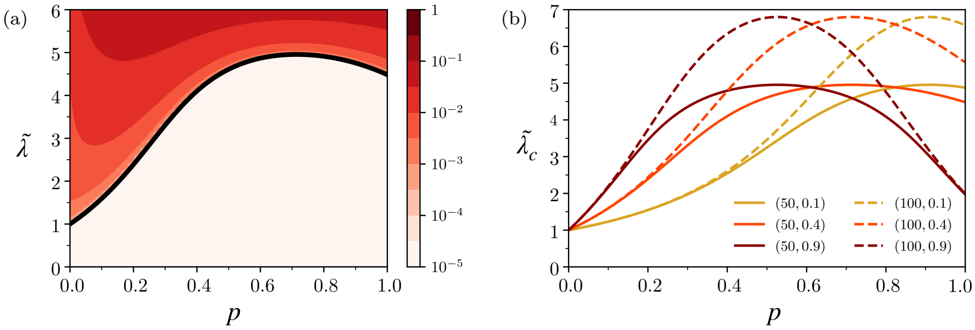


Figure 5.12: Dependence of the epidemic threshold on the mobility parameter p. The underlying metapopulation is the star-like configuration with $\kappa = 10$ leaves. All patches are uniformly populated ($\alpha = 1$) and the average connectivity of both hub and leaves is fixed to $\langle k \rangle_h = \langle k \rangle_l = 5$, so that $\beta = 1$. (a) Comparison of the theoretical epidemic threshold obtained using Eq. (5.59) (solid line), scaled by its value for $p = 0$, and the steady values of the epidemic size ρ (color code) for $(k_{\max}, \delta) = (50, 0.4)$. (b) Rescaled epidemic threshold $\tilde{\lambda}_c$ for different configurations $(k_{\max}, \delta)$, shown in the legends, with solid and dashed lines for $k_{\max} = 50$ and 100, respectively.

note that the exact value of the peak of the epidemic threshold $\lambda(p^*)$ is the same regardless of the position of the peak. Indeed, its magnitude is shaped by the local heterogeneities, leading to an increase of the peak as the parameter k_{max} increases, i.e., as the hub becomes more vulnerable with respect to the leaf. Mathematically, these observations entail the components of the eigenvector are fully-determined by the mobility network whereas the associated eigenvalues depend on how different the mixed subpopulations are.

Finally, we abandon the assumption of uniformly distributed populations and study the effect of the population asymmetry α on the peak of the epidemic threshold. Apart from the parameter β, let us define a new parameter γ to control global heterogeneities so that $\langle k^2 \rangle_l = \gamma \langle k^2 \rangle_h$. Note that the specific degree distributions of both hubs and leaves impose a constraint on these values, which must fulfill that

$$\gamma = \frac{\beta^2 \langle k \rangle_h^2}{\langle k \rangle_h (k_{\max} + 1) - k_{\max}}. \tag{5.60}$$

Fig. 5.13 represents the value of the rescaled epidemic threshold at the peak as a function of the population asymmetry α and the parameters controlling the heterogeneities, β in panel (a) and γ in panel (b). The first striking result in both panels is that the epidemic peak is considerably reduced as the population is asymmetrically distributed, approaching $\tilde{\lambda}_c(p^*) = 1$ when $\alpha \to 0$. This behavior is observed because, as the leaves' population becomes negligible with respect to the hub's one, the agents from the leaves hardly alter the degree

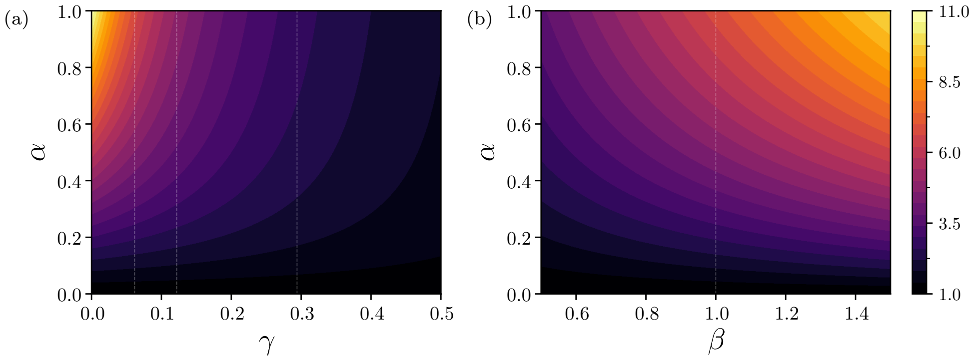


Figure 5.13: Peak of the rescaled epidemic threshold $\tilde{\lambda}_c(p^*)$ as a function of α, β, and γ, with $\langle k \rangle_h = 5$. In (a), all patches have the same average connectivity, with $\beta = 1$ while the local heterogeneity of the hub is modulated by γ. Dashed lines correspond to the values of γ for $k_{\max} = 100$, 50, and 20, from left to right. The plot in (b) considers a fixed value of $\gamma \approx 0.0617$, corresponding to $k_{\max} = 100$ when $\beta = 1$, tuning the connectivity of the leaves with β. The population asymmetry is modulated by α for all cases.

distribution of the residents in the hub.

Furthermore, note that, if we fix β, decreasing γ leads to an increase of the peak, thus highlighting a higher beneficial effect of the mobility. This effect is rooted in the fact that both subpopulations become more different among each other because of the higher k_{max} value needed to decrease γ. Likewise, the constraint posed by Eq. (5.60) makes the epidemic threshold increase with β as a result of the higher k_{max} value to keep γ constant.

In a nutshell, the results here shown provide solid evidence on the microscopic origin of the epidemic detriment caused by recurrent mobility. To link the epidemic detriment with the results obtained when studying epidemic spreading in contact networks, we must recall that there are different studies proving that the endemic states close to the epidemic threshold are mainly driven by the most connected nodes and the ties connecting them [19, 259]. In our model, the mixing between the different subpopulations promoted by mobility reduces the probability for them being connected among each other, thus preventing the emergence of localized endemic states and, as a consequence, increasing the threshold.

5.4 Conclusions

The interplay between recurrent mobility and disease spreading in the MIR model produces an unexpected phenomenon: the mixing between different subpopulations, rather than promoting epidemic spreading, sometimes may be beneficial to keep an ongoing outbreak under control. This chapter of the thesis has been devoted to disentangle the mechanisms responsible for such surprising result. To do so, we have constructed a toy configuration, the star-like metapopulation, which allow for systematically tuning different structural features characterizing the metapopulation such as the population distribution or the flow architecture and analytically quantify their impact on the onset of diseases.

First, we have studied the MIR model proposed in Chapter 4 on top of the star-like metapopulation. By tuning the aforementioned features, we have recovered the three different types of metapopulations observed so far. Namely, those for which no epidemic detriment appears and mobility favors epidemic spreading (type I cities) and those for which the epidemic detriment partially (type II) or totally (type III) emerges. Qualitatively, we have characterized a region of the parameters' space for which the epidemic detriment emerges as a result of the homogenization of the population driven by mobility. In contrast, this kind of macroscopic arguments based on the redistribution of the population are not always valid, for the epidemic detriment also emerges when mobility tends to accumulate population in those vulnerable areas.

The limitations of these macroscopic arguments have pushed us to go further and seek the microscopic/mesoscopic roots of the detriment. For this purpose, we have proposed an extension of the MIR model by introducing tunable permanence times or trips duration. Interestingly, we have found that extending the trip duration reduces the epidemic threshold in some scenarios when the size of the interacting populations are constant and independent of this time scale. A thorough analysis has revealed that the epidemic detriment emerges because the back-and-forth movements allow infectious agents to stay away from the infection hotspots, thus reducing the potential number of contagions made during their infectious window.

Finally, we have extended the MIR model to accommodate heterogeneous contact distributions of the population. Regarding the interplay between mobility and the number of contacts, unlike the previous approaches, in this new model we have assumed that agents preserve their number of interactions regardless of the population gathered in their destination. Based on macroscopic arguments, this fact should suppress any dependence of the epidemic threshold on the mobility; however, the epidemic detriment also occurs in this scenario, which manifests its microscopic nature. Specifically, recalling some concepts in network science, we conclude that mobility dismantles the localized endemic states by weakening the ties linking the most connected and, hence, vulnerable agents.

Overall, the epidemic detriment constitutes an emergent phenomena arising from the variation of the contact structures of the population. Interestingly, a similar phenomenon has been reported when studying epidemic spreading across temporal networks [260]. From a theoretical point of view, the next step is the formulation of a framework mapping the flows observed across different locations onto the time-varying connections of temporal networks. This formalism would enable to apply all the machinery recently developed to characterize dynamics on these configurations [261–264], thus improving our understanding of the impact of human mobility on epidemic spreading.

Chapter 6

Applications of the MIR model

> *During an outbreak, we rarely see the full picture at first, and this is where mathematics is essential.*
>
> A. Kucharski [249].

Mathematical models are nowadays indispensable tools to assess the stage of an ongoing outbreak, forecast its short-term evolution and guide policy-makers in the design of control policies to reduce its impact on the society. In this sense, driven by the important theoretical advances made over the last decades, epidemic frameworks have become general enough to accommodate different factors such as age-structured populations [222, 265, 266], simulations at the individual level [267–269] or detailed compartmental schemes to reflect the complex course of some real diseases [270, 271].

However, there is no free lunch, and the increased realism of epidemic models comes hand in hand with an increased complexity that makes it difficult to identify, with simple physical arguments, the key ingredients behind the observed spatio-temporal spreading patterns. In this chapter of the thesis, we seek to optimize this trade-off existing between the applicability of theoretical models to specific real outbreaks and the possibility of getting physical insights into the role of the different mechanisms involved in their spread. To do so, we propose different extensions of the MIR model which allow us to tackle different epidemic scenarios on real systems when not losing the analytical power provided by its equations and the theoretical knowledge acquired thus far.

In Section 6.1, we apply the MIR model to a large dataset of 163 cities distributed across 4 countries around the globe. Our goal is to find a better understanding, rooted in the architecture of human flows, of the influence of daily mobility in these cities despite the differences in terms of land coverage, urbanization or population existing among them. To do so, we use a dimensionality reduction technique relying on the LouBar method [226] which allows classifying the set of a patches inside a given city into two categories according to their population density. We reveal an universal trend dictating how the flows between the most dense areas shape the epidemic threshold. Based on the nature of the epidemic detriment, we propose different interventions modifying the architecture of the mobility network to increase the cities' resilience without the need to massively confine the population.

In Section 6.2, we generalize the MIR model to tackle the modeling of the course of COVID-19 pandemic across Spain. The theoretical challenge posed by COVID-19 is two-fold. On the one hand, the complex natural course of the disease along with its highly different outcome as a function of the age of the patient demand to introduce much more complex compartmental schemes. More importantly, capturing COVID-19 trajectory requires real-time epidemic modeling and formulating versatile enough formalisms to capture the time-varying nature of different factors determining its evolution, such as mobility restrictions or lockdowns. After validating the model with data in Spain, we simplify it and obtain an analytical expression for the interplay between non-pharmaceutical interventions (NPIs) and the effective reproduction number, which sheds light into the factors determining their efectivenness.

Finally, in Section 6.3, we study how the MIR model should be modified to capture the contagion mechanisms of vector-borne diseases such as the Dengue or Chikungunya viruses. We show that the qualitative dependence of the epidemic threshold on the mobility strongly depends on the clinical features of the disease such as the number of asymptomatic patients. As a case study, we derive a theoretical indicator to quantify the vulnerability of the different neighborhoods within an endemic area for Dengue, the city of Cali in Colombia, showing that it captures very accurately the spatial distribution of cases reported from 2015 to 2016.

6.1 Preparing for the next pandemic

COVID-19 constitutes the most recent and clear example of how vulnerable our current society is to the irruption of highly contagious pathogens. This high vulnerability is the outcome of the combination between the existence of a vastly interconnected mobility network, removing any spatial barrier containing a localized outbreak, and the growing phenomenon of urbanization, which tends to accumulate more and more of the population in large cities, to the extent that in 2007 it exceeded the population living in rural environments worldwide [272, 273]. The interplay of globalization and urbanization turns large cities into contagion centers accelerating the reaction phase of the disease [274]. In this sense, disentangling the roots behind such acceleration is essential to be prepared for new forthcoming pathogens.

In the same spirit as different works within the so-called science of cities [275–278], we here seek to explore hierarchies and regularities of cities' structures to explain the factors shaping their vulnerability to epidemic outbreaks. To succeed in this endeavor, we rely on the LouBar method proposed in [226], which allows a dimensionality reduction of the mobility networks. This way, the LouBar method enables to encapsulate superficial cultural, structural or historical differences existing between different cities and focus on the mesoscopic structures governing the flows and agents' distribution across them. In some sense, this method resembles the renormalization group widely studied in statistical physics [279], which is able to classify dynamics of different nature within the same universality class.

6.1.1 LouBar method

The LouBar method establishes a hierarchical structure within the different geographical areas inside a city as a function of a given attribute of interest. This attribute may be related with demographic features such as the population, density or socio-economic indicators such as the GDP per capita. In practical terms, the LouBar method is a non-parametric method which allows classifying the patches in a metapopulation in different levels according to their contribution to the Lorenz curve of the attribute of interest.

If the variable of interest x is continuous and governed by a given probability function $f(x)$ and cumulative distribution $F(x)$, the Lorenz curve $L(F(x))$ is defined as

$$L(F(x)) = \frac{1}{\langle x \rangle} \int\limits_{x_{min}}^{x} x' f(x') dx' \ , \tag{6.1}$$

where the factor $\langle x \rangle$ is included to ensure that $L(1) = 1$. In economic terms, the value $L(x)$ represents the proportion of the total wealth of a system shared between those whose income is lower than x . Given this definition, the Lorenz curve has been used to quantify economic inequalities via the Gini coefficient. This coefficient is defined as the area between the Lorenz curve and the curve corresponding to an egalitarian economic distribution provided by $L(F(x)) = F(x)$.

Note that our variables are discrete and correspond to different attributes of the patches. Assuming that the values are sorted so that $x_1 < \ldots < x_{N_P}$,

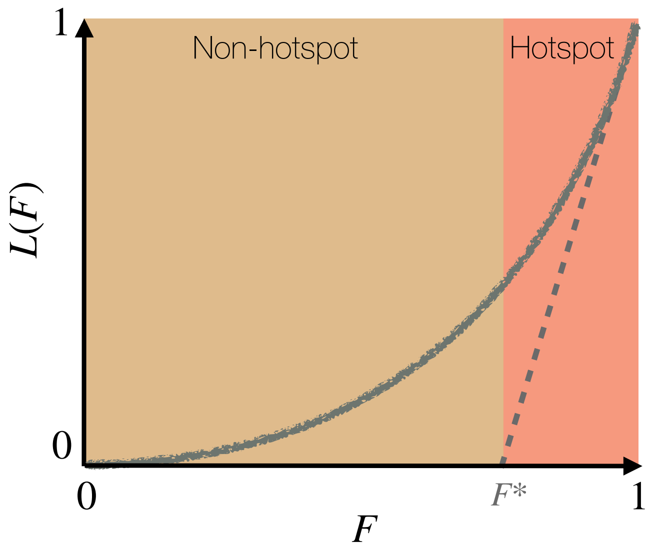


Figure 6.1: LouBar method for classifying patches according to their contribution to the cumulative distribution F of a given attribute x across the metapopulation.

the Lorenz curve is computed as

$$L\left(F(i)\right) = \frac{\sum_{j=1}^{i} x_j}{\sum_{j=1}^{N_P} x_j} \,, \tag{6.2}$$

with

$$F(i) = \frac{i}{N_P} \,. \tag{6.3}$$

Once the Lorenz curve $L(F)$ is constructed parametrically, the LouBar method draws the tangent line to the curve at $(1,1)$ and finds the intersection point of the latter with the x-axis, setting a threshold denoted by $F*$. From Eq. (6.2), the intersection point is computed as

$$F^* = 1 - \frac{\langle x \rangle}{x_n} \,. \tag{6.4}$$

The intersection point F^* allows discriminating those areas contributing to the inequalities of the distribution. In particular, for egalitarian distributions, i.e. $x_n = \langle x \rangle$, the intersection point is $F^* = 0$ and, consequently, all the patches are equivalent. As the value x_n grows with respect to the average, the intersection point F^* adopts intermediate values separating the patches into two subsets. In this sense, we identify a patch i as a hotspot when $F(i) > F^*$. An scheme of the LouBar method is depicted in Fig. 6.1.

In a nutshell, the LouBar method allows coarse-graining the patches in a metapopulation as a function of their similarities according to a given property. One application of this method was proposed by Bassolas et al. in [231]. In this manuscript, the attribute of interest is the number of flows departing from each area. Applying the method iteratively, the authors embed the complex nature of the mobility network into a hierarchical structure formed by the different hotspot levels and their aggregated connections. Interestingly, they derive an indicator, the hierarchical flow, quantifying to what extent the flows connect hotspots within the same level in different cities, and find that there exist a correlation between the latter indicator and different environmental and social features such as the level of pollution or the share of public transport.

6.1.2 MIR model with hotspots

Here we want to check whether the LouBar method can be applied to shed some light on the interplay between mobility and the epidemic threshold in real cities. The dataset here analyzed comprises 163 urban areas distributed across four different countries: Italy, Australia, South Africa and United States. To construct the mobility network of each city, we use the Google COVID-19 Aggregated Mobility Research Dataset, which contains anonymized mobility flows aggregated over users who have turned on the Location History setting, which is off by default. We use data collected weekly from November 3rd 2019 to February 29th 2020 to avoid being influenced by lockdowns. These flows are obtained at the S2 cells level 13, whose typical size is 1.27 km^2 [280].

Providing that the definition of cities is not univocal in all these countries, the spatial resolution of the metapopulation and its patches vary from one country to another. Moreover, apart from the mobility flows, we need to extract the population and the area of each specific patch. In this sense, depending on the availability of population data per country, the mobility flows between S2 cells are aggregated to patch size of comparable resolution between countries. As a result, we construct the following metapopulations:

- United States: 50 Urban areas with zip code patches [281].
- Italy: 49 Communes with S2 cell patches [282].
- South Africa: 31 Municipalities with ward patches [283].
- Australia: 33 Statistical Areas Level 4, with Level 2 patches [284].

Regarding the epidemic model, we use the original (and simplest) version of the MIR model introduced in Chapter 4. Recall that in this model the

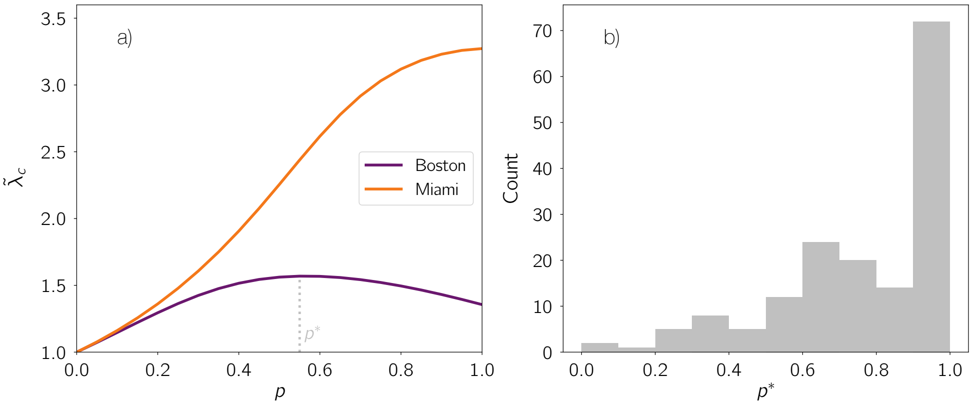


Figure 6.2: Panel a: Rescaled epidemic threshold $\tilde{\lambda}_c$ as a function of the mobility p for Boston (purple line) and Miami (orange line), with the localization of the peak of the epidemic threshold p^* pinpointed with a dashed line. Note that this value allows distinguishing between type II cities ($p^* < 1$) like Boston and type III cities ($p^* = 1$) like Miami. Panel b: Histogram of the p^* values for each of the 163 cities contained in the dataset.

epidemic threshold in this model is given by

$$\lambda_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M})} , \tag{6.5}$$

with

$$M_{ij} = n_j \left[\delta_{ij}(1-p)^2 \frac{f_j}{\tilde{n}_j} + p(1-p) \left(\frac{R_{ij} f_j}{\tilde{n}_j} + \frac{R_{ji} f_i}{\tilde{n}_i} \right) + p^2 \sum_{l=1}^{N_P} \frac{R_{il} R_{jl} f_l}{\tilde{n}_l} \right] . \tag{6.6}$$

Given its relevance for the spread of airborne diseases, we assume that the function f governing the number of contacts inside each patch is a monotonically increasing function with the population density. In particular, for the sake of simplicity, we choose $f(i) = \tilde{n}_i / a_i$, being $\tilde{n}_i$ the effective population gathered inside patch i and a_i its area.

Before applying the LouBar method, let us make a qualitative classification of the cities in our dataset according to the critical regimes described in Chapter 5. For this purpose, we study the evolution of the rescaled epidemic threshold $\tilde{\lambda}_c$ on the mobility. Within the 163 cities, we find 109 type II cities, for which a non-monotonic epidemic detriment is observed and 54 type III cities characterized by a monotonic epidemic detriment; examples of both cities types are shown in Fig 6.2a. Likewise, for the sake of completeness, we represent a histogram with the localization of the peak of the epidemic threshold p^* in Fig. 6.2b.

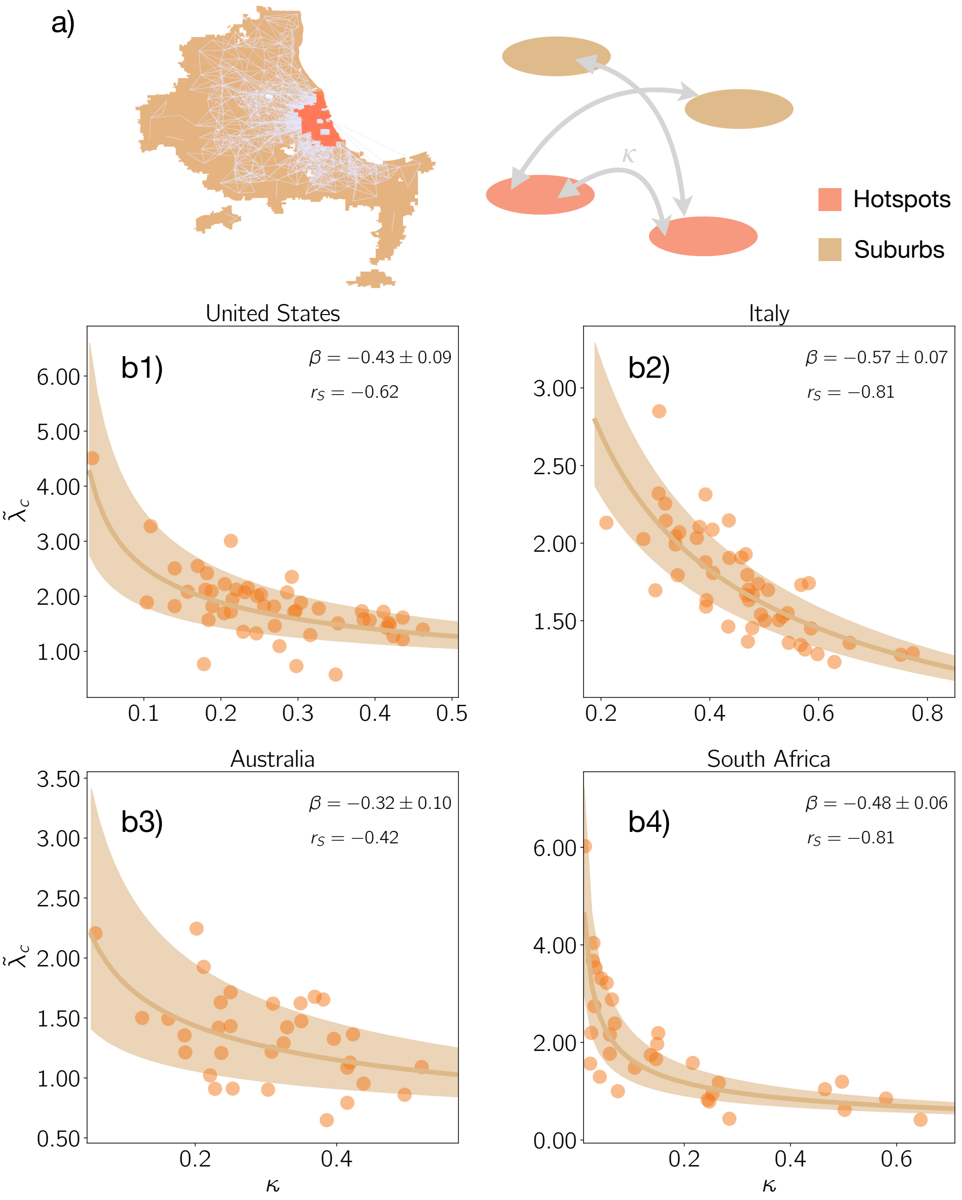


Figure 6.3: Panel a): Mobility network of Chicago in the United States and its associated low-dimensional representation. The brown areas correspond to suburbs whereas the red areas identify the hotspots defined according to the population density. Panels b1-b4: Rescaled epidemic threshold $\tilde{\lambda}_c$ as a function of the hotspot flow concentration κ for each of the cities analyzed in United States (b1), Italy (b2), Australia (b3) and South Africa (b4). In all these panels, the metric β is the exponent of the power-law function to which data are fitted in each country (see text for details) and r_S is the Spearman rank correlation coefficient relating the rescaled epidemic threshold values to the hotspots flow concentration. The shadowed region shows 95% confidence interval for the fitted dependencies.

Beyond the qualitative description, here we want to connect the metapopulation features with the quantitative value of the threshold. To do so, we apply the LouBar method to find the hierarchies existing within each city and extract the population density hotspots; the remaining patches are denoted by suburbs. Motivated by the crucial role of contacts' variability in shaping the onset of diseases, we define a new indicator κ_k as the share of the flows in a given city k connecting hotspots. Assuming that $H(k)$ encodes the set of hotspots within city k and that T_{ij}^k denotes the flows going from patch i to patch j inside city k, this indicator is given by

$$\kappa_k = \frac{\sum\limits_{i,j \in H_k} T_{ij}^k}{\sum\limits_{i,j} T_{ij}^k}. \tag{6.7}$$

Note that this metric lies in the range $0 \leq \kappa \leq 1$. An illustrative scheme of the dimensionality reduction here proposed is represented in Fig 6.3a.

