Universitat de les Illes Balears

MASTER'S THESIS

Quantifying mobility responses to COVID-19 containment strategies in Spain

Mar Cuevas Blanco

Master's Degree in Physics of Complex Systems

(Specialization/Pathway in Complex Systems)

Centre for Postgraduate Studies

Institute for Cross-Disciplinary Physics and Complex Systems (IFISC)

Academic Year 2021-2022

Advisor: Dr. Sandro Meloni

Quantifying mobility responses to COVID-19 containment strategies in Spain

Mar Cuevas Blanco

Master's Thesis

Centre for Postgraduate Studies

University of the Balearic Islands

Academic Year 2021-2022

Keywords:
Computational Epidemiology, Mobility, Complex Networks, Complex Systems, COVID-19

Thesis Supervisor's Name: Dr. Sandro Meloni

To my grandfather, I miss you everyday.

Acknowledgments

First and foremost, I would like to thank Dr. Sandro Meloni, the best supervisor that one can ask for. Without his assistance, guidance and encouragement, this Master's Thesis would not have been possible. Moreover, I also would like to thank him for his closeness and for his constant willingness to help me at any time. I especially thank him for his patience and for having read, with detail and humour, the endless versions of this project.

In addition, my parents and sister, Lara, deserve my sincere thanks as well. They have always been my biggest supporter and my academic and personal references. I would like to especially thank them for the support they have given me in the development of this Master's Thesis and for not getting tired of me talking about it non-stop for over a year. My sister and best friend Lara, I would also like to thank her for all her support, affection, and advice that allowed me to get to where I am today. I know you can accomplish anything you put your mind to, but keep in mind that the place you want to go must be where you are happiest.

From the bottom of my heart, I would like to thank Gorka for all the support and advice he has provided me throughout this project. He is always by my side though good and bad times. Sharing life with him, both personally and academically, is the greatest gift I have ever received.

It gives me great pleasure to acknowledge all of my friends and colleagues, both those who have always been here and those who have entered my life in recent years. Specifically, I wish to thank my colleagues at IFISC for their support and encouragement in the final stages of drafting this project.

Last but not least: https://youtu.be/wGRF3GQ4Wdk.

Mar Cuevas Blanco

Abstract

The onset of COVID-19 in late 2019 had a wide and profound effects on our lives. Some of the outstanding impacts of the pandemic are on the way people interact and travel. In response to the sanitary crisis, many countries have implemented containment policies that have proven effectiveness in controlling and mitigating the spread of the disease. However, the increase in the frequency of epidemics observed in recent years, underlines the importance of knowledge on the effects that restrictive policies on human mobility have on epidemic spreading. This is useful not only for understanding and predicting the dynamics of COVID-19 infection, but it is rather essential to better cope with similar scenarios in the future. The purpose of this Master's Thesis is to assess how the human mobility network, on a country-wide basis, has evolved in response to restrictive measures and how these changes affect the ability of the network to support diffusion. For that purpose, we use movement data, of mobile phone users, that account for the number of trips between each pair of locations in Germany and Spain. In Germany, we focus on the first outbreak, while in Spain, we extend the coverage period to more than a year later, allowing us to identify long-lasting changes. The study of mobility patterns in both countries showed that traffic was effectively reduced. In Germany, pre-pandemic values were re-established in early June, 2020, whereas in Spain they are no longer reached. Furthermore, this reduction did not occur homogeneously in the network, with long-distance flows showing a greater reduction than short-distance flows. Furthermore, in Spain, we find that long-distance travel has been one of the primary drivers of the epidemic across the country during the early stages of the pandemic. To better understand these results, we go one step further and, for Spain, we study and characterise weekly mobility networks, using the tools provided by Network Science. Our analyses reveal that, since the implementation of the containment policies, local connections have a higher probability of being retained and hence paths are generally longer, since more local steps have to be included. Moreover, we find evidence of profound structural changes in the networks that remain present in the long term. Such changes cannot be explained by a uniform reduction of mobility alone. The resulting mobility networks are less dense, more clustered and local, and hence, more homogeneous. Finally, we study how these changes affect the spread of epidemics. To this end, we implement an epidemiological metapopulation model that takes into account different containment scenarios. We find that lockdown policies have a significant impact on the spread of the outbreak. In addition, a reduction in long-distance travel, reduces the geographical spread of the disease. This effect is primarily due to structural changes in mobility networks. All of this suggests that specific mobility restrictions, that target long-distance connections, should be used to limit the spread of a disease. We hope these findings will be of great assistance in mitigating the current COVID-19 pandemic and lead to a better preparedness for similar future scenarios.

List of Acronyms and Terms

AI Cumulative Incidence
CCAES Center for the Coordination of Health Alerts and Emergencies
CNE National Epidemiology Center
COVID-19 Coronavirus Disease, 2019
ICU Intensive Care Unit
INE Statistics National Institute
MERS Middle East respiratory syndrome
MITMA Ministry of Transport, Mobility and Urban Agenda
OSLOM Order Statistics Local Optimization Method
RENAVE National Epidemiological Surveillance Network
RT-PCR Reverse Transcription – Polymerase Chain Reaction
SAR Secondary Attack Rate
SARS Severe Acute Respiratory Syndrome
S, E, I, R Susceptible Exposed Infected Recovered
SI Supplementary Information
SiViES National Surveillance Network Epidemiological of Spain
SPL Shortest Path Length
USA United States of America
WHO World Health Organization

Contents

List of Captions

List of Tables

Chapter 1

Introduction

In December, 2019, a series of cases of hospitalized patients with a new disease characterized by pneumonia and respiratory failure were reported in Wuhan, in the Chinese province of Hubei. This disease was caused by a new virus (SARS-CoV-2)[1] that is part of the coronavirus family, which includes common viruses that cause a variety of diseases from headache or chest colds to more severe diseases like the severe acute respiratory syndrome (SARS) and the Middle East respiratory syndrome (MERS).

The word "corona" means crown and it is so called by the appearance that coronaviruses have with the pointed proteins that protrude from them. These spike proteins are important as they attach to human cells to infect them, allowing it to replicate within the cell and spread to other cells [2]. On February 11, 2020, the World Health Organization (WHO) named this etiological agent as **COVID-19** (Coronavirus Disease, 2019). It is a highly contagious disease and has quickly spread around the world.

In order to control the spread of COVID-19, many countries have implemented various containment measures. These measures affect mobility and this, in turn, can have effects on the disease transmission. In this Master's Thesis we study mobility trends in Spain and Germany during the COVID-19 pandemic and address the question of to what extent these changes in mobility have caused deep long-term structural changes in the mobility networks. In addition, we analyze how these changes can affect spreading processes. To do this, we approach this problem using a Complex Systems approach and, in particular, using Network Science.

1.1 Biology of Spreading

Virologists are constantly monitoring the mutations that cause changes in the spike protein through a process called genomic surveillance [2]. As genetic changes occur in the virus over time, the SARS-CoV-2 virus begins to form genetic lineages. These different genetic lineages can change how quickly the virus spreads, the severity of the disease it causes, or the effectiveness of treatments against it. Scientists call viruses with these changes "variants" [2]. To help with public discussions on variants, the WHO convened a group of scientists, who currently constitute the so-called Technical Advisory Group on Virus Evolution [3], and who have recommended the use of letters from the Greek alphabet, i.e., Alpha, Beta, Gamma, Delta, which will be easier and more practical to be discussed by non-scientific audiences.

There are a series of epidemiological parameters, summarized below, the meaning of which must be understood, and for which it is necessary to know average values for the case of COVID-19, to be

able to assess statistics and information on its evolution. In our case this basic information has been extracted from the Spanish Ministry of Health[1] [4].

The **median incubation period**[2] for COVID-19 is 5.1 days, and at 11.7 days 95% of symptomatic cases have already developed their symptoms. This parameter is used to calculate, for a case contact, the quarantine time that must be carried out to avoid transmitting the infection to other people. The **mean serial interval**[3] of COVID-19, in numerous epidemiological observations, was shorter than the incubation period, in such a way that it was possible to determine that the transmission of the infection began 1-2 days before the onset of symptoms. These values may differ from variant to variant. The **latency period**[4] for SARS-CoV-2 is 1-2 days shorter than the incubation period, that is, it would be around 3-4 days. The average time from the onset of symptoms to recovery is 2 weeks when the disease has been mild, and 3-6 weeks when it has been severe or critical. The time between the onset of symptoms and the appearance of severe symptoms, such as hypoxemia, is 1 week, and 2-8 weeks until death occurs. However, although this description corresponds to the norm, there have been many cases of people who report prolonged and recurrent symptoms, for weeks or months, and that in some contexts it has been called persistent COVID-19 or "Long COVID". The **secondary attack rate** (SAR) expresses the number of cases of an infection that occur among contacts within the incubation period, after exposure of a primary case, in relation to the total number of contacts that have been exposed. It is a measure of contagiousness and is useful in evaluating control measures. It varies according to the conditions in which transmission occurs: the type and duration of exposure, the closed or open environment, the relationships between the more or less close contact and patients zero, the amount of viable virus capable of being transmitted through respiratory secretions, the absence of preventive measures, etc.

In case studies on close contacts published in scientific journals, the SAR for COVID-19 has been highly variable, ranging from 0.7% to 75% [4, 5]. In some cases, it has been observed that short-term events (work meetings, meals, sporting events etc.) can lead to high SARs, since they have a great capacity to transmit the virus to other people. There is evidence that human-to-human transmission has occurred among close contacts since the middle of December, 2019 in Wuhan, China [6].

The vulnerability of the contacts, together with a closed environment and close contact between people, create the ideal conditions for the transmission of SARS-CoV-2. This combination has been observed in all countries in nursing homes with SAR among residents and workers above 50% [4]. The **basic reproduction number**, $\mathcal{R}_0$, is the average of secondary cases produced from a case and varies according to social contacts [4]. Two reviews collecting 32 studies that used different methodologies, estimate $\mathcal{R}_0$ values in the range 1.5-6.5 during the epidemic in Wuhan and in Italy, observing how the public health and physical distancing measures taken have had a direct impact on the decrease in $\mathcal{R}_0$. The **effective reproductive number** ($\mathcal{R}_e$) is the estimate of how many people on average have been infected each day from the existing cases observed during an epidemic (at the time they are notified). Unlike $\mathcal{R}_0$, which would be an averaged and theoretical calculation, $\mathcal{R}_e$ is a value that takes into account the real-time observation of the epidemic and allows its dynamic evolution to be followed [4]. If $\mathcal{R}_e > 1$, the number of cases will increase, such as at the start of an epidemic. When $\mathcal{R}_e$ <1 there will be a decline in the number of cases.

The period in which a case can transmit the infection to another person can be inferred by detecting viable virus in clinical samples, although cell culture is a technique that can have a relatively low sensitivity. The real time reverse transcription–polymerase chain reaction (real time RT–PCR), is one of the most accurate laboratory methods for detecting, tracking and studying the COVID-19 coronavirus and viruses in general. However, there is some controversy regarding its use due to

[1] Spanish Ministry of Health: Ministerio de Sanidad (https://www.sanidad.gob.es/).

[2] The incubation period of a disease is defined as the duration between initial exposure and the onset of disease symptoms.

[3] The serial interval is defined as the time between the onset of the disease in the primary case and the onset of the disease in the secondary case.

[4] The latency period is the time that elapses from exposure to the virus until the moment when the person can transmit the disease, that is, it is the period that immediately precedes the infectious period.

its ability to detect viral RNA[5] for very long periods that cannot always be related to viruses with infective capacity [4]. Using the RT-PCR technique, it has been observed that most of the infected people present a high viral load (between 105 and 108 genome copies per nasopharyngeal or saliva sample) before the onset of symptoms. In patients who have a mild infection, the peak of the viral load in nasal and oropharyngeal samples occurs during the first 5-6 days after the onset of symptoms and practically disappears on the tenth day. In some patients, virus is detected beyond day 10, but the viral load is of the order of 100-1000 times lower, which suggests that the transmission capacity is progressively decreasing. In addition, it has been possible to demonstrate the absence of infective virus (no growth of the virus in cultures) with viral loads below 105 RNA copies. This seems to indicate that in people with mild symptoms, beyond the first week after the onset of symptoms, the probability of transmitting the infection to others would be very low, even when the virus is still detectable by PCR. These statistics of results obtained from RT-PCR tests have been assessed in the Spanish Ministry of Health report [4].

A very important aspect to understand the evolution of COVID-19, its relationship with protection measures, how are the use of masks, and prevention (vaccination campaigns, physical distancing, and restrictions on mobility), is to know how the SARS-CoV-2 coronavirus is transmitted.

During the first few months of the COVID-19 pandemic in early 2020, the WHO and other public health agencies downplayed the airborne or aerosol transmission route, recognizing it as a potential route for transmission only during certain medical procedures such as intubation [7].

The erroneous assumption that transmission through proximity involves large respiratory droplets was historically used for decades to deny airborne transmission of tuberculosis and measles, becoming medical dogma. Direct measurements of aerosols[6] and droplets were ignored, revealing defects in medical assumptions, such as the overwhelming number of aerosols produced in respiratory activities and the arbitrary limit on $5\mu m$ particle size between aerosols and droplets, rather than the correct limit of $100\mu m$ [8]. It is sometimes argued that since respiratory droplets are larger than aerosols, they must contain more virus. However, in diseases in which pathogen concentrations have been quantified by particle size, smaller aerosols showed higher pathogen concentrations than droplets when both were measured [8].

All people, when they speak and breathe, emit aerosols from their airways of different sizes ranging from nanometers to hundreds of microns, and people with an active respiratory infection can emit aerosols containing pathogens, called bioaerosols[7]. Depending on the sizes and density of these aerosols, the aerodynamic behavior is different. Only particles larger than $100\mu m$ are considered to have "ballistic" behavior. These large particles are normally deposited on the ground and other surfaces in a few seconds because of gravity, in some cases being able to travel a maximum distance of $2m$ from the person who emits them. These particles could reach a susceptible person who is nearby, impacting somewhere (eyes, mouth, nose) from which they could cause infection. Any other particle smaller than $100\mu m$ is considered an aerosol, since it is suspended in the air for a long time, ranging from seconds to hours and that can be inhaled at a distance greater than $2m$ from the emitter or even in the absence of a emitter, if they are still suspended in the air. Depending on their size or aerodynamic behavior, aerosols between $15\mu m$ and $100\mu m$ reach the upper respiratory tract, aerosols from $5\mu m$ to $15\mu m$ can reach the trachea and main bronchi, and aerosols less than or equal to $5\mu m$ have ability to reach the alveoli [9]. The aerosols produced by breathing or speaking are, more than 80-90%, very small ($<2.5\mu m$), coming from the lungs and, on average, more than 500 are generated per litre of exhaled air. Coughing produces 3,000 particles and sneezing 40,000, mostly small (1-$10\mu m$), from the upper respiratory tract. In the case of maximum viral-loading derived from experimental data of COVID-19 patients, the minimum size of a respiratory particle that can contain SARS-CoV-2 is calculated to be approximately $4.7\mu m$ [10].

[5] RNA: Ribonucleic Acid.

[6] Aerosols: set of solid and liquid particles suspended in the air that can be observed and measured, with sizes between 0.001 and $100\mu m$.

[7] Bioaerosols: aerosols of biological origin.

The deposition of inhaled particles in the lungs occurs mainly by the following mechanisms: impact by inertia, sedimentation by gravity and Brownian motion (random movement that is observed in the particles that are in a fluid medium, as a result of collisions with the molecules of said fluid) [9].