Our hypothesis based on the physics found in Chapter 5 is that the detrimental effect of the mobility should be hindered in those cities where the flow is mainly concentrated around hotspots. In this scenario, mobility does not allow the population to escape from the most vulnerable areas, therefore reducing the beneficial effect of contacts' variability. To have a fair comparison of the effect of the mobility in two cities with disparate population densities, we use the rescaled epidemic threshold $\tilde{\lambda_c}$ at $p = 1$, which is defined as

$$\tilde{\lambda_c} = \frac{\lambda_c(p=1)}{\lambda_c(p=0)} = \frac{d_{max}}{\Lambda_{\max}(\mathbf{M_{p=1}})}\ , \tag{6.8}$$

where d_{max} represent the largest population density observed within the patches of each city.

Fig. 6.3b represents the rescaled epidemic threshold as a function of the κ value of each city in each of the four countries analyzed here. There we confirm our hypothesis, finding a moderate negative correlation between both variables. As a proof of concept, we have fitted the curve of each country to a power-law decay $\tilde{\lambda_c} \sim A\kappa^{\beta}$ by using least-squares regression, obtaining negative values in all the cases for the exponent β. The different values observed in these exponents may arise as a consequence of the different spatial resolutions with which the mobility networks in each country are constructed or obey further structural features here not considered.

6.1.3 Interventions on the mobility matrix

To confirm the negative effect of the concentration flow around hotspots for cities' vulnerability, we perform an intervention on the mobility network consisting in removing all the flows connecting different hotspots and distributing them among the suburbs neighboring areas, preserving the proportions as dictated by the original matrix $\mathbf{R}$. In the following, this intervention is referred to as *reshuffling* intervention. The effect of such intervention is captured in Fig. 6.4a, which presents a histogram of the values for the ratio between the normalized threshold after and before the intervention, $\tilde{\lambda}_c^{\mathrm{MOD}}$ and $\tilde{\lambda}_c$ respectively. Note that most cities fall into the beneficial domain for which removing the flow between hotspots increases city' resilience. In contrast, there are a few cities, around 9% of the cities studied, for which the intervention is detrimental and accelerates epidemic spreading. We believe that this phenomenon is rooted in those details of the underlying density distribution which are overlooked when coarse-graining the flow between hotspots in κ. A breakdown of the cities classified according to the effect of the *reshuffling* intervention and their country is made in Fig. 6.4b.

Despite the clear beneficial effect of removing the flows between hotspots, this intervention is not doable as a short-term reaction to an ongoing outbreak due to the long times needed to achieve the desired redistribution of the population across the neighboring suburbs. Instead, a more realistic intervention would be that of targeting specific individuals and forcing them to stay in their associated patches. Let us stress that this intervention does not entail the lockdown of this population at home but only prevents the circulation of the targeted individuals across the metapopulation. We explore four different scenarios in which the flows connecting two categories of patches are deleted and converted into self-loops. The affected subpopulations are:

- *Intervention I*: Individuals moving from hotspots to suburbs.
- *Intervention II*: Individuals moving from suburbs to hotspots.
- *Intervention III*: Individuals moving from hotspots to hotspots.
- *Intervention IV*: Individuals moving from suburbs to suburbs.

Fig. 6.4c summarizes the effects of deploying each of the interventions described in this section for the United States (US) cities. The first striking result in this figure is the asymmetric outcome of closing the flows involving individuals moving between hotspots and suburbs. Specifically, we observe that preventing residents in the suburbs from visiting the hub increases the city's resilience whereas removing the routes from hotspots to suburbs makes the city more vulnerable. To understand this asymmetry, we must recall that hotspots behave as contagion centers; hence, every intervention aimed at reducing the time spent by the population in these areas will be beneficial to stop the propagation of the disease. For this reason, intervention I works and intervention II

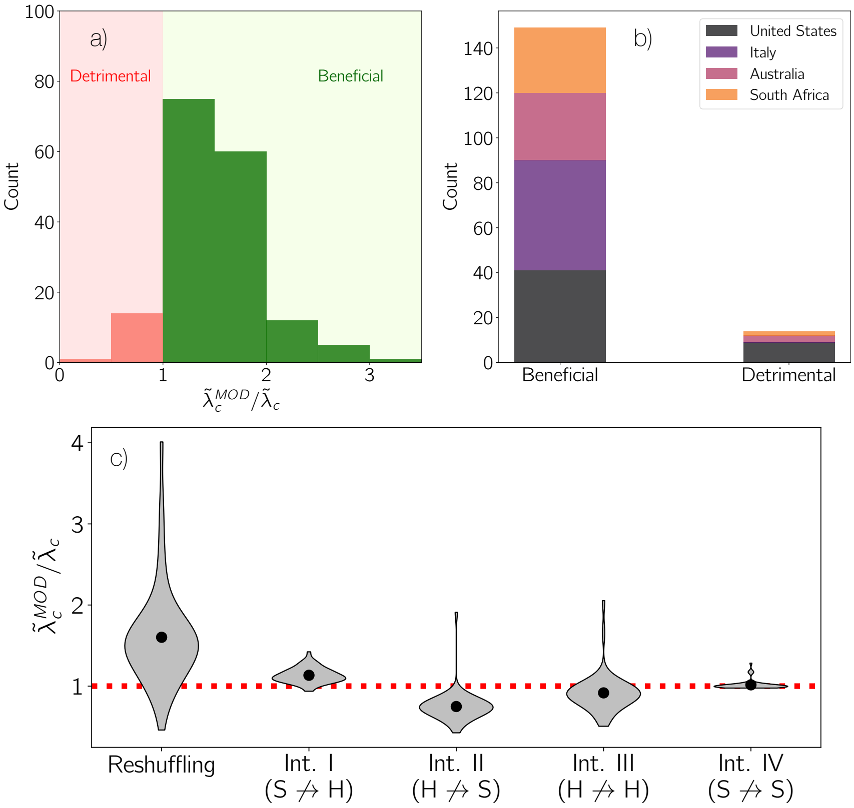


Figure 6.4: Panel a): Histogram for the ratio between the rescaled epidemic threshold after the reshuffling intervention $\tilde{\lambda}_c^{MOD}$ and the original threshold $\tilde{\lambda}_c$, highlight two different types of cities: those for which this intervention is beneficial and those for which is detrimental. Panel b): Breakdown of the cities belonging to each country (color code) as a function of the outcome of the reshuffling intervention. Panel c): Distribution of the ratios following different intervention schemes for the US cities. The black dots represent the average of each distribution. For the sake of clarity, we explicitly indicate the flows affected in each intervention. For example, $S \nrightarrow H$ removes the flows from suburbs to hotspots and turns them into self-loops. The red dashed line highlights the case in which the interventions do not have any effect on the threshold.

fails, because in the latter case the spatial isolation of the outbreak is made at the expense of a higher exposure of the residents in the hub to the contagion sources. Finally, we observe that, on average, acting on the flows connecting patches from the same category (namely interventions III and IV) does not alter the threshold because of the internal similarities within each category.

6.1.4 Application to COVID-19 in United States

Up to now, we have analyzed how the structural features a city, i.e. the population distribution and its mobility patterns, are able to predict the impact of mobility on the onset of diseases across a city and how we can improve the resilience of the city by altering the mobility network. Now we are interested in determining whether the dimensionality reduction scheme here proposed can characterize the actual vulnerability of a city to the emergence of outbreaks upon the arrival of an index case. In particular, due to data availability, we focus on the impact of COVID-19 pandemic, caused by the spread of the SARS-CoV-2 virus, on different US cities. In this sense, although our approach is based on a SIS model, we think that our findings about the role of the structure of the metapopulation are general enough to remain valid when more complex compartmental schemes are introduced [265, 285–287].

To quantify cities' vulnerability, we extract the infection curves from data collected at the county level from different sources [288, 289]. There are different sophisticated methods to estimate the speed at which an epidemic evolves from the infection curves. Most of them rely on the estimation of the effective reproductive number, which requires to know some features of the disease such as its generation time [102, 290, 291]. As we want to keep our approach as simple and general as possible, we decide to follow a more basic approach and fit the infection data to an exponential curve, $I_k \sim e^{bt}$, which correspond to the expected behavior for the early stage of the disease according to the standard compartmental models [29], and define the empirical vulnerability of the city as its exponent, denoted in the following by b^{exp}.

The exponential fitting procedure should be carefully made to avoid some practical difficulties associated with the data. First, we use the Savitzky-Golay filter [292] to reduce the inherent noisy nature of the data by fitting intermediate windows [293] with low-order polynomials. Moreover, as we seek to explore the connection between structure and vulnerability, the vulnerability metric should not be affected by any external factor such as those NPIs promoted in the different regions. For this purpose, we only use for fitting purposes data corresponding to two weeks after the first 100 cases were reported in a given area.

Fig. 6.5a illustrates the fitting procedure for the city of Miami. To connect the hotspot flow concentration κ with the empirical vulnerability b, we must realize that the former does not include any information about the overall population density, but only accounts for the variation of the density profile of a city driven by mobility. Providing the crucial role of population density for SARS-CoV-2 transmission, a combination of both factors should be included

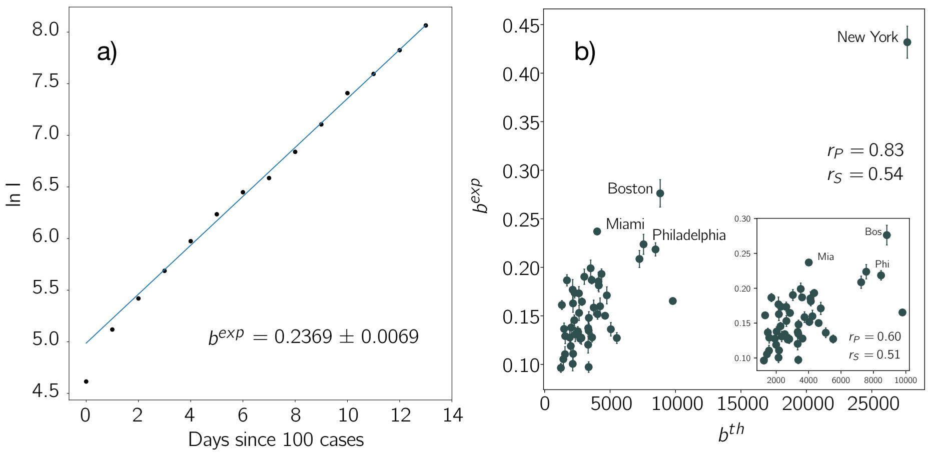


Figure 6.5: Panel a): Cumulative number of cases in logarithmic scale as a function of the days after the first 100 cases were reported in Miami. The black line shows the fit to an exponential function $I(t) = Ae^{b^{exp}t}$, where b^{exp} corresponds to the empirical city vulnerability. Panel b): Empirical city vulnerability b^{exp} as a function of the theoretical vulnerability b^{th} obtained as the combination between population density information and the concentration of flows between hotspots (see text for details). The correlation between both variables is quantified with the Pearson and Spearman correlation coefficientes, denoted by r_P and r_S respectively. The inset panel removes New York city to obtain a more representative value of the Pearson coefficient.

to provide a reasonable theoretical estimation for cities' vulnerability. Specifically, we choose $b^{th} = \langle d^{hot} \rangle \kappa^{0.43}$, where $\langle d^{hot} \rangle$ represents the average density within hotspots and the scaling corresponds to the exponent of the power law fit obtained in Fig. 6.3b. Fig. 6.5b represents the theoretical and empirical estimation for cities' vulnerability. The positive correlation existing among both variables indicates that the combination between the population density and the flows' architecture of real cities provides a plausible explanation for the heterogeneous growth rates observed at the early stages of COVID-19 across the United States.

6.2 COVID-19 compartmental model

In the last part of Section 6.1, we have proven that simple arguments based on the structure of the underlying metapopulation are useful to characterize the very early stages of COVID-19 outbreaks. Nonetheless, COVID-19 represents the most important humanitarian, social and economic crisis humankind has faced since 1918 Influenza pandemic, with more than 200 million of cases and 4 million of deaths reported as of September . Over this period, the course of

COVID-19 has been shaped by different social, political, economical and demographic aspects which have demanded to evolve the traditional frameworks towards more tailored mathematical models [28, 181, 294–298].

These models usually contain more elaborate compartmental schemes to capture the particularities concerning the spread of SARS-CoV-2 and the heterogeneous clinical features of COVID-19 patients. In particular, the large volume of contagions made during the presymptomatic stage (around 40% [99]) requires accounting for the existence of infectious individuals who have not developed symptoms yet. Moreover, COVID-19 has overburdened the health system and taken a huge death toll in many countries, which reflect the need for including some compartments capturing the clinical course of the disease. In its turn, the severity of COVID-19 strongly varies as a function of the age of the patient, which prevents the use of mean-field theories in favor of including age-structured populations in the model.

Apart from the additional epidemiological and clinical compartments, modeling COVID-19 represents an important theoretical challenge, for models should be general enough to provide real-time assessment on the expected impact of different NPIs [92, 299–302], which in the absence of vaccines, were essential to keep the pandemic under control worldwide. Therefore, the equations of the model should include the mechanisms needed to account for these interventions and evaluate their outcome.

In this section, we propose an extension of the multiplex MIR model presented in Chapter 4 to accommodate the specific features associated with the spread of SARS-CoV-2 virus and COVID-19 disease. After a general presentation of the assumptions of the model and its equations, we particularize the model to capture the course of COVID-19 in Spain. Finally, we derive a general expression for the effective reproduction number which allows for readily assessing the impact of different interventions.

6.2.1 Model equations

Let us assume that our system is composed of N_P patches and N_G layers, which can be identified with different types of individuals showing unequal COVID-19 impact such as age-groups [303, 304] or ethnicities [305, 306]. The movements of the population are captured in the multiplex mobility network $\mathbf{R}$, whose elements R_{ij}^{g} encodes the probability that an individual of class g and residing inside i moves to j. Furthermore, we assume that each patch i is populated by n_i individuals, which are broken down into n_i^g individuals belonging to each class g. Denoting as N the total number of individuals across the system and as n^g the total number of individuals of class g, the following equation is fulfilled:

$$N = \sum_{g=1}^{N_G} \sum_{i=1}^{N_P} n_i^g = \sum_{i=1}^{N_P} n_i = \sum_{g=1}^{N_G} n^g \ . \tag{6.9}$$

The compartmental scheme adopted in the model is summarized in Fig. 6.6. Namely, the model can be splitted into two well-differentiated parts, the epidemiological compartments determining the spread of the virus and the clinical compartments governing the outcome of the disease:

- The epidemiological part of the model is a variant of a SEIR model which distinguishes between asymptomatic and symptomatic individuals. Therefore, the compartments correspond to: susceptible (S), exposed (E), asymptomatic infectious (A), symptomatic infectious (I).

- Regarding the clinical part of the model, the compartments included are: to be admitted in ICU[1] (pre-hospitalized in ICU, P_H), fatal prognosis (pre-deceased, P_D), admitted in ICU that will recover (H_R) or decease (H_D), recovered (R) and deceased (D). Compartments P_H and P_D are introduced to account for the delays observed for individuals before they are admitted in ICU or before they die outside ICU, respectively. Similarly, the compartments H_R and H_D represent the individuals in ICU, but separated according to their final state (recovered or deceased, respectively), which is reached at different rates.

Our system is completely characterized by variables $\rho_i^{m,g}(t)$ which account for the probabilities that individuals of class g assigned to patch i are in state m

---

[1] Intensive Care Units.

at time t. The temporal evolution of these quantities is given by

$$\rho_i^{S,g}(t+1) = \rho_i^{S,g}(t)\,(1-\Pi_i^g(t)), \tag{6.10}$$

$$\rho_i^{E,g}(t+1) = \rho_i^{S,g}(t)\,\Pi_i^g(t) + (1-\eta^g)\,\rho_i^{E,g}(t), \tag{6.11}$$

$$\rho_i^{A,g}(t+1) = \eta^g\,\rho_i^{E,g}(t) + (1-\alpha^g)\,\rho_i^{A,g}(t), \tag{6.12}$$

$$\rho_i^{I,g}(t+1) = \alpha^g\,\rho_i^{A,g}(t) + (1-\mu^g)\,\rho_i^{I,g}(t), \tag{6.13}$$

$$\rho_i^{P_D,g}(t+1) = \mu^g\,\theta^g\,\rho_i^{I,g}(t) + (1-\zeta^g)\,\rho_i^{P_D,g}(t), \tag{6.14}$$

$$\rho_i^{P_H,g}(t+1) = \mu^g\,(1-\theta^g)\,\gamma^g\,\rho_i^{I,g}(t) + (1-\lambda^g)\,\rho_i^{P_H,g}(t), \tag{6.15}$$

$$\rho_i^{H_D,g}(t+1) = \lambda^g\,\omega^g\,\rho_i^{P_H,g}(t) + (1-\psi^g)\,\rho_i^{H_D,g}(t), \tag{6.16}$$

$$\rho_i^{H_R,g}(t+1) = \lambda^g\,(1-\omega^g)\,\rho_i^{P_H,g}(t) + (1-\chi^g)\,\rho_i^{H_R,g}(t), \tag{6.17}$$

$$\rho_i^{D,g}(t+1) = \zeta^g\,\rho_i^{P_D,g}(t) + \psi^g\,\rho_i^{H_D,g}(t) + \rho_i^{D,g}(t), \tag{6.18}$$

$$\rho_i^{R,g}(t+1) = \mu^g\,(1-\theta^g)\,(1-\gamma^g)\,\rho_i^{I,g}(t) + \chi^g\,\rho_i^{H_R,g}(t) + \rho_i^{R,g}(t). \tag{6.19}$$

The flow of the compartmental models for individuals belonging to class g and patch i is as follows: Susceptible individuals contract the disease at time t with probability $\Pi_i^g(t)$ and becomes Exposed, spending in this compartment $(\eta^g)^{-1}$ time steps corresponding with the latent period. Afterwards, they become asymptomatic infectious during $(\alpha^g)^{-1}$ time steps, eventually developing symptoms and entering the infectious symptomatic where they remain during $(\mu^g)^{-1}$ time steps. The clinical part contains three different patches. The first one is death without being hospitalized in ICU, with probability θ^g, after a latency period governed by rate ζ^g. Otherwise, with probability γ^g the individuals are hospitalized in ICU, which are reached at rate λ^g, while with probability $1-\gamma^g$ they recover. Individuals in ICU have a fatality probability ω^g, which is reached at a rate ψ^g, whereas the recovery is reached at a rate χ^g.

Following the multiplex MIR model, the expression for $\Pi_i^g(t)$ reads as

$$\Pi_i^g(t) = (1-p^g)\,P_i^g(t) + p^g \sum_{j=1}^{N_P} R_{ij}^g\,P_j^g(t)\,, \tag{6.20}$$

where p^g represents the degree of mobility of class g and $P_i^g(t)$ corresponds to the probability that a susceptible individual from class g contract the disease inside i. To calculate the latter probability, we include a refinement in the interaction stages with respect to the original MIR model and the extensions explained throughout this thesis. Namely, we include a more realistic function shaping the dependence of the number of contacts on the density of each patch. In particular, following [307], we choose

$$f_i = 2 - e^{-\xi \tilde{n}_i/a_i}\;, \tag{6.21}$$

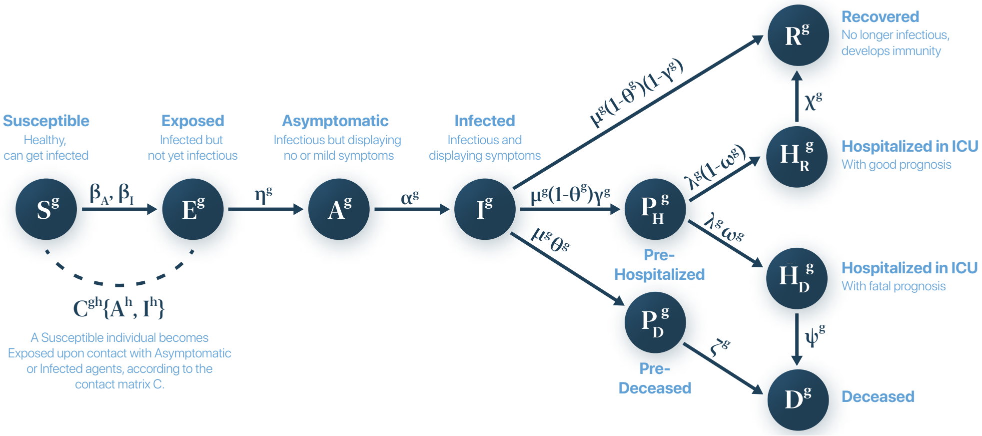


Figure 6.6: Compartmental scheme of the model tailored for COVID-19. The available compartments are: susceptible (S^g), exposed (E^g), asymptomatic infectious (A^g), symptomatic infectious (I^g), pre-hospitalized in ICU (P_H^g), pre-deceased (P_D^g), in ICU before recovery (H_R^g), in ICU before death (H_D^g), deceased (D^g), and recovered (R^g), where g denotes the class of individual. The arrows indicate the transition probabilities.

where a_i corresponds with the area of patch i and $\tilde{n}_i$ with its effective population. Taking into account the mobility scheme, the effective population inside patch i is given by

$$\tilde{n}_i = \sum_{g=1}^{N_G} \tilde{n}_i^g \ , \tag{6.22}$$

with

$$\tilde{n}_i^g = \sum_{j=1}^{N_P} M_{ji}^g n_j^g \ . \tag{6.23}$$

For the sake of readability, we define in the latter expression M_{ij}^g as the probability that one individual from class g and residing in i is located at j. Specifically,

$$M_{ij}^g = (1 - p^g)\delta_{ij} + p^g R_{ij}^g \ . \tag{6.24}$$

The form of Eq. (6.21) makes the number of contacts made in densely packed areas at most twice those occurring in less populated patches. In addition, we define the average number of contacts made by the individual inside class g as $\langle k^g \rangle$ and assume that the mixing between different classes of individuals is governed by the matrix $\mathbf{C}$, where C^{gh} encodes the fraction of the contacts made by agents of class g involving agents of class h. Under the latter constraints, we obtain

$$P_i^g(t) = 1 - \prod_{h=1}^{N_G} \prod_{j=1}^{N_P} \prod_{m \in \{A,I\}} (1-\beta_m)^{z^g \, \langle k^g \rangle \, f_i \, C^{gh} \, \frac{n_{j\to i}^{m,h}(t)}{\tilde{n}_i^h}} \ , \tag{6.25}$$

which includes different infectiousness for both symptomatic and asymptomatic individuals. In the latter expression, z^g is a needed factor to satisfy the constraint posed by the average number of contacts of each stratum g. In particular, assuming that the number of interactions for these individuals in one patch i is given by $x^g f_i$, the following condition must be fulfilled:

$$\sum_{i=1}^{N_P} \tilde{n}_i^g x^g f_i = n^g \langle k^g \rangle \ , \tag{6.26}$$

so the scaling factor reads as

$$x^g = z^g \langle k^g \rangle \ , \tag{6.27}$$

with

$$z^g = \frac{n^g}{\sum\limits_{i=1}^{N_P} f_i \tilde{n}_i^g} \ . \tag{6.28}$$

Finally, $n_{j\to i}^{m,h}(t)$ refers to the number of individuals from class h in state m residents in patch j that are placed in i. Therefore,

$$n_{j\to i}^{m,h}(t) = n_j^h \, \rho_j^{m,h}(t) \, M_{ji}^h. \tag{6.29}$$

6.2.2 Modeling confinement measures

The model previously described allows for characterizing COVID-19 impact when the SARS-CoV-2 virus spreads uncontrolledly. Nonetheless, the alarming numbers associated with COVID-19 forced the authorities of many countries to implement control policies to mitigate the outbreaks and reduce the pressure to the health systems. In the absence of vaccines, different NPIs have been promoted ranging from prophylactic measures such as the use or masks or hands hygiene to population-wide lockdowns or strict contact tracing policies. Given its ubiquity worldwide, here we focus on the impact of lockdowns and social distancing measures.

To incorporate confinement policies in our formalism, we consider that a fraction of the population of class g, $\kappa_0^g(t)$, is under lockdown at time t. Note that we do not assume this parameter to be constant to capture the time-varying nature of this type of interventions [308]. In this sense, $\kappa_0^g = 1$ corresponds to a total lockdown scenario in which all activity is paralyzed whereas

$\kappa_0^g = 0$ reflects a less strict policy not affecting the population. Moreover, we introduce the effects of social distancing measures taken by the population as a reduction δ of the number of contacts made by the non-confined population. We assume the social distancing δ to be constant in time.

Let t_c denote the time at which the NPIs are initially implemented. First, lockdowns and social distancing measures alter the number of contacts made by the population. Therefore, the average number of interactions of each degree-class should be time-dependent. Specifically,

$$\langle k^g \rangle(t) = (1 - \Theta(t - t_c))\, \langle k^g \rangle + \Theta(t - t_c)\, \langle k_c^g \rangle(t) \ , \tag{6.30}$$

where $\Theta(x)$ is the Heaviside function, that is 1 if $x \geqslant 0$ and 0 otherwise. To estimate the number of interactions of the confined population k_c, we leverage the contact matrices presented in [221] and break down the daily interactions according to the environment where they occur, distinguishing between contacts at home, schools, workplace or in other locations. In particular, we assume a total school closure and that population under lockdown restrict their interactions to the contacts made at home whereas the rest of the population has only additional contacts in their respective workplaces. Denoting both contributions as k^h and k^w respectively, we obtain

$$\langle k_c^g \rangle(t) = \kappa_0^g(t)\, k_{\mathrm{h}}^g + (1 - \kappa_0^g(t))(1 - \delta)\, k_{\mathrm{h+w}}^g \ . \tag{6.31}$$

Likewise, the confinement of the population leads to a reduction of the mobility of the population. Accordingly, we set

$$p^g(t) = (1 - \kappa_0^g(t)\, \Theta(t - t_c))\, p^g \ . \tag{6.32}$$

Finally, we recall that lockdowns have been typically associated with the creation of social bubbles involving the members of each household. These social bubbles are crucial to control the extent of the outbreaks, for completely isolating households with susceptible individuals from the society prevents them from being reached by infectious peers. Unfortunately, social bubbles are not perfectly shielded due to sporadic mixing events between members of different households while covering essential needs such as buying groceries, drugs etc. To quantify the effectiveness of such isolation, we introduced the permeability ϕ governing the degree of mixing between different households and assumed to be constant in time.

The creation of social bubbles entails a depletion of the pool of susceptible individuals. Assuming that the household size is given by σ, the probability of finding a household without infectious members at time t is given by

$$CH_i(t) = \left(\frac{1}{n_i} \sum_{g=1}^{N_G} \left(1 - \rho_i^{E,g}(t) - \rho_i^{A,g}(t) - \rho_i^{I,g}(t) \right) n_i^g \right)^{\sigma} \ , \tag{6.33}$$

After introducing the social bubbles mechanism, Eqs. (6.10)-(6.11) turn into

$$\begin{aligned}
\rho_i^{S,g}(t+1) &= \left(\rho_i^{S,g}(t) + \rho_i^{CH,g}(t)\right)(1-\Pi_i^g(t)) \times \\
&\quad (1-\Theta(t-t_c)\,(1-\phi)\,\kappa_0^g(t)\; CH_i(t))\,, \qquad (6.34) \\
\rho_i^{E,g}(t+1) &= \left(\rho_i^{S,g}(t) + \rho_i^{CH,g}(t)\right)\,\Pi_i^g(t) \times \\
&\quad (1-\Theta(t-t_c)\,(1-\phi)\,\kappa_0^g(t)\; CH_i(t)) + \\
&\quad (1-\eta^g)\,\rho_i^{E,g}(t)\,, \qquad (6.35) \\
\rho_i^{CH,g}(t+1) &= \left(\rho_i^{S,g}(t) + \rho_i^{CH,g}(t)\right) \times \\
&\quad (1-\phi)\,\kappa_0^g(t)\; CH_i(t)\,\Theta(t-t_c)\,, \qquad (6.36)
\end{aligned}$$

where we have added Eq. (6.36) governing the occupation of a new CH compartment, gathering those individuals living in fully susceptible households, which form perfectly isolated social bubbles from the rest of the society. Summarizing, the lockdown policies are captured in the formalism by the time series describing confinement $\kappa_0^g(t)$, the social distancing measures taken by the active population governed by δ, and the mixing between different social bubbles governed by the household permeability ϕ.

6.2.3 Model validation: COVID-19 in Spain

Here we study how the formalism presented can be applied to study the evolution of COVID-19 across Spain. The first step is to construct the multiplex metapopulation by defining the groups conforming each layer, their demographic distribution and their mobility patterns. Due to the relevance of the age of the patient for the clinical course of COVID-19, we split the population into three different groups: the young population Y, spanning up to 25 years old, the adult population M, whose ages ranges from 25 years to 64 years old and the elderly population O gathering the rest of the population.

The mobility data were provided by the Instituto Nacional de Estadística (INE), and consist in extensive surveys made in 2011 [309] capturing the daily flow of active population between municipalities with more than 100 habitants. Therefore, the resolution of the data forces us to identify each patch with a municipality, considering all the individuals inside it as statistically equivalent agents. As most of these data are associated with the adult population, we incorporate them in the adult layer whereas we force the young and elderly population to stay in their respective municipalities, i.e. $R_{ij}^{Y,O} = \delta_{ij}$. In its turn, the average number of contacts $\langle k \rangle$ and the mixing matrices governing the interactions between different age strata were obtained from an exhaustive study revealing the social structure of 152 countries distributed worldwide [221]. These parameters are reflected in Table D.1.

The next step is to incorporate the clinical parameters, which are summarized in Table D.2. These parameters are extracted from existing literature addressing COVID-19 spread in other countries [310–312] or from official reports published by national health institutes in Spain [313, 314]. Regarding the epidemiological parameters, we fix the incubation period $\eta^{-1} + \alpha^{-1} = 5.2$ days [315] and set $\beta_A = 0.5\beta_I$ following [265, 267], but leave the rest of the parameters for the model calibration in light of the lack of consensus on their values. In general, these parameters are independent of the age group but, to reflect the high number of asymptomatic patients within young population, we force the latter to stay inside the A compartment by assuming that

$$(\alpha^Y)^{-1} = \alpha^{-1} + \mu^{-1} - 1 \ , \tag{6.37}$$

$$(\mu^Y)^{-1} = 1 \ . \tag{6.38}$$

The confinement measures are introduced into the model from March 14 onward, matching the implementation of the first nationwide lockdown. We obtained the time series $\kappa_0^g(t)$ from official reports [316] by estimating the work mobility reduction with respect to a baseline day, November 19th, 2019. We assume that these time series are equal for all age groups and refer in the following to $\kappa_0^{Y,A,O}(t)$ as $\kappa_0(t)$. Likewise, the household permeability ϕ and the social distancing δ are also left for model calibration.

In a nutshell, despite the complex compartmental scheme and parameter space here introduced, most of the information needed can be straightforwardly incorporated from the literature. Indeed, the free parameters which are tuned to calibrate the model are: the infectiousness per contact β_I, the asymptomatic infectious period α^{-1}, the symptomatic infectious period μ^{-1}, the household permeability ϕ and the social distancing δ. Moreover, to set the initial conditions, we arbitrarily fix the initial date as February 9$^{\text{th}}$, 2020 and assume that A_0 individuals are initially infected, whose spatial distribution matches the share of cases reported by March 3$^{\text{rd}}$, 2020 in the official data. A complete description of the parameters and their values can be found in Tables D.1-D.4.

The calibration of the model is made by using the Approximate Bayesian Computation (ABC) method [317]. The main advantage of this computational method with respect to other Bayesian approaches is that one does not have to define the functional form for the posterior distribution of the parameters. Instead the method relies on inferring this distribution from the set of parameters generating the trajectories which are accepted according to the goodness of the fit. In our case, we use the logarithmic least squares error between the number of fatalities reported up to April 8$^{\text{th}}$ and those predicted by the

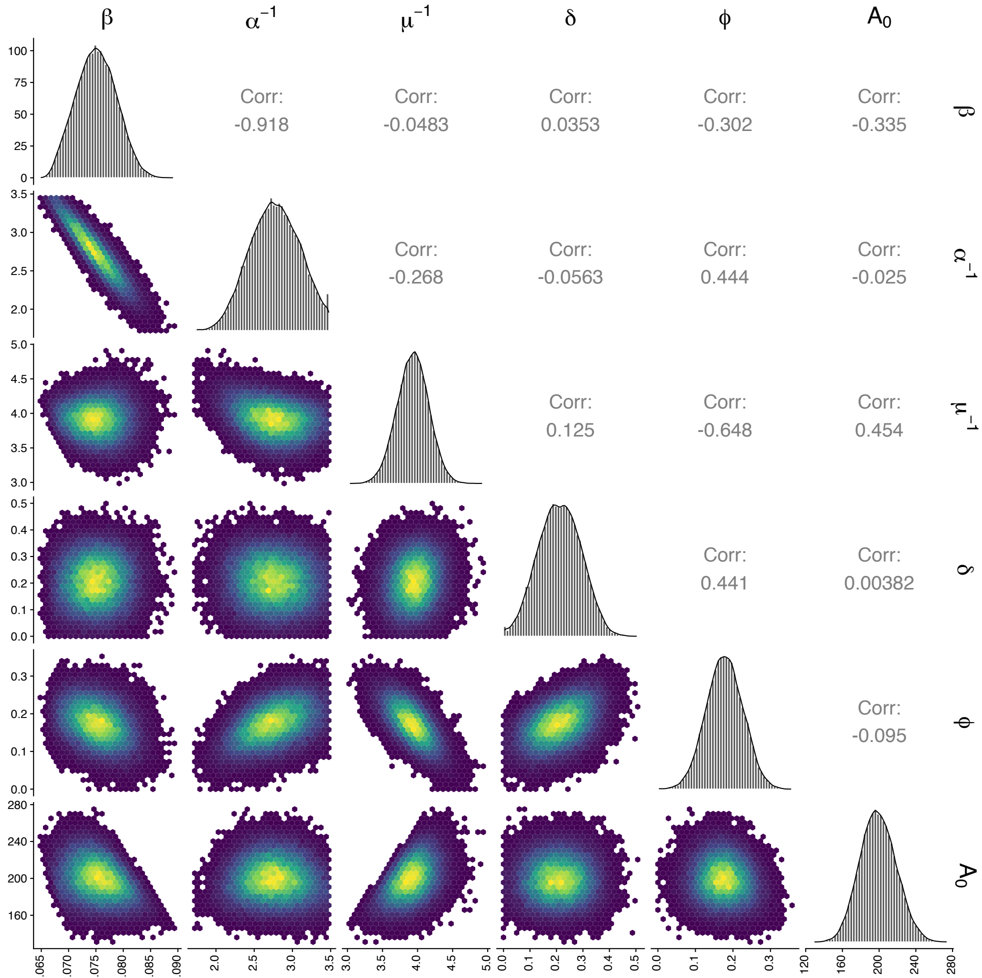


Figure 6.7: Calibration analysis: the diagonal of the table contains the posterior distributions for the parameters obtained after model calibration. The lower triangular part shows the scatter plot for every pairwise combination of parameters present in the posterior parameter space whereas the upper triangular one represent the entries of the covariance matrix for the parameters involved in the calibration.

equations as an objective function for the ABC algorithm and keep the first percentile of the simulations. The posterior distribution obtained along with the covariance matrix between the variables are represented in Fig 6.7.

In Fig 6.8a we show that the model is calibrated to correctly capture the daily number of fatalities in Spain and provides accurate forecasts of their evolution beyond the dates used for calibration purposes, which validate the formalism. To guide the readers' eye, we have introduced the course of the

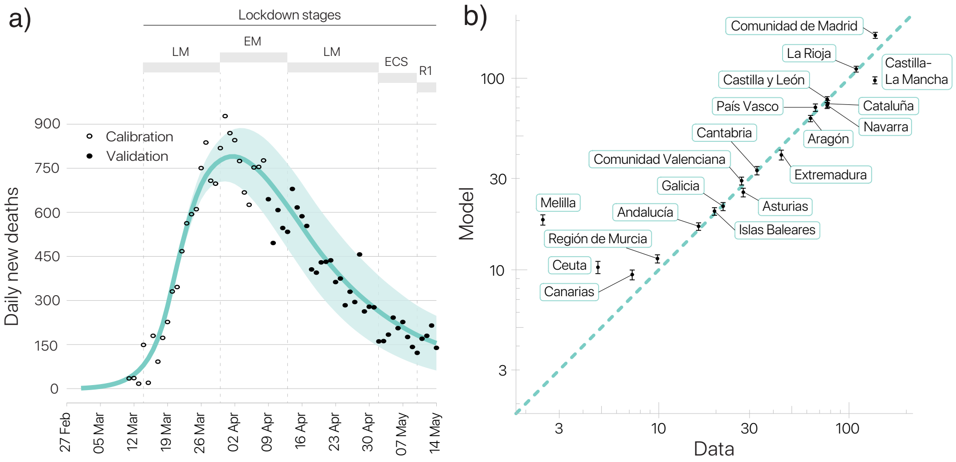


Figure 6.8: Model validation and spatiotemporal propagation of COVID-19 across Spain. Panel a): Solid lines show model predictions for the daily fatalities whereas dots correspond to real data. The shadowed areas represent the 95% prediction interval. Panel b): Correlation between the accumulated number of deaths ($r_P = 0.98$) per 100000 inhabitants observed by May 15 and those predicted by the model. The plot shows this correlation at the level of Comunidades Autónomas of Spain. Error bars cover the 95% prediction interval.

major interventions promoted by Spain authorities. The LM phase corresponds with the population lockdown where work mobility was allowed. The EM phase identifies the most strict lockdown restricting all economic activities except the essential ones. Finally, during the ECS and L1 the authorities started relaxing the policies by allowing some outdoors and commercial activities respectively.

To fully leverage the resolution of the equations, we also assess whether the formalism is able to capture the heterogeneous impact of the disease across the highest administrative units in Spain, referred to as Comunidades Autónomas. We observe that the introduced seeds are enough to predict the fatalities observed in some areas but fails in capturing the evolution of COVID-19 in other regions, possibly due to initial underreporting of cases there. To solve this discordance, we introduce the minimal set of infectious individuals in those largest populated cities of each Comunidad Autónoma to ensure a fair correspondence between data and model, which is optimized by using the Nelder-Mead method. After such correction, which does not alter the curves presented in Fig 6.8a, we observe a remarkable agreement between the predictions of the model and the data. Interestingly, some regions like Ceuta or Melilla does not present any seeds, which highlights the crucial role of case exportation from other regions driven by mobility [37, 318].

6.2.4 Effective reproduction number

Apart from reproducing the course of COVID-19 disease across Spain or forecast its evolution, our model allows assessing the effectiveness of the mobility restrictions implemented to keep the pandemic under control. In this sense, the usual avenue to quantify the stage of an ongoing outbreak is the use of the effective reproduction number $\mathcal{R}$, which quantifies the expected number of contagions made by one infected individual at some point of the outbreak. Note that this indicator differs from $\mathcal{R}_0$ for the latter only remains valid at the beginning of the outbreak, where the index case is surrounded by a fully-susceptible population. There are two different interpretations [290] for defining the effective reproduction number $\mathcal{R}(t)$ at a given time t:

- Instantaneous reproduction number $\mathcal{R}^{\text{inst}}(t)$ [291]: Expected number of secondary contagions made by infectious individuals at time t.
- Case reproduction number $\mathcal{R}^c(t)$ [102]: Expected number of secondary contagions made by one individual contracting the disease at time t.

When the parameters governing the contagions and the epidemic curves evolve smoothly, both approaches yield the same result. Nevertheless, in presence of strong NPIs, this behavior cannot be expected and the information provided by them is different. In particular, $\mathcal{R}^{\text{inst}}(t)$ constitutes a backward projection which quantifies the potential harm caused at a given time t by previously infected individuals. Therefore, this indicator is extremely useful in real-time modeling to anticipate the growth or mitigation of an outbreak. In contrast, $\mathcal{R}^c(t)$ looks forward and predict the potential contagions made by individuals infected at time t. This indicator is less affected by the noisy nature of the data and provide a better retrospective assessment of the impact of the different interventions at the expense of losing the real-time picture of the outbreak.

Due to the retrospective approach of this work, let us study the evolution of the case reproduction number, so in what follows we assume that $\mathcal{R}(t) = \mathcal{R}^c(t)$. It is worth remarking that the case reproduction number was already used in Chapter 2 when addressing the impact of contact tracing policies on the spread of diseases across contact networks. To derive its expression in this model, we propose an incremental rationale starting from basic first principles and adding the different layers of complexity of the model to eventually compute the most accurate estimation. Following a mean-field approach to the standard SIR model and assuming constant epidemiological conditions, the effective reproduction number reads:

$$\mathcal{R}(t) = (\beta_A \tau_A + \beta_I \tau_I)\langle k \rangle \, \rho_S(t) \,, \tag{6.39}$$

where β_m corresponds to the infectiousness of individual in the state m, τ_m is the time spent in each compartment and $\langle k \rangle$ denotes the average number of contacts. Note that Eq. (6.39), yet simple, identifies the different contributions to the reproduction number on which we can act to control the outbreak. For instance, the infectiousness β can be reduced by prophylactic measures, the average infectious period τ is shortened via the implementation of efficient test-tracing and isolate policies, the number of contacts $\langle k \rangle$ is reduced via lockdown policies whereas the pool of susceptible individuals ρ^S can be depleted by the vaccination of the population.

Despite the insights provided by the mean field approach, it overlooks all the specific details concerning the structure of contacts and the mobility patterns of the population, which prevents us from getting a fair assessment on the impact of the interventions. To do so, let us define the effective reproduction number associated with individuals from class g and patch i as the expected number of contagions made by one of them contracting the disease at time t over the infectious period. This quantity, denoted in the following by $\mathcal{R}_i^g$, is given by

$$\mathcal{R}_i^g(t) = \left(\beta_A^g \tau_A^g + \beta_I^g \tau_I^g\right) \sum_{j=1}^{N_P} \sum_{h=1}^{N_G} k_{ij}^{gh}(t)\, \tilde{\rho}_j^{S,h}(t)\,. \tag{6.40}$$

In the previous expression, $k_{ij}^{gh}(t)$ represents the expected number of contacts made by the infected individual inside patch j with agents from group h whereas $\tilde{\rho}_j^{S,h}$ corresponds to the fraction of these agents in the susceptible state. Taking into account the mobility network, the latter is given by

$$\tilde{\rho}_j^{S,h}(t) = \frac{\displaystyle\sum_{k=1}^{N_P} n_k \rho_k^{S,h}(t) M_{kj}^h(t)}{\displaystyle\sum_{k=1}^{N_P} n_k M_{kj}^h(t)}\,. \tag{6.41}$$

To compute $k_{ij}^{gh}(t)$, let us recall that the number of contacts made by agents from class g inside each patch j is governed by $z^g \langle k^g \rangle(t) f_j$ and the matrix $\mathbf{C}$, determining the mixing of agents from different groups. Incorporating both factors, we obtain that

$$k_{ij}^{gh}(t) = M_{ij}^g z^g(t)\, \langle k^g \rangle(t)\, f_j C^{gh}\,. \tag{6.42}$$

In a nutshell, Eq. (6.40) allows for accounting for the effect of the non-trivial contact and mobility patterns governing contagions. Nonetheless, analogously to Eq. (6.39), it assumes a static picture by considering that some epidemiological parameters such as the fraction of susceptible population or the average number of contacts remain constant over the entire infectious window of agents contracting the disease. In presence of strong NPIs, this assumption does not hold and it is necessary to reformulate the equations to take into account the system evolution as the external policies are enforced. In particular,

$$\begin{aligned}\mathcal{R}_i^g(t) = \sum_{s=t}^{\infty} \left(\zeta^{A,g}(s-t)\,\beta_A + \zeta^{I,g}(s-t)\,\beta_I\right) \\ \times \sum_{j=1}^{N_P} \sum_{h=1}^{N_G} k_{ij}^{gh}(s)\,\tilde{\rho}_j^{S,h}(s)\,, \end{aligned} \tag{6.43}$$

where $\zeta^{m,g}(s)$ is the probability that one agent from group g is in the infectious state m after s time steps since contracting the disease. The evolution of these probabilities is given by

$$\zeta^{E,g}(t+1) = (1-\eta^g)\,\zeta^{E,g}(t)\,, \tag{6.44}$$

$$\zeta^{A,g}(t+1) = \eta^g\,\zeta^{E,g}(t) + (1-\alpha^g)\,\zeta^{A,g}(t)\,, \tag{6.45}$$

$$\zeta^{I,g}(t+1) = \alpha^g\,\zeta^{A,g}(t) + (1-\mu^g)\,\zeta^{I,g}(t)\,. \tag{6.46}$$

As the individual enters the E compartment right after contracting the disease, the initial conditions are given by $\zeta^{E,g}(0) = 1$ and $\zeta^{A,g}(0) = \zeta^{I,g}(0) = 0$. Accordingly, solving explicitly the recursive Eqs. (6.44) to (6.46), we find

$$\zeta^{E,g}(t) = (1-\eta^g)^t\,, \tag{6.47}$$

$$\zeta^{A,g}(t) = \frac{\eta^g}{\alpha^g - \eta^g}\left[(1-\eta^g)^t - (1-\alpha^g)^t\right]\,, \tag{6.48}$$

$$\begin{aligned}\zeta^{I,g}(t) = {} & \frac{\eta^g\,\alpha^g}{(\alpha^g-\eta^g)\,(\mu^g-\eta^g)\,(\mu^g-\alpha^g)} \\ & \times \left[(\mu^g-\alpha^g)\,(1-\eta^g)^t + (\eta^g-\mu^g)\,(1-\alpha^g)^t \right. \\ & \left. + (\alpha^g-\eta^g)\,(1-\mu^g)^t\right]\,. \end{aligned} \tag{6.49}$$

Once we have characterized the individual reproduction number for each patch and group, we can aggregate them to characterize the reproduction number of each group g, $\mathcal{R}^g$ or the global reproduction number of the system $\mathcal{R}$. To do so, we must compute a weighted average of the individual reproduction numbers by considering the distribution of contagions across ages and patches

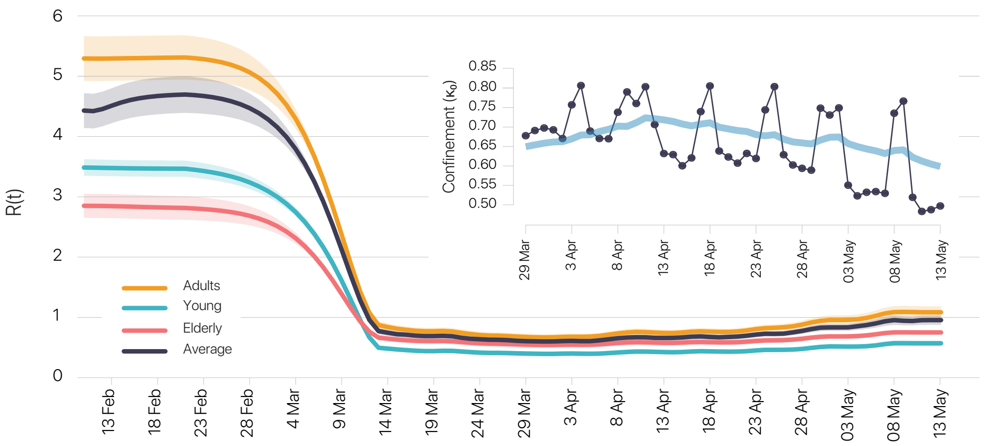


Figure 6.9: Time evolution of the effective reproduction number $\mathcal{R}^g(t)$ for each age stratum and its global value $\mathcal{R}(t)$. Inset: The black line represents the temporal evolution of the degree of confinement $\kappa_0(t)$ extracted from data daily updated by INE [316]. The blue line shows the average over a 7 days time window to smooth out fluctuations observed in the data. To improve the visibility of the inset, degrees of confinement from March 14 until March 28 are not represented.

at time t. Accordingly,

$$\mathcal{R}^g(t) = \frac{\displaystyle\sum_{m\in\{S,CH\}} \sum_{i=1}^{N_P} \left(\rho_i^{m,g}(t-1) - \rho_i^{m,g}(t)\right) n_i^g \, \mathcal{R}_i^g(t)}{\displaystyle\sum_{m\in\{S,CH\}} \sum_{i=1}^{N_P} \left(\rho_i^{m,g}(t-1) - \rho_i^{m,g}(t)\right) n_i^g}, \tag{6.50}$$

$$\mathcal{R}(t) = \frac{\displaystyle\sum_{m\in\{S,CH\}} \sum_{i=1}^{N_P} \sum_{g=1}^{N_G} \left(\rho_i^{m,g}(t-1) - \rho_i^{m,g}(t)\right) n_i^g \, \mathcal{R}_i^g(t)}{\displaystyle\sum_{m\in\{S,CH\}} \sum_{i=1}^{N_P} \sum_{g=1}^{N_G} \left(\rho_i^{m,g}(t-1) - \rho_i^{m,g}(t)\right) n_i^g} \, . \tag{6.51}$$

Fig. 6.9 represents the time evolution of the reproduction number for each group g, with $g \in \{Y, M, O\}$ and the global reproduction number along with the time series characterizing the mobility reduction observed in Spain from March 14th onward. There we check that the reproduction number is highly correlated with the mobility reduction, which suggest the relevance of the latter intervention to mitigate the first epidemic wave in Spain. Remarkably, the global reproduction number reflects a subcritical scenario $\mathcal{R} \leq 1$ while reopening the economy despite the reproduction number for the adult population corresponds to a supercritical scenario $\mathcal{R}^M \geq 1$.

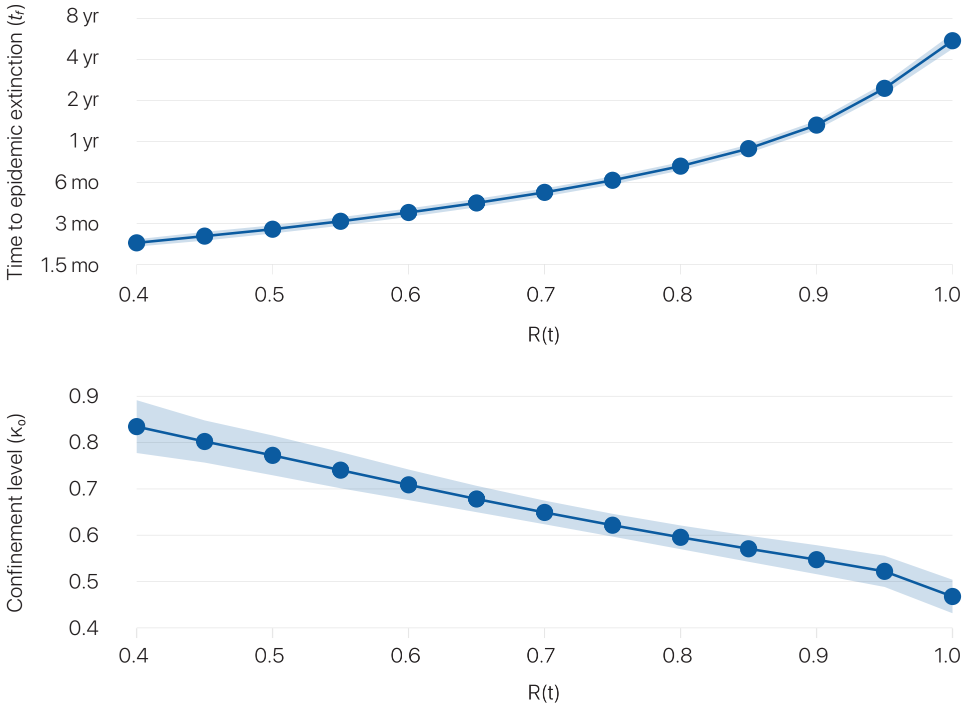


Figure 6.10: Top: Time to epidemic extinction from May 15 as the reproduction number is fixated from this day onwards. The reproduction number is kept constant by adjusting the confinement level through time. We define epidemic extinction as a daily incidence of less than 10 cases. Bottom: Average confinement level necessary to keep the reproduction number constant. At each time-step, we correct the confinement level by steps of 0.1% until the effective reproduction number $\mathcal{R}(t)$ differs less than 1% from the envisaged value.