An investigation by Allen and Ibrahim [11] concluded that larger droplets fall to the ground by gravity (representing 99% of the emitted liquid mass), but the smallest aerosols ($<5\mu m$) (emitted both when speaking, breathing and coughing) can remain in the environment for times greater than half an hour. These minor aerosols, upon emission, may reach distances of up to $2m$ from the emission point, which makes certain distances between people insufficient.

Most SARS-CoV-2 outbreaks involving three or more people have been associated with time spent indoors. Evidence confirms that far-field airborne transmission (defined as within-room but beyond $2m$) of SARS-CoV-2 is occurring [11]. Increasing air changes per hour and air filtration are simple strategies that could be deployed to help reduce risk from within-room, far-field airborne transmission of SARS-CoV-2. Healthy building controls, like higher ventilation and enhanced filtration, are a fundamental, but often overlooked, part of risk reduction strategies that could have benefit to control the transmission of SARS-CoV-2.

In view of the evidence found to date, it can be established that: 1) Aerosols generated by people with active SARS-CoV-2 infection contain viable viruses; 2) The viruses contained in aerosols have the capacity to generate infection, especially in certain circumstances: in proximity to the index case for a long time and in closed and poorly ventilated spaces; 3) Target tissues are accessible, for aerosols of any size with entry ports anywhere in the respiratory tract.

For all the above, there is consistent, strong evidence that SARS-CoV-2 spreads by airborne transmission. Although other routes can contribute, the airborne route is likely to be dominant [8].

The predominant form of spread of SARS-CoV-2, through aerosols, which involves close contact between infected people and potential hosts, highlights the importance of the mobility of people at different time and spatial scales to understand the evolution of COVID-19 pandemic.

1.2 Relationship between infectious disease spreading and mobility

There are numerous studies that emphasize the importance of mobility in the spread of an epidemic [12, 13, 14, 15]. For geographical spread to take place and hence the outbreak of an epidemic to occur, the combination of human interactions that can give rise to contagion must occur along with human mobility at different spatial scales [16]. Infectious contacts are responsible for the proliferation of the disease, but it is the large-scale human mobility that is responsible for its geographical spread. Both the dynamics and the control and evolution of a disease are therefore intimately connected with the joint dynamics of epidemiological and mobility processes that take place on different spatial, temporal and organizational scales [17]. It is then the mobility of people at all scales, from displacement at the level of provinces or cities to mobility between countries and continents, which favors the spread of the disease on a large scale and therefore the appearance of an epidemic. Hence, controlling large-scale mobility is critical to prevent the spreading from accelerating.

Consequently, due to the recent appearance and global spread of COVID-19, many countries have decided to implement containment policies in order to try to reduce the transmission of the disease. Spain began to implement measures from March, 2020. For this reason, in this work we focus on analyzing movement data in Spain since February, 2020 and covering a period of more than a year.

It has been shown that mobility, thanks to the different containment measures, was significantly reduced. Nevertheless, it remains unclear whether these reductions have caused deep, long-term structural changes in the mobility network. As well as the effect of these structural changes in the network's ability to support diffusion such as the spread of epidemics. Therefore, thanks to the availability of this mobility data, we have the opportunity to understand more deeply what the role of human mobility in an epidemic is as well as to identify what type of interventions and policies that concern mobility are a priority to control possible similar future scenarios as carried out in [18, 19, 20], where they study the impact of travel limitations.

1.3 Situation in Spain

According to information collected from Equipo Covid-19[8] [21] and other official sources, a summarized chronology of COVID-19 evolution In Spain is given below. The Equipo Covid-19 report contains the analysis of COVID-19 cases notified by the Autonomous Communities to the National Epidemiological Surveillance Network (RENAVE) through the SiViES platform (System for Surveillance in Spain), which is managed by the National Epidemiology Center (CNE). In accordance with the criteria established in the surveillance protocol at all times from the start of the pandemic. Six periods are indicated in this report to characterize the evolution of the COVID-19 pandemic in Spain. However, since the data analyzed in this Master's Thesis covers from February 21, 2020 to April 9, 2021 included, we will only refer to the first four periods that cover this time frame:

- **First period**: From the start of the pandemic to June 21, 2020, when the first wave of the COVID-19 epidemic and its corresponding State of Alarm ended.
- **Second period**: From June 22 to December 6, 2020, the inflection point of the 14-days cumulative incidence (AI) of COVID-19 cases, between the second and third epidemic waves.
- **Third period**: From December 7, 2020 to March 14, 2021, the inflection point of the AI to 14 days of COVID-19 cases, between the third and fourth waves.
- **Fourth period**: From March 15 to June 19, 2021, the inflection point of the AI to 14 days of COVID-19 cases, between the fourth and the fifth epidemic waves. Notice that only the first part of the fourth period is covered in this study.

Moreover, there are some noteworthy aspects in each period that deserve to be highlighted.

First period: The first case of COVID-19 in Spain was registered on January 31, 2020. It was a German patient admitted to the island of La Gomera (Canary Islands). Nine days later, a second case was detected in Palma de Mallorca (Balearic Islands), also imported. On February 24, the first cases appeared in continental Spain, with a subsequent exponential community circulation of the virus, mainly affecting the Community of Madrid, the Basque Country and Catalonia. As these data indicate, the onset of COVID-19 infections in Spain occurred as a result of direct contact with foreign tourists, suggesting long-distance mobility as a potential cause in triggering the pandemic in Spain. On March 4, the first fatality related to COVID-19 was known in Spain. On March 9, when there were 999 positive cases and 16 deaths in Spain, the autonomous governments of the most affected regions began to take measures to prevent the spread. On March 10, the Council of Ministers agreed to ban all direct flights between Italy and Spain and to suspend events involving more than a thousand people in Madrid, La Rioja and Vitoria. On March 14, when in Spain there were around 6,000 cases and 200 deaths, the Council of Ministers declared a State of Alarm throughout the national territory with the aim of curbing the health emergency, introducing a package of measures that significantly restricted mobility. It was initially established for a period of 15 calendar days by

[8]Equipo Covid-19: organization belonging to the Carlos III Health Institute whose objective is scientific-technical support for the National Health System, health control and research in health sciences.

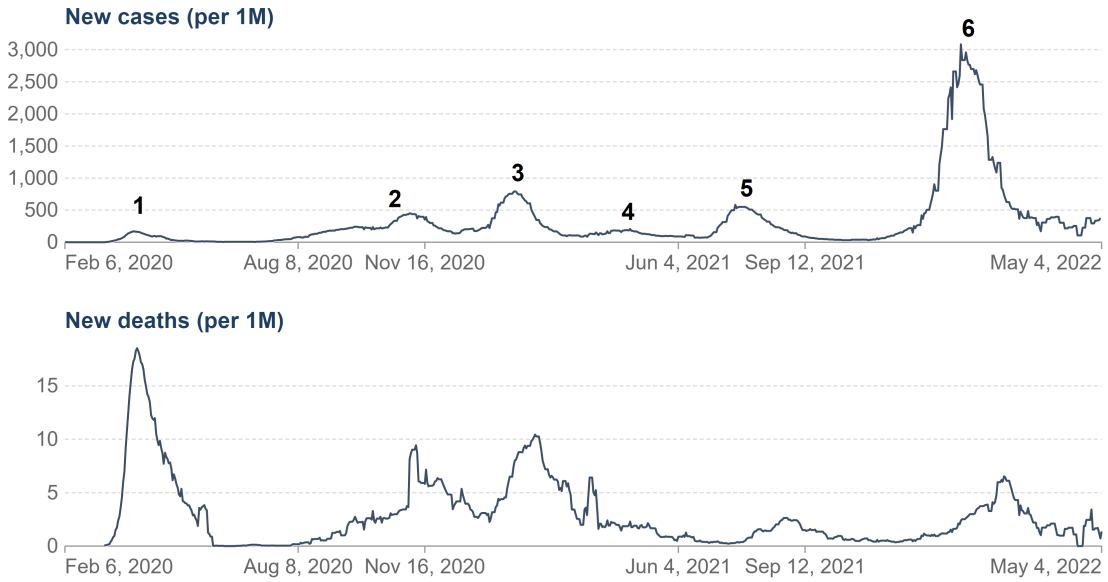


Figure 1.1: Evolution of daily cases and daily deaths from the COVID-19 pandemic in Spain. This data has been provided by the Carlos III Health Institute obtained from the individualized declaration of COVID-19 cases to the National Epidemiological Surveillance Network (RENAVE) through the SiViEs computer application. Figure provided by [22]. The six waves of infections that have occurred in Spain are marked with numbers. Data updated until May 4, 2022.

Royal Decree 463/2020. During the first weeks of the State of Alarm, the number of infections and deaths increased at a very high rate. The escalation of infections in March, 2020 was exponential. The curves, represented in the graphs of Figure 1.1 show that the number of known infections and daily deaths, followed a very sharp ascending line. The lack of diagnostic tests that existed at that time makes us think in perspective that the figures of the pandemic at this time were, in reality, much higher than those we know. Some research even suggests that in March, 2020 only 1 in 10 cases of COVID-19 was detected. Despite the entry into force on March 14 of the State of Alarm, hospitals collapsed and the lack of protective equipment reported to our health workers the highest rate of infections in the world. Nursing homes for the elderly became sources of contagion, and on many occasions our most vulnerable group of elderly people could not be cared for due to the lack of resources, both human and material. According to the data provided by the Ministry of Health, this first wave resulted in 21,418 deaths and 130,749 infections.

Thanks to the confinement of the population, a progressive decrease in the effective reproductive number ($\mathcal{R}_e$) was achieved, from > 2.0, to 0.98 on April 4, while the infection rate decreased from 40% (prior to confinement) to 3, 2% on April 6 [4].

After forty-four days of confinement, on April 28, 2020, the Council of Ministers approved a de-escalation plan in four phases to gradually reduce confinement. The passage to each of the successive phases would be carried out differently in each province or island according to its epidemiological situation, with a minimum period of two weeks between phases.

The "**Phase 1 - Initial**" allowed social gatherings at home, the opening of small businesses, hotel terraces with 30% capacity, the opening of the hotel sector, the reactivation of the agri-food and fishing sector and of the open-air markets. The museums were opened to 30% of the capacity and the holding of shows with up to 30 spectators in closed places and 200 in open spaces was allowed. Medium sports training for professionals and non-professional sports activity without physical contact and without the use of changing rooms was allowed. Places of worship were opened to 30% of their capacity and wakes could be held with a limited number of attendees.

The "**Phase 2 - Intermediate**" contemplated the opening of all commercial premises regardless of their size (including shopping centres and cinemas), table service was allowed inside bars and restaurants, the capacity in premises and public spaces was increased and the beaches were opened.

The "**Phase 3 - Advanced**" continued with the gradual removal of restrictions for the different economic and socio-cultural sectors and for the general mobility of citizens, increasing the activities allowed and the planned capacity, but trying to avoid crowds and strictly maintaining the social distance between people.

On May 21, the use of masks in public spaces became mandatory from the age of 6.

The end of the lockdown met its goal and significantly stopped new cases of COVID-19. The first wave of the pandemic is used as a benchmark to compare those that followed it. None of them have repeated terrible data such as the peak of deaths from COVID-19 registered on March 30, 2020: 913.

Second period: The second wave of infections took place, on June 21, 2020, after 98 days and six extensions, the first State of Alarm ended and the entire country officially entered the "new normal" phase. However, the relaxation of certain measures during the summer translated into a progressive increase in infections, that triggered the accumulated incidence and the scenario of hospitalizations, ICUs and deaths was repeated. On October 21, 2020, Spain exceeded one million infected. On October 25, the central government again decreed a State of National Alarm to face the second wave of the pandemic, but without establishing a new general confinement of the population. A curfew was imposed throughout the national territory from 11 at night to 6 in the morning, allowing the autonomous communities to advance or delay it one hour in their territory, and meetings of more than six people were prohibited. On October 29, the Congress of Deputies approved to extend the State of Alarm until May, 2021. The data of deaths of the second wave, in many reports, overlap with that of the first. Until December 6, 2020, according to data made public by the Ministerio de Sanidad [23], 46,646 people died from COVID-19 in Spain. Most of the deaths occurred during the first wave, although some autonomous communities registered higher peaks of deaths in the second. Between the months of March and May, 2020, the death toll from COVID-19 was 28,985.

Third period: At the end of February, 2021, experts began to see the end of the third wave of the pandemic. The curve began to rise on December 11, 2020. On December 27, the vaccination campaign began. On January 7, 2021, the two million infected officially confirmed were exceeded, and on February 9, the three million infected were exceeded. In this third phase of the pandemic, 19,200 people died from COVID-19 between December, 2020 and February, 2021 according to data published by RENAVE.

Fourth period: The second State of Alarm ended on May 8, 2021. The fourth wave was held in April, 2021 of the pandemic was marked by the massive vaccination campaign against COVID-19. The fourth wave was, in a way, atypical compared to the previous ones, which came to be described by the director of the Centre for the Coordination of Health Alerts and Emergencies (CCAES), Fernando Simón, as a “small ripple”. In terms of mortality, the fourth wave represents an improvement compared to the previous ones. The fact that around 10% of the population already had the complete vaccination schedule, mostly the most vulnerable groups, reduced the impact of this wave.

Already out of the fourth period, on July 13, 2021, the four million officially confirmed infected were exceeded. On July 19, more than half of the Spanish population was vaccinated with the complete schedule. On September 1, 2021, 70% of the population was fully vaccinated.

The different restrictions implemented over time in Spain throughout the evolution of the pandemic summarized above, have affected mobility. These changes not only affect the flow of trips, which

translate in a reduction of the global mobility, but also have effects on the structure of the mobility network. These changes could be deep enough to remain in the long term, and as we anticipated, are of great importance in the spread of an epidemic. Therefore, they are object of study in this work.

1.4 Objectives of the Master's Thesis

This Master's Thesis is intended to study how human mobility has changed in the different waves of the COVID-19 epidemic in Spain. In addition we asses how the mobility network have responded to the imposed containment measures and their impact on the spread of the epidemic. For this, the period studied ranged from the beginning of the first wave to the first stages of the fourth.

Chapter 2

State of the art and Motivation

Development and use of mathematical models to predict the evolution of the COVID-19 epidemic and the impact of the measures for its control, today constitutes a research priority worldwide. Understanding human mobility changes and spatial interaction patterns are crucial for understanding and forecasting COVID-19 dynamics [24].

Most of the present models used in epidemics are based on a relatively simple model, the SIR (Susceptible-Infected-Recovered) model, that was formulated almost a century ago, in 1927, by Kermack-McKendrick in [25]. The spread of disease on networks has attracted considerable recent attention by the physics community, who has used the SIR models to deal with the explanation of the spread of different diseases on a wide variety of networks [26]. The SIR classifies the population into the following three groups: Susceptible (S), state in which the individual can be infected by another person who is infected; Infected (I), state during which the individual is infected and can also infect others; and Recovered (R), or cured, a state during which the individual cannot even be infected by having acquired immunity (temporary or permanent) or infect (by having recovered or have passed the contagious stage of the disease).

Non-pharmaceutical interventions have been shown to be critical to delaying and containing the COVID-19 pandemic. These include testing and tracing, ban travels and large gatherings, non-essential business and school and university closures, and restrictions on domestic mobility and physical isolation [27]. Maier and Brockmann [28] explained the distinctive subexponential increase of confirmed cases during the early phase of the epidemic, contrasting with an initial exponential growth expected for an unconstrained outbreak, as a direct consequence of containment policies that effectively deplete the susceptible population.