The latter fact opens the door to age-specific control policies aiming at controlling outbreaks when minimizing the socio-economic collateral damage of the interventions at play. For this purpose, we study in Fig. 6.10 the minimum degree of confinement needed to keep $\mathcal{R}$ fixed to a given value. Note that the generality of the formalism allows exploring other different goals such as maintaining the ICU occupation at convenient levels [319] or reducing the magnitude of the epidemic peak [320]. For the sake of completeness, we also represent the extension of lockdown policies needed to suppress the disease as a function the lockdown strength in Fig. 6.10, observing that small variations in the confined population entail significant changes in the control of an outbreak.

6.2.5 Phase diagram for the impact of control policies

Eq. (6.51) tracks the evolution of the effective reproduction number but its high complexity prevents us from getting any clear insight into the role of each mechanism involved in the NPIs. For this purpose, let us now consider the existence of a single intervention at t_c characterized by the mobility reduction κ_0, the household permeability ϕ and the social distancing δ. Likewise, we neglect the heterogeneities existing among the different patches and neglect the variations in the pool of susceptible individuals during the intervention. This assumption enables to retrieve Eq. (6.39), which at time t_c is given by

$$\mathcal{R}(t_c) = \left(\beta_A \langle \tau_A \rangle + \beta_I \langle \tau_I \rangle\right) k(t_c) \rho^S(t_c) \ . \tag{6.52}$$

Taking into account the expressions for the fraction of susceptible population and the contacts at $t = t_c$ provided by Eq. (6.34) and Eq. (6.31) respectively, we obtain

$$\begin{aligned} \mathcal{R}(t_c) = \rho^S(t_c) \left(\frac{\beta^A}{\langle \alpha \rangle} + \frac{\beta^I}{\langle \mu \rangle} \right) \left(1 - \kappa_0 \left(1 - \phi\right) CH\right) \\ \times \left(\kappa_0 \left\langle k_{\mathrm{h}} \right\rangle + \left(1 - \delta\right) \left(1 - \kappa_0\right) \left\langle k_{\mathrm{h+w}} \right\rangle\right) , \end{aligned} \tag{6.53}$$

being $\langle \cdot \rangle$ the average of the different quantities over the different age groups g. In Fig. 6.11a we fix the social distancing to the median calibrated value $\delta = 0.207$ and compute the effective reproduction number as a function of the mobility reduction κ_0 and the household permeability ϕ. There we observe that both reducing the mobility of the population or the mixing between different households favor the control of the disease. It is also worth noticing that the phase diagram is divided into two regions: the supercritical and the subcritical regimes. Both regions are separated by the line $\mathcal{R}(t_c) = 1$ corresponding with a scenario in which the daily incidence, i.e. the number of new cases, remains constant.

The expression provided by Eq. (6.53) allows estimating the minimum degree of confinement which would have been required to keep COVID-19 under control in Spain, with $R(t_c) = 1$. In particular, after some algebra, we have that

$$A \left(\kappa_0^c\right)^2 - B \, \kappa_0^c + C = 0 \, , \tag{6.54}$$

with

$$A = (1 - \phi) \left\langle CH \right\rangle \left(\left\langle k_{\mathrm{h+w}} \right\rangle (1 - \delta) - \left\langle k_{\mathrm{h}} \right\rangle \right) \ , \tag{6.55}$$

$$B = \left\langle k_{\mathrm{h+w}} \right\rangle (1 - \delta) \left(1 + \left\langle CH \right\rangle (1 - \phi)\right) - \left\langle k_{\mathrm{h}} \right\rangle , \tag{6.56}$$

$$C = \left\langle k_{\mathrm{h+w}} \right\rangle (1 - \delta) - \frac{1}{\rho^S \left(\frac{\beta^A}{\langle \alpha \rangle} + \frac{\beta^I}{\langle \mu \rangle} \right)} \, . \tag{6.57}$$

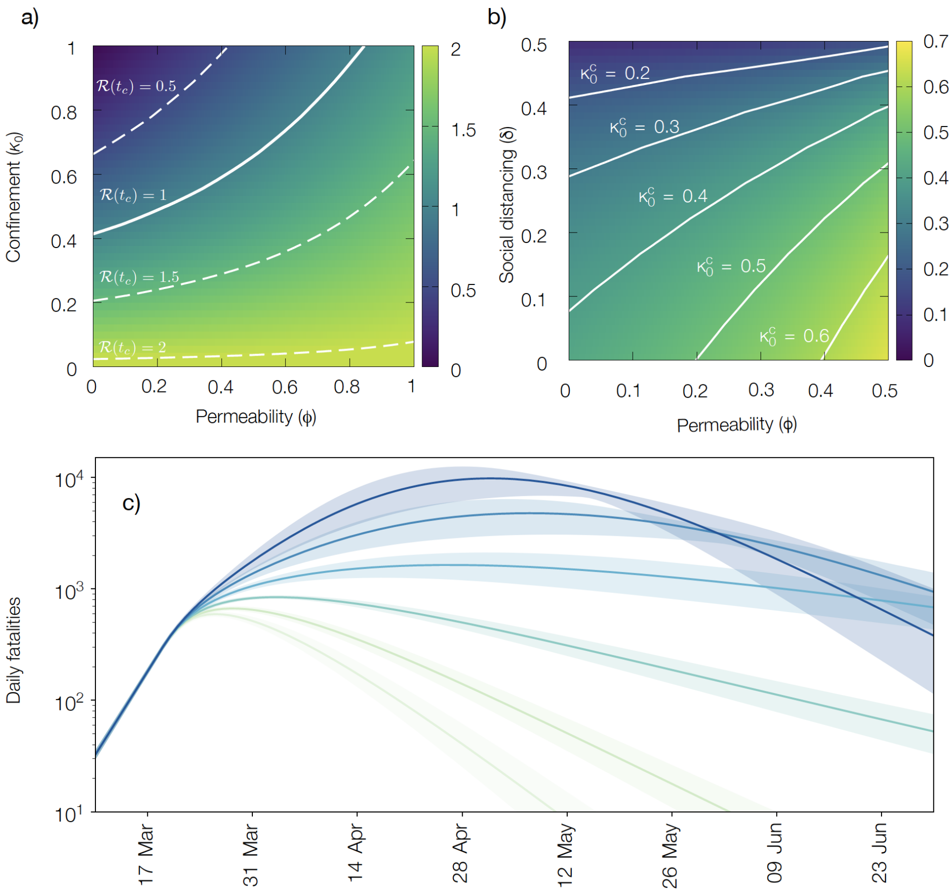


Figure 6.11: Panel a): Effective reproduction number when mobility restrictions are imposed $\mathcal{R}(t_c)$ (color code) as a function of the confinement κ_0 and the household permeability ϕ computed from Eq. (6.53). The white solid line reflects the condition $\mathcal{R}(t_c) = 1$ separating the different regimes for which enforcing the confinement leads to the flattening or bending of the epidemic curve, respectively. The dashed white lines represent contour lines for different values of $\mathcal{R}(t_c)$. The social distancing value is fixed to $\delta = 0.207$. Panel b): Critical confinement κ_0^c computed from Eq. (6.58) needed to ensure that the system reaches the bending regime, i.e. $\mathcal{R} = 1$, as a function of the household permeability ϕ and the social distance δ. Panel c): Temporal evolution of the daily fatalities as a function of the confined population κ_0 in a single intervention taking place on March 14 while fixing $(\delta, \phi) = (0.207, 0.174)$.

Therefore, the critical value κ_0^c is given by

$$\kappa_0^c = \frac{B \pm \sqrt{B^2 - 4\,A\,C}}{2\,A}\,. \tag{6.58}$$

Note that we must retain the physical meaningful solution for which $\kappa_0^c \in [0, 1]$. We represent in Fig. 6.11b the dependence of the critical confinement on both the social distancing and the household permeability. Remarkably, we observe the relevance of establishing social bubbles by precluding the mixing between different households, for decreasing the permeability leads to substantial reduction in the critical confinement values.

To complete the picture, we study the impact of the confined population on the dynamics. For this purpose, we set $(\delta, \phi) = (0.209, 0.174)$, i.e. $\kappa_0^c = 0.44$ and study how the time evolution of the daily fatalities evolves as κ_0 is varied. We observe two clearly differentiated behaviors: when $\kappa_0 > \kappa_0^c$ reducing mobility *flattens* the curve by delaying and reducing its peak. In contrast, when $\kappa_0 > \kappa_0^c$, this intervention *bends* the epidemic curve by anticipating and reducing the peak. The qualitative differences between both behaviors resemble the expected outcome of reducing the reproduction number in supercritical (flattening) and subcritical (bending) regimes for SIR-like models.

6.3 Vector-borne diseases

COVID-19 represents a single example within the vast realm of communicable diseases circulating around the globe. Well before COVID-19 emergence, communicable diseases already represented one of the leading causes of deaths worldwide, specially when referring to low-income countries [321]. In this sense, traditional SIS and SIR models has been adapted to accommodate the specific clinical features of some of the deathliest diseases transmitted among humans such as Tuberculosis [270, 322] or the HIV/AIDS [323, 324]. Nonetheless, these models need to be rethought when modeling the spread of vector-borne diseases (VBD), whose spread is not driven by the close interaction between hosts, but instead requires the participation of an external agent carrying the pathogen from the infectious individual to the susceptible one.

VBD transmitted by mosquitoes like Dengue, Chikungunya or Malaria are responsible for million of deaths worldwide, being especially harmful in those tropical areas whose humidity and temperature ranges favors the proliferation of the vectors driving its transmission [325]. In this sense, the concerns on the impact of these diseases have been continuously growing as a byproduct of vectors adaptation [326] and climate change [327–330], which have progressively turn formerly hostile areas for mosquitoes survival into endemic regions. Proofs of this phenomenon are the increasing transmission of West-Nile virus or the first Zika local cases acquired in Europe [331, 332].

Mathematical modeling of the spread of VBD [78, 333–339] is a very active field in light of their current relevance and their potential harmfulness in the short and medium term. The theoretical foundations for most of the formulated frameworks were established by the seminal Ross-Macdonald model proposed in 1911 by Ronald Ross [51] and revisited in 1957 by George Macdonald [340].

6.3.1 Ross-Macdonald model

Let us consider that our system is composed of N_H hosts and N_M vectors. The Ross-Macdonald (RM) model assumes that both vectors and hosts can be either Susceptible or Infected. Regarding the mechanisms driving contagions, the RM model assumes that each vector makes β bites per time step. Moreover, susceptible hosts (vectors) contract the disease when contacting with infectious vectors (hosts) with probability λ^{MH} (λ^{HM}). Finally, infected hosts are assumed to recover at a given rate μ^I and become susceptible again whereas infected vectors die at a rate μ^M, being replaced by newborn susceptible ones. The basic processes described above are schematized in Fig. 6.12.

Following a mean field approach, the dynamics is completely characterized by the time evolution of the fraction of hosts' and vectors' populations in the infected state, denoted in what follows by ρ^H and ρ^M. Namely,

$$\dot{\rho}^H = -\mu^H \rho^H + (1-\rho^H)\beta\gamma\lambda^{MH}\rho^M \ , \tag{6.59}$$

$$\dot{\rho}^M = -\mu^M \rho^M + (1-\rho^M)\beta\lambda^{HM}\rho^H \ , \tag{6.60}$$

where γ denotes the ratio between vectors' abundance and host' population, i.e. $\gamma = N_M/N_H$. Conceptually, the RM model resembles a double SIS model where contagions take place among vectors and hosts. In this sense, there exists a phase transition between a disease-free state and an endemic equilibrium taking place at $\mathcal{R}_0 = 1$. To derive the basic reproduction number, we cast Eqs. (6.59) and (6.60) into

$$\dot{\vec{\rho}} = (\mathbf{F} - \mathbf{V})\vec{\rho} \ , \tag{6.61}$$

with

$$F = \begin{pmatrix} 0 & \beta\gamma\lambda^{MH} \\ \beta\lambda^{HM} & 0 \end{pmatrix} \ , \tag{6.62}$$

$$V = \begin{pmatrix} \mu^H & 0 \\ 0 & \mu^M \end{pmatrix} \ . \tag{6.63}$$

The NGM approach introduced in Chapter 5 allows computing $\mathcal{R}_0$ as

$$\mathcal{R}_0 = \Lambda_{\max}(\mathbf{F}\mathbf{V^{-1}}) = \sqrt{\frac{\beta^2\gamma\lambda^{MH}\lambda^{HM}}{\mu^H\mu^M}} \ . \tag{6.64}$$

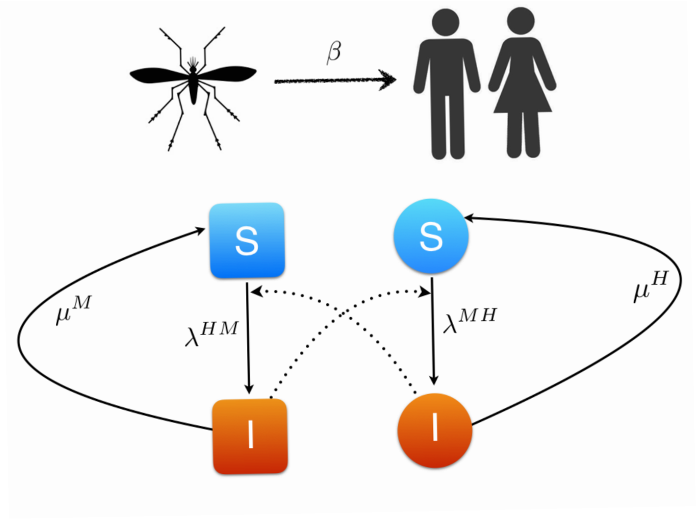


Figure 6.12: Schematic representation of the processes described in the RM model. Relevant parameters are: *(i)* the probability that an infected vector transmits the disease to a healthy individual, λ^{MH}, *(ii)* the probability that an infected human transmits the disease to a healthy vector, λ^{HM}, *(iii)* the feeding rate of vectors β, *(iv)* the probability that an infected human recovers, μ^H, and *(v)* the mortality rate of vectors, μ^H.

6.3.2 MIR model for vector-borne diseases

The mean field approach introduced in the original RM model cannot explain the heterogeneous impact of VBD in urban areas [341, 342]. Furthermore, an important control mechanism to keep VBD such as Dengue in subcritical regimes is the introduction of Wolbachia bacteria in the mosquitoes population [343–345]. When resources are limited, it is crucial to determine those areas triggering contagions, where the introduction of Wolbachia should be promoted. All these shortcomings of the original RM model call for the addition of a spatial layer to the model. To do so, we extend the MIR model proposed in Chapter 4 to accommodate the mechanisms of the RM model in the reaction phase of our framework.

Let us assume that our system corresponds to a metapopulation composed of N_P patches. Each patch of the metapopulation i is characterized by its number of residents n_i and the vectors' abundance m_i. We assume that mobility is dependent on the dynamical state of the individual. This assumption is rooted in the fact that symptoms associated with VBD are usually severe and limit the mobility of the infected individuals. To capture this fact, we assume that susceptible individuals move with probability p whereas infected population moves with probability αp, with $\alpha \in [0, 1]$. Note that, if symptoms are severe enough to completely hinder mobility, α might be also interpreted as the fraction of asymptomatic infectious individuals. Finally, we force vectors to stay in their associated areas, for their typical flying ranges, ranging from 100 to 400 meters [346, 347], are negligible with respect to the usual spatial resolution of the patches determined by data availability.

Given the nature of the RM model, the system's dynamics is completely described by $2 \times N$ equations governing the time evolution of the fraction of population and vectors in the infected state in each patch i, which are denoted by ρ_i^H and ρ_i^M respectively. Specifically, these equations read as

$$\rho_i^H(t+1) = \rho_i^H(t)(1-\mu^H) + (1-\rho_i^H(t))\Pi_i^H(t), \tag{6.65}$$

$$\rho_i^M(t+1) = \rho_i^M(t)(1-\mu^M) + (1-\rho_i^M(t))\Pi_i^M(t) \ . \tag{6.66}$$

$\Pi_i^H(t)$ ($\Pi_i^M(t)$) corresponds to the probability than one susceptible individual (vector) associated with i gets infected. The former is given by

$$\Pi_i^H(t) = (1-p)P_i^H(t) + p\sum_{j=1}^{N} R_{ij}P_j^H(t) \ , \tag{6.67}$$

where R_{ij} encodes the entries of the OD matrix and $P_i^H(t)$ is the probability that an agent placed in population i at time t is infected. This probability can be written as

$$P_i^H(t) = 1 - \left(1 - \lambda^{MH}\rho_i^M \frac{1}{\tilde{n}_i(t)}\right)^{\beta m_i} \ , \tag{6.68}$$

where βm_i represents the total number of bites made by vectors associated with patch i. Note that, as a result of coupling mobility and dynamics, the effective population in patch i, $\tilde{n}_i$ is a time-dependent variable. Namely,

$$\begin{aligned}\tilde{n}_i(\alpha, p, \vec{\rho}(t)) = {} & \left[1 - p\left(1-(1-\alpha)\rho_i^H(t)\right)\right] n_i \\ & + p\sum_{j=1}^{N} R_{ji}\left(1-(1-\alpha)\rho_j^H(t)\right) n_j \ .\end{aligned} \tag{6.69}$$

The probability of infection for vectors is much easier, for they always remain in their associated patch. In particular,

$$\Pi_i^M(t) = 1 - \left(1 - \lambda^{HM}\frac{\tilde{I}_i(t)}{\tilde{n}_i(t)}\right)^{\beta} \ , \tag{6.70}$$

where $\tilde{I}_i(t)$, which represents the effective infected population of patch i at time t, is given by

$$\tilde{I}_i(t) = (1-\alpha p)n_i\rho_i^H(t) + \alpha p\sum_{j=1}^{N} R_{ji}n_j\rho_j^H(t) \ . \tag{6.71}$$

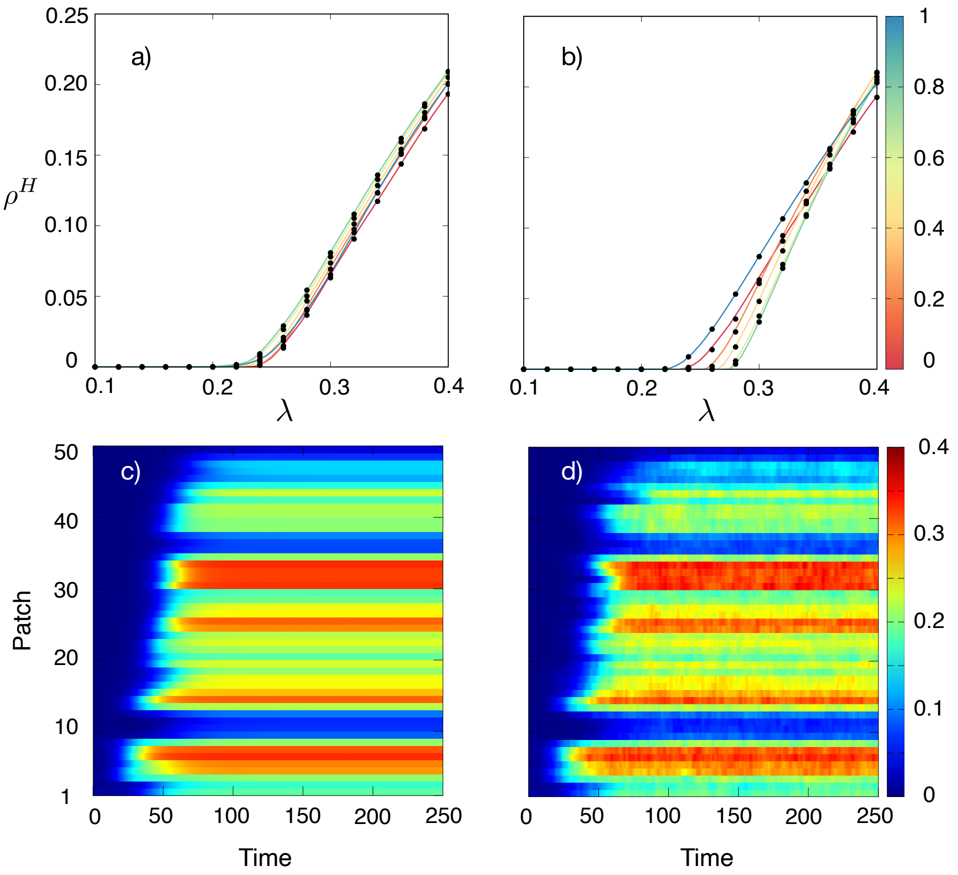


Figure 6.13: Top: Epidemic size ρ_H as a function of the contagion rate between vectors and humans $\lambda^{MH} = \lambda^{HM} = \lambda$ and the human mobility p (color code). Dots correspond to results of averaging 50 stochastic realizations whereas solid lines represent the theoretical predictions obtained by iterating the equations of the formalism. Model parameters have been fixed to $(\beta, \mu^H, \mu^M) = (1, 0.3, 0.3)$. The values assumed for the restriction of the mobility of infected agents are (a) $\alpha = 0$ and (b) $\alpha = 1$. Bottom: Temporal evolution of the fraction of infected agents (color code) according to the equations of the formalism (Panel c) and by tracking the individual state of each agent while performing a single Monte Carlo simulation (Panel d). Infectivity is fixed to $\lambda = 0.3$ whereas mobility parameter is set to $p = 0.3$. The rest of parameters are the same as in Panel a).

Eqs. (6.65) and (6.66) allows tracking the spatio-temporal evolution of VBD in networked metapopulation systems. As a sanity check, we validate them by comparing their predictions with results from mechanistic Monte Carlo simulations in a synthetic metapopulation, a BA network with $N_P = 50$ patches and $\langle k \rangle = 4$. We assume that all the patches are homogeneously populated by $n = 1000$ inhabitants and draw vectors' abundance from a uniform distribution spanning the range $m_i \in [300, 1700]$.

Let us first study the epidemic curves relating the epidemic size ρ^H to the infectivities λ^{HM} and λ^{MH}. For the sake of simplicity, we fix $\lambda^{HM} = \lambda^{MH} = \lambda$. Recall that the epidemic size gives information about the state of the system in the stationary regime via

$$\rho^H = \frac{\sum\limits_{i=1}^{N_P} n_i \rho_i^H(\infty)}{\sum\limits_{i=1}^{N_P} n_i} \ . \tag{6.72}$$

Figs. 6.13a-b show that the equations fairly reproduce the impact of mobility p on the epidemic curves for $\alpha = 0$ (panel a) and $\alpha = 1$ (panel b). Interestingly, the behavior of the system close to the epidemic threshold is highly influenced by the mobility of the infected population α. In particular, when mobility is independent of the host state ($\alpha = 1$) we observe an epidemic detriment similar to the one reported for diseases transmitted among humans, whereas this behavior is lost when neglecting the mobility of the infected population ($\alpha = 0$).

To further validate the equations, we check whether the time evolution of the epidemic trajectories can be retrieved with our model. To do so, in Figs. 6.13c-d, we fix the λ value and start a localized outbreak in a single patch, letting the system evolve according to the Markovian equations (panel c) or the microscopic processes occurring in a single Monte Carlo realization (panel d). There we check that the equations are able to reflect the infection pathways observed in the metapopulation, despite the noisy nature of the simulation induced by the stochasticity of the underlying processes.

Fig. 6.13 reveals a change in the behavior of the epidemic threshold as the new parameter α is tuned. We now aim at deriving an analytical expression of the epidemic threshold to capture this phenomenon. For this purpose, we apply the NGM method. Recall that close to the epidemic threshold, the size of the infected populations for both vectors and hosts are negligible. Therefore, we can assume

$$\rho_i^H(t) = \epsilon_i^H(t) \ll 1 \tag{6.73}$$

$$\rho_i^M(t) = \epsilon_i^M(t) \ll 1 \ . \tag{6.74}$$

Both assumptions allow us to linearize the equations governing the dynamics. Specifically, Eq. (6.70) now reads as

$$\Pi_i^M(t) \simeq \frac{\lambda^{HM}\beta}{\tilde{n}_i^0} \sum_{j=1}^{N_P} \left[\delta_{ij}(1-\alpha p) + \alpha p R_{ji}\right] n_j \epsilon_j^H(t) \ , \tag{6.75}$$

where we have defined $\tilde{n}_i^0 \equiv \tilde{n}_i(\alpha, p, 0)$. In its turn, Eq. (6.68) can be approximated as

$$P_i^H(t) \simeq \lambda^{MH}\beta\frac{m_i}{\tilde{n}_i^0}\epsilon_i^M(t) \ . \tag{6.76}$$

Introducing Eqs. (6.73)-(6.76) into Eqs. (6.65)-(6.66) and keeping only linear terms, we obtain:

$$\begin{aligned}
\dot{\epsilon}_i^H &= -\mu^H\epsilon_i^H(t) + \lambda^{MH}\beta\sum_{j=1}^{N_P}\left(pR_{ij}\frac{m_j}{\tilde{n}_j^0} + (1-p)\delta_{ij}\frac{m_i}{\tilde{n}_i^0}\right)\epsilon_j^M, && (6.77)\\
\dot{\epsilon}_i^M &= -\mu^M\epsilon_i^M(t) + \lambda^{HM}\beta\sum_{j=1}^{N_P}\left(\alpha pR_{ji}\frac{n_j}{\tilde{n}_i^0} + (1-\alpha p)\delta_{ij}\frac{n_i}{\tilde{n}_i^0}\right)\epsilon_j^H, && (6.78)
\end{aligned}$$

where we have defined $\dot{\epsilon}_i^x = \epsilon_i^x(t+1) - \epsilon_i^x(t)$, with $x = \{H, M\}$. In terms of overall populations rather than fractions, the evolution of the system is given by

$$\begin{aligned}
\dot{\tilde{\epsilon}}_i^H &= -\mu^H\tilde{\epsilon}_i^H(t) + \lambda^{MH}\beta\sum_{j=1}^{N_P}\underbrace{\left(pR_{ij}\frac{n_i}{\tilde{n}_j^0} + (1-p)\delta_{ij}\frac{n_i}{\tilde{n}_i^0}\right)}_{M_{ij}}\tilde{\epsilon}_j^M, && (6.79)\\
\dot{\tilde{\epsilon}}_i^M &= -\mu^M\tilde{\epsilon}_i^M(t) + \lambda^{HM}\beta\sum_{j=1}^{N_P}\underbrace{\left(\alpha pR_{ji}\frac{m_i}{\tilde{n}_i^0} + (1-\alpha p)\delta_{ij}\frac{m_i}{\tilde{n}_i^0}\right)}_{\tilde{M}_{ij}}\tilde{\epsilon}_j^H, && (6.80)
\end{aligned}$$

where $\tilde{\epsilon}_i^H = n_i\epsilon_i^H$ and $\tilde{\epsilon}_i^M = m_i\epsilon_i^M$. Recall that in the NGM method, the tensor **F** contains all the contagions of the susceptible population whereas the tensor **V** encodes the transitions affecting the infected agents. Following these definitions, the former system of equations can be casted into

$$\dot{\tilde{\epsilon}}_l^x = \left(F_{ij}^{xl} - V_{ij}^{xl}\right)\tilde{\epsilon}_j^l \ , \tag{6.81}$$

with

$$\begin{aligned}
F_{ij}^{xl} &= \delta^{xH}\delta^{lM}\beta\lambda^{MH}M_{ij} + \delta^{xM}\delta^{lH}\beta\lambda^{HM}\tilde{M}_{ij} \ , && (6.82)\\
V_{ij}^{xl} &= \delta^{xH}\delta^{lH}\mu^H + \delta^{xM}\delta^{lM}\mu^M \ . && (6.83)
\end{aligned}$$

Given Eq. (6.81), the NGM allows computing the basic reproduction number $\mathcal{R}_0$ as

$$\mathcal{R}_0 = \Lambda_{\max}(\mathbf{F}\mathbf{V}^{-1}) = \sqrt{\frac{\beta^2\lambda^{HM}\lambda^{MH}}{\mu^M\mu^H}}\sqrt{\Lambda_{\max}(\mathbf{M}\tilde{\mathbf{M}})} \ . \tag{6.84}$$

Note that, for $p = 0$, we have that $M_{ij} = \delta_{ij}$ and $\tilde{M}_{ij} = \gamma_i \delta_{ij}$, with $\gamma_i = m_i/n_i$. Therefore, in the static case, the basic reproduction number is

$$\mathcal{R}_0(p=0) = \max\left\{ \sqrt{\frac{\beta^2 \lambda^{HM} \lambda^{MH} \gamma_i}{\mu^M \mu^H}} \right\} , \tag{6.85}$$

i.e. the largest individual reproduction number computed by Eq. (6.64) within the same of patches of the metapopulation. In general, the value of the basic reproduction number depends on a nontrivial combination of the mobility parameters p and α, the OD matrix $\mathbf{R}$ and the distribution of vectors and individuals across the system, $\vec{m}$ and $\vec{n}$ respectively. To estimate the value of the epidemic threshold, we fix $\lambda^{HM} = \lambda^{MH} = \lambda$ and recall that $\mathcal{R}_0(\lambda_c) = 1$. Imposing this constraint into Eq. 6.84, we finally obtain

$$\lambda_c = \sqrt{\frac{\mu^H \mu^M}{\beta^2 \Lambda_{\max}(\mathbf{M}\tilde{\mathbf{M}})}} . \tag{6.86}$$

Fig. 6.14 reveals the accuracy of the estimation for the epidemic threshold, for it perfectly captures the boundaries between the disease-free and the endemic equilibria when representing the phase diagrams of the system. Beyond this agreement, there we confirm that the impact of the mobility p on the threshold is highly determined by the α value. Specifically, we observe a wide range of behaviors, ranging from a non-monotonic epidemic detriment for $\alpha = 1$ to a non-monotonic reduction of the epidemic threshold for $\alpha = 0$, passing through an oscillatory regime for $\alpha = 0.5$. The absence of epidemic detriment in the case of $\alpha = 0$ can be associated with the negative effect of removing the mobility of infectious agents located in the contagion hotspots to the rest of the metapopulation, reported in Section 6.1.

Regardless of the qualitative behavior of the threshold, the three scenarios explored share a common feature: there is a critical mobility value p_c for which an abrupt change in the slope of the epidemic threshold occurs. This kind of behavior has already been characterized in Chapter 4 and encodes a change in the localization of the leading eigenvector and therefore, a change in the localization of the epidemic outbreaks. To get more insights into the localization of the most affected areas, we represent in Fig 6.14 the components of the leading eigenvector of the matrix $\mathbf{M}\tilde{\mathbf{M}}$ and highlight the p_c values. Remarkably, for $\alpha = 0$, we observe a second critical point located at $p'_c \simeq 0.92$.

6.3.3 Dengue in Cali

From a practical point of view, the previous results have profound implications for the effectiveness of control policies targeting specific areas such as the

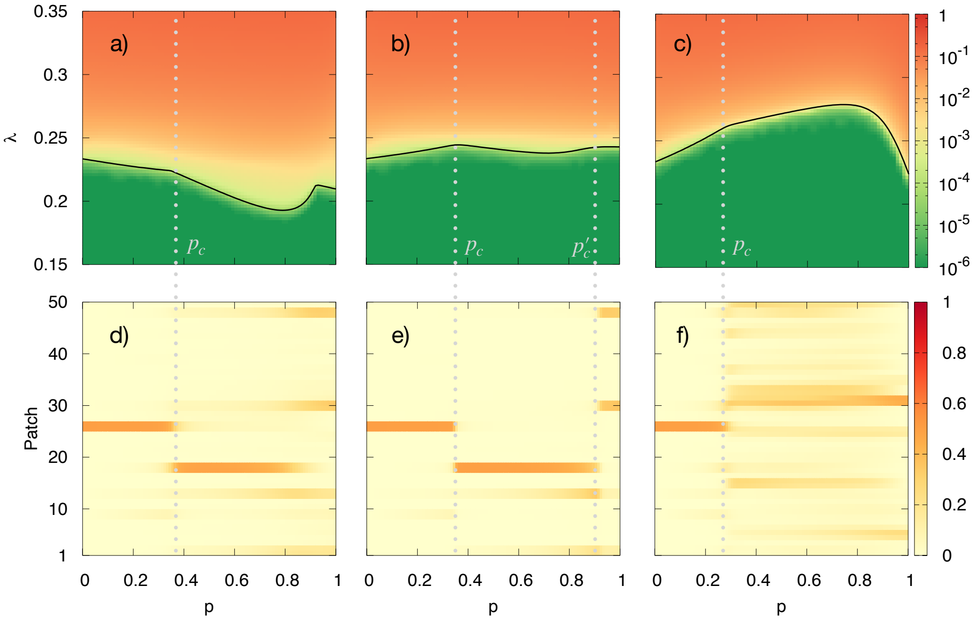


Figure 6.14: Panels (a)-(c) in the top show three epidemic diagrams $\rho^H(p, \lambda)$ in a synthetic metapopulation of $N = 50$ patches. Each panel corresponds to a different re-scaling value (α) for the mobility of infected humans, namely: (a) $\alpha = 0$, (b) $\alpha = 0.5$ and (c) $\alpha = 1$. In addition, we have set $\lambda^{HM} = \lambda^{MH} = \lambda$ while the rest of the RM parameters are: $\mu^H = 0.3$, $\mu^M = 0.3$, and $\beta = 1.0$. The color code shows the incidence ρ^H as obtained from agent-based simulations while the solid curve represents the prediction for the epidemic threshold, λ_c, calculated from Eq. (6.86). The bottom panels (d)-(f) show the evolution, as a function of p, of the N_P components of the eigenvector of matrix $\mathbf{M}\tilde{\mathbf{M}}$ corresponding to the maximum eigenvalue $\Lambda_{\max}(\mathbf{M}\tilde{\mathbf{M}})$.

introduction of the Wolbachia bacteria. In this section, we apply the machinery developed above to study a real case: the spread of Dengue across Cali, which constitutes an endemic region due to the weather conditions facilitating the proliferation of the carriers of this disease, the *Aedes aegypti* and *Aedes albopictus* mosquitoes. Specifically, here we aim at deriving a data-driven indicator, hereinafter referred to as epidemic risk ER, which informs about the vulnerability of the different areas across Cali. To do so, we exploit the physical meaning of the elements of the critical matrices $\mathbf{M}$ and $\tilde{\mathbf{M}}$.

Specifically, M_{ij} corresponds to the expected number of interactions between a single infected vector from j and the population from patch i. In its turn, $\tilde{M}_{ij}$ contains the expected interactions between one individual from patch i and the vector population in patch j. We define the epidemic risk of one area i, ER_i, as the total number of indirect interactions of the residents there with the rest of the population in the system, mediated by the participation of a

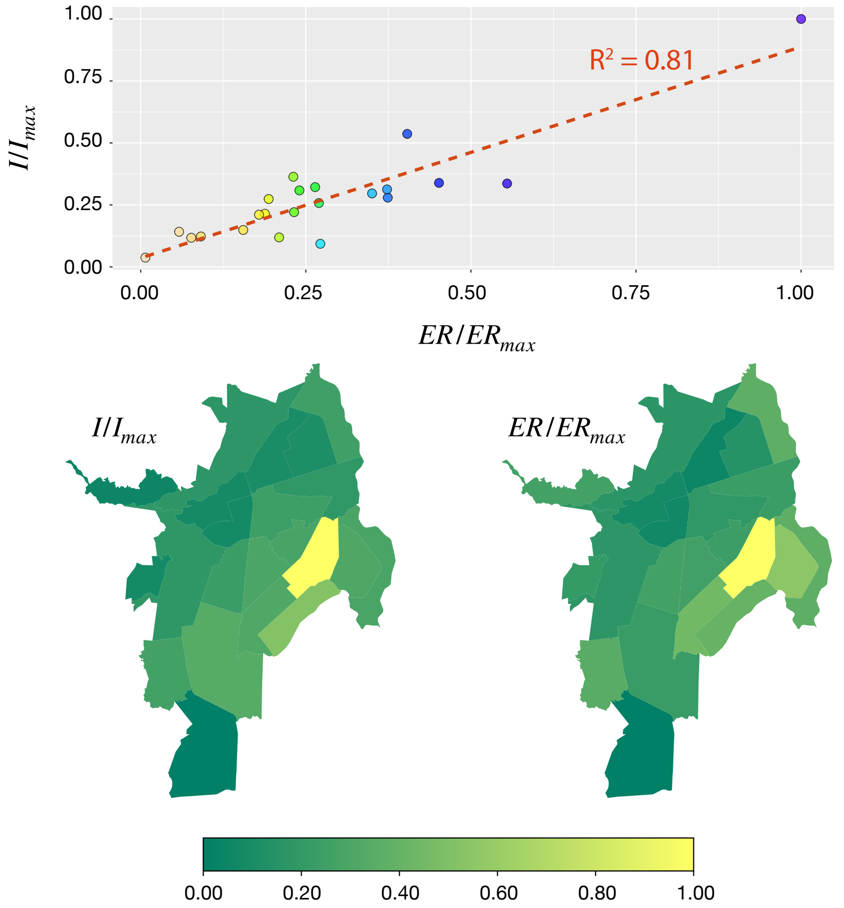


Figure 6.15: Top: Normalized epidemic risk (ER/ER_{max}) and normalized Dengue prevalence (I/I_{max}) for each of the 22 districts of Cali. Color encodes the Epidemic Risk, from the lower (yellow) to the highest (blue).The correlation between the two variables yields a coefficient of determination of $R^2 = 0.81$. Bottom: Spatial distributions of the normalized Dengue prevalence in the city of Cali (left) and the normalized epidemic risk (right) according to Eq. (6.87). The parameters concerning agents' mobility have been set to $(p, \alpha) = (0.36, 0.75)$.

vector. In terms of the variables of the model,

$$ER_i = \sum_{j,k=1}^{N_P} n_j \tilde{M}_{jk} M_{ki} \ . \tag{6.87}$$

To test the performance of the new data-driven indicator, we must add the information about the spatial distribution of the vectors on top of the Cali metapopulation constructed in Chapter 4. For this purpose, we rely on reports yearly released by the local authorities [348] containing different entomological indicators at the level of *comuna*, which constitutes the definition of patch in our metapopulation. In particular, we assume that the ratio between mosquitoes and residents in each area i, γ_i, is proportional to the *recipient index*, which is computed as the proportion of containers deployed across the different areas containing *Aedes* pupae. Regarding the mobility, we fix $\alpha = 0.75$, which reflects the number of asymptomatic Dengue cases [349, 350] and leave p as a free parameter to be calibrated.

To get an empirical estimation of the exposure of the different areas, we aggregate the Dengue cases reported from 2015 to 2016 in the epidemic surveillance reports published by the local authorities [351, 352]. To estimate the mobility p, we seek to optimize the correlation between both the empirical and the theoretical risk indicators, yielding $p = 0.36$. Interestingly, this value constitutes a reasonable estimation, for it implies that individuals spent 36% of their time inside their workplaces. Finally, we compare in Fig 6.15 the spatial distribution of cases according to the data with the epidemic risk indicator, finding a remarkable agreement between both of them ($R^2 = 0.81$). Therefore, we prove that simple metapopulation models as the one proposed here, comprising demographic, social and entomological features of a given city, are complex enough to provide reliable information about the spatial dissemination of VBD and guide the design of control policies.

6.4 Conclusions

From the birth of mathematical epidemiology in the XVIII century, epidemic models have been continuously evolving towards two complementary directions. On the one hand, most of the theoretical work has been devoted to add layers of complexity to the original mean-field theories and study how the new information alters progressively the phase diagrams and the critical properties of the models. On the other hand, leveraging the increasing complexity of the theoretical frameworks, different data-driven models have been proposed to address real epidemic scenarios with the aim of providing reliable estimations on the progress of an outbreak and test interventions to mitigate it. In this sense, while Chapters 4 and 5 are restricted to the theoretical domain, in this chapter we have shown how the models proposed throughout this thesis can represent useful tools to study real outbreaks.

To further characterize the interplay between daily urban flows and epidemic spreading in real metapopulations, we have first quantified the dependence of the epidemic threshold on the mobility for 163 cities distributed worldwide. The theoretical insights obtained when analyzing the origin of the epidemic detriment for the simple star-like metapopulation in Chapter 5 have pushed us to perform a dimensionality reduction of the mobility network and extract an universal trend: those cities with a large volume of commuters across densely populated areas are more vulnerable to the spread of diseases. Based on the latter results, we have proposed realistic interventions on cities' architecture to increase their resilience to future epidemic outbreaks, finding that cities must evolve towards more egalitarian distribution of facilities, thus hindering the crowding of individuals inside their hotspots. Finally, we have shown that a combination of the population density and the distribution of flows across the city provides a plausible mechanism to explain the heterogeneous growth rates for COVID-19 cases during the first wave in the United States.

Despite the information provided by the former analysis, the complex features characterizing the spread of SARS-CoV-2 and the clinical course of COVID-19 call for more sophisticated compartmental models to forecast its evolution and the impact of the different NPIs promoted to cut down its transmission worldwide. Here we have tackled this challenge and generalized the multiplex MIR model introduced in Chapter 4 to model COVID-19 evolution in Spain during the first epidemic wave. The proposed model reproduces very accurately the evolution of the daily fatalities in Spain across both national and regional scales. To estimate the impact of the different NPIs, we have derived a numerical estimation for the evolution of the effective reproduction number, revealing how relevant the different lockdown phases implemented in Spain were to control the pandemic. Finally, we have performed a mean-field analysis on the latter expression, which has enabled to derive the critical confinement needed to ensure a proper control of the disease. In this sense, the inverse dependence of the critical confinement on the mixing of individuals from different households reveals the relevance of the creation of social bubbles to control an outbreak when trying to minimize the collateral socio-economic damage of the restrictions in place.

To round off this chapter, we have extended the original MIR model to accommodate the contagion mechanisms of VBD. Remarkably, we have found that the qualitative behavior of the epidemic threshold as mobility varies is strongly influenced by how the symptoms associated with the disease affects the mobility of the host. In addition, we have found that abrupt changes in the localization of epidemic outbreaks occur when tuning mobility, which constitutes a relevant result providing most policies promoted to control VBD target specific areas via the introduction of Wolbachia bacteria. Finally, we have addressed the spread of Dengue across the city of Cali and derived a theoretical indicator, the epidemic risk, from the matrices determining the epidemic threshold. This indicator incorporates demography, mobility, vectors' abundance and their complex interplay and captures very accurately the spatial distribution of cases reported in Cali from 2015 to 2016.

The MIR model provides the basic building blocks to construct realistic models allowing us to characterize the spread of real diseases. In this sense, the generality of the recurrent mobility scheme introduced in the MIR model makes it easily extendable to address not only other infectious diseases but also spreading phenomena of different nature. For instance, the modification of the rules governing the reactions of the agents could make the MIR model a suitable framework to explore diverse problems such as the diffusion of cultural traits among different hunter-gatherers societies [246, 247] or the complex interactions governing ecosystems' stability [353].

Chapter 7

Conclusions

Finally, we want to conclude this thesis by giving a global summary of our most relevant findings and discuss possible future research directions arising from the models here proposed. The main objective of this thesis has been to abandon the study of individual isolated dynamics to account for the adaptability of complex systems and characterize the phenomena resulting from intertwining different complexity levels.

In the first part of the thesis, we have focused on the study of interacting dynamical processes across static contact networks, assuming that the microscopic rules governing the evolution of each dynamics depend on both the macroscopic state of the system and the evolution of other coeval processes. Within this scope, we have tackled four different problems: the spread of interacting pathogens, the efficiency of control policies based on contact tracing strategies, the irruption of social movements and the onset of corrupt societies. Depending on the nature of the problem, we have considered either cooperative or competitive schemes as a function of whether both dynamics are mutually enhanced or hindered as a function of their interaction.

Despite the diversity of the topics addressed, some common conclusions can be drawn on the role of the different interaction schemes. Namely, cooperation leads to a change in the nature of the phase transitions undergone by the system, turning the second-order smooth transitions characteristic from synchronization or epidemic models into first-order explosive ones. These transitions emerge naturally from the coupling between different dynamics and provide a plausible explanation for the sudden emergence of simultaneous important outbreaks across history or the astonishing speed with which social movements evolves from small initial critical masses to involve an important share of the population. In its turn, the introduction of competing dynamics yields a surprising result: the spreading advantage of heterogeneous contact networks, typically reported for individual dynamics, is hindered or even reversed, giving rise to a faster diffusion in homogeneous configurations. This result is a direct consequence of the higher vulnerability of heterogeneous networks to targeted attacks, which in this case is associated with the simultaneous and detrimental presence of the competing dynamic process.

Concerning the specificities of each problem analyzed, we have proven that the competition between two interacting pathogens can lead either to a coexistence regime where both survive endemically or to a dominance regime where the weakest virus becomes extinct. The most clear example of such competition is the existence of mutually exclusive strains of the same disease circulating simultaneously across a given population. The formalism here proposed could constitute the starting point towards a more sophisticated framework incorporating the natural evolution of the virus to capture the complex landscape of genomic variants observed for different diseases such as Influenza, Dengue or COVID-19.

We also explored the microscopic roots of the successful application of contact tracing to fight COVID-19 pandemic in some countries. Specifically, we have reveal that contact tracing is able to anticipate and reduce the epidemic peak by targeting the highly connected nodes, showing a qualitative different behavior from other control policies such as symptomatic detections. Nevertheless, our results do not shed light onto the viability of contact tracing strategies in real scenarios, for we assume a homogeneous penetration of the application in the population, independent of demographic variables such as age or socio-economic level, and neglect the contribution of other control policies. We think that incorporating the latter realistic features into our model is worthy of future research and would increase the robustness of our findings.

From a social point of view, we have shown how social movements can emerge abruptly as a consequence of the positive coupling between the diffusion of ideas and the consensus raised on the population about their veracity. Moreover, the model proposed also reveals that, in absence of global consensus, doubtful information is better propagated in scarcely connected populations. This phenomenon arises as a consequence of the limited access to information, which increases the likelihood that a person observes an uniform opinion among his/her acquaintances. Studying the spread of *fake news* in different real contexts would be helpful to confirm this surprising phenomenon highlighting the relevance of contrasting information sources to avoid misinformation. Moreover, another straightforward extension of the model is to include the time varying nature of social interactions to capture the emergence of polarized states and echo-chambers, which are frequently observed in social platforms and arise from the confrontation between two opposing opinions on a given problem.

Finally, we have also illustrated how epidemic models can be adapted to capture social dynamics such as the propagation of corruption. Remarkably, the punishment mechanism driven by the delation of corrupt individuals by honest peers leads to an emergence of a corrupt equilibrium, whose stability is reinforced when social coercion is introduced in the model. The evident difficulties to gather data on the topic makes difficult introducing further refinements in the model to better capture the evolution of corruption. Nonetheless, we hope this model inspires future research on the application of compartmental models in different fields beyond mathematical epidemiology.

In the second part of the thesis, we have studied the impact of mobility on epidemic spreading. The prominent role of human movements, across urban, national and international scales, maps the evolution of an outbreak onto a reaction-diffusion process where the reaction refers to the contagion and recovery processes shaping the transmission of the pathogen among hosts and diffusion corresponds to their spatial dissemination. Among the different flavors of mobility, we have focused on studying the interplay between daily mobility in cities and community transmission, with the aim of determining the features turning urban environments into contagion centers accelerating the local spread of global pandemics.

We have proposed a new model which accounts for the recurrent nature of commuting mobility and allows one to readily incorporate the mobility network of the residents in a city without any information loss. The most important result is that the effect of tuning mobility is not universal and strongly depends on the specificities concerning pathogen transmission dynamics, the spatial distribution of the population and their mobility patterns. Therefore, the systematic deployment of the same control strategies in different cities can be counterproductive. Along this line, we have discovered a hitherto unreported phenomenon emerging in some scenarios, the so-called epidemic detriment, by which mobility, rather than fostering epidemic spreading, is detrimental to the onset of diseases. This finding has profound implications for policy making purposes since it reveals that the spatial isolation of affected areas cannot be as efficient as thought to mitigate an ongoing outbreak.

The epidemic detriment is just one of the examples that highlight the need to follow systemic approaches to characterize complex phenomena in nature such as the interaction between mobility and epidemics. To explain the roots of the epidemic detriment, we have extended the aforementioned model to separate the contribution of the different features shaping the dynamics such as the recurrent nature of daily mobility or the existence or heterogeneous populations showing different vulnerabilities to disease spreading. From our analysis, we conclude that the impact of mobility on epidemic spreading is the outcome of the trade-off between two antagonistic effects: on the one hand, mobility tends to accelerate the spatiotemporal spread of diseases by mixing different subpopulations and, on the other hand, mobility allows infected individuals to stay away from the most vulnerable areas, thus reducing their potential number of contagions. Therefore, the detrimental effect of human mobility on epidemic spreading emerges in those scenarios where the latter mechanism gains more relevance. Leveraging this phenomenon, we have explored different containment strategies based on the modification of the mobility networks observed in real cities. We obtain that an efficient policy substantially increasing cities' resilience consists in an asymmetric isolation of vulnerable areas forbidding the entrance of individuals from different parts of the cities but allowing residents living there to follow their usual commuting patterns.

We think that our model provides solid theoretical foundations to construct more refined frameworks tailored for specific real diseases. As a proof of concept, we have explored the insights provided by our model into the evolution of real epidemic scenarios. Specifically, we have proven how the average population density and the flows' distribution provide a plausible explanation for the different growth rates observed across United States cities during the early stage of the COVID-19 pandemic. Moreover, we have proposed a further theoretical extension of the model to accommodate different features shaping the evolution of the pandemic, such as the existence of presymptomatic contagions, the age-dependent nature of the clinical and epidemiological features of the disease and the non-pharmaceutical interventions promoted to keep it under control. This model has allowed us to reproduce the evolution of the fatalities in Spain during the first epidemic wave and characterize under which conditions different interventions such as social distancing policies, the creation of social bubbles or population lockdowns are able to mitigate the pandemic. Beyond COVID-19, we have also studied the spread of Dengue across the city of Cali, in Colombia. After adapting the model to capture the contagion mechanisms of vector-borne diseases, we have derived a risk indicator, compiling demographical features of the city and information about the abundance of vectors and the mobility of its residents, which matches the spatial distribution of cases reported from 2015 to 2016 in Cali.

In a nutshell, this thesis constitutes our little contribution to the continuous search for the microscopic origin of the unexpected collective phenomena observed in nature. The lack of isolation of real systems and the intricate relationship between different dynamics running across diverse complexity scales demand the formulation of unifying frameworks which take into account their simultaneous propagation and the possible interdependencies existing among them. We hope that the few examples shown in this thesis will inspire future research on this topic and enable to keep advancing in our knowledge about both the foundations and applications of complex systems.

Conclusiones

Por último, queremos concluir esta tesis recapitulando nuestros hallazgos más relevantes y discutiendo las posibles direcciones de investigación que se abren a partir de los modelos propuestos en esta tesis. El objetivo principal de esta tesis ha sido abandonar el estudio de dinámicas individuales aisladas para dar cuenta de la adaptabilidad de los sistemas complejos y caracterizar los fenómenos resultantes del intercambio de información entre diferentes dinámicas y escalas de complejidad presentes en sistemas reales.

En la primera parte de la tesis, nos hemos centrado en el estudio de procesos dinámicos interactuantes en redes de contacto estáticas, asumiendo que las reglas microscópicas que rigen la evolución de cada dinámica dependen tanto del estado macroscópico del sistema como del progreso del otro proceso coevolutivo. Bajo este enfoque, hemos abordado cuatro problemas diferentes: la propagación acoplada de diferentes patógenos, la eficacia de las políticas de control basadas en estrategias de rastreo de contactos, la irrupción de movimientos sociales y la aparición de la corrupción en la sociedad. Dependiendo de la naturaleza del problema, hemos considerado un esquema de cooperación o competición en función de si los dos procesos dinámicos se potencian o perjudican mutuamente como resultado de su interacción.