Social distancing measures are one of the most effective non-pharmaceutical methods governments have counted on to control the spread of COVID-19. The social distancing measures have led to significant changes in mobility behavior, while the mobility change has also inevitably affected the dynamics of the disease [24, 29].

The COVID-19 pandemic has been the subject of numerous studies and an unprecedented number of scientific publications. Many of the studies have been focused on different aspects of mobility. Evaluating the impact of lockdowns, forced by COVID-19, on population mobility is important to help characterizing the changes in social dynamics that affected viral diffusion [30]. In all these studies, large anonymous mobility datasets from mobile phone networks (i.e., [15, 31, 32, 33]), or several social networks ([34]) have been used.

Some of these studies have addressed the relationship between mobility reductions, lockdowns and the transmission of COVID, such as those carried out in [20, 29, 31, 35, 36]. Other studies have

tried to provide epidemiological mathematical models to project the impact that travel limitations have had on the national and international spread of the epidemic (i.e. [18, 19, 20]), focusing, in some cases, efforts on the implementation of short-term modeling forecasts that can provide time-critical information for decisions on containment and mitigation strategies ([35, 37]). [32, 37, 38] analyzed how people in different counties and states in USA reacted to the social distancing guidelines, including mobility reduction. Other relatively novel studies have used a different approach, consisting of assessing whether the mobility limitations imposed by COVID-19 have been able to produce profound structural changes in mobility networks ([15, 33, 34, 39]), and how these changes could, in turn, impact on the spread of the epidemic ([15, 34]).

This latter is the approach that we have chosen in this work. A preliminary work has consisted in replicating the results published in [15], based on the methodology and the data used in it (which corresponds to Germany), since they are freely available. In a second phase, the same methodology for mobility trends has been implemented to the mobility data in Spain, but extending the study period until the fourth wave of the epidemic in April, 2021 and giving a comparison of the results obtained in the two countries. In a third phase, we have studied how these changes in mobility have caused structural changes in the mobility networks. Lastly, in a fourth phase, we study the impact of these structural changes on the spread of epidemics.

2.1 Structure of the Thesis

The rest of this Master's Thesis is structured as follows. Chapter 3 presents the different methodologies used. This chapter is structured as follows. First Sections, 3.1 and 3.2, summarize the different datasets used in the study as well as further data processing carried out prior to the study. It is followed by Section 3.3 where the methodology used to construct the mobility networks is presented. 3.4 summarizes the different measures carried out in order to study mobility trends as well as those measures implemented to characterize the mobility networks. Finally, Section 3.5 details the model implemented for studying how these changes affect the spread of the epidemic. For this, the SIR model is explained first, then the extension to a metapopulation model, to finally show the details of the metapopulation SIR model with containment used.

In Chapter 4 the main results of this Master's Thesis are presented. We first study mobility trends in Germany and Spain. Section 4.1.1 focuses on reproducing the results of the mobility trends in Germany studied in [15]. Hereunder an analogous study of mobility trends is made for Spain mobility in Section 4.1.2, with an extension in the period under study in order to include until the fourth wave of the epidemic in April, 2021. Moreover, it is included an additional detailed analysis of the periods of interest. Afterward, in Section 4.2, the results obtained in the characterization of mobility networks in Spain are shown. It follows Section 4.3 where the results of the implementation of the epidemiological model presented above are presented. Finally, Chapter 5 summarizes the main conclusions of the work and suggests future research lines.

Chapter 3

Methodologies

3.1 Data collection

3.1.1 Mobility datasets

In this study, we first focus, on reproducing the mobility trends observed in **Germany** studied in [15]. The dataset, which is publicly accessible [40], provides daily mobility flows in Germany from January 1, 2020 to June 10, 2020, including also the data corresponding to March, 2019, which will be used as a comparative baseline as in [15]. Originally, the mobility data were provided by Telefónica[1] and later added by the company Teralytics[2]. These sources provide mobility flows in Germany, obtained from mobile data, in the study period. The trips are collected from telephone towers, defining that a trip begins when a device leaves its current tower and ends when it becomes stationary, for at least 15 minutes, in another or the same tower. The dataset takes into account the around 43.6 million Telefónica users (2019), equivalent this, during times with normal mobility flows (March, 2019), to a total of 176 million daily trips or equivalent to 3.8 trips per user and per day. These trips are spatially aggregated to the level of the 401 counties in Germany and temporarily by days, also taking into account the internal flows of each location. The dataset is public in an anonymized form. The element F_{ji} quantifies the total number of trips from i to j and has dimension 401$\times$401. First, a threshold of 5 trips is established, and daily flows below this value are removed for privacy reasons. Also, the identity of the counties is anonymized by assigning to each one an integer and furthermore, a small noise is added to the distances between each pair of counties. This distances, calculated from the coordinates of the centroid, are multiplied by a random number drawn from a normal distribution with mean 1 and a standard deviation of 0.02.

To carry out the main study of this Master's Thesis, inspired by [15] for **Spain**, we have used data publicly accessible on the MITMA[3] website [41]. The dataset used, at the time of access, offers mobility data in Spain during the period of the COVID-19 pandemic at the national level. Trips are collected mainly from mobile phone positioning data. The period covered is from February 24, 2020 until April 9, 2021, also including the data corresponding to the reference period from February 14 to 20, 2020. In our case, we decided to use the same reference period as in [15] (from March 2 to 8, 2020). We justified the choice of this week as baseline in Section 4.1.2. This dataset collects the number of trips between each pair of locations, by days and hour. Likewise, the distance between the origin and destination is indicated, grouped by blocks, comprising distances between 0.5 and 2km up to distances >100km. The dataset is available at the district level and at the municipality level. In this study, we have used the data at the municipality level, comprising these trips between all pairs of the 8,131 possible locations (municipalities) in Spain or, clusters of these in the case

[1] Telefónica: Spanish multinational telecommunications company. It is the most important telecommunications company in Europe.

[2] Teralytics: software company in Zurich, Switzerland.

[3] MITMA: Ministry of Transport, Mobility and Urban Agenda.

that the area is sparsely populated. The identity of the municipalities is public and the shapefile of the zoning used is provided. For manageability reasons, the study on mobility networks has been conducted at the provincial level. That is, we have aggregated the data at the level of the 52 provinces of the country, where Ceuta and Melilla have been included as provinces.

3.1.2 Other datasets

We have used free access shapefiles at the Spanish province level on the Opendatasoft[4] page [42]. Data on the total number of cases of COVID-19 infected in Spain at the provincial level were obtained from the Instituto de Salud Carlos III[5], available through its website [43]. Finally, regarding the census population data used for the implementation of the SIR model used in Section 3.5, public data of the total census by provinces in 2020 in Spain are downloaded from the INE[6] website [44].

3.2 Data processing

This Section explains the further processing of the mobility datasets, performed prior to the study.

When studying mobility trends in **Germany**, we focus on daily mobility changes $\Delta n(t)$. These daily mobility changes quantify the relative difference between the total number of trips per day, compared to the number of trips of the same weekday in the baseline period. To do this, the average daily flow between all pairs of locations corresponding to the trips in March, 2019 is used as the baseline. This means that our baseline contains for each day of the week $t_0 \in \{1, 2, 3, 4, 5, 6, 7\}$, the average number of trips between each pair of possible locations $\langle F_{ji}(t_0)\rangle$. Then, we compare the mobility of each day of 2020, included in the studied period, with the average mobility corresponding to the same day of the week. We use March, 2019 as the reference period since it is assumed that structurally it is the closest to the same month of 2020 where the vast majority of restrictions and containment policies were imposed in Germany, which consequently resulted in a drastic reduction of the mobility flows. In the original work [15], it was taken into account that the choice of this baseline does not consider or explain general changes in mobility patterns, which may be due, for example, to seasonal effects or federal holidays. Consequently, and to be also consistent with the original study, when reproducing the mobility trends in Germany, during the first months of the COVID-19 pandemic, we excluded the dates corresponding to federal holidays. In general, the days that are national holidays have a lower mobility, compared to the same day of the week in the baseline. As when making the comparison with the baseline, we cannot know if this drastic reduction is common on holidays, it was decided to remove them to avoid distorting the data.

Regarding our study of mobility trends in **Spain**, first, the original dataset shows us the number of trips between all municipalities in the country and by hour. Given the large amount of data, we aggregate the data in such a way that we obtain daily flow matrix, where, again, each element $F_{ji}(t)$ represents the total number of trips from municipality i to municipality j taking into account each possible pairs for locations, including flows within municipalities ($F_{ii}(t)$). In addition, when studying changes in mobility networks as well as the effect on spreading processes, we conducted our study at the provincial level. Therefore we aggregated the data at the level of the $n = 52$ provinces of Spain. In this way, our daily flow matrix, $F(t)$, is reduced from having dimension 8,131×8,131 to 52×52, which is much more manageable. To carry out this aggregation, it is taken into account that in the dataset, each municipality is identified with a code of letters and numbers whose first two digits, two integers, correspond to the province to which the municipality belongs

[4]Opendatasoft: french software company.
[5]Instituto de Salud Carlos III: Spanish autonomic public organism.
[6]INE: Instituto Nacional de Estadística. Autonomous Spanish organism.

to. In addition, this provincial code is common in the different datasets used.

Our study, as we have already advanced, once the part of reproducing the changes in the mobility trend in Germany is performed, it is then, focused on studying the changes in mobility that took place in **Spain**. Therefore, we started by carrying out a study similar to the one made for Germany. In this case, we calculate daily mobility changes $\Delta n(t)$ with the selected reference period (March 2 to 8, 2020) (Section 4.1.2 provides further arguments for the choice of this reference period). Thus, in our reference period, we have a dataset with the total daily trips for each day of the week. This choice presents the same problems as the one chosen for the study carried out in Germany. In addition, these effects can be increased as the reference period is not an average made with the data of a whole month. To be consistent in our study, and to try to minimize the effects of choosing this reference period, for the same reasons as in the study of Germany, those dates corresponding to national holidays were removed.

3.3 Mobility Networks

3.3.1 Weekly Mobility Networks

In our study on mobility in Spain we built **weekly mobility networks**, G_T, in the same way as in [15]. These are constructed using nodes that represent the $n = 52$ locations (Spanish provinces, see Section 3.2) and as edges, the average daily flow during the week T. Therefore, it is a directed and weighted network. In our case, the average flow between each pair of locations (nodes) is not the same in both directions. In other words, there is not the same average flow of people who travel from Madrid to Barcelona as they do from Barcelona to Madrid. On the other hand, our network is weighted since the strength of the interactions is not always the same. Edge weights, w_{ji}, represent the average daily flow along the edge during the week considered, T.

When building G_T, the weekly mobility network of calendar week T, we used all the trips that have been made between the locations throughout the week. We call $\mathcal{D}_T$ the set of days in calendar week T and calculate the edge weights using the daily mobility matrix of the dataset $F(t)$ (see Section 3.2). Therefore, to obtain the average weekly flow of the edge $i \to j$, w_{ji}, we summed the daily flow of people who have traveled from i to j each day of the week and we average the number of days that the week contains [15]:

$$w_{ji} = \frac{1}{|\mathcal{D}_T|} \sum_{t' \in \mathcal{D}_T} F_{ji}(t'), \tag{1}$$

where $|\mathcal{D}_T|$ denotes the number of days in calendar week T. In this way, we obtain a network whose nodes are the same as for daily mobility networks, that is, the $n = 52$ locations, but whose edge weights represent the average daily number of trips between each possible pair of locations ji.

3.3.2 Rescaled Weekly Mobility Networks

When studying how the global reduction in mobility affects the networks, compared to structural changes in the network, we built **weekly rescaled mobility networks** $G^R_{T_0}(T)$. To do this, we use the structure of the weekly mobility network of a given week T_0, chosen as the baseline, and rescale its flows, $w_{ji}(T_0)$, by to those of calendar week T. Specifically, we scale the flows of our baseline network $w_{ji}(T_0)$ by a scale factor that is given by the total flow during calendar week T, divided by the total flow during the baseline week T_0:

$$w^R_{ji}(T) = w_{ji}(T_0) \times \frac{\sum_{i,j=1}^{n} w_{ji}(T)}{\sum_{i,j=1}^{n} w_{ji}(T_0)}. \tag{2}$$

With this rescaling we obtain a network that is structurally similar to the baseline weekly mobility network but has the same total amount of trips as the corresponding weekly mobility network of calendar week T. Consequently, we are able to isolate the effects of the global reduction in mobility.

3.3.3 Thresholding procedure

Now we explain the filtering that has been performed in Spain's weekly mobility networks. In the original work [15], for Germany, they impose a threshold of 5 trips and all edges whose weight is less than this amount are omitted. This filtering is carried out to be consistent with the dataset where there are no total flows between the locations below this threshold, for privacy reasons. For our study in Spain, we decided to impose a threshold of 200 trips in the weekly mobility networks and in the rescaled ones. To understand this choice, different from that of Germany, we must take into account how the data has been aggregated. In the original work the locations correspond to the 401 counties of Germany. Each county (in 2020) had an average population of about 2×10^5 people. In contrast, in our study for Spain, we aggregate the data at the level of the 52 Spanish provinces. Each province had an average population of 9×10^5 inhabitants, slightly more than 4 times that of a county.

Filtering flows of less than 5 trips in Germany allows the authors to study structural changes in mobility networks. But in Spain, we observe that it is not enough to properly study these changes since we have networks that are practically fully connected for the different calendar weeks. Therefore, we chose to impose a threshold of w_{min}=200 trips for our weekly mobility networks and flows below this value are not considered, which removes redundant trips. We have tested different thresholds and found that the results do not change substantially.

3.4 Measures and Metrics

In this Section, we explain how the different observables studied in this Master's Thesis have been obtained.

As we advanced in Section 3.1, the mobility dataset used in this study include, for each pair of locations, the geographical distance between them. In the case of Germany, the distance between the centroids is given for each pair of counties. In the case of Spain, it appears in the dataset to which distance range each length of the flow considered belongs. For our study, we need both formats.

In the first place, to study mobility trends in Germany and Spain, in addition to the change in the daily mobility of the total flow, we study how it changes depending on the distance traveled. To do this, we group the daily mobility flows in Germany by distance ranges summarized in Table 3.1.

D	Range (km)
D_1	$d \leq 10$
D_2	$10 < d \leq 50$
D_3	$50 < d \leq 200$
D_4	$200 < d \leq 600$
D_5	$600 < d$

Table 3.1: Distance ranges considered for the study of mobility trends in Germany.

Finally, for different measures explained, we need to know the geographical distance between each pair of locations. To do this, we use dataset [42] that provides the coordinates of the centroids of each province. We determine the distance between both locations from these coordinates, as follows.

To calculate the distance between locations i and j, we denote lat_i, lon_i to the latitude and longitude of centroid i and lon_j, lat_j to those of j. Therefore the differences between latitudes and longitudes are given by:

$$\Delta_{lat} = lat_j - lat_i. \tag{3}$$
$$\Delta_{lon} = lon_j - lon_i. \tag{4}$$

In this way, we can obtain the geographical distance in kilometers as in Equation 3 from [45]:

$$d_{ij} = 2r \sin^{-1}\left(\sqrt{\sin\left(\frac{\Delta_{lat}}{2}\right)^2 + \cos\left(lat_i\right) \times \cos\left(lat_j\right) \times \sin\left(\frac{\Delta_{lon}}{2}\right)^2}\right), \tag{5}$$

where r is the mean radius of the earth that we have considered as r = 6,367km . It has to be taken into account that angles need to be in radians.