A pesar de la diversidad de los temas tratados, se pueden extraer algunas conclusiones comunes sobre el papel de los distintos esquemas de interacción. En particular, la cooperación entre procesos dinámicos da lugar a un cambio en la naturaleza de las transiciones de fase que experimenta el sistema, convirtiendo las transiciones suaves de segundo orden características de los modelos de sincronización o epidemias en transiciones explosivas de primer orden. Estas transiciones surgen de forma natural del acoplo entre diferentes dinámicas y proporcionan una explicación plausible para diferentes fenómenos tales como la aparición repentina de importantes brotes epidémicos simultáneos a lo largo de la historia o la asombrosa velocidad con la que los movimientos sociales evolucionan desde pequeñas masas críticas iniciales hasta involucrar a una parte importante de la población. Asimismo, la introducción de dinámicas competitivas nos ha permitido hallar un resultado sorprendente: la mayor capacidad difusiva de las redes heterogéneas observada tradicionalmente en dinámicas individuales se ve claramente reducida o incluso se revierte, dando lugar a una difusión más rápida en redes homogéneas. Este resultado es una consecuencia directa de la mayor vulnerabilidad de las redes heterogéneas a los ataques dirigidos, que en este caso se corresponden con la presencia simultánea de dinámicas que se perjudican mutuamente.

Restringiéndonos a la fenomenología propia de cada problema estudiado, hemos demostrado que la competencia entre dos patógenos puede conducir a un régimen de coexistencia en el que ambos sobreviven de forma endémica o a un régimen de dominación en el que el virus más débil se extingue. El ejemplo más claro de esta competencia es la existencia de cepas mutuamente excluyentes de la misma enfermedad que circulan simultáneamente en una población determinada. El formalismo aquí propuesto podría constituir el punto de partida hacia un marco más sofisticado que incorpore la evolución natural del virus para capturar el complejo ecosistema de variantes genómicas observado para diferentes enfermedades como la gripe, el dengue o el COVID-19.

Asimismo, hemos explorado las raíces microscópicas de la aplicación exitosa del rastreo de contactos para combatir la pandemia de COVID-19 en algunos países. En concreto, hemos revelado que el rastreo de contactos es capaz de anticipar y reducir el pico epidémico mediante la detección temprana de los nodos altamente conectados de la red. Este efecto es completamente diferente al de otras políticas de control como las detecciones sintomáticas. Sin embargo, nuestros resultados no permiten extraer ninguna conclusión sobre la viabilidad de las estrategias de rastreo de contactos en escenarios reales, ya que asumimos una penetración homogénea de la aplicación en la población, independiente de variables demográficas como la edad o el nivel socioeconómico, y despreciamos la contribución de otras políticas de control. Creemos que la incorporación de estas últimas características realistas a nuestro modelo debe ser investigada en el futuro y aumentaría la robustez de nuestros resultados.

Desde el punto de vista social, hemos demostrado cómo los movimientos sociales pueden surgir de forma abrupta como consecuencia del acoplamiento positivo entre la difusión de ideas y el consenso suscitado en la población sobre su veracidad. Además, el modelo propuesto también revela que, en ausencia de consenso global, la información dudosa se propaga mejor en poblaciones escasamente conectadas. Esto es debido al efecto negativo del acceso a pocas fuentes de información, lo que aumenta la probabilidad de que un individuo asuma de manera incorrecta que una idea esté aceptada en la sociedad. El estudio de la propagación de *fake news* en diferentes redes reales sería útil para confirmar este sorprendente fenómeno que pone de manifiesto la relevancia de contrastar las fuentes de información para evitar la desinformación. Además, otra extensión directa del modelo es incluir la posible variación temporal de las interacciones sociales con el objetivo de reproducir la aparición de estados polarizados y cámaras de eco observados frecuentemente en redes sociales como resultado de la confrontación entre dos posturas opuestas ante el mismo problema.

Por último, hemos ilustrado cómo los modelos epidémicos pueden adaptarse para captar fenómenos inherentemente sociales como la propagación de la corrupción. En este sentido, la introducción de un mecanismo de castigo impulsado por la delación de individuos corruptos por parte de individuos honestos conduce a la aparición de un equilibrio corrupto, cuya estabilidad se refuerza cuando se introduce la coacción social en el modelo. En este caso, la evidente dificultad para recopilar datos reales con los que contrastar los modelos aquí introducidos dificulta su posible refinamiento con el objetivo de capturar la evolución de la corrupción. No obstante, esperamos que el modelo introducido inspire futuras investigaciones sobre cómo los modelos compartimentales pueden ser aplicados en diferentes campos más allá de la epidemiología matemática.

En la segunda parte de la tesis, hemos estudiado el impacto de la movilidad en la propagación de las epidemias. La gran relevancia de los movimientos humanos, a escala local, nacional e internacional, convierte la evolución de un brote en un proceso de reacción-difusión en el que los contagios y recuperaciones determinan los procesos de reacción mientras que la difusión está gobernada por la movilidad. Entre los diferentes tipos de movilidad, nos hemos centrado en el estudio del impacto de la movilidad diaria en núcleos urbanos para la transmisión comunitaria de un virus, con el objetivo de determinar las características que convierten las ciudades en centros de contagio, acelerando la propagación local de las pandemias globales.

Hemos propuesto un nuevo modelo que tiene en cuenta la naturaleza recurrente de la movilidad cotidiana y permite incorporar fácilmente la red de movilidad de los residentes de una ciudad sin pérdida alguna de información. El resultado más importante es que el efecto de la variación de la movilidad no es universal y depende en gran medida de las características relacionadas con la transmisión del agente patógeno, la distribución espacial de la población y sus patrones de movilidad. Por lo tanto, la imposición sistemática de las mismas estrategias de control en diferentes ciudades puede ser contraproducente. En esta línea, hemos descubierto un fenómeno hasta ahora inédito que surge en algunos escenarios, el llamado detrimento epidémico, por el que la movilidad, lejos de favorecer la propagación de enfermedades, es perjudicial para la aparición de brotes epidémicos. Este hallazgo tiene profundas implicaciones para la elaboración de políticas, ya que pone en entredicho la eficacia del aislamiento espacial de las zonas afectadas para combatir una epidemia.

El detrimento epidémico es sólo uno de los ejemplos que ponen de manifiesto la necesidad de proponer enfoques sistémicos para caracterizar la aparición de fenómenos complejos de la naturaleza como la interacción entre movilidad y epidemias. Para explicar las causas del detrimento epidémico, hemos ampliado nuestro modelo y discriminado la contribución de las diferentes rasgos asociados a la dinámica tales como la naturaleza recurrente de la movilidad diaria o la existencia de poblaciones heterogéneas con diferentes vulnerabilidades a la propagación de una enfermedad. A partir de nuestro análisis, concluimos que el impacto de la movilidad en la propagación de la epidemia es el resultado del balance entre dos mecanismos opuestos: por un lado, la movilidad tiende a acelerar la propagación espacio-temporal de las enfermedades al mezclar diferentes subpoblaciones y, por otro lado, la movilidad permite que los individuos infectados se alejen de las zonas más vulnerables, reduciendo así el número potencial de contagios que pueden provocar. Por tanto, el efecto perjudicial de la movilidad para la propagación de epidemias es dominante en aquellos escenarios en los que este último mecanismo adquiere mayor relevancia. Aprovechando este fenómeno, hemos explorado diferentes estrategias de contención basadas en la modificación de las redes de movilidad en ciudades reales. Obtenemos que una política eficiente que aumenta sustancialmente la resiliencia de las ciudades consiste en un aislamiento asimétrico de las zonas vulnerables, prohibiendo la entrada de individuos a dichas zonas pero permitiendo que sus residentes sigan sus patrones habituales de desplazamiento.

Creemos que nuestro modelo proporciona una base sólida teórica para construir marcos más refinados adaptados a enfermedades reales. Como prueba de ello, hemos explorado cómo el modelo propuesto nos permite obtener información acerca de la evolución de escenarios epidémicos reales. En concreto, hemos comprobado cómo la densidad media de la población y la distribución de los flujos proporcionan una explicación plausible de las diferentes velocidades de expansión observadas en las ciudades de Estados Unidos durante la fase inicial de la pandemia de la COVID-19. Además, hemos propuesto una ampliación teórica del modelo para incluir diferentes rasgos característicos de dicha pandemia, como la existencia de contagios presintomáticos, la alta variación de las características clínicas y epidemiológicas de la enfermedad en función de las características socio-demográficas de la población y las intervenciones no farmacéuticas promovidas para mantenerla bajo control. Este modelo nos ha permitido reproducir la evolución de las víctimas mortales en España durante la primera ola epidémica y caracterizar bajo qué condiciones diferentes intervenciones como las políticas de distanciamiento social, la creación de burbujas sociales o los confinamientos poblacionales son capaces de mitigar la pandemia. Además de la COVID-19, hemos estudiado también la propagación del dengue en la ciudad de Cali, en Colombia. Tras adaptar el modelo para capturar los mecanismos de contagio de las enfermedades transmitidas por vectores, hemos derivado un indicador de riesgo, que incluye variables demográficas, entomológicas y asociadas con la movilidad, que permite reproducir la distribución espacial de los casos reportados entre 2015 y 2016 en la ciudad de Cali.

Como conclusión final, esta tesis constituye nuestra pequeña contribución a la búsqueda continua del origen microscópico de los fenómenos complejos observados en la naturaleza. La falta de aislamiento de los sistemas reales y la compleja relación entre diferentes dinámicas involucrando diversas escalas de complejidad exigen la formulación de marcos teóricos que integren los posibles efectos derivados de su propagación simultánea y las interdependencias existentes entre ellas. Esperamos que los pocos ejemplos mostrados en esta tesis inspiren futuras investigaciones a este respecto y permitan seguir avanzando en nuestro conocimiento sobre los sistemas complejos.

Appendices

A Dominance threshold for competing diseases

The analysis of the epidemic curves for the spread of competing diseases reveals the existence of a second threshold λ_c' destabilizing the dominance solution in which one pathogen prevails over the other one in favor of the coexistence solution where both diseases survive in the population. To compute its value, it is convenient to rewrite Eqs. (2.21-2.24) to capture the evolution of the epidemic size ρ and the relative abundance of each specie $\Delta\rho$. In this way, for mutually exclusive diseases ($q = 0$), the evolution of these variables read

$$\rho_i^{t+1} = (1-\mu)\rho_i^t + (1-\rho_i^t)[\Pi^{SS}]_i^t \ , \tag{A.1}$$

$$\Delta\rho_i^{t+1} = (1-\mu)\Delta\rho_i^t + (1-\rho_i^t)[\Pi^{SS}]_i^t \left([f^{IS}]_i^t - [f^{SI}]_i^t\right) \ . \tag{A.2}$$

Eq. (A.1) implies that the prevalence is independent of the microscopic configuration of the population. Therefore, the phase transition between both solutions is completely characterized by the $\vec{\Delta\rho}$ values. Specifically, when $\vec{\Delta\rho} \simeq \vec{\rho}$, the dominance solution is stable and the weak pathogen becomes extinct. In contrast, $\vec{\Delta\rho} = \vec{0}$ characterizes the stabilization of the coexistence solution. To find the dominance threshold, let us first assume that the dynamics has reached the stationary state so that $\rho_i^{t+1} = \rho_i^t = \rho_i^*$ and $\Delta\rho_i^{t+1} = \Delta\rho_i^t = \Delta\rho_i^* \, \forall i$. In this case, Eq. (A.2) turns into:

$$\Delta\rho_i^* = \rho_i^* \left([f^{IS}]_i^* - [f^{SI}]_i^*\right) \ . \tag{A.3}$$

Incorporating Eqs. (2.26-2.27), the previous expression becomes:

$$\Delta\rho_i^* = \rho_i^* \frac{[g^{IS}]_i^* - [g^{SI}]_i^*}{[g^{IS}]_i^* + [g^{SI}]_i^* - [g^{IS}]_i^*[g^{SI}]_i^*} \ , \tag{A.4}$$

where the individual probabilities of getting each pathogen are given by

$$[g^{IS}]_i^* = [g^{IS}]_i\left(\vec{\rho}^*, \vec{\Delta\rho}^*\right) = 1 - \prod_j^N \left(1 - \lambda A_{ij}\frac{\rho_j^* + \Delta\rho_j^*}{2}\right) \tag{A.5}$$

$$\left[g^{SI}\right]_i^* = [g^{IS}]_i\left(\vec{\rho}^*, \vec{\Delta\rho}^*\right) = 1 - \prod_j^N \left(1 - \lambda A_{ij}\frac{\rho_j^* - \Delta\rho_j^*}{2}\right) \ . \tag{A.6}$$

In the coexistence regime, $\vec{\Delta\rho} = \vec{0}$ so that $[g^{IS}]_i = \left[g^{SI}\right]_i^* = g_i^* \, \forall i$. To compute the dominance threshold, we assume that the system in the coexistence regime and introduce a small perturbation in one pathogen such that the perfect coexistence amongthem is broken. Mathematically, this implies that $\vec{\Delta\rho}^* = \vec{\epsilon} \ll 1$ which allows us to linearize Eq. (A.4), obtaining

$$\begin{aligned} \epsilon_i^* &= \sum_{l=1}^{N} \left[\rho_i^* \frac{\lambda A_{il} \left(1 - g_i^*\right)}{\left(1 - \lambda A_{il} \rho_l^* / 2\right) \left(1 - \left(1 - g_i^*\right)^2\right)} \right] \epsilon_l^* \\ &= \sum_{l=1}^{N} \mathcal{M}_{il}(\vec{\rho^*}; \lambda) \epsilon_l^* \ . \end{aligned} \tag{A.7}$$

Eq. (A.7) constitutes an eigenvalue equation with N possible solutions. In particular, the coexistence solution will be unstable if the perturbations introduced to the system are able to grow and stable otherwise. This implies that all the eigenvalues of matrix $\mathcal{M}$, Λ_i, must satisfy $\Lambda_i < 1$. In particular, the destabilization of the coexistence solution will take place when $\Lambda_{\max}(\mathcal{M}) = 1$. Therefore, to find the expression for the second epidemic threshold λ_c', one must iterate the individual SIS dynamics for the dominant pathogen, obtain the stationary state $\vec{\rho}^*$, compute the elements of the matrix $\mathcal{M}$ and check its largest eigenvalue. In particular, the dominance threshold λ_c' satisfies that

$$\lambda_c' := \left\{ \min \lambda \mid \Lambda_{\max}(\mathcal{M}(\vec{\rho^*}; \lambda)) = 1 \right\} \ . \tag{A.8}$$

B Derivation of $\mathcal{R}(t)$ with contact tracing

We start by defining the effective reproduction number $\mathcal{R}(t)$ as the expected number of contagions that one individual becoming exposed at time t produces over her infectious period. Denoting the number of contagions made by agent i after contracting the disease at time $t-1$ by $\mathcal{I}_i(t)$, the effective reproduction number reads

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

To get the expression for $\mathcal{I}_i(t)$ we must contemplate all the possible combinations of the epidemiological states of agent i before entering into the compartment D or R, for afterwards they are not able to produce any further contagion. For this purpose, let us define $\mathcal{P}_i(\tau_E, \tau_P, \tau_{I_A}|t)$ ($\mathcal{P}_i(\tau_E, \tau_P, \tau_{I_S}|t)$) as the joint probability that an asymptomatic (symptomatic) patient remains in the exposed compartment during τ_E time steps, in the presymptomatic state during τ_P steps and in the asymptomatic (symptomatic) infectious compartment for τ_{I_A} (τ_{I_S}) time steps. Incorporating the joint probabilities, it is straightforward to prove that

$$\begin{aligned}\mathcal{I}_i(t) = p \sum_{\tau_E=1}^{\infty}\sum_{\tau_P=1}^{\infty}\sum_{\tau_S=0}^{\infty} \mathcal{P}_i(\tau_E, \tau_P, \tau_S|t) &\left[\sum_{s=t+\tau_E+1}^{t+\tau_E+\tau_P} \sum_{j=1}^{N} A_{ij}\lambda_{I_A}\rho_j^S(s) \right. \\ &\left. + \sum_{s=t+\tau_E+\tau_P+1}^{t+\tau_E+\tau_P+\tau_{I_S}} \sum_{j=1}^{N} A_{ij}\lambda_{I_S}\rho_j^S(s) \right] \\ + (1-p) \sum_{\tau_E=1}^{\infty}\sum_{\tau_P=1}^{\infty}\sum_{\tau_{I_A}=0}^{\infty} \mathcal{P}_i(\tau_E, \tau_P, \tau_{I_A}|t) &\left\{ \sum_{s=t+\tau_E+1}^{t+\tau_E+\tau_P+\tau_{I_A}} \sum_{j=1}^{N} A_{ij}\lambda_{I_A}\rho_j^S(s) \right\} .\end{aligned} \tag{B.2}$$

The derivation of the joint probabilities is immediate when contact tracing is not at work. In this particular case, the probabilities of staying inside each of the possible states become independent of each other and, more importantly, independent of the dynamics. In particular

$$\mathcal{P}_i(\tau_E, \tau_P, \tau_{I_A}|t) = P(\tau_E)P(\tau_P)P(\tau_{I_S}) \ , \tag{B.3}$$

$$\mathcal{P}_i(\tau_E, \tau_P, \tau_{I_A}|t) = P(\tau_E)P(\tau_P)P(\tau_{I_A}) \ , \tag{B.4}$$

with

$$
\begin{aligned}
P(\tau_E) &= (1-\eta)^{\tau_E - 1}\eta \ , \\
P(\tau_P) &= (1-\alpha)^{\tau_P - 1}\alpha \ , \\
P(\tau_{I_A}) &= (1-\mu)^{\tau_{I_A} - 1}\mu \ , \\
P(\tau_{I_S}) &= [(1-\mu)(1-\delta)]^{\tau_{I_S} - 1}\left(1-(1-\delta)\left(1-\mu\right)\right) \ .
\end{aligned} \tag{B.5}
$$

The inclusion of contact tracing complicates the computation of the joint probabilities for two reasons: *i)* the probability of staying in each infectious compartment is dependent on the dynamical state of the system, so we need the epidemic trajectory and *ii)* the probability of entering into an advanced stage of the disease depends on not having been detected earlier because of the contact tracing mechanism. Taking into account both factors, in the case of symptomatic patients, the joint probabilities read:

$$
\begin{aligned}
\mathcal{P}_i(\tau_E, \tau_P, \tau_S | t) = {}& (1-\eta)^{\tau_E - 1}\eta \\
& \times \Bigg\{ \delta_{\tau_S,0} \cdot \Pi_i^{CT}(t+\tau_E+\tau_P) \qquad\qquad\qquad\qquad\qquad\qquad \text{(B.6)} \\
& + (1-\delta_{\tau_S,0}) \cdot \alpha \left(1 - \Pi_i^{CT}(t+\tau_E+\tau_P)\right) \Bigg\} \\
& \times \prod_{s=t+\tau_E+1}^{t+\tau_E+\tau_P-1} (1-\alpha)\left[1 - \Pi_i^{CT}(s)\right] \\
& \times \Bigg\{ \delta_{\tau_S,0} + (1-\delta_{\tau_S,0}) \prod_{s=t+\tau_E+\tau_P+1}^{t+\tau_T-1} (1-\mu)(1-\delta)\left[1 - \Pi_i^{CT}(s)\right] \\
& \times \left\{ \delta + (1-\delta)\left[\Pi_i^{CT}(t+\tau_T) + \left(1 - \Pi_i^{CT}(t+\tau_T)\right)\mu\right] \right\} \Bigg\}
\end{aligned} \tag{B.7}
$$

while, for the asymptomatic individuals, we obtain

$$
\begin{aligned}
\mathcal{P}_i(\tau_E, \tau_P, \tau_{I_A}|t) = (1-\eta)^{\tau_E - 1}\eta & \\
\times \Bigg\{ \delta_{\tau_{I_A},0} \cdot \Pi_i^{CT}(t+\tau_E+\tau_P) & \qquad \text{(B.8)} \\
+ (1-\delta_{\tau_{I_A},0}) \cdot \alpha \left(1 - \Pi_i^{CT}(t+\tau_E+\tau_P)\right) \Bigg\} & \\
\times \left[\prod_{s=t+\tau_E+1}^{t+\tau_E+\tau_P-1} (1-\alpha)\left[1-\Pi_i^{CT}(s)\right] \right] & \\
\times \Bigg\{ \delta_{\tau_{I_A},0} + (1-\delta_{\tau_{I_A},0}) \prod_{s=t+\tau_E+\tau_P+1}^{t+\tau_T-1} (1-\mu)\left[1-\Pi_i^{CT}(s)\right] & \\
\times \left[\Pi_i^{CT}(t+\tau_T) + \left(1-\Pi_i^{CT}(t+\tau_T)\right)\mu\right] \Bigg\} & \qquad \text{(B.9)}
\end{aligned}
$$

In the former expressions, we use τ_T to encapsulate the whole infectious cycle which corresponds with $\tau_T = \tau_E + \tau_P + \tau_{I_S}$ and $\tau_T = \tau_E + \tau_P + \tau_{I_A}$ for symptomatic and asymptomatic individuals respectively. In its turn, we also include $\Pi_i^{CT}(t)$ which is the probability that an agent i is detected at time t via the contact tracing of her detected acquaintances. Therefore

$$
\Pi_i^{CT}(t) = 1 - \prod_{j=1}^{N} \left[1 - f^2 A_{ij} \rho_j^D(t)\right] \ . \qquad \text{(B.10)}
$$

C MIR model with household contacts

One of the limitations of the MIR model is that it neglects the contacts made with the household members. Here we explore how to modify the equations to accommodate the latter interactions and study whether the epidemic detriment persists in this new scenario. Let us consider a SIS disease spreading over a metapopulation with N_P patches. Recall that the course of the disease is completely characterized by the time evolution of the fraction of the population living in each patch in the infected state ρ_i, which is given by:

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

where Π_i accounts for the probability that one individual from patch i contracts the disease. This probability reads

$$\Pi_i(t) = (1-p)P'_{ii}(t) + p\sum_{j=1}^{N_P} R_{ij}P'_{ij}(t) \ . \tag{C.2}$$

$P'_{ij}(t)$ denotes the probability that one individual from patch i and visiting j at time t becomes infected. Taking into account the household contacts, we compute this probability as

$$P'_{ij}(t) = \left[1-\left(1-P_i^H(t)\right)(1-P_j(t))\right] \tag{C.3}$$

where $P_i^H(t)$ corresponds to the probability of getting infected when contacting with the household members and $P_j(t)$ encodes the probability of contracting the disease in the workplace j as dictated by Eq. 4.15. Finally, assuming a mean field approach within each patch i so that the number of contacts inside every household is given by k_i^C, the probability $P_i^H(t)$ is given by:

$$P_i^H(t) = 1-(1-\lambda\rho_i(t))^{k_i^C} \ . \tag{C.4}$$

The expression of the epidemic threshold is slightly modified in this new model. Specifically, following a similar procedure to the one explained in the main text, consisting in the linearization of the equations governing the steady state of the dynamics, we arrive to the following expression:

$$\lambda'_c = \frac{\mu}{\Lambda_{\max}(\mathbf{M}')} \ , \tag{C.5}$$

being the elements of the new critical matrix

$$M'_{ij} = \delta_{ij}\left(k_i^C + n_j(1-p)^2\frac{f_j}{\tilde{n}_j}\right) + n_j p(1-p)\left(\frac{R_{ij}f_j}{\tilde{n}_j} + \frac{R_{ji}f_i}{\tilde{n}_i}\right) + n_j p^2\sum_{l=1}^{N_P}\frac{R_{il}R_{jl}f_l}{\tilde{n}_l} \ . \tag{C.6}$$

Recalling the elements of the original matrix $\mathbf{M}$, we can rewrite the latter expression as

$$M'_{ij} = M_{ij} + \delta_{ij} k_i^C \ . \tag{C.7}$$

The new term does not contain any contribution with p because agents interact with cohabitants, regardless of whether they move or not. As a consequence, the sign of the linear correction in the perturbative analysis remains unchanged, so the epidemic detriment with mobility is preserved.

D COVID-19 model in Spain

We now describe the choice of our parameters to study the epidemic outbreak in Spain, whose values are enumerated in tables D.1 to D.4. Specifically, each table contains the values of the different components conforming our formalism: the features of the underlying multiplex metapopulation (table D.1), the clinical (table D.2) and epidemiological (table D.3) features of the disease and the parameters governing lockdown policies (table D.4).

Regarding epidemiological parameters, the incubation period has been reported to be $\eta^{-1} + \alpha^{-1} = 5.2$ days [315] on average which, in our formalism, must be distributed into the exposed and asymptomatic compartments. In principle, if one does not expect asymptomatic transmissions, most of this time should be spent inside the exposed compartment, thus being the asymptomatic infectious compartment totally irrelevant for disease spreading. However, along the line of some works [354–356], we have found that the unfolding of COVID-19 cannot be explained without accounting for infections from individuals not developing any symptoms previously. In particular, after calibrating the model, we have estimated $\alpha^{-1} = 2.756$ (95% CrI: 2.135 – 3.377) days as asymptomatic infectious period. In turn, the infectious period while being symptomatic is calibrated to $\mu^{-1} = 3.915$ days (95% CrI: 3.470 – 4.360), except for the young stratum, for which we have reduced it to 1 day, assigning the remaining days as asymptomatic; this is due to the reported mild symptoms in young individuals, which may become unnoticed [357]. Furthermore, we assume asymptomatic individuals are half as infectious as symptomatic ones, $\beta_A = 0.5\beta_I$, as in [265, 267].

Regarding the clinical parameters, we fix the fatality rate $\omega = 30\%$ for ICU patients according to official reports [314]. We incorporate from previously published studies the typical time from ICU admission to death as $\psi^{-1} = 7$ days [310] and the stay in ICU for those overcoming the disease as $\chi^{-1} = 21$ days [311]. In turn, the parameters γ^g and θ^g, controlling the infected population requiring ICU beds and dying without occupying any of them respectively, are estimated by correcting the real observed values [314] with the estimated underreporting from [313]. Likewise, the parameters ζ^g and λ^g are computed by assuming that times between onset of symptoms and ICU occupation or fatal outcome are 8 and 11 days [314] on average, respectively.

Table D.1: Parameters of the model related to geographic and population data, including mobility, and their values for Spain. See Appendix D for a detailed explanation.