Once again, to study how mobility trends changes over time depending on the distance traveled, we group the daily mobility flows in Spain by distance ranges, D, shown in Table 3.2.

D	**Range (km)**
D_1	$0.5 < d \leq 2$
D_2	$2 < d \leq 5$
D_3	$10 < d \leq 50$
D_4	$50 < d \leq 100$
D_5	$100 < d$

Table 3.2: Distance ranges considered for the study of mobility trends in Spain.

The difference between the distance ranges considered, for both countries, lies in the scale at which the data is available. In Spain, the data is provided at the municipal level, a smaller scale than the counties in Germany. This enabled us to conduct a more detailed study of short-haul routes. As for the long ones, for greater clarity, all trips whose distance traveled is greater than 100km are grouped together, since we consider that these trips are, most likely, not made in a single day. In Germany, the distance ranges chosen by the authors of [15] are respected. Even so, and for comparison reasons, in both countries, we have considered long-distance travels those whose distance travelled is greater than 10km.

3.4.1 Measures to study Mobility Trends

We now explain the different measures that are made in order to study mobility trends both in Germany and Spain. We base our analysis on daily mobility flows $F(t)$ and focus on daily mobility changes $\Delta n(t)$.

- **Daily mobility change:** We compare the mobility of a certain date in the study period $t \in \mathcal{T}$, with the mobility in a reference period $\mathcal{T}_0$, before the mobility restrictions are imposed.

Therefore, for a certain t we calculate the mobility change $\Delta n(t)$ by comparing the total flow of t, $N(t)$, with the expected one, $N_0(t)$ as in [15]:

$$\Delta n(t) = \left(\frac{N(t)}{N_0(t)}\right) - 1. \tag{6}$$

Given that mobility is dependent on the day of the week, the expected flow with which we compare is the total during the reference period corresponding to the same day or period of the week. In the case of Germany, this is an average flow since the reference period covers one month. On the other hand, in the case of Spain, it corresponds to the total flow for each day or days of the corresponding week in the reference period.

We also study the change in mobility for a single location (provinces) in Spain, $\Delta n^{(i)}(t)$. To carry this out, we calculate, in a analogous way to that used in Germany for the counties [15], the change in the mobility of a province. To do this, we use the same procedure but only take into account the flow of trips originating in the location i:

$$N^{(i)}(t) = \sum_{j=1}^{n} F_{ji}(t). \tag{7}$$

Where we are summing flows originated in i over all possible destinations. We consider different calendar weeks as t and hence $F(t)$ represents the average daily flow during the considered week along each connection. This weekly mobility networks are obtained as shown in Section 3.3.

- **Daily mobility change per distance:** To understand whether mobility has been reduced in a homogeneous way, we study the dependence on the distance traveled in the reduction of mobility as in [15]. We calculate the distance-dependent mobility change $\Delta n_D(t)$. We proceed in the same way as for the daily mobility change, but considering only the trips included in range $D = \{d : d_{min} < d \leq d_{max}\}$. Therefore, for a given t and a certain range D, the distance-dependent mobility change is obtained as:

$$N_D(t) = \sum_{(i,j)\in\Phi_D} F_{ji}(t), \tag{8}$$

where Φ_D is the set of pairs of locations $(i.j)$ whose distance, d_{ji}, is included in the interval D, $\Phi_D = \{(i,j) : d_{ji} \in D\}$.

- **M-days centered moving average:** In addition to calculating the daily mobility change, given that mobility shows strongly dependence on the weekday, we isolate the fluctuations within the week by calculating the weekly centered moving average. For M odd, the M-day centered moving average of a given date t is the unweighted arithmetic mean of the $(M-1)/2$ days before and after t, including also t in the average:

$$N_M(t) = \langle F(t)\rangle_M = \frac{1}{M} \sum_{t'=-\frac{M-1}{2}}^{t'=\frac{M-1}{2}} F(t+t'), \tag{9}$$

in our case we choose $M = 7$. In this way, we isolate the fluctuations due to the weekly internal variation and we are able to study its general trend.

- **Pearson correlation coefficient:** With the aim of assessing the effect of large-scale social interaction in the spreading of COVID-19 in Spain, we study the correlation between mobility patterns in different distance ranges, based on the distance traveled, and the number of

infected individuals registered in the county. This study, carried out as a first approach to the problem, is inspired by the work carried out in the analysis of mobility patterns and the transmission of the virus in the USA by [46]. To do this, we calculated the Pearson correlation coefficient between both magnitudes as:

$$\rho_{X,Y} = \frac{cov\,(X,Y)}{\sigma_X \sigma_Y}, \tag{10}$$

where cov is the covariance and σ the standard deviation [47]. It is taken into account that, for different reasons, such as incubation time, the need for large-scale infectious encounters, etc., an increase in mobility flows leads to an increase in the number of cases with a certain offset. Therefore, we calculate the correlation coefficient for different time lags.

3.4.2 Measures to study structural changes in the Mobility Networks

In this Section we explain the different measurements that we have carried out to characterize the topology of the weekly mobility networks. Among others, we have studied the degree and strength distributions, path lengths, clustering coefficients and other magnitudes that allow us to study the Small-World Effect that we would introduce later on, as well as the capacity of the network to support diffusion.

Firstly, we focused on the identification of the most **important nodes**. Of course, there are different definitions of importance and for each one its corresponding magnitude [26, 48]. These measures are formalized in network theory and are known as **centrality measures** [48, 49].

- **Degree:** A key magnitude, and perhaps more intuitive, is the **degree** or connectivity of a node [26, 48, 50]. This magnitude quantifies the number of links that the node has. The main idea behind is that a node is more central or important than another if it has a greater degree than the second [49]. We call the degree of the node i of the network, k_i[49, 50].

  In the case of undirected and unweighted networks, the total number of links, L, is given by:

$$L = \frac{1}{2}\sum_{i=1}^{n} k_i, \tag{11}$$

  where the factor $1/2$ is included since each link is counted twice in the sum and n is the total number of nodes in the network. Once we know the degree of the nodes in our network, then we can calculate another important property, which is the **average degree** of the network [50]. Again, in the case of a undirected network, this is given by:

$$\langle k \rangle = \frac{1}{n}\sum_{i=1}^{n} k_i = 2\frac{L}{n}. \tag{12}$$

  In our case, our mobility networks are directed and weighted. We firstly take into account the directionality of the network. In directed networks, each node has two degrees, k_i^{in} and k_i^{out} the **incoming** and **outgoing** degree [48, 50]. Directed links, $i \to j$, have a source, i, the node from which they leave, and a target, j which is the destination node. Then k_i^{in} counts the number of links that target the node i and k_i^{out} the number of links that leave i. We can then define the total degree of node i as the sum of these two quantities:

$$k_i = k_i^{in} + k_i^{out}. \tag{13}$$

Analogously to the case of undirected networks, we can obtain the total number of links as:

$$L = \sum_{i=1}^{n} k_i^{in} = \sum_{i=1}^{n} k_i^{out}, \tag{14}$$

where in this case it is not necessary to include the factor $1/2$ since each link is only counted once, either as an in-coming or as an out-coming link. Again, we get the average degree, in this case as:

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i^{in} = \frac{1}{n} \sum_{i=1}^{n} k_i^{out} = \frac{L}{n}. \tag{15}$$

Thanks to the degree we can have an idea of the immediate effects that take place in the network. In the case at hand, the spread of epidemics, the spread to other nodes, to a first approximation, will be proportional to the degree of each one [49].

Another important measure is the **degree distribution**, p_k. This measure tells us the probability that a randomly selected node has degree k [49, 50, 51]. Since it is a probability, it must be normalized $\sum_{k=0}^{\infty} p_k = 1$. For a given network of n nodes, p_k is the normalized histogram, so that we have: $p_k = n_k/n$, where n_k represents the number of nodes with degree k. In this way we can calculate the average degree as:

$$\langle k \rangle = \sum_{k=0}^{\infty} k p_k. \tag{16}$$

In the last 30 years is has been discovered that, actually, in most real networks the degree follows a heavy-tailed distribution, in most cases a power-law one, where nodes with highly different degrees coexist [50, 51]. This property has, among others, effects on the spread of epidemics, for example by decreasing the epidemic threshold or affecting the speed with which the epidemic spreads. Due to this heterogeneity, knowing the average degree is not always enough information [52]. In these cases, it is better to study the degree distribution.

- **Strength:** As we mentioned before, until now we have taken into account the directionality of our links but not their weight. To account for this, we calculate the **strength** in addition to the degree [50, 53]. In our case, we remember that the weight of an edge corresponds, for the weekly mobility networks, to the average daily flow along the connection. To calculate this quantity, we do it in a similar way to how we have calculated the degree of each node, but instead of counting the number of links, we weigh each link with its corresponding flow. In terms of the adjacency matrix[7], we obtained the in- and out-going degree of node i as:

$$k_i^{in} = \sum_{j=1}^{n} a_{ij}, \qquad k_i^{out} = \sum_{j=1}^{n} a_{ji}, \tag{17}$$

So, we get the strength of node i, s_i, as:

$$s_i^{in} = \sum_{j=1}^{n} a_{ij} w_{ij}, \qquad s_i^{out} = \sum_{j=1}^{n} a_{ji} w_{ji}. \tag{18}$$

In the same way, we calculate $\langle s \rangle$, the **average strength**, and, p_s, the **strength distribution**.

[7] The adjacency matrix is a mathematical representation of a network. For a directed network with n nodes it has dimension $n \times n$. The element a_{ji} is 1 if there is a link pointing from i to j and zero otherwise [50].

■ **Shortest path lengths:** Physical systems are characterized by physical distances in an obvious way. But, in the case of networks, this relationship is not so immediate since the physical distance ceases to be relevant in many cases. In networks, the physical distance is therefore replaced by the path length[8] [50].

Important properties of the network can be inferred from the **average shortest path length**. The shortest path between two nodes (i, j) tells us what is the minimum number of edges that we must be traveled to get from i to j or vice versa. This quantity, denoted as $d(i, j)$, is known as the distance between nodes [49]. If there is not a path that connects both nodes, the distance is set to infinity [49, 51, 54]: $d(i, j) := \infty$. Since we want to calculate the average over all possible pairs of nodes in the network, in order to have finite distances, the graph or subgraph considered must form a single connected component[9].

If we extend this measure to the case of directed networks [49], the procedure is analogous, but in this case, we may find that the distance from i to j is not the same as the distance to go from j to i, so, in general, $\overrightarrow{d}(i, j) \neq \overrightarrow{d}(j, i)$. Or even, it may be the case that there is a path from i to j and not from j to i.

If we know the path length between all possible pairs of nodes, we can average over all nodes to obtain the average shortest path length for directed or undirected networks [49, 50]:

$$\langle d \rangle = \frac{1}{n(n-1)} \sum_{i,j=1}^{n} \overrightarrow{d}(i, j). \tag{19}$$

Generalizing this measure to the case of weighted networks is not so straightforward [49, 51, 54]. In some cases, the weight of the edges can be interpreted as the length of the edge. These lengths can correspond to euclidean distances, such as the geographical distance between two points. In our case, we make the same choice as in [15], which corresponds to a usual choice, in which larger weights are related to smaller distances. Hence, edge weights, w_{ji}, represent the flow or the average flow along the connection during a time t. We define, therefore, the length of the edge as the inverse of its corresponding flow, $L_{ji} = 1/w_{ji}$. Using this metric, we can calculate the shortest path length (SPL) between two nodes using Dijkstra's algorithm [55].

By repeating the algorithm for all the different possible sources in our graph, we obtain the average SPL as the average over all possible sources of the average shortest path length of each source:

$$L = \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j \neq i} L_{ji}. \tag{20}$$

We are also interested in studying how the average SPL varies with geographic distance. In [15] they calculate the expected shortest path length $L_d(T)$ at a distance d. Inspired by this study we calculate the average shortest path length at a distance range d as follows. First, we must choose the intervals in which we are going to discretize the range of distances between provinces in Spain. Using the method explained at the beginning of Section 3.4 we obtain an upper bounding for the distances r_{max}, given by the maximum geographical distances that separates the centroid of two provinces. Our goal is to build an array R whose first component corresponds to $0km$ and it's last to one r_{max}. The remaining components are selected such in the interval of distances given by two consecutive components of R there are a significant number of pairs of provinces whose distance belongs to. Hence for every two consecutive components of R the interval of distances is then given by $d = (R_x, R_{x+1}]$. For a given interval

[8] A path is defined as a route within the network and its length corresponds to the number of edges it contains [50].
[9] Connected component: set of vertices in a graph connected to each other by paths.

d we calculate the average shortest path including every pair of location whose geographical distance $r(j,i)$ falls in the interval as:

$$L_d(T;r) = \frac{1}{n_d(n_d-1)} \sum_{i=1}^{n_d} \sum_{j \neq i, r(i,j) \in d} L_{ji}, \tag{21}$$

where L_{ji} is obtained as the shortest path between j and i if the distance between centroids $r(j,i)$ falls in the interval $(R_x, R_{x+1}]$, regardless of the shortest path distance, being n_d the number of locations that fulfill the distance constraint.

Additionally, for studying structural changes in the mobility networks we study the **Small-World Effect**. Considering the definition that was previously given for the distance in a network one can define the mean shortest distance from each node i to every other node in the network [56]. From this quantity one can define the mean distance between all nodes in the network by averaging over all the possible pairs of nodes ji. To to have finite values for this quantity one sometimes needs to define $\mathcal{C}_m$, the components of a network, where $m = 1, 2, ...$ and hence we average over paths within components:

$$L = \frac{\sum_m \sum_{ij \in \mathcal{C}_m} L_{ij}}{\sum_m n_m^2}, \tag{22}$$

where n_m is the number of nodes that belong to the component $\mathcal{C}_m$. This measure can be extended for directed networks by taking into account that the distance from i to j may not be the same as the distance from j to i.

The Small-World Effect is, mathematically, the hypothesis that the average path length scales logarithmically with the size of the network. This effect can have important implications in the system. For instance, in the spreading of a disease, different nodes or locations will get infected faster as the number of steps from an infected site are lower.

- **Relaxation and mixing time:** There are some additional measurements that we can perform to understand more in depth the structural changes in the network according to the phenomenology of random walks. The propagator of an edge-centered random walk is defined by the non-normalized Laplacian given by:

$$\mathcal{L}_{ji} = \delta_{ji} \sum_k \omega_{ki} - \omega_{ji}. \tag{23}$$

The Perron Frobenius theorem guarantees us that the eigenvalues of the Laplacian fulfill $0 \leq \lambda_1 < \lambda_2 \leq ... \leq \lambda_n$ for the case of directed networks formed by a single component [57]. When our network has more than a single component, we will have a zero eigenvalue for each of them. The smallest non-zero eigenvalue defines the **relaxation time**. If we call λ_{rlx} to this first non-zero eigenvalue, the relaxation time is then defined as $t_{rlx} = 1/\lambda_{rlx}$ (*SI*[10] *from* [15] and [58]). This magnitude quantitatively measures the time scale in which a random walk approaches the uniform distribution at equilibrium.

For node-centered random walks, the propagator is analogously defined with the transition matrix:

$$p_{ji} = \frac{w_{ji}}{\sum_k w_{ki}}. \tag{24}$$

[10] SI: supplementary Information

The operator p_{ji} is a stochastic matrix whose largest eigenvalue, for a single component, is $\pi_m = 1$. If we call π_{mix} to the second greater eigenvalue of p different than one, the **mixing time** is obtained as: $t_{mix} = 1/(1 - \pi_{mix})$. Similarly to relaxation time, the mixing time quantifies the time scale on which a node-center random walk approaches the non-uniform equilibrium (*SI from* [15] and [57]).

- **Clustering coefficient:** Another relevant property that distinguishes real networks from random networks is the tendency to form links between neighboring nodes [59]. To study this trend, we measure the **clustering coefficient** of the nodes. This measures the probability that two neighbors of a node are neighbors to each other [48, 50, 51]. It is therefore a measure of the local density of the network.

  For an undirected and unweighted network, the local clustering coefficient of a node i with degree k_i is obtained as [48, 60]:

$$C_i = \frac{2T(i)}{k_i(k_i - 1)}, \tag{25}$$

  where $T(i)$ is the number of triangles involving node i. Therefore, if we have $C_i = 0$, this implies that none of the neighbors of node i link to each other. On the contrary, if we have $C_i = 1$, all the neighbors of node i are connected to each other.

  Following an analogous procedure as for the degree, we can average over all the nodes in the network:

$$\langle C \rangle = \frac{1}{n} \sum_{i=1}^{n} C_i. \tag{26}$$

  In this way, we obtain the probability that two randomly selected nodes in the network have a link between them [50]. We can also write this relation in terms of the adjacency matrix as [59]:

$$C_i(A) = \frac{\frac{1}{2} \sum_{j \neq i} \sum_{h \neq (i,j)} a_{ij} a_{ih} a_{jh}}{\frac{1}{2} k_i(k_i - 1)} = \frac{\left(A^3\right)_{ii}}{k_i(k_i - 1)}, \tag{27}$$

  where (and from now on) the sum is over $j \neq i$ and $h \neq (i, j)$.

  In order to be able to measure this magnitude in our mobility networks, we need a generalization of this measure to the case of directed and weighted networks. To account for edge weights, the local clustering coefficient must take into account how much weight is present in the neighborhood of the vertex, compared to a limiting case [59]. There are different proposals, but in our case, we will use Equation 9 from [61] where it is defined as the geometric average of the subgraphs edge weights [60]. According to this definition, the clustering coefficient of node i is obtained as:

$$\tilde{C}_i(\tilde{W}) = \frac{1}{k_i(k_i - 1)} \sum_{j,h} \left(\tilde{w}_{ij} \tilde{w}_{jh} \tilde{w}_{hj}\right)^{\frac{1}{3}} = \frac{\left(\tilde{W}^{[\frac{1}{3}]}\right)^3_{ii}}{k_i(k_i - 1)}, \tag{28}$$

  where flows are normalized by the maximum flow of the network, $\tilde{w}_{ij} = w_{ij} / \max(w_{ij})$ [60] and $W^{[\frac{1}{k}]} = \{w_{ij}^{\frac{1}{k}}\}$ [59]. If we define I_i as the normalized average of the intensities of the triangles containing node i:

$$I_i(\tilde{W}) = \frac{1}{2T(i)} \left(\tilde{W}^{[\frac{1}{3}]}\right)^3_{ii}, \tag{29}$$

we can rewrite Equation 28 as [60]:

$$\tilde{C}_i(\tilde{W}) = C_i I_i. \tag{30}$$

According to this definition, those triangles whose edges are equal to the maximum will contribute $I_i = 1$. Also, those triangles with one link whose weight is negligible, will have a negligible contribution to the sum as well [60].

It is important to note that in Equations 25, 27 28 and 30, if the degree of the considered node is <2, a clustering coefficient of 0 is assigned. In addition, up to now and in the future, self-edges or loops are not considered [60].

Finally, this result can be generalized for directed-weighted networks [59]. For this, it is considered that given 3 nodes, in a directed network, the maximum number of triangles increases. Up to 8 triangles can be formed since we have two possible options for each pair of nodes *SI from* [15] and [59].

In the case of directed networks, if we consider a node i, only those edges that have i as source or target can form triangles. To do this, we use the definition given by Equation 10 of [59], which corresponds to the one used by [15], formulated by generalizing the result of Equation 27:

$$\tilde{C}_i^D(\tilde{W}) = \frac{\frac{1}{2}\sum_{j,h}(\tilde{w}_{ij} + \tilde{w}_{ji}) + (\tilde{w}_{ih} + w_{hi}) + (\tilde{w}_{jh} + w_{hj})}{[k_i^{tot}(k_i^{tot} - 1) - 2k_i^{\leftrightarrow}]} = \frac{\left[\tilde{W}^{[\frac{1}{3}]} + \left(\tilde{W}^{\intercal}\right)^{[\frac{1}{3}]}\right]^3_{ii}}{2[k_i^{tot}(k_i^{tot} - 1) - 2k_i^{\leftrightarrow}]}, \tag{31}$$

where k_i^{tot} is the sum of the incoming and outgoing degree of node i, and $k_i^{\leftrightarrow}$ is its reciprocal degree[11] of i. The additional term in the numerator is included, since any product of the form $\tilde{w}_{ij}\tilde{w}_{jh}\tilde{w}_{hj}$ only accounts for one of the 8 possible triangles that can be formed due to the directionality of the network. Furthermore, the term that corresponds to the reciprocal degree in the denominator is included to avoid counting false triangles formed by a pair of directional edges pointing towards the same node. There are $k_i^{\leftrightarrow}$ of these occurrences and for each one we have counted two false triangles. Hence, the denominator corresponds to the total number of possible triangles that node i can form [59].

In this way, the weighted clustering coefficient for a directed network quantifies the probability of the triangles that may exist, as well as its average intensity, taking into account the incoming and outgoing flow of the given focal node. Again, this local clustering coefficient can be averaged in an analogous manner to that introduced above for an unweighted and undirected network clustering in Equation 26.

■ **Communities:** Moreover, regarding the internal structure and organization of the network, we study how the meso-scale of the network evolves over the course of the pandemic. The detection of **communities** shows the main structural characteristics of the networks, providing us with information about its internal organization. However, there is no single definition of a community. We can understand a community as a set of nodes that are structurally more

[11] Provided that we do not account for self-interactions, the reciprocal degree of node i is defined, in terms of the adjacency matrix, as: $k_i^{\leftrightarrow} = \sum_{j \neq i} a_{ij}a_{ji}$ [59].

connected within each other than with the rest of the network.

The problem of finding cohesive groups of nodes in networks is known as community detection. It is one of the most active areas in Network Science nowadays. Unfortunately, community detection is in fact a challenging task, but many methods have been recently developed that obtains good results for practical situations. In the present work we use OSLOM[12] [62], implemented in a freely available software that can be accessed at http://www.oslom.org.

3.5 Epidemic spreading model

In an attempt to answer the question of to what extent containment measures help mitigate the spread of epidemics, the authors of [15] developed and implemented a SIR metapopulation model that includes containment measures. The possible applications of this model go beyond COVID-19, thus it can be applied to similar future scenarios. In our case, we employed the model, whose code is publicly available, to carry out a similar study to that carried out in Germany, in Spain.

Before presenting the details of this model, we will explain the basics of the SIR model on which our model is based.

3.5.1 SIR Model

The classical SIR model or Kermack-McKendrick model is a compartmental model where the population under study is divided in three compartments [25]. In particular, the total population, considered constant N, is divided into three different classes: **Susceptible** (S), who can be infected by the disease, **Infected** (I), the population that are infected and can spread the disease, and **Removed** or **Recovered** (R), those who for different reasons (being immune, being isolated, having passed the disease and having recovered, having died, etc.) do not belong to either (S) or (I) [63]. If we call $S(t)$, $I(t)$ and $R(t)$ the number of individuals in each compartment as a function of time t, the total population of our system is given by [64]:

$$N = S(t) + I(t) + R(t). \tag{32}$$

The model makes different assumptions regarding the transmission and infection processes can be summarized as follows [65, 66]:

1. An average individual belonging to the population makes enough contact to be able to transmit the infection with other with βN individuals per unit of time, where β is >0.
2. Each infected individual remains in this state for a time equal to $1/\gamma$, where γ is a constant and positive parameter. Therefore, the infected leave the infected compartment at a rate γI.
3. A susceptible individual who has contact with an infected one is considered immediately infected. That is, the incubation period is short enough to be negligible.
4. Since the total population, N, of the system is considered constant, the only way to go from one class to another is through the processes described above. Births and deaths are not taken into account which is only valid as long as the temporal scales of the disease are much longer than the demographic ones.

[12] OSLOM: Order Statistics Local Optimization Method.

Taking into account 1, and under the assumption that the size of the different compartments is large enough so that the mixing between the different classes is homogeneous (well mixed assumption), we obtain that the number of new infected per unit of time and per infected is $(\beta N)\,(S/N)$ [66]. This result follows from the fact that the probability that a random encounter between an infected individual and a susceptible one is S/N, resulting in the ratio of new infections being given by $(\beta N)\,(S/N)\,I = \beta SI$. Hence the SIR model is characterized by the following set of ordinary differential equations [64]:

$$\begin{cases} \dot{S} &= -\beta SI. \\ \dot{I} &= \beta SI - \gamma I. \\ \dot{R} &= \gamma I. \end{cases} \tag{33}$$

Where the initial conditions are given by $S(0) = S_0$, $I(0) = I_0$ and $R(0) = R_0$ with S_0, I_0, R_0>0 [65]. Generally, this set of equations cannot be solved analytically, but we can still get a lot of information by following an approximate approach. For this, it is taken into account that $S(t)$, $I(t)$ and $R(t)$ are positive functions bounded by N since they represent populations [66]. Therefore, from the equation for the susceptible dynamics we can see that $\dot{S} < 0,\ \forall t$ and hence $S(t)$ is a monotonically decreasing positive function [64]:

$$\lim_{t\to\infty} S(t) = S_\infty \text{ with } S_\infty \in (0, S_0]. \tag{34}$$

Following an analogous reasoning for the recovered ones, we see that $\dot{S} < 0,\ \forall t$, we see that $R(t)$ is a positive and monotonically increasing function [64]. The number of removed individuals is bounded by the total population size, hence:

$$\lim_{t\to\infty} R(t) = R_\infty \text{ with } R_\infty \in [R_0, N]. \tag{35}$$

Regarding the behavior of the infected individuals, the reasoning is not so straightforward [65]. If we rewrite the equation of $\dot{I}$ as $\dot{I} = I\,(\beta S - \gamma)$, we can see that $I(t)$ can have two different behaviors, depending on the sign of the factor that multiplies it. It can be increasing until reaching a maximum and then decreasing or monotonically decreasing. Therefore, the dynamics is given by the initial condition of $S(t)$:

$$\left.\frac{dI}{dt}\right|_{t=0} = I_o\,(\beta S_0 - \gamma) \longrightarrow \begin{cases} \left.\frac{dI}{dt}\right|_{t=0} & > 0 \text{ if } S_0 > \frac{\gamma}{\beta}. \\ \left.\frac{dI}{dt}\right|_{t=0} & < 0 \text{ if } S_0 < \frac{\gamma}{\beta}\gamma I. \end{cases} \tag{36}$$

From Equation 36, we can see that the SIR model gives rise to a threshold phenomenon [65]. If $S > S_c = \gamma/\beta$ an epidemic will occur, while if $S < S_c = \gamma/\beta$ it won't. This threshold defines the **basic reproduction number**:

$$\mathcal{R}_0 = \frac{\beta S_0}{\gamma}. \tag{37}$$

Thus, if $\mathcal{R}_0 > 1$ the infection spreads and an epidemic will occur, while if $\mathcal{R}_0 < 1$ the infection dies out and there will be no epidemic. Epidemiologically, $\mathcal{R}_0$ quantifies the number of secondary infections arising from a given individual in a population composed solely of susceptible individuals [64]. This interpretation is based on considering a single finished subject, $I = 1$, then the number of new cases of infection per unit of time produced by that individual is βS. If we only have susceptible individuals in our population, this quantity is βS_0. Since one infective individual remains infectious $1/\gamma$, the number of secondary cases that will occur as a consequence of this first event is $\beta S_0/\gamma$, which corresponds to the expression of $\mathcal{R}_0$.

In addition, there are multiple variations of this model that can be used depending on the characteristics of the problem considered [66]. Some examples are the SIS model, which describes a disease without immunity against reinfection, or the SEIR or SEIS models, which considers an exposure time between infection and the infective state.

3.5.2 SIR Metapopulation Model

There are numerous extensions of the compartmental models presented above in order to understand and study the spread of a disease across a large geographic region [67]. This is the case of the COVID-19 pandemic, but also of other recent scenarios such as Ebola or Zika Fever, in which long-distance travels play a crucial role in driving the spread of the epidemic [15, 67]. In order to carry out a real-world study, it is necessary to couple the actual mobility data with the epidemiological model. It is therefore necessary to choose the movement mechanism to be used. To address this problem, we use the so-called **metapopulation models**. These models describe interacting subpopulations that are spatially structured in some level. Hence, each subpopulation i occupies a location and individuals within each location are divided into the usual compartments that denote their status with respect to the modeled disease [52]. For the SIR model, for each location one has:

$$N_i = S_i + I_i + R_i. \tag{38}$$

As a consequence, the total system on n population is constant:

$$N = \sum_{i=1}^{n} N_i. \tag{39}$$

In our case, these populations correspond to the census of each province. Disease transmission is considered local, but infected individuals based in one place can cause infections in other places with a rate that is proportional to the flow between locations [15]. This flow being, in our case, the average daily flow of a given calendar week.

Two of the most common choices for the mechanistic representation of motion are the random diffusion motion model and the recurrent recurrent mobility model [67]. The first specifies the rates at which the inhabitants of a town visit another town. Also, individual behavior is not tracked. That is, once an individual travels, the system forgets their previous location. In the model developed in [15] they choose to use the recurrent mobility model where individual behavior is not tracked. This category includes those models where the frequency with which a host travels away from its location before returning is specified. To do this, a home location is assigned to each host and it is described how it temporarily travels to other locations before returning to the origin. Therefore, visitors can interact with, and thus infect, residents of other locations during travel. In addition, these travelers will return to their place of origin after each visit.

3.5.3 SIR Metapopulation Model with Containment

Finally, we introduce the model used in our study. As we advanced, this metapopulation model with containment was developed and implemented in [15].

The model, whose code is freely available in [68], is an extension of a well-known commuter-dynamics SIR metapopulation model presented in [69]. In particular, it accounts for a drastic reduction in the total mobility, a substantial part of the change in mobility patterns due to the different containment measures. They consider two possible scenarios that can lead to the reduction

in mobility: the **distancing** scenario, and a stricter one, which is **isolation**.

The system is divided into n subpopulations, in our case, $n = 52$ provinces of Spain. Each subpopulation has a population given by $N_i^{(pop)}$, that we set to the census, and is divided into three compartments: Susceptibles (S), Infected (I) and Recovered or Removed (R).

For a given time t we have the flow matrix $F(t)$, where the element $F_{ji}(t)$ quantifies the total number of trips from location i to location j. In particular, our flow matrices are the average weekly mobility networks for the different calendar weeks, where each element represents the average daily number of trips between the two locations.