Symbol	Description	Estimates for $g \in \{Y, M, O\}$ in Spain
n_i^g	Region population	Data provided by INE
s_i	Region surface	Data provided by INE
R_{ij}^g	Mobility matrix	Data provided by INE
$\langle k^g \rangle$	Average total number of contacts	$(11.8, 13.3, 6.8)$
$\langle k_h^g \rangle$	Average number of contacts at home	$(3.1, 3.2, 3.3)$
$\langle k_w^g \rangle$	Average number of contacts at work	$(1.8, 5.2, 0.0)$
C^{gh}	Contacts-by-age matrix	$\begin{pmatrix} 0.5980 & 0.3849 & 0.0171 \\ 0.2440 & 0.7210 & 0.0350 \\ 0.1919 & 0.5705 & 0.2376 \end{pmatrix}$
ξ	Density factor	0.01 km^2
p^g	Mobility factor	$(1.0, 1.0, 1.0)$
σ	Average household size	2.5

Table D.2: Clinical parameters of the model and their estimations with 95% credible intervals for COVID-19 in Spain, with strata separation $g \in \{Y, M, O\}$ when appropriate. See Appendix D for a detailed explanation.

Symbol	Description	States	COVID-19 Spain estimates	References
θ^g	Direct death probability	$I \rightarrow P_D$	$(0.0, 0.008, 0.047)$	[313, 314]
γ^g	ICU probability	$I \rightarrow P_H$	$(0.0003, 0.003, 0.026)$	[313, 314]
ζ^g	Pre-deceased rate	P_D	7.084 days^{-1} 95% CrI: 6.640 – 7.537	[314]
λ^g	Pre-hospitalized in ICU rate	P_H	4.084 days^{-1} 95% CrI: 3.640 – 4.537	[314]
ω^g	Fatality probability in ICU	$P_H \rightarrow H_D$	0.3	[314]
ψ^g	Death rate in ICU	H_D	7 days^{-1}	[358]
χ^g	ICU discharge rate	H_R	21 days^{-1}	[311]

Table D.3: Epidemic parameters of the model and their estimations with 95% credible intervals for COVID-19 in Spain, with strata separation $g \in \{Y, M, O\}$ when appropriate. See Appendix D for a detailed explanation.

Symbol	Description	States	COVID-19 Spain estimates	Assignment
β_I	Infectivity of symptomatic	$S \to E$	0.075 95% CrI: 0.068 – 0.082	Calibrated
β_A	Infectivity of asymptomatic	$S \to E$	$0.5\,\beta_I$	Assumed
η^g	Exposed rate	E	2.444 days^{-1} 95% CrI: 1.823 – 3.064	Calibrated
α^g	Asymptomatic rate	A	$(5.671, 2.756, 2.756)$ days^{-1} 95% CrI^Y: 5.046 – 6.356 95% CrI^M: 2.135 – 3.377 95% CrI^O: 2.135 – 3.377	Calibrated
μ^g	Infectious rate	I	$(1, 3.915, 3.915)$ days^{-1} 95% CrI: 3.470 – 4.360	Y: Assumed M, O: Calibrated

Table D.4: Parameters describing the confinement and their estimations with 95% credible intervals for COVID-19 in Spain, with strata separation $g \in \{Y, M, O\}$ when appropriate. See Appendix D for a detailed explanation.

Symbol	Description	COVID-19 Spain estimates	Assignment	Epidemiological significance
κ_0^g	Mobility reduction	Time-varying	Estimated from INE [316]	Decreases population mobility Decreases contacts Increases household isolation
ϕ	Household permeability	0.174 95% CrI: 0.079 – 0.269	Calibrated	Mixing among households Decreases household isolation
δ	Social distancing	0.207 95% CrI: 0.053 – 0.359	Calibrated	Reduces contacts of the non-confined population

Acknowledgements

"Es de bien nacido ser agradecido." Para no fallar al refrán, me gustaría dedicar unas palabras de agradecimiento por todo el apoyo y el cariño de la gente que me ha rodeado durante los últimos años que han sido fundamentales para dar lugar a esta tesis.

En primer lugar, me siento un afortunado por haber tenido dos directores de tesis como Jesús y Mario. Jesús, gracias por confiar en mí desde el primer momento y por haberme formado tanto durante los últimos años. Me has enseñado muchas, infinitas cosas tanto a nivel profesional y personal y no se me ocurre mejor ejemplo a seguir que tu pasión, trayectoria y esfuerzo para progresar en mi carrera científica con el objetivo de formar el binomio en el futuro del que tanto hemos hablado. Mario, gracias por estar siempre tan disponible para hablar sobre temas científicos (y no científicos) y por cómo te implicas en la investigación; trabajar contigo es un verdadero placer y las conversaciones en tu despacho me han alegrado muchos días aciagos durante estos años. Por último, quiero mencionar al que a todos los efectos ha sido mi tercer director, Álex. He disfrutado muchísimo todos los proyectos juntos y quiero agradecerte la gran acogida que me has dado siempre, desde la maratón que nos dimos hace cuatro años hasta hacerme sentir parte de la familia que tenéis en vuestro grupo de investigación.

Esta tesis no habría sido la misma sin contar con las innumerables discusiones científicas, conversaciones, cervezas, cenas y momentos únicos con mis compañeros en el grupo de investigación GOTHAM, Adriana, Alessio, Francho, Hugo, Pablo y Sergio. Habéis hecho que el día a día de la tesis haya sido mucho más llevadero y espero que sigáis formando parte de mi vida en el futuro. Asimismo, me gustaría agradecer a todos los miembros de nuestro particular COVID-19 task force, por crear un ambiente único de trabajo mezclando incansables horas de trabajo y diversión a partes iguales. En especial, me gustaría reconocer el sudor, sufrimiento, celebraciones y palizas al FIFA que viví con Benja y Joan, con los que durante dos meses hablé más que con mi propia familia (a pesar de estar confinado). No me puedo olvidar tampoco de Lluís, eres un fenómeno y me alegro de que el azar nos pusiera al lado en la sesión de pósters de Sevilla, ya que ahí conocí a un gran compañero de trabajo y también amigo. Por otro lado, me gustaría agradecer al grupo de cafés de becarios del departamento por tantas conversaciones, momentos de descanso y anécdotas que hacían que deseara que llegaran las 10:30 a diario.

Likewise, I would like to say a few words to the international students, Wesley and Mozhgan, who have visited us over the last years. Unfortunately, the COVID-19 pandemic did not allow us to make plans, travel and fully exploit your time here in Zaragoza, but I am really glad to have met both of you. Specially, thanks Wesley for all the enlightening scientific discussions mingled with our usual bad jokes and Mozhgan to show me that the passion for research can overcome all obstacles. Thanks also to the professor Gourab Ghoshal and all the members of the group led by him for making me feel like at home during my research stay at University of Rochester. In particular, I would like to thank Sayat for all the moments we spent together and for making my life much easier there.

Además de en el plano profesional, no me he podido sentir más arropado por mi familia durante los últimos años. En especial, quiero darle las gracias a mi madre, por ser una mujer única, una auténtica campeona, y por enseñarme cómo se puede luchar contra viento y marea para sacar adelante a una familia y, a pesar de las dificultades, mostrarte siempre una sonrisa cuando la necesitas. Gracias también a mi hermana, por tener ese vínculo tan especial de mellizos, por ser mi apoyo diario, escucharme siempre y por tu altruismo con todos.

Por último, quiero darle las gracias a todos mis amigos por todo el apoyo que me han dado y por representar la mejor vía de escape en los momentos más duros de la tesis. Blasco, Caba, Dani, Neila y Pablo, gracias por estar al pie del cañón siempre y por los todos los momentos que hemos vivido y viviremos. Ricardo y Marina, gracias por todas nuestras conversaciones existenciales, las risas tontas que hemos echado y por esa cena en la pizzería que tuvimos; os puedo asegurar que representó un antes y después en la tesis. Ricardo, te mereces un monumento por la paciencia que has tenido conmigo estos meses de escritura y me alegro de haber establecido por costumbre los míticos días de tenis, piscina y croqueta. A Jaime, gracias por ser mi hermano mayor desde que éramos unos canijos y cuidar tan bien de mí siempre. Por último, gracias Loren por todas las innumerables aventuras que hemos pasado juntos y por ser una persona esencial en mi vida.

Seguramente me dejaré a mucha gente en estos agradecimientos. Gracias a todos por vuestra ayuda y por estar siempre ahí.

Bibliography

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

[2] H. H. Heng, Jama **300**, 1580 (2008).

[3] J. Dupré, *The disorder of things: Metaphysical foundations of the disunity of science* (Harvard University Press, 1993).

[4] S. Weinberg, *Dreams of a final theory* (Vintage, 1994).

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

[6] B. D. Malamud, G. Morein, and D. L. Turcotte, Science **281**, 1840 (1998).

[7] Z. Olami, H. J. S. Feder, and K. Christensen, Physical review letters **68**, 1244 (1992).

[8] N. W. Watkins, G. Pruessner, S. C. Chapman, N. B. Crosby, and H. J. Jensen, Space Science Reviews **198**, 3 (2016).

[9] M. Serafino, G. Cimini, A. Maritan, A. Rinaldo, S. Suweis, J. R. Banavar, and G. Caldarelli, Proceedings of the National Academy of Sciences **118**, e2013825118 (2021).

[10] I. Voitalov, P. van der Hoorn, R. van der Hofstad, and D. Krioukov, Physical Review Research **1**, 033034 (2019).

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

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

[13] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, EPJ Data Science **8**, 1 (2019).

[14] S. P. Fraiberger, R. Sinatra, M. Resch, C. Riedl, and A.-L. Barabási, Science **362**, 825 (2018).

[15] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, Science **354**, aaf5239 (2016).

[16] A. Fornito, A. Zalesky, and M. Breakspear, Nature Reviews Neuroscience **16**, 159 (2015).

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

[18] A. S. Mata and S. C. Ferreira, Physical Review E **91**, 012816 (2015).

[19] S. C. Ferreira, R. S. Sander, and R. Pastor-Satorras, Physical Review E **93**, 032314 (2016).

[20] D. H. Silva, S. C. Ferreira, W. Cota, R. Pastor-Satorras, and C. Castellano, Physical Review Research **1**, 033024 (2019).

[21] J. G. Orlandi, J. Soriano, E. Alvarez-Lacalle, S. Teller, and J. Casademunt, Nature Physics **9**, 582 (2013).

[22] E. Bullmore and O. Sporns, Nature Reviews Neuroscience **10**, 186 (2009).

[23] M. P. Menden, D. Wang, M. J. Mason, B. Szalai, K. C. Bulusu, Y. Guan, T. Yu, J. Kang, M. Jeon, R. Wolfinger, et al., Nature Communications **10**, 1 (2019).

[24] V. Domínguez-García, V. Dakos, and S. Kéfi, Proceedings of the National Academy of Sciences **116**, 25714 (2019), ISSN 0027-8424.

[25] S. Pilosof, M. A. Porter, M. Pascual, and S. Kéfi, Nature Ecology & Evolution **1**, 1 (2017).

[26] P. Landi, H. O. Minoarivelo, Å. Brännström, C. Hui, and U. Dieckmann, Population Ecology **60**, 319 (2018).

[27] L. Danon, A. P. Ford, T. House, C. P. Jewell, M. J. Keeling, G. O. Roberts, J. V. Ross, and M. C. Vernon, Interdisciplinary perspectives on infectious diseases **2011** (2011).

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

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

[30] C. Castellano, S. Fortunato, and V. Loreto, Reviews of Modern Physics **81**, 591 (2009).

[31] L. Weng, F. Menczer, and Y.-Y. Ahn, Scientific Reports **3**, 1 (2013).

[32] Y. Moreno, M. Nekovee, and A. F. Pacheco, Physical review E **69**, 066130 (2004).

[33] W. Souma, Y. Fujiwara, and H. Aoyama, Physica A: Statistical Mechanics and its Applications **324**, 396 (2003).

[34] M. Y. Korniyenko, M. Patnam, R. M. del Rio-Chanon, and M. A. Porter, *Evolution of the global financial network and contagion: A new approach* (International Monetary Fund, 2018).

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

[36] G. S. Costa, W. Cota, and S. C. Ferreira, Physical Review Research **2**, 043306 (2020).

[37] A. Aleta and Y. Moreno, BMC Medicine **18**, 157 (2020).

[38] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Physical Review Letters **124**, 048301 (2020).

[39] P. Guerra, W. Meira Jr, C. Cardie, and R. Kleinberg, in *Proceedings of the International AAAI Conference on Web and Social Media* (2013), vol. 7.

[40] E. Estrada, D. J. Higham, and N. Hatano, Physica A: Statistical Mechanics and its Applications **388**, 764 (2009).

[41] K. Guruharsha, J.-F. Rual, B. Zhai, J. Mintseris, P. Vaidya, N. Vaidya, C. Beekman, C. Wong, D. Y. Rhee, O. Cenaj, et al., Cell **147**, 690 (2011).

[42] V. Spirin and L. A. Mirny, Proceedings of the national Academy of sciences **100**, 12123 (2003).

[43] R. Acuna-Soto, C. Viboud, and G. Chowell, PLOS ONE **6**, e23467 (2011).

[44] E. Lofgren, N. H. Fefferman, Y. N. Naumov, J. Gorski, and E. N. Naumova, Journal of virology **81**, 5429 (2007).

[45] A. Bruns, T. Highfield, and J. Burgess, American behavioral scientist **57**, 871 (2013).

[46] I. Perugorría and B. Tejerina, Current Sociology **61**, 424 (2013).

[47] N. Carney, Humanity & Society **40**, 180 (2016).

[48] R. Ross and H. P. Hudson, Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character **93**, 212 (1917).

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

[50] W. H. Hamer, *Epidemic disease in England: the evidence of variability and of persistency of type* (Bedford Press, 1906).

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

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

[53] W. O. Kermack and A. G. McKendrick, Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character **115**, 700 (1927).

[54] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society of London. Series A, containing papers of a mathematical and physical character **138**, 55 (1932).

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

[56] O. Diekmann, J. Heesterbeek, and J. Metz, Journal of Mathematical Biology **28**, 365 (1990).

[57] K. Dietz, Statistical methods in medical research **2**, 23 (1993).

[58] B. F. Nielsen, L. Simonsen, and K. Sneppen, Physical Review Letters **126**, 118301 (2021).

[59] F. Wong and J. J. Collins, Proceedings of the National Academy of Sciences **117**, 29416 (2020).

[60] D. C. Adam, P. Wu, J. Y. Wong, E. H. Lau, T. K. Tsang, S. Cauchemez, G. M. Leung, and B. J. Cowling, Nature Medicine **26**, 1714 (2020).

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

[62] C. Castellano and R. Pastor-Satorras, Physical Review X **7**, 041024 (2017).

[63] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Whang, Physics Reports **424**, 175 (2006).

[64] V. Latora, V. Nicosia, and G. Russo, *Complex Networks* (Cambridge University Press, 2017).

[65] E. Estrada, *The structure of complex networks: theory and applications* (Oxford University Press, 2012).

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

[67] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, 2008).

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

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

[70] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, ACM Transactions on Information and System Security **10**, 1 (2008).

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

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

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

[74] W. H. Organization, *Global tuberculosis report 2013* (World Health Organization, 2013).

[75] N. G. Reich, S. Shrestha, A. A. King, P. Rohani, J. Lessler, S. Kalayanarooj, I.-K. Yoon, R. V. Gibbons, D. S. Burke, and D. A. T. Cummings, Journal of The Royal Society Interface **10**, 20130414 (2013).

[76] R. G. Webster, W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka, Microbiology and molecular biology reviews **56**, 152 (1992).

[77] A. S. Lauring and E. B. Hodcroft, JAMA (2021).

[78] I. Rodriguez-Barraquer, F. Costa, E. J. M. Nascimento, N. Nery, P. M. S. Castanha, G. A. Sacramento, J. Cruz, M. Carvalho, D. D. Olivera, J. E. Hagan, et al., Science **363**, 607 (2019).

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

[80] N. Azimi-Tafreshi, Physical Review E **93**, 042303 (2016).

[81] W. Cai, L. Chen, F. Ghanbarnejad, and P. Grassberger, Nature Physics **11**, 936 (2015).

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

[83] L. Chen, F. Ghanbarnejad, and D. Brockmann, New Journal of Physics **19**, 103041 (2017).

[84] P.-B. Cui, F. Colaiori, and C. Castellano, Physical Review E **99**, 022301 (2019).

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

[86] P.-B. Cui, F. Colaiori, and C. Castellano, Physical Review E **96**, 022301 (2017).

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

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

[89] C. Poletto, S. Meloni, A. V. Metre, V. Colizza, Y. Moreno, and A. Vespignani, Scientific Reports **5**, 1 (2015).

[90] F. Pinotti, É. Fleury, D. Guillemot, P.-Y. Böelle, and C. Poletto, PLOS Computational Biology **15**, e1006530 (2019).

[91] J. Gómez-Gardeñes and Y. Moreno, Physical Review E **73**, 056124 (2006).

[92] S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, et al., Nature **584**, 257 (2020).

[93] V. J. Lee, C. J. Chiew, and W. X. Khong, Journal of travel medicine **27**, taaa039 (2020).

[94] E. Han, M. M. J. Tan, E. Turk, D. Sridhar, G. M. Leung, K. Shibuya, N. Asgari, J. Oh, A. L. García-Basteiro, J. Hanefeld, et al., The Lancet (2020).

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

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

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

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

[99] Center for disease control, *Covid-19 pandemic planning scenarios* (2020).

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

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

[102] J. Wallinga, American Journal of Epidemiology **160**, 509 (2004).

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

[104] G. Cencetti, G. Santin, A. Longa, E. Pigani, A. Barrat, C. Cattuto, S. Lehmann, M. Salathe, and B. Lepri, Nature Communications **12**, 1 (2021).

[105] M. Zald and J. McCarthy, *The dynamics of social movements: Resource mobilization, social control, and tactics* (Winthrop Publishers Cambridge, Mass., 1979).

[106] D. P. Maki and M. Thompson, *Mathematical models and applications: with emphasis on the social life, and management sciences* (Prentice Hall, 1973).

[107] D. J. Daley and D. G. Kendall, Nature **204**, 1118 (1964).

[108] D. Centola and M. Macy, American Journal of Sociology **113**, 702 (2007).

[109] M. L. Markus, Communication research **14**, 491 (1987).

[110] J. Coleman, E. Katz, and H. Menzel, Sociometry **20**, 253 (1957).

[111] D. McAdam and R. Paulsen, American journal of sociology **99**, 640 (1993).

[112] K.-D. Opp and C. Gern, American sociological review **58**, 659 (1993).

[113] M. Granovetter, American Journal of Sociology **83**, 1420 (1978).

[114] D. J. Watts, Proceedings of the National Academy of Sciences **99**, 5766 (2002).

[115] W. Wang, M. Tang, P. Shu, and Z. Wang, New Journal of Physics **18**, 013029 (2016), ISSN 1367-2630, URL `http://dx.doi.org/10.1088/1367-2630/18/1/013029`.

[116] J. Gómez-Gardeñes, L. Lotero, S. N. Taraskin, and F. J. Pérez-Reche, Scientific Reports **6**, 19767 (2016).

[117] I. Iacopini, G. Petri, A. Barrat, and V. Latora, Nature Communications **10**, 1 (2019).

[118] S. Galam, Y. Gefen, and Y. Shapir, Journal of Mathematical Sociology **9**, 1 (1982).

[119] P. Clifford and A. Sudbury, Biometrika **60**, 581 (1973).

[120] S. Galam, The European Physical Journal B-Condensed Matter and Complex Systems **25**, 403 (2002).

[121] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Advances in Complex Systems **3**, 87 (2000).

[122] R. Hegselmann, U. Krause, et al., Journal of artificial societies and social simulation **5**, 1 (2002).

[123] A. Pluchino, V. Latora, and A. Rapisarda, The European Physical Journal B-Condensed Matter and Complex Systems **50**, 169 (2006).

[124] A. Pikovsky, J. Kurths, M. Rosenblum, and J. Kurths, *Synchronization: a universal concept in nonlinear sciences*, 12 (Cambridge university press, 2003).

[125] Z. Neda, E. Ravasz, Y. Brechet, T. Vicsek, and A.-L. Barabási, Nature (London) **403**, 849 (2000).

[126] E. Codrons, N. F. Bernardi, M. Vandoni, and L. Bernardi, PLOS ONE **9**, e107538 (2014).

[127] Y. Kuramoto, in *International symposium on mathematical problems in theoretical physics* (Springer, 1975), pp. 420–422.

[128] A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou, Physics Reports **469**, 93 (2008).

[129] F. A. Rodrigues, T. K. D. Peron, P. Ji, and J. Kurths, Physics Reports **610**, 1 (2016).

[130] J. Gómez-Gardeñes, Y. Moreno, and A. Arenas, Physical Review Letters **98**, 034101 (2007).

[131] J. Gómez-Gardeñes, S. Gómez, A. Arenas, and Y. Moreno, Physical Review Letters **106**, 128701 (2011).

[132] S. Boccaletti, J. Almendral, S. Guan, I. Leyva, Z. Liu, I. Sendiña-Nadal, Z. Wang, and Y. Zou, Physics Reports **660**, 1 (2016).

[133] R. M. D. Souza, J. Gómez-Gardeñes, J. Nagler, and A. Arenas, Advances in Physics **68**, 123 (2019).

[134] R. Axelrod, Journal of Conflict Resolution **41**, 203 (1997).

[135] F. Battiston, V. Nicosia, V. Latora, and M. S. Miguel, Scientific Reports **7**, 1 (2017).

[136] M. M. Danziger, I. Bonamassa, S. Boccaletti, and S. Havlin, Nature Physics **15**, 178 (2018).

[137] V. Nicosia, P. S. Skardal, A. Arenas, and V. Latora, Physical Review Letters **118**, 138302 (2017).

[138] S. Vosoughi, D. Roy, and S. Aral, Science **359**, 1146 (2018).

[139] A. Puglisi, A. Baronchelli, and V. Loreto, Proceedings of the National Academy of Sciences **105**, 7936 (2008).

[140] A. Baronchelli, M. Felici, V. Loreto, E. Caglioti, and L. Steels, Journal of Statistical Mechanics: Theory and Experiment **2006**, P06014 (2006).

[141] R. Amato, L. Lacasa, A. Díaz-Guilera, and A. Baronchelli, Proceedings of the National Academy of Sciences **115**, 8260 (2018).

[142] I. Iacopini, S. Milojević, and V. Latora, Physical review letters **120**, 048301 (2018).

[143] I. Iacopini, G. Di Bona, E. Ubaldi, V. Loreto, and V. Latora, Physical Review Letters **125**, 248301 (2020).

[144] M. Bonaventura, V. Ciotti, P. Panzarasa, S. Liverani, L. Lacasa, and V. Latora, Scientific Reports **10**, 1 (2020).

[145] J. Poncela-Casasnovas, M. Gutiérrez-Roig, C. Gracia-Lázaro, J. Vicens, J. Gómez-Gardeñes, J. Perelló, Y. Moreno, J. Duch, and A. Sánchez, Science advances **2**, e1600451 (2016).

[146] A. Aleta and Y. Moreno, EPJ Data Science **8**, 22 (2019).

[147] C. P. Roca, J. A. Cuesta, and A. Sánchez, Physical Review E **80**, 046106 (2009).

[148] M. Perc, J. Gómez-Gardenes, A. Szolnoki, L. M. Floría, and Y. Moreno, Journal of the royal society interface **10**, 20120997 (2013).

[149] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, 1998).

[150] J. W. Weibull, *Evolutionary game theory* (MIT press, 1997).

[151] H. Gintis, *Game theory evolving* (Princeton university press, 2009).

[152] P. Hammerstein and R. Selten, Handbook of game theory with economic applications **2**, 929 (1994).

[153] B. Drossel, Advances in physics **50**, 209 (2001).

[154] D. Friedman, Econometrica: Journal of the Econometric Society pp. 637–666 (1991).

[155] O. Cadot, Journal of public economics **33**, 223 (1987).

[156] K. Basu, S. Bhattacharya, and A. Mishra, Journal of public Economics **48**, 349 (1992).

[157] J.-H. Lee, Y. Iwasa, U. Dieckmann, and K. Sigmund, Proceedings of the National Academy of Sciences **116**, 13276 (2019).

[158] V. N. Kolokoltsov and O. A. Malafeyev, Dynamic Games and Applications **7**, 34 (2015).

[159] V. N. Kolokoltsov, International Journal of Statistics and Probability **1**, 77 (2012).

[160] J.-H. Lee, K. Sigmund, U. Dieckmann, and Y. Iwasa, Journal of Theoretical Biology **367**, 1 (2015).

[161] J.-H. Lee, M. Jusup, and Y. Iwasa, Journal of Theoretical Biology **428**, 76 (2017).

[162] P. Verma and S. Sengupta, PLOS ONE **10**, e0133441 (2015).

[163] P. Verma, A. K. Nandi, and S. Sengupta, Scientific Reports **7**, 42735 (2017).

[164] P. Verma, A. K. Nandi, and S. Sengupta, Journal of Theoretical Biology **450**, 43 (2018).

[165] A. Cardillo, J. Gómez-Gardeñes, D. Vilone, and A. Sánchez, New Journal of Physics **12**, 103034 (2010).

[166] C. Granell, S. Gómez, and A. Arenas, Physical review letters **111**, 128701 (2013).

[167] C. Granell, S. Gómez, and A. Arenas, Physical Review E **90**, 012808 (2014).

[168] B. Steinegger, A. Cardillo, P. D. L. Rios, J. Gómez-Gardeñes, and A. Arenas, Physical Review E **97**, 032308 (2018).

[169] B. Steinegger, A. Arenas, J. Gómez-Gardeñes, and C. Granell, Physical Review Research **2**, 023181 (2020).

[170] Z. Wang, Q. Guo, S. Sun, and C. Xia, Applied Mathematics and Computation **349**, 134 (2019).

[171] A. Z. Kaminski, Journal of Theoretical Politics **1**, 77 (1989).

[172] L. P. Markowitz, Comparative Politics **50**, 103 (2017).

[173] A. L. Osipian, European Education **40**, 27 (2008).