We now assume that each individual has a home location and a work location they commute to (it can also be the same location) every day. With this assumption, we can identify the columns of our flow matrix as the the living places and the rows as the working places. From this flow matrix F we construct our probability matrix p by normalizing this fluxes. If F_{ji} is the flux of people that leave in i and work in j, then p_{ji} is the probability that the people that live in i go to work at j:

$$p_{ji} = \frac{F_{ji}}{\sum_j F_{ji}}. \tag{40}$$

That is, p_{ji} is given by the flux of people that live in i and work in j divided by the flux of all possible locations where people that live in i can go to work. Then:

$$\sum_j p_{ji} = 1, \tag{41}$$

this is a simplified mechanistic approach where a Markovian assumption is made. At each time step the movement of a given individual is obtained according to a probability matrix. Hence, individuals are not labeled, and at every time step the same traveling probability applies to every individuals in the location [52]. From this matrix we can obtain the number of individuals that live in i and work in j by multiplying the census of i by the probability of living in i and working in j:

$$N_{ji}^{(pop)} = p_{ji} N_i^{(pop)}. \tag{42}$$

Again, the total population of i is then given the population of individuals that live in i and work in j summed over all possible work locations j:

$$N_i^{(pop)} = \sum_j N_{ji}^{(pop)}. \tag{43}$$

We now address the infection dynamics. Initially all individuals are set to be susceptible $S_{ji} = N_{ji}^{(pop)}$, except for an infection origin province i^*, where we distribute I_0 infected individuals. If this infected origin province we have n compartments, corresponding to the different locations where people can go to work, we distribute this initial number of infected individuals, in different compartments of this province, proportionally to the population that lives in province i^* and works in province j:

$$I_{ji^*} = \frac{N_{ji^*}}{N_{i^*}} I(t=0). \tag{44}$$

Then at $t = 0$ we have that the susceptibles are:

$$S_{ji}(t=0) = \begin{cases} N_{ji}^{(pop)} & \text{if } i \neq i^*. \\ N_{ji^*}^{(pop)} - I_{ji^*} & \text{if } i = i^*. \end{cases} \tag{45}$$

Two possible ways in which a susceptible of compartment S_{ji} can get infected are considered:

1. While being at home in location i and getting infected by the other people that live in i and work at any other location k.
2. While working in j and getting infected by people that work in j and live at any other location k.

We assume that people spend half of the time at home and half of the time in the location where they work. Then the dynamics equations for the susceptibles are:

$$\frac{dS_{ji}}{dt} = -S_{ji}\lambda_{ji} = -S_{ji}\left(\lambda_i^{home} + \lambda_j^{work}\right) = -S_{ij}\left(\frac{\beta}{2}\frac{\sum_k^n I_{ki}}{\sum_k^n N_{ki}^{(pop)}} + \frac{\beta}{2}\frac{\sum_k^n I_{jk}}{\sum_k^n N_{jk}^{(pop)}}\right). \tag{46}$$

The evolution of the susceptible that live in i and work in j is a decreasing function (we do not add susceptible to the population) proportional to the number of susceptible that again live in i and work in j multiplied by the rates of infection. As we have discussed, there are two possible ways of getting infected, one at home and one at work. The rate of infection while being at home is $\beta/2$, since we spend half of the time at home, multiplied by the density of infected that live in i. This density is given by the number of infected people that live in i (and work anywhere else), divided by the total population that lives in i (and work anywhere else). In the same way we obtain the rate at which we get infected at work.

The dynamic for the Infected is obtained analogously with an additional recovery term:

$$\frac{dI_{ji}}{dt} = -\mu I_{ji} + S_{ji}\left(\lambda_i^{home} + \lambda_j^{work}\right), \tag{47}$$

where μ is the inverse of the recovery time. Since the population size is constant, the third equation is:

$$\frac{dR_{ji}}{dt} = \mu I_{ji}. \tag{48}$$

Now we address what concerns the stochastic simulation. The algorithm used is a stochastic binomial sampling algorithm. We have that the probability that an individual in compartment S_{ji} becomes infected in the interval $[t, t+\Delta t]$ is:

$$P(\Delta t; \lambda_{ji}) = 1 - e^{-\lambda_{ji}\Delta t}. \tag{49}$$

If we chose Δt sufficiently small such that we can consider λ_{ij} constant in the interval then the number of susceptible individuals that get infected after Δt given a total transmission force λ_{ji} is:

$$\Delta(S_{ji} \to I_{ji}) \sim \text{Binom}(S_{ji}(t), P(\Delta t; \lambda_{ji})). \tag{50}$$

Similarly, the amount of infected that recover during the interval:

$$\Delta(I_{ji} \to R_{ji}) \sim \text{Binom}(I_{ji}(t), P(\Delta t; \mu)). \tag{51}$$

Finally, in what concerns the model, we explain how the changes in mobility are included. The model that we have explained until now, only considers relative changes in the flow, since it uses the commute probability matrix instead of the flow matrix. It is then important to consider the trip reduction due to containment κ:

$$\kappa = \frac{N(T_L)}{N(T_0)}. \tag{52}$$

Since it is expected that a reduction in the number of trips affects the average number of close-proximity contacts k, which affects the basic reproduction number:

$$\mathcal{R}_0 = \frac{\widetilde{\beta}k}{\mu}, \tag{53}$$

where $\widetilde{\beta}$ is the infection rate per contact between a single infected and a single susceptible. In the above description we have:

$$\beta = \widetilde{\beta} k. \tag{54}$$

However, finding the precise relation between the trip reduction and the number of close-proximity contacts κ is an open problem. The authors provide arguments to show that the relation between κ and k results in, at least, a corresponding linear decrease in the basic reproduction number and, at most, a quadratic decrease. They then implement the lockdown scenarios with linear scaling, which is the most "pessimistic" case.

Two possible scenarios for the decrease of the total number of trips are considered. They are both based on the assumption that each individual of a subpopulation of size $N_{ji}^{(pop)}$ contributes with a constant average number of trips f_{ji} to the flow:

$$F_{ji} = N_{ji}^{(pop)} f_{ji}. \tag{55}$$

Then the trip reduction due to containment κ:

$$\kappa_i = \frac{\sum_k \left(F_{ki}(T_L) + F_{ik}(T_L)\right)}{\sum_k \left(F_{ki}(T_0) + F_{ik}(T_0)\right)}, \tag{56}$$

where we are saying that the trip reduction in the population i is equal to the in and outgoing flow of i during lockdown, divided by the same quantity in normal times.

The first scenario, **isolation**, implies that a fraction $1 - \kappa_i$ of the individuals at location i isolates entirely and hence they are effectively removed from the transmission process while the remaining individuals do not change their behavior:

$$F_{ji}(T_L) = [\kappa_i N_{ji}^{(pop)}(T_L)] f_{ji}. \tag{57}$$

In the second scenario, **distancing**, mobility reduction leads to a proportional reduction in the average number of contacts, which translates into a homogeneous reduction of the individual trips count:

$$F_{ji}(T_L) = N_{ji}^{(pop)}(T_L)] \left[\kappa_i f_{ji}\right]. \tag{58}$$

This leads to a reduction in the transmission rate $\beta_i = \kappa_i \beta$ and the force of interaction become:

$$\begin{cases} \lambda_i^{home} & = \frac{\beta_i}{2} \frac{\sum_k^m I_{ki}}{\sum_k^n N^{(pop)}}. \\ \lambda_i^{work} & = \frac{\beta_j}{2} \frac{\sum_k^m I_{jk}}{\sum_k^n N_{jk}^{(pop)}}. \end{cases} \tag{59}$$

3.5.4 Details on the Simulation

We now address the selection of the different parameters for the simulation. It is important to note that our objective with this study is to address the question of to what extent the different containment measures imposed in Spain to control the spread of COVID-19 affect the spread of the epidemic in general. We intend to study how these structural differences impact the spread of an infectious disease. Therefore, as stated in [15], we use epidemiological parameters similar to those of COVID-19 but with no intention of reproducing the actual spread of the disease in Spain.

As we advanced in 1.1, there are different studies that estimate the value of the basic reproductive number $\mathcal{R}_0$ in the range 1.5-6.5. Consequently, we use $\mathcal{R}_0$=3 in our simulations. We also introduced that the probability of contagion after one week of infection is very low in patients with mild symptoms, therefore we used a recovery rate of μ=1/8 days. In addition, in the place of origin of

the outbreak (which corresponds to a Spanish province) we initially introduce I_0=100 infected individuals. This choice of epidemiological parameters corresponds to that used in [15], which allows us to adequately compare the results obtained in the original study in Germany with those obtained in this work for Spain.

We implement the model to carry out different studies. First, we study the dependence of the model on the origin outbreak source by comparing the final size of the epidemic, $R\,(t \to \infty)$, selecting as origin outbreak three different provinces and repeating the simulation for the different weeks on which we focus our study. Furthermore, in each case, the simulation is carried out for the three possible mobility scenarios, without restrictions, distancing and isolation. We chose three provinces whose census corresponds to the lowest census, one with a medium census and finally the one with the highest census. Since mobility varies over time, the provinces whose net flow is lower, medium or higher, varies. We therefore chose to use the census to determine this dependence but although taking into account that there is probably a relationship between higher census and higher net flow. In this way we do not obtain strictly those provinces with the highest, lowest and intermediate flow, but we hope that this study is adequate enough to be able to evaluate the dependency of the model on the origin of the outbreak. The results have been averaged over 100 realizations.

Moreover, for different measures, we repeated the above simulations using a single random outbreak source and averaging again over 100 realizations. Likewise, we also obtain the results in the case of a randomly selected outbreak at each realization. For these results, the provinces of Santa Cruz de Tenerife and Las Palmas have been excluded from among the possible locations where the epidemic can start. This is done because in the weeks, where the vast majority of containment polices where imposed, and consequently the drastic reduction on mobility is observed, after filtering, these two provinces are disconnected from the Iberian Peninsula and, therefore, the epidemic cannot propagate. Finally, the dependence of $R\,(t \to \infty)$ with $\mathcal{R}_0$ is studied. For this, we carry out for the different mobility networks, the simulation in the three possible scenarios for different values of $\mathcal{R}_0$, with a random outbreak source and averaging the results over 50 realizations. In addition we also perform this simulations for the rescaled mobility networks.

Chapter 4

Results

This Section aims to characterize, with different measures and perspectives, mobility in Spain since the start of the COVID-19 pandemic. For this, the period of study covers from March, 2020 to April, 2021. Specifically, our main goal is to study whether the changes in mobility due to the different restrictions imposed have caused deep structural changes in the mobility networks in the long-term. We start by studying **mobility trends** 4.1. To ensure the precision of the calculations, we first reproduced the results of mobility trends in Germany obtained by [15] and then we carry out an analogous study for Spain. Next, we studied the **mobility networks** in Spain 4.2 inspired by the work made by [15] for Germany. We have included some additional measures that allow us to study in more depth how the structure of the networks has changed over time. Finally, we studied how these changes in mobility affect the **spread of the epidemic** in Section 4.3 by implementing the epidemiological model presented in Section 3.5.3.

4.1 Mobility Trends in Germany and Spain

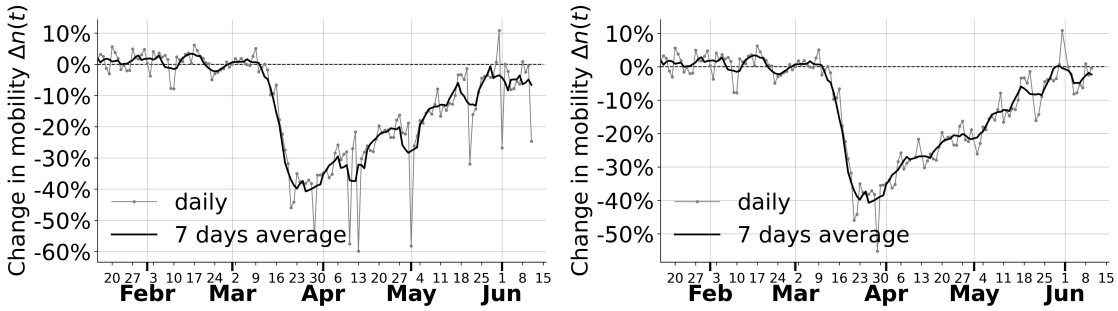
In this Section, we focus on studying mobility trends in Germany and Spain. These results aim to analyze the changes in mobility that have taken place since the beginning of the COVID-19 pandemic.

4.1.1 Mobility Trends in Germany

In this Section we focused on reproducing the mobility trends observed in Germany and studied in [15]. The period studied ranges from the end of February, 2020 to mid-June of the same year.

For that purpose, we study the mobility change, $\Delta n(t)$, obtained as the relative difference between the total number of trips compared to the reference period (see Section 3.4.1). Moreover, to reduce the effect of daily fluctuations in mobility we consider the 7 days moving average (see Section 3.4.1). These results, presented in Figure 4.1, correspond to Figure 1a from [15]. However, in Figure 4.1(a) we see some points characterized by an extremely low mobility, corresponding to federal holidays in Germany. Removing them, we obtain Figure 4.1(b) showing exactly the same trend as in [15]. According to what was observed in the original work, until the beginning of the implementation of containment measures in early March, we found levels of mobility very similar to those of the pre-pandemic ones. Furthermore, we observe that mobility in Germany is substantially reduced during the pandemic. The greatest reduction occurred in mid-March, this period corresponding to the implementation of the vast majority of the containment policies. Over the course of 3 weeks, the mobility trend falls to 40% below the baseline. This decline was followed by an upturn that begins in early April. This increase takes place progressively over the following months, recovering pre-pandemic values at the beginning of June. In addition, it

took place in small bursts originated at the same time as relaxations in mobility restrictions occurred.



(a) (b)

Figure 4.1: Change in total movements $\Delta n(T)$ during the first months of the COVID-19 pandemic, in Germany (January 11, 2020-June 10, 2020), relative to March, 2019: 4.1(a) including federal holidays and 4.1(b) excluding federal holidays. Reproduction of Figure 1A from [15].

In what follows, we studied in detail how mobility changed over the considered period and which trips were more affected. To do this, we studied the dependence of the change in mobility on the distance traveled. We calculated how the change in mobility per distance, $\Delta n_D(t)$ (see Section 3.4.1), has varied over time. For that, we divide the total flow of trips into 5 ranges of distances travelled, D. Figure 4.2 shows the result obtained by doing the 7-day moving average of this amount for each D. Once again, in accordance with the original work, we observe that the reduction in mobility in Germany has not occurred homogeneously. We found that the reduction is greater in long-distance trips, which is consistent with the fact that long-distance trips were more affected by containment measures.

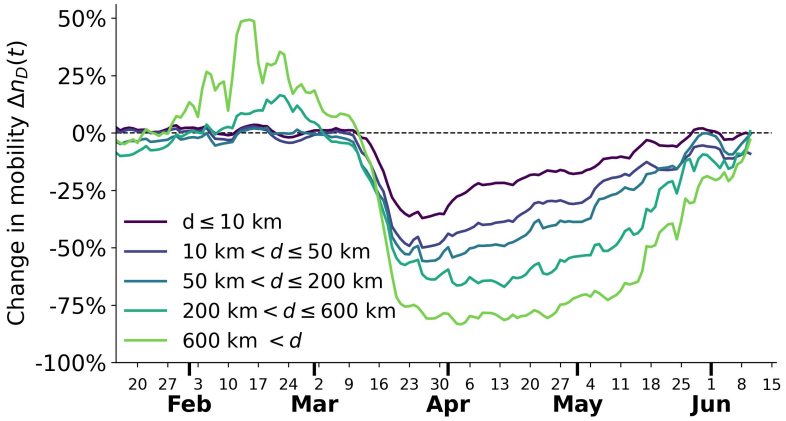


Figure 4.2: Relative mobility changes $\Delta n_D(t)$ for different distance ranges D (7-d moving average) during the first months of the COVID-19 pandemic, in Germany (January 11, 2020-June 10, 2020), relative to March, 2019. Reproduction of Figure 2A from [15].

In a complementary way, we calculated the difference between the short-distance mobility change (which includes those trips whose distance traveled is $\leq 10km$) and the long-distance change (trips whose distance traveled is $>10km$). Again, we present the 7-day moving average of this quantity over time in Figure 4.3, reproducing Figure 2b from [15]. This difference between short- and long-distance mobility change allows us to identify unusual mobility patterns. Although, as we have seen in Figure 4.1, average mobility recovers normal values at the beginning of June in terms of

total flow, long-distance trips have not yet recovered pre-pandemic values. This result indicates a long-lasting structural change in the mobility network.

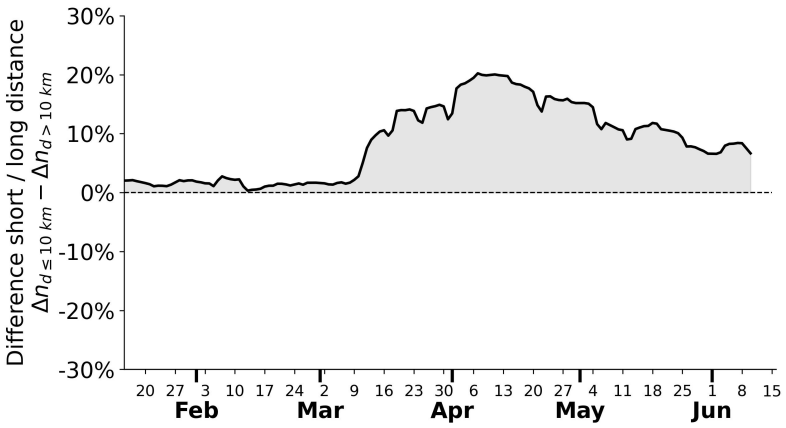


Figure 4.3: Difference between short-distance mobility change $\Delta n_{d \leq 10km}(t)$ and long-distance mobility change $\Delta n_{d>10km}(t)$ during the first months of the COVID-19 pandemic, in Germany (January 11, 2020-June 10, 2020), relative to March, 2019. Reproduction of Figure 2B from [15].

4.1.2 Mobility Trends in Spain

Now we carried out a study analogous to the one presented in the previous Section for mobility in Spain. We extended the period covered until April 9, 2021 in order to study differences between the different phases (see Section 1.3) of the pandemic.

We began by reproducing Figure 4.1 for Spain, being the results obtained shown in Figure 4.4. In a complementary manner, we show in the same figure the total number of infected individuals at the national level, which allows us to identify more clearly the effects of these changes in mobility in the spread of the disease. Vertical bars mark milestones concerning policy issues.

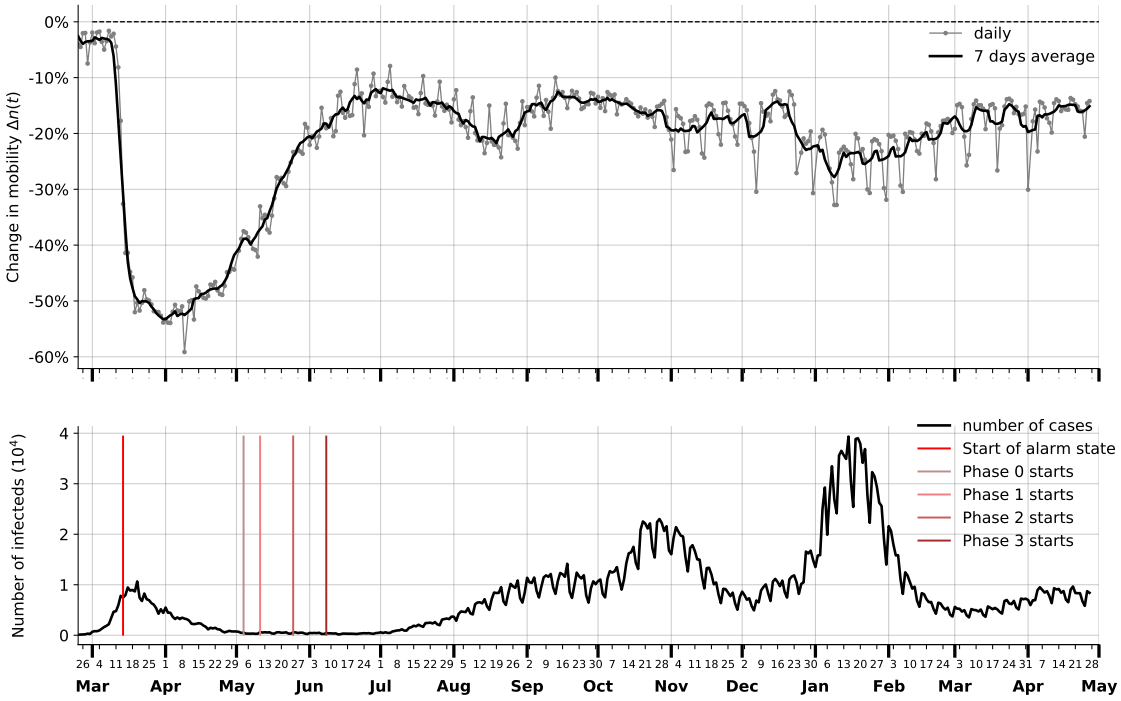


Figure 4.4: Evolution of the total mobility in Spain relative to a reference period of time, compared to the total number of infected of COVID-19 registered in the country (February 24, 2020-April 9, 2021). Upper panel: Change in total movements $\Delta n(T)$ relative to March, 2020 (2-8). Lower panel: Evolution of the total number of infected individuals in Spain. Vertical bars mark milestones in policy measures.

We show that mobility has effectively been reduced as a consequence of the containment policies imposed in Spain. It presents a large descent at the end of March where values of 50% below the reference values are recorded. This fall is followed by an immediate and slow rebound in mobility where, in mid-June, values of 10% below the baseline are registered. After this increase, mobility stabilized in the range [-10, -30]%, experiencing smaller fluctuations, although never recovering the values prior to the pandemic. The fluctuations in mobility have clear consequences in the total number of cases registered, with approximately one month of delay.

More in depth, the greatest decrease in mobility, where values of 50% below the baseline are reached, is recorded in the first weeks after the implementation of the State of Alarm. The minimum is recorded at the beginning of April, 2020. This first drop begins before the containment-measures implementation date since some activities, such as face-to-face classes, had previously been banned. Consequently, approximately one month after the start of this decline, the number of registered cases drops significantly. Starting in mid-April, mobility began to slowly recover, reaching just 10% below the baseline in early July. Thus, again a month later, an increase in the number of cases begins to be registered, which corresponds to the so-called second wave. This rise of the total mobility is explained by the relaxation of the confinement measures and the increase in trips on account of summer holidays. Relative mobility maintains values between -10% and -20% until the beginning of November. The second wave lasts until the beginning of December. During this period, relative mobility shows a slight decrease at the end of August, which translates into a small drop in the total number of cases at the end of September and the beginning of October. After a slight decrease, we find a new relative mobility maximum in mid-December that triggers a rebound in the number of cases, with this third wave reaching its maximum in mid-January, 2021. This maximum is followed by a new decrease in mobility after the end of holidays resulting in the end of the third wave approximately one month later. From the end of January, mobility increased again, following a positive trend until the end of the period under consideration, that we also observed in

the number of cases leading to the fourth wave.

Once again, studying the mobility reduction as a function of distance, allows us to understand how this change has occurred and what types of trips have experienced a greater reduction in the different periods. For each pair of possible locations (municipalities), this distance travelled corresponds to the geographic distance between the origin and the destination. In Figure 4.5 we show the same results as in Figure 4.2 but for Spain. Moreover, in a complementary way, we present the difference between the short-distance mobility change and the long-distance mobility change in Figure 4.6.

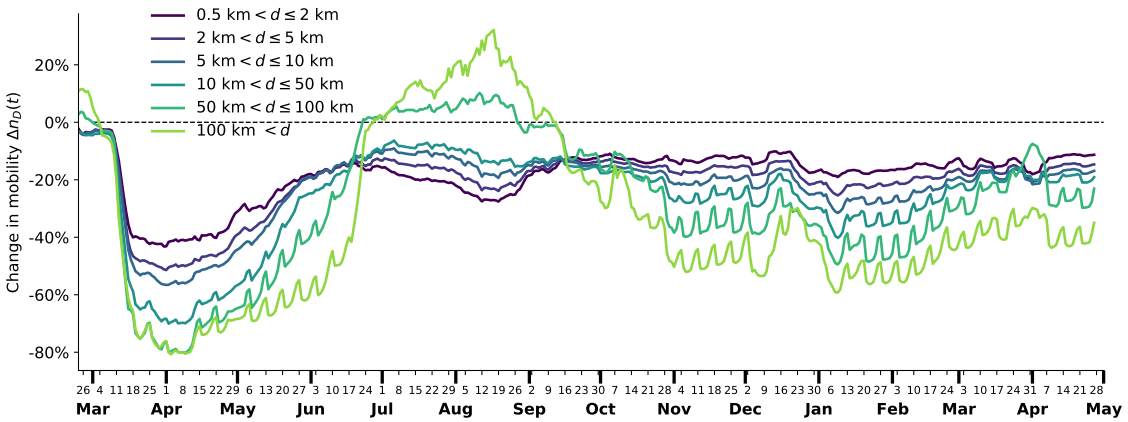


Figure 4.5: Relative mobility changes $\Delta n_D(t)$ for different distance ranges D (7-d moving average) in Spain (February 24, 2020-April 9, 2021), relative to March, 2020 (2-8).

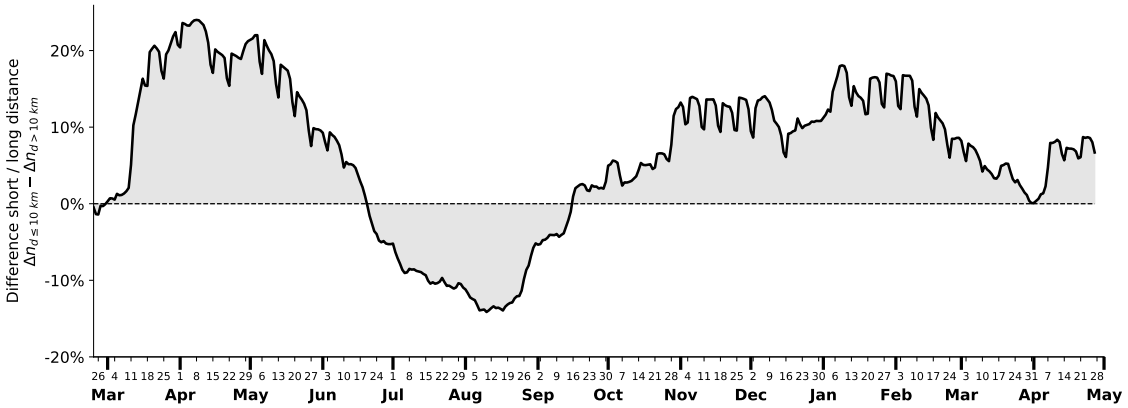


Figure 4.6: Difference between short-distance mobility change $\Delta n_{d \leq 10km}(t)$ and long-distance mobility change $\Delta n_{d>10km}(t)$ in Spain (February 24, 2020-April 9, 2021), relative to March, 2020 (2-8).

We found that the reduction in mobility has not occurred in a homogeneous manner. Excluding summer 2020, long-distance trips show a greater reduction compared to short-haul. This effect, originated by the initial containment measures, which were strongly aimed to reduce long-distance trips, linger until May, 2021 regardless of the measures in force at any given time. Once mobility partially recovers after the initial decline, trips whose distance traveled is $<50km$ maintain practically constant values throughout. The rise in long-distance trips observed during the summer, Christmas and Easter is explained by the holiday and vacation periods as well as with the current policies.

The importance of this lies in the fact that we can relate the different upswings in mobility that we placed as part of the origin of the successive waves in the epidemic, with a much greater relative increase in long-distance trips than in short-distance ones. Mobility, or more specifically long-distance travel, are not the only triggers for the various outbreaks of the epidemic. It is mobility combined with other factors, such as the easing of containment measures, that determine the development of the epidemic. All this leads us to think that the epidemic can be controlled when we have local outbreaks, and it is the mixing of people that is responsible for the large outbreaks in the epidemic. Hence we infer that mobility affects the spreading of the epidemic, but long-distance travel are responsible of the overall modulation.

In a complementary way to what was mentioned above, we have calculated the Pearson correlation coefficient between $\Delta n_D(t)$ for the different distances D, and the total number of infected individuals in the country, following the study carried out by [46] for the USA. These results are shown for the first months of the second period in Figure 4.7. Correlation is found to be significant, at a 91%, for the longest trips after a lag of approximately 47 days. Additionally the Pearson correlation stays significant (ρ>0.7) for time lags greater than 14 days. Strong and statistically significant correlation between long-scale travels, and reduction of COVID-19 case growth exists. This study is a first approximation since only a linear dependence between both magnitudes is quantified. But, as it is well known, many other factors influence this process. As the epidemic progresses, this linear relationship becomes less clear as other factors, such as vaccination, play an important role. That is why, for later periods, we have not obtained such good results. In addition, a greater correlation can be expected during the first phase of the epidemic, where mobility and its patterns play a leading role, but since mobility data prior to March, 2020 are not available, it has not been possible to carry out the study.