[174] D. Achlioptas, R. M. D. Souza, and J. Spencer, Science **323**, 1453 (2009).

[175] N. Perra, B. Gonçalves, R. Pastor-Satorras, and A. Vespignani, Scientific Reports **2**, 1 (2012).

[176] O. J. Benedictow and O. L. Benedictow, *The Black Death, 1346-1353: the complete history* (Boydell & Brewer, 2004).

[177] *The World of Air Transport in 2018* (2018).

[178] V. Colizza, A. Barrat, M. Barthélemy, and A. Vespignani, BMC Medicine **5**, 34 (2007).

[179] M. Tizzoni, P. Bajardi, C. Poletto, J. Ramasco, D. Balcan, B. Gonçalves, N. Perra, V. Colizza, and A. Vespignani, BMC Medicine **10**, 165 (2012).

[180] M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, O. C.-. D. W. Group†, L. du Plessis, N. R. Faria, R. Li, et al., Science **368**, 493 (2020).

[181] M. Gilbert, G. Pullano, F. Pinotti, E. Valdano, C. Poletto, P.-Y. Boalle, E. D. Ortenzio, Y. Yazdanpanah, S. P. Eholie, M. Altmann, et al., The Lancet **395**, 871 (2020).

[182] L. Kramer, *The normal heart and the destiny of me: two plays* (Grove Press, 2000).

[183] R. Guimera, S. Mossa, A. Turtschi, and L. A. N. Amaral, Proceedings of the National Academy of Sciences **102**, 7794 (2005).

[184] E. Estrada and J. Gómez-Gardeñes, Physica D: Nonlinear Phenomena **323**, 57 (2016).

[185] R. Levins, Bulletin of the Entomological Society of America **15**, 237 (1969).

[186] P. Opdam, Landscape ecology **5**, 93 (1991).

[187] J. V. Wells and M. E. Richmond, Wildlife Society Bulletin pp. 458–462 (1995).

[188] A. Terui, N. Ishiyama, H. Urabe, S. Ono, J. C. Finlay, and F. Nakamura, Proceedings of the National Academy of Sciences **115**, E5963 (2018).

[189] L. Laikre, F. Olsson, E. Jansson, O. Hössjer, and N. Ryman, Heredity **117**, 279 (2016).

[190] C. Thomas, Conservation biology **8**, 373 (1994).

[191] I. González-García, R. V. Solé, and J. Costa, Proceedings of the National Academy of Sciences **99**, 13085 (2002).

[192] D. Balcan, B. Gonçalves, H. Hu, J. J. Ramasco, V. Colizza, and A. Vespignani, Journal of Computational Science **1**, 132 (2010).

[193] Q. Zhang, K. Sun, M. Chinazzi, A. P. y Piontti, N. E. Dean, D. P. Rojas, S. Merler, D. Mistry, P. Poletti, L. Rossi, et al., Proceedings of the National Academy of Sciences **114**, E4334 (2017).

[194] D. J. Watts, R. Muhamad, D. C. Medina, and P. S. Dodds, Proceedings of the National Academy of Sciences **102**, 11157 (2005).

[195] S. C. Kramer, S. Pei, and J. Shaman, PLOS Computational Biology **16**, e1008233 (2020).

[196] L. Wang and X. Li, Chinese Science Bulletin **59**, 3511 (2014).

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

[198] G. K. Zipf, American sociological review **11**, 677 (1946).

[199] X. Li, H. Tian, D. Lai, and Z. Zhang, International journal of environmental research and public health **8**, 3134 (2011).

[200] P. McCullagh and J. A. Nelder, *Generalized linear models* (Routledge, 2019).

[201] S. A. Stouffer, American sociological review **5**, 845 (1940).

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

[203] V. Colizza, R. Pastor-Satorras, and A. Vespignani, Nature Physics **3**, 276 (2007).

[204] V. Colizza and A. Vespignani, Journal of Theoretical Biology **251**, 450 (2008).

[205] J. D. Noh and H. Rieger, Physical review letters **92**, 118701 (2004).

[206] D. Balcan, V. Colizza, B. Goncalves, H. Hu, J. J. Ramasco, and A. Vespignani, Proceedings of the National Academy of Sciences **106**, 21484 (2009).

[207] M. C. González, C. A. Hidalgo, and A.-L. Barabási, Nature **453**, 779 (2008).

[208] L. Alessandretti, P. Sapiezynski, V. Sekara, S. Lehmann, and A. Baronchelli, Nature human behaviour **2**, 485 (2018).

[209] L. Alessandretti, P. Sapiezynski, S. Lehmann, and A. Baronchelli, PLOS ONE **12**, e0171686 (2017).

[210] A. Wesolowski, N. Eagle, A. J. Tatem, D. L. Smith, A. M. Noor, R. W. Snow, and C. O. Buckee, Science **338**, 267 (2012).

[211] L. Lotero, R. G. Hurtado, L. M. Floría, and J. Gómez-Gardeñes, Royal Society Open Science **3**, 150654 (2016).

[212] M. Macedo, L. Lotero, A. Cardillo, H. Barbosa, and R. Menezes, in *Complex Networks XI* (Springer, 2020), pp. 269–281.

[213] D. Balcan and A. Vespignani, Nature Physics **7**, 581 (2011).

[214] D. Balcan and A. Vespignani, Journal of Theoretical Biology **293**, 87 (2012).

[215] V. Belik, T. Geisel, and D. Brockmann, The European Physical Journal B **84**, 579 (2011).

[216] V. Belik, T. Geisel, and D. Brockmann, Physical Review X **1**, 011001 (2011).

[217] H. Coşkun, N. Yıldırım, and S. Gündüz, Science of the Total Environment **751**, 141663 (2021).

[218] D. W. Wong and Y. Li, PLOS ONE **15**, e0242398 (2020).

[219] N. Kadi and M. Khelfaoui, Bulletin of the National Research Centre **44**, 1 (2020).

[220] Eurostat, *Passenger mobility statistics* (2021).

[221] K. Prem, A. R. Cook, and M. Jit, PLOS Computational Biology **13**, e1005697 (2017).

[222] D. Mistry, M. Litvinova, A. P. y Piontti, M. Chinazzi, L. Fumanelli, M. F. Gomes, S. A. Haque, Q.-H. Liu, K. Mu, X. Xiong, et al., Nature Communications **12**, 1 (2021).

[223] Alcaldía de Cali, *Cali en cifras, 2013*.

[224] A. Wesolowski, C. O. Buckee, D. K. Pindolia, N. Eagle, D. L. Smith, A. J. Garcia, and A. J. Tatem, PLOS ONE **8**, e52971 (2013).

[225] United States Census. Bureau.

[226] T. Louail, M. Lenormand, O. G. Cantu Ros, M. Picornell, R. Herranz, E. Frias-Martinez, J. J. Ramasco, and M. Barthelemy, Scientific Reports **4**, 5276 (2014).

[227] S. Phithakkitnukoon, Z. Smoreda, and P. Olivier, PLOS ONE **7**, e39253 (2012).

[228] M. Tizzoni, P. Bajardi, A. Decuyper, G. Kon Kam King, C. M. Schneider, V. Blondel, Z. Smoreda, M. C. González, and V. Colizza, PLOS Computational Biology **10**, 1 (2014).

[229] B. Hawelka, I. Sitko, E. Beinat, S. Sobolevsky, P. Kazakopoulos, and C. Ratti, Cartography and Geographic Information Science **41**, 260 (2014).

[230] S. Jiang, J. Ferreira, and M. C. González, IEEE Transactions on Big Data **3**, 208 (2017).

[231] A. Bassolas, H. Barbosa-Filho, B. Dickinson, X. Dotiwalla, P. Eastham, R. Gallotti, G. Ghoshal, B. Gipson, S. A. Hazarie, H. Kautz, et al., Nature Communications **10**, 4817 (2019), ISSN 2041-1723.

[232] R. A. Marcus, The Journal of Physical Chemistry A **105**, 2612 (2001).

[233] F. Ahmed, N. Ahmed, C. Pissarides, and J. Stiglitz, The Lancet Public Health **5**, e240 (2020).

[234] M. J. Townsend, T. K. Kyle, and F. C. Stanford, *Outcomes of covid-19: disparities in obesity and by ethnicity/race* (2020).

[235] J. T. Chen and N. Krieger, Journal of Public Health Management and Practice **27**, S43 (2021).

[236] S. Shen, W. Koech, J. Feng, T. M. Rice, and M. Zhu, BMJ open **7**, e015780 (2017).

[237] L. Gauvin, M. Tizzoni, S. Piaggesi, A. Young, N. Adler, S. Verhulst, L. Ferres, and C. Cattuto, Humanities and Social Sciences Communications **7**, 1 (2020).

[238] M. Adeel, A. G. Yeh, and F. Zhang, Transportation **44**, 1519 (2017).

[239] Y. Xu, A. Belyi, I. Bojic, and C. Ratti, Computers, Environment and Urban Systems **72**, 51 (2018).

[240] H. Barbosa, S. Hazarie, B. Dickinson, A. Bassolas, A. Frank, H. Kautz, A. Sadilek, J. J. Ramasco, and G. Ghoshal, Scientific Reports **11**, 1 (2021).

[241] M. D. Domenico, A. Solé-Ribalta, E. Cozzo, M. Kivelä, Y. Moreno, M. A. Porter, S. Gómez, and A. Arenas, Physical Review X **3** (2013).

[242] M. Kivela, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter, Journal of Complex Networks **2**, 203 (2014).

[243] L. Lotero, A. Cardillo, R. Hurtado, and J. Gómez-Gardeñes, in *Understanding Complex Systems* (Springer International Publishing, 2016), pp. 149–164.

[244] M. McPherson, L. Smith-Lovin, and J. M. Cook, Annual review of sociology **27**, 415 (2001).

[245] S. Gómez, A. Díaz-Guilera, J. Gómez-Gardeñes, C. J. Pérez-Vicente, Y. Moreno, and A. Arenas, Physical Review Letters **110** (2013).

[246] A. B. Migliano, F. Battiston, S. Viguier, A. E. Page, M. Dyble, R. Schlaepfer, D. Smith, L. Astete, M. Ngales, J. Gomez-Gardenes, et al., Science Advances **6**, eaax5913 (2020).

[247] A. B. Migliano, A. E. Page, J. Gómez-Gardeñes, G. D. Salali, S. Viguier, M. Dyble, J. Thompson, N. Chaudhary, D. Smith, J. Strods, et al., Nature Human Behaviour **1**, 1 (2017).

[248] N. Varghese, *Globalization of higher education and cross-border student mobility* (Citeseer, 2008).

[249] H. W. Eves, *Return to Mathematical Circles: A Fifth Collection of Mathematical Stories and Anecdotes*, vol. 5 (Brooks/Cole, 1988).

[250] M. M. Norton, N. Tompkins, B. Blanc, M. C. Cambria, J. Held, and S. Fraden, Physical review letters **123**, 148301 (2019).

[251] C. Poletto, M. Tizzoni, and V. Colizza, Scientific Reports **2**, 1 (2012).

[252] C. Poletto, M. Tizzoni, and V. Colizza, Journal of Theoretical Biology **338**, 41 (2013).

[253] J. T. Matamalas, M. D. Domenico, and A. Arenas, Journal of The Royal Society Interface **13**, 20160203 (2016).

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

[255] J. A. P. Heesterbeek, Acta biotheoretica **50**, 189 (2002).

[256] A. Apolloni, C. Poletto, J. J. Ramasco, P. Jensen, and V. Colizza, Theoretical Biology and Medical Modelling **11**, 3 (2014).

[257] P. Bosetti, P. Poletti, M. Stella, B. Lepri, S. Merler, and M. De Domenico, Proceedings of the National Academy of Sciences **117**, 30118 (2020).

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

[259] C. Castellano and R. Pastor-Satorras, Physical Review Letters **105** (2010).

[260] E. Valdano, M. R. Fiorentin, C. Poletto, and V. Colizza, Physical Review Letters **120** (2018).

[261] N. Masuda and P. Holme, eds., *Temporal Network Epidemiology* (Springer Singapore, 2017).

[262] P. Holme and J. Saramäki, eds., *Temporal Network Theory* (Springer International Publishing, 2019).

[263] A. Koher, H. H. Lentz, J. P. Gleeson, and P. Höevel, Physical Review X **9** (2019).

[264] E. Valdano, L. Ferreri, C. Poletto, and V. Colizza, Physical Review X **5**, 021005 (2015).

[265] L. Di Domenico, G. Pullano, C. E. Sabbatini, P.-Y. Boëlle, and V. Colizza, BMC Medicine **18**, 240 (2020).

[266] L. Pellis, S. Cauchemez, N. M. Ferguson, and C. Fraser, Nature Communications **11**, 1 (2020).

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

[268] N. Hoertel, M. Blachier, C. Blanco, M. Olfson, M. Massetti, M. S. Rico, F. Limosin, and H. Leleu, Nature medicine **26**, 1417 (2020).

[269] E. Pepe, P. Bajardi, L. Gauvin, F. Privitera, B. Lake, C. Cattuto, and M. Tizzoni, Scientific data **7**, 1 (2020).

[270] S. Arregui, M. J. Iglesias, S. Samper, D. Marinova, C. Martin, J. Sanz, and Y. Moreno, Proceedings of the National Academy of Sciences **115**, E3238 (2018).

[271] G. Giordano, M. Colaneri, A. Di Filippo, F. Blanchini, P. Bolzern, G. De Nicolao, P. Sacchi, P. Colaneri, and R. Bruno, Nature Medicine pp. 1–6 (2021).

[272] *United Nations, Department of Economic and Social Affairs, Population Division (2019). World Urbanization Prospects 2018: Highlights (ST/ESA/SER.A/421)*, accessed: 2019-01-30.

[273] E. Alirol, L. Getaz, B. Stoll, F. Chappuis, and L. Loutan, The Lancet. Infectious diseases **11**, 131 (2011).

[274] B. Rader, S. V. Scarpino, A. Nande, A. L. Hill, B. Adlam, R. C. Reiner, D. M. Pigott, B. Gutierrez, A. E. Zarebski, M. Shrestha, et al., Nature medicine **26**, 1829 (2020).

[275] W. Pan, G. Ghoshal, C. Krumme, M. Cebrian, and A. Pentland, Nature Communications **4**, 1961 (2013).

[276] M. Batty, *The new science of cities* (MIT press, 2013).

[277] M. Barthelemy, *The structure and dynamics of cities* (Cambridge University Press, 2016).

[278] M. Barthelemy, Nature Reviews Physics **1**, 406 (2019).

[279] M. Kardar, *Statistical physics of fields* (Cambridge University Press, 2007).

[280] *s2geometry* (accessed 2020- 08-17), URL `https://github.com/google/s2geometry`.

[281] *Us census bureau* (accessed 2020- 08-27), URL `https://www.census.gov`.

[282] *Facebook* (accessed 2020- 08-27), URL `https://data.humdata.org/organization/facebook`.

[283] *Statistics South Africa*, `http://www.statssa.gov.za`.

[284] *Australian Bureau of Statistics*, `https://www.abs.gov.au/`.

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

[286] B. Rader, S. V. Scarpino, A. Nande, A. L. Hill, B. Adlam, R. C. Reiner, D. M. Pigott, B. Gutierrez, A. E. Zarebski, M. Shrestha, et al., Nature Medicine (2020).

[287] M. Gatto, E. Bertuzzo, L. Mari, S. Miccoli, L. Carraro, R. Casagrandi, and A. Rinaldo, Proceedings of the National Academy of Sciences **117**, 10484 (2020).

[288] *New york times*, `https://github.com/nytimes/covid-19-data`.

[289] *Usa facts*, `https://usafacts.org/visualizations/coronavirus-covid-19-spread-map`.

[290] K. M. Gostic, L. McGough, E. B. Baskerville, S. Abbott, K. Joshi, C. Tedijanto, R. Kahn, R. Niehus, J. A. Hay, P. M. De Salazar, et al., PLOS Computational Biology **16**, e1008409 (2020).

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

[292] A. Savitzky and M. J. Golay, Analytical chemistry **36**, 1627 (1964).

[293] *Smoothing in Python* (accessed 2020- 08-17), URL `https://plotly.com/python/smoothing/`.

[294] M. Chinazzi, J. T. Davis, M. Ajelli, C. Gioannini, M. Litvinova, S. Merler, A. P. y Piontti, K. Mu, L. Rossi, K. Sun, et al., Science **368**, 395 (2020).

[295] J. T. Wu, K. Leung, and G. M. Leung, The Lancet **395**, 689 (2020).

[296] G. Pullano, F. Pinotti, E. Valdano, P.-Y. Boëlle, C. Poletto, and V. Colizza, Eurosurveillance **25** (2020).

[297] J. Zhang, M. Litvinova, W. Wang, Y. Wang, X. Deng, X. Chen, M. Li, W. Zheng, L. Yi, X. Chen, et al., The Lancet Infectious Diseases **20**, 793 (2020).

[298] G. N. Wong, Z. J. Weiner, A. V. Tkachenko, A. Elbanna, S. Maslov, and N. Goldenfeld, Physical Review X **10**, 041033 (2020).

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

[300] N. G. Davies, A. J. Kucharski, R. M. Eggo, A. Gimma, W. J. Edmunds, T. Jombart, K. O'Reilly, A. Endo, J. Hellewell, E. S. Nightingale, et al., The Lancet Public Health **5**, e375 (2020).

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

[302] H. Tian, Y. Liu, Y. Li, C.-H. Wu, B. Chen, M. U. Kraemer, B. Li, J. Cai, B. Xu, Q. Yang, et al., Science **368**, 638 (2020).

[303] C. Bonanad, S. Garcia-Blas, F. Tarazona-Santabalbina, J. Sanchis, V. Bertomeu-González, L. Facila, A. Ariza, J. Nunez, and A. Cordero, Journal of the American Medical Directors Association **21**, 915 (2020).

[304] M. Monod, A. Blenkinsop, X. Xi, D. Hebert, S. Bershan, S. Tietze, M. Baguelin, V. C. Bradley, Y. Chen, H. Coupland, et al., Science **371**, eabe8372 (2021).

[305] D. Pan, S. Sze, J. S. Minhas, M. N. Bangash, N. Pareek, P. Divall, C. M. Williams, M. R. Oggioni, I. B. Squire, L. B. Nellums, et al., EClinicalMedicine **23**, 100404 (2020).

[306] K. Mackey, C. K. Ayers, K. K. Kondo, S. Saha, S. M. Advani, S. Young, H. Spencer, M. Rusek, J. Anderson, S. Veazie, et al., Annals of internal medicine **174**, 362 (2021).

[307] H. Hu, K. Nigmatulina, and P. Eckhoff, Mathematical Biosciences **244**, 125 (2013).

[308] K. Leung, J. T. Wu, and G. M. Leung, Nature Communications **12**, 1 (2021).

[309] *Instituto Nacional de Estadistica. Spain*, `https://www.ine.es`, accessed: 2020-02-26.

[310] N. Wilson, A. Kvalsvig, L. T. Barnard, and M. G. Baker, Emerging Infectious Diseases **26** (2020).

[311] P. K. Bhatraju, B. J. Ghassemieh, M. Nichols, R. Kim, K. R. Jerome, A. K. Nalla, A. L. Greninger, S. Pipavath, M. M. Wurfel, L. Evans, et al., New England Journal of Medicine **382**, 2012 (2020).

[312] X. Yang, Y. Yu, J. Xu, H. Shu, J. Xia, H. Liu, Y. Wu, L. Zhang, Z. Yu, M. Fang, et al., The Lancet Respiratory Medicine (2020), URL `https://doi.org/10.1016/S2213-2600(20)30079-5`.

[313] M. Pollán, B. Pérez-Gómez, R. Pastor-Barriuso, J. Oteo, M. A. Hernán, M. Pérez-Olmeda, J. L. Sanmartín, A. Fernández-García, I. Cruz, N. F. de Larrea, et al., The Lancet **396**, 535 (2020).

[314] Red Nacional de Vigilancia Epidemiológica, Tech. Rep., Instituto de Salud Carlos III (2020).

[315] Q. Li, X. Guan, P. Wu, X. Wang, L. Zhou, Y. Tong, R. Ren, K. S. Leung, E. H. Lau, J. Y. Wong, et al., New England Journal of Medicine **382**, 1199 (2020).

[316] *Instituto Nacional de Estadistica. Mobility reduction.*, `https://www.ine.es/en/covid/covid_movilidad_en.htm`, accessed: 2020-06-17.

[317] T. Toni, D. Welch, N. Strelkowa, A. Ipsen, and M. P. Stumpf, Journal of The Royal Society Interface **6**, 187 (2008).

[318] M. Mazzoli, D. Mateo, A. Hernando, S. Meloni, and J. J. Ramasco, MedRxiv (2020).

[319] L. Lacasa, R. Challen, E. Brooks-Pollock, and L. Danon, PLOS ONE **15**, e0241027 (2020).

[320] D. H. Morris, F. W. Rossine, J. B. Plotkin, and S. A. Levin, Communications Physics **4**, 1 (2021).

[321] A. Boutayeb, Handbook of disease burdens and quality of life measures p. 531 (2010).

[322] M. Tovar, S. Arregui, D. Marinova, C. Martín, J. Sanz, and Y. Moreno, Nature Communications **10**, 1 (2019).

[323] H.-F. Huo, R. Chen, and X.-Y. Wang, Applied Mathematical Modelling **40**, 6550 (2016).

[324] I. Cremin, R. Alsallaq, M. Dybul, P. Piot, G. Garnett, and T. B. Hallett, Aids **27**, 447 (2013).

[325] J. M. Reinhold, C. R. Lazzari, and C. Lahondère, Insects **9**, 158 (2018).

[326] S. Sherpa, M. G. Blum, and L. Després, Evolution **73**, 1793 (2019).

[327] J. C. Semenza, J. E. Suk, V. Estevez, K. L. Ebi, and E. Lindgren, Environmental Health Perspectives **120**, 385 (2012).

[328] S. J. Ryan, C. J. Carlson, E. A. Mordecai, and L. R. Johnson, PLOS Neglected Tropical Diseases **13**, e0007213 (2019).

[329] J. Rocklöv and R. Dubrow, Nature immunology **21**, 479 (2020).

[330] K. D. Lafferty, Ecology **90**, 888 (2009).

[331] O. J. Brady and S. I. Hay, The Lancet **394**, 1991 (2019).

[332] T. Bakonyi and J. M. Haussig, Eurosurveillance **25**, 2001938 (2020).

[333] G. España, J. Grefenstette, A. Perkins, C. Torres, A. C. Carey, H. Diaz, F. de la Hoz, D. S. Burke, and W. G. van Panhuis, Scientific Reports **8** (2018).

[334] Q. Zhang, K. Sun, M. Chinazzi, A. P. y Piontti, N. E. Dean, D. P. Rojas, S. Merler, D. Mistry, P. Poletti, L. Rossi, et al., Proceedings of the National Academy of Sciences **114**, E4334 (2017).

[335] A. Allard, B. M. Althouse, S. V. Scarpino, and L. Hébert-Dufresne, Proceedings of the National Academy of Sciences **114**, 8969 (2017).

[336] A. Allard, B. M. Althouse, L. Hébert-Dufresne, and S. V. Scarpino, PLOS Pathogens **13**, e1006633 (2017).

[337] E.-E. Ooi, K.-T. Goh, and D. J. Gubler, Emerging Infectious Diseases **12**, 887 (2006).

[338] D. L. Chao, I. M. Longini, and M. E. Halloran, PLOS ONE **8**, e76044 (2013).

[339] E. Massaro, D. Kondor, and C. Ratti, Scientific Reports **9**, 1 (2019).

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

[341] V. Vanlerberghe, H. Gómez-Dantés, G. Vazquez-Prokopec, N. Alexander, P. Manrique-Saide, G. Coelho, M. E. Toledo, C. B. Ocampo, and P. V. d. Stuyft, Revista Panamericana de Salud Pública **41**, e16 (2018).

[342] R. F. Flauzino, R. Souza-Santos, C. Barcelllos, R. Gracie, M. d. A. F. M. Magalhães, and R. M. d. Oliveira, Revista de saude publica **43**, 1035 (2009).

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

[344] A. A. Hoffmann, B. Montgomery, J. Popovici, I. Iturbe-Ormaetxe, P. Johnson, F. Muzzi, M. Greenfield, M. Durkan, Y. Leong, Y. Dong, et al., Nature **476**, 454 (2011).

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

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

[347] P. F. Verdonschot and A. A. Besse-Lototskaya, Limnologica **45**, 69 (2014).

[348] C. Mora-Aguirre, G. Perlaza, S. Duque-Rodriguez, M. Rueda, G. Martinez, and J. H. Ruiz, *Asis report* (2016).

[349] T. J. Schaefer, P. K. Panda, and R. W. Wolford, StatPearls [Internet] (2020).

[350] R. Hilgenfeld and S. G. Vasudevan, *Dengue and Zika: Control and Antiviral Treatment Strategies*, vol. 1062 (Springer, 2018).

[351] Secretaría de Salud del Valle del Cauca., *Boletin epidemiologico. valle del cauca,semana epidemiologica n. 52* (2015).

[352] Secretaría de Salud del Valle del Cauca., *Boletin epidemiologico. valle del cauca,semana epidemiologica n. 39* (2016).

[353] H. J. Miller, S. Dodge, J. Miller, and G. Bohrer, International Journal of Geographical Information Science **33**, 855 (2019).

[354] R. Li, S. Pei, B. Chen, Y. Song, T. Zhang, W. Yang, and J. Shaman, Science (2020), ISSN 0036-8075.

[355] H. Nishiura, N. M. Linton, and A. R. Akhmetzhanov, International Journal of Infectious Diseases **93**, 284 (2020).

[356] T. Ganyani, C. Kremer, D. Chen, A. Torneri, C. Faes, J. Wallinga, and N. Hens, Eurosurveillance **25** (2020).

[357] Q. Bi, Y. Wu, S. Mei, C. Ye, X. Zou, Z. Zhang, X. Liu, L. Wei, S. A. Truelove, T. Zhang, et al., The Lancet Infectious Diseases **20**, 911 (2020).

[358] R. J. Wilson, C. Y. Zhang, W. Lam, D. Desfontaines, D. Simmons-Marengo, and B. Gipson, Proceedings on Privacy Enhancing Technologies **2020**, 230 (2020).