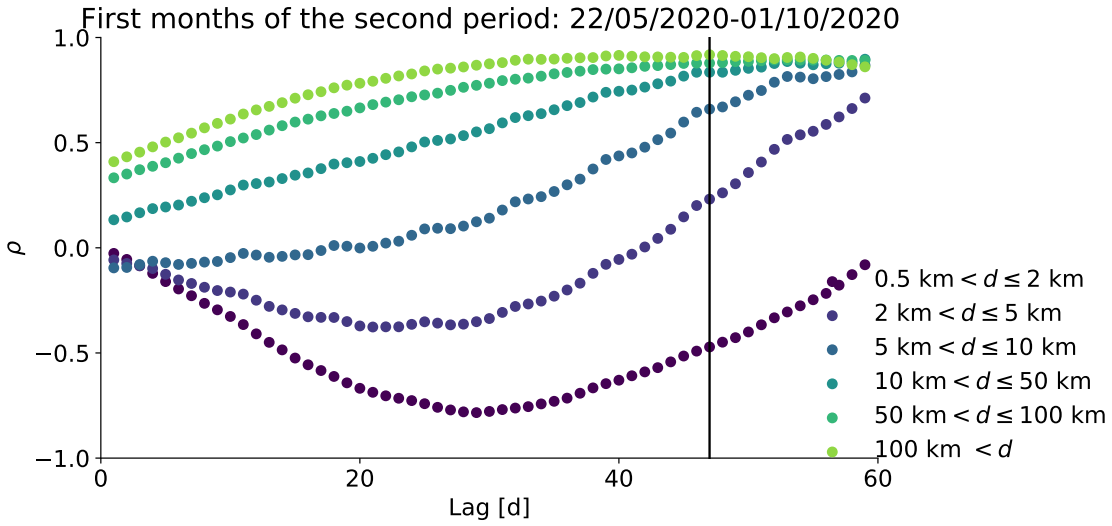
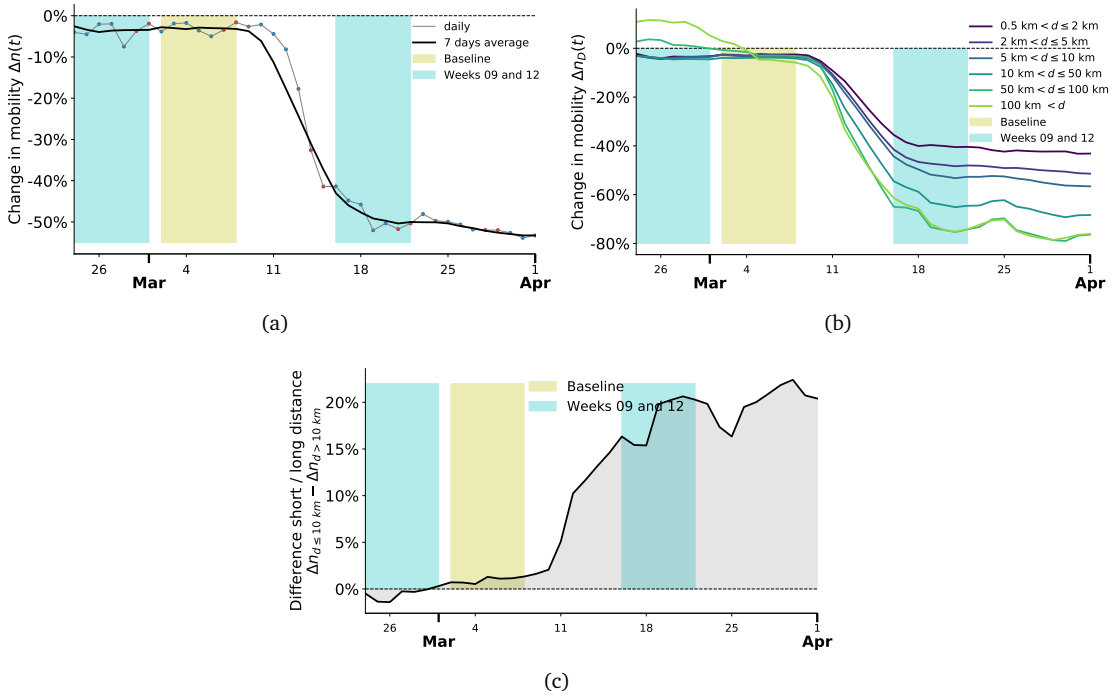


Figure 4.7: Relationship between the relative mobility change $\Delta n_D(t)$ for different distance ranges D (7-d moving average), relative to March, 2020 (2-8) and the number of infected registered during the first months of the second period (March 22, 2020-October 1, 2020). This relationship is obtained by calculating the Pearson correlation coefficient ρ for different time lags.

In the following figures we zoom in on the results presented in Figures 4.4, 4.5 and 4.6 in different periods. We consider this set to be representative of the different periods and waves of the epidemic in Spain for the reasons that we detailed below. In Figures 4.8, 4.9 and 4.10 we present the different zooms and highlight in blue those weeks that will be further studied. Additionally, the following figures allow us to make a more detailed analysis of mobility at each moment.

First, in Figure 4.8, we present a zoom for March, 2020. Additionally, we highlight in yellow calendar week 10 (March 2 to 8, 2020) that, as we advanced earlier, is used as our reference baseline period. This selection is based on the fact that this calendar week is the last to present normal mobility prior to the implementation of containment policies. Therefore, we consider that it is the most appropriate pre-pandemic reference in order to study the changes that take place during the first wave. As we can see, mobility presents normal values in the different distance ranges, when comparing with pre-pandemic period. Nevertheless, longer trips present slighter lower relative values but this difference is small enough in order to consider this week as representative of the "normal" mobility. Moreover, since the selection of calendar week 10 as baseline is the same as in [15] it allows us to more adequately compare the results obtained with those found in Germany.



(a) (b)

(c)

Figure 4.8: Zoom of figures 4.4 (4.8(a): change in total movements), 4.5 (4.8(b): change in total movements for different distances) and 4.6 (4.8(c): difference between short- and long-distance change) for the reduction in mobility observed in the first wave (March 1, 2020-April 1, 2020), relative to March, 2020 (2-8). Yellow bar marks calendar week 10 selected as baseline. Blue bars highlight calendar weeks 09 and 12 that are further analyzed throughout this Master's Thesis. In Figure 4.8(a) blue dots indicate weekdays and red dots correspond to weekends.

Week 09 (from February 24, 2020 to March 1 of the same year) and week 12 (from March 16, 2020 to 22 of the same month) are also highlighted in Figure 4.8. We use calendar week 09 as the representative week of the network prior to the implementation of restrictive measures. In this way, we reinforce that our baseline is representative of the mobility we had before the changes in mobility began. Moreover, with the aim of studying the effect of the most drastic reduction in mobility observed at the end of March and the beginning of April, we selected calendar week 12. Additionally, we include calendar week 16 (from March 13 to 19, 2020). In order to study the second period, we have further studied calendar weeks 30 and 34, which run from July 20 to 26, 2020 and from August 17 to 23, 2020, respectively. If we again zoom into this period, as shown in Figure 4.9, we see that the chosen weeks are at a maximum and minimum of the total change in mobility, respectively. Nevertheless, both are in the period in which the total relative flows of the longest distances are maximum. For characterizing the third period, we chose calendar week 52

(December 21 to 27, 2020), coinciding with the relative maximum of long-distance trips during this period, as can be seen in Figure 4.10. Finally, we also included in this set calendar week 66 (counting from January 1, 2020), which includes those days between March 29 and April 4, 2021. This week is representative of the long-term mobility network. It allows us, therefore, to study which changes both in the flow and in the structure of the network can be considered as deep enough to remain present regardless of the restrictive measures adopted.

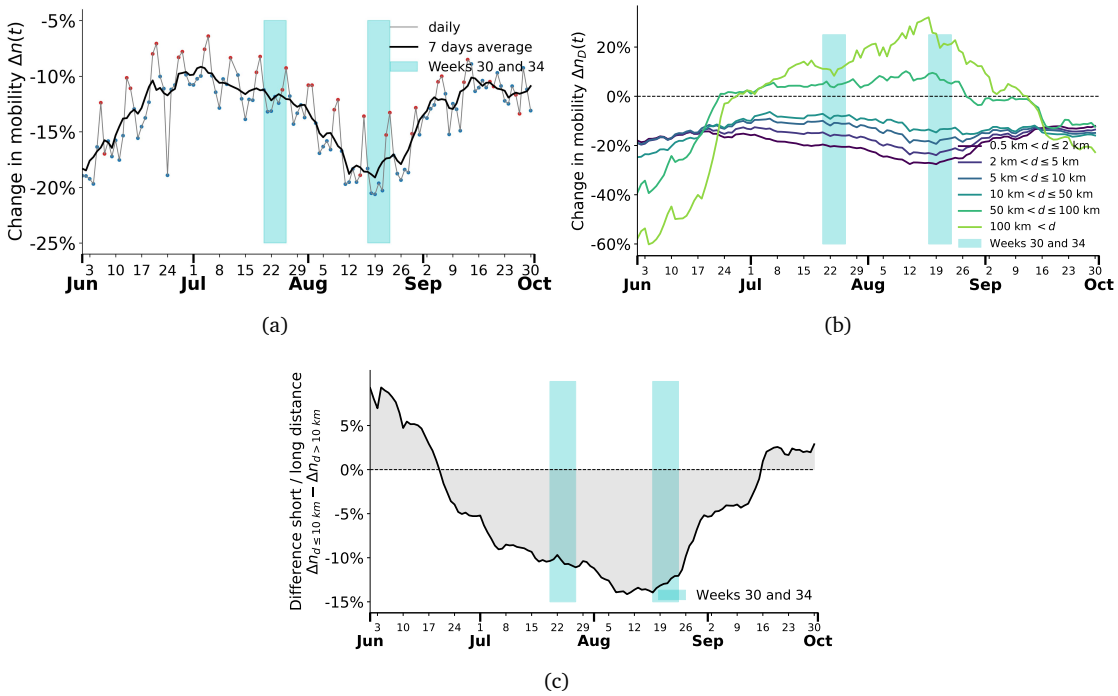


Figure 4.9: Zoom of figures 4.4 (4.9(a): change in total movements), 4.5 (4.9(b): change in total movements for different distances) and 4.6 (4.9(c): difference between short- and long-distance change) for the reduction in mobility after summer of 2020 corresponding to the second epidemic wave (June 1, 2020-September 30, 2020), relative to March, 2020 (2-8). Blue bars mark calendar weeks 30 and 34 that are further analyzed throughout this Master's Thesis. In Figure 4.9(a) blue dots indicate weekdays and red dots correspond to weekends.

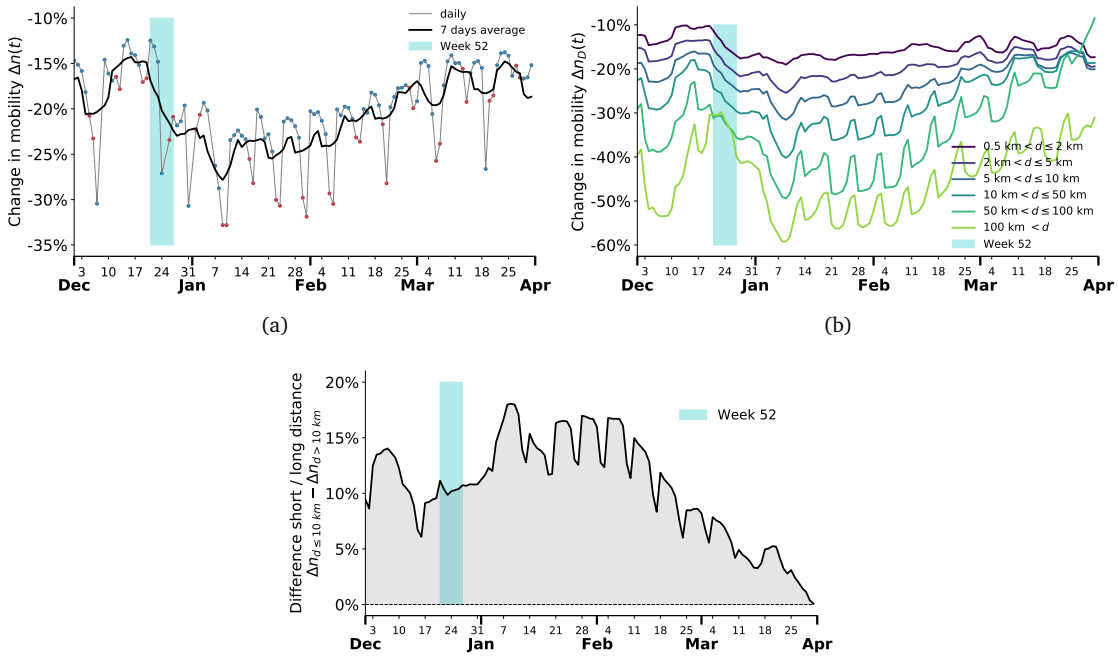


Figure 4.10: Zoom of figures 4.4 (4.10(a): change in total movements), 4.5 (4.10(b): change in total movements for different distances) and 4.6 (4.10(c): difference between short- and long-distance change) for the reduction in mobility after Christmas of 2020 corresponding to the third wave (December 1, 2020-March 31, 2021), relative to March, 2020 (2-8). Blue bars mark calendar week 52 that are further analyzed throughout this Master's Thesis. In Figure 4.10(a) blue dots indicate weekdays and red dots correspond to weekends.

Additionally, the results presented in Figures 4.8, 4.9 and 4.10 allow us to identify some differences in the mobility patterns observed in the first three waves. As we have discussed, after the first big drop in mobility at the end of March, mobility quickly recovered to re-reach levels of -20% within 2 months. Despite the fact that the total flow begins to recover, long-distance trips have a slower growth, always maintaining this relative flow below that of short-distances, as can be seen in Figures 4.4. This slow growth is also observed after Christmas 2020, where total mobility reaches a maximum in mid-December that is maintained until the end of the year. However, these values are not reached again until mid-March 2021, as can be seen in Figure 4.10(a). Always keeping below the total relative flow of long-distances Figure 4.10(b). On the other hand, as shown in Figure 4.9(a), the recovery after the summer months is much faster. After the minimum observed at the end of August, 2020, mobility returns to similar values to those of mid-July in mid-September. Another key difference is that, while total mobility decreases to a minimum and then increases again, the flow of long-distance trips continues to grow, as can be seen in Figure 4.9(b). These reached a maximum in mid-August, coinciding with the minimum observed in short-distance flows (Figure 4.5), where values above 20% with respect to the reference period are registered. All these changes might be related to the re-enforcement of containment measures when the different epidemic waves begin.

Lastly for this comparison, we address the weekly internal variability of the total change in mobility. For this, in Figures 4.8(a), 4.9(a) and 4.10(a) we have represented the daily change of the days corresponding to weekdays in blue and those corresponding to weekends in red. For this comparison, we took into account that each day is compared with the corresponding day of the week in the reference period. We see that at the beginning of the pandemic, mobility hardly shows any internal variability during the week. The trend obtained with the moving average fits well to the daily fluctuations during the first epidemic wave, while this is not the case for the second and third

epidemic waves. Although the global trend in mobility is marked by weekdays, weekends show an offset from the average, being positive during the second epidemic wave, and negative during the third epidemic wave. The fact that the offset has different signs in the second and third waves is probably result of the different severity in restrictive measures applied in each epidemic wave combined with the social awareness at each moment and climatic factors, among others. However, the explanation of this fact would require additional information and a more detailed analysis, which is outside the scope of this Master's Thesis.

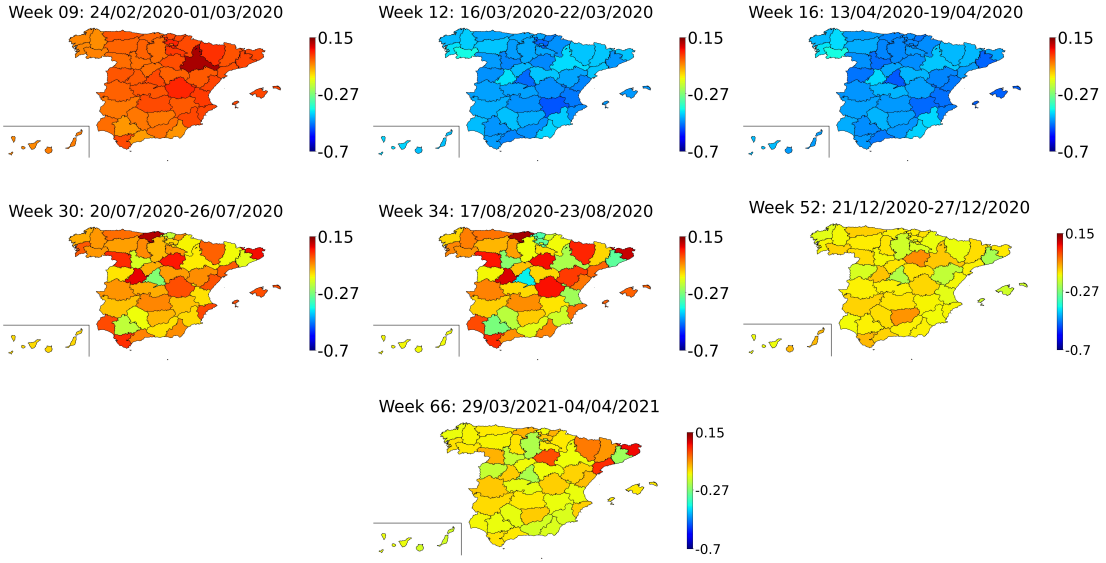


Figure 4.11: Mobility change $\Delta n^{(i)}(t)$ in Spain provinces for different calendar weeks (weeks 09, 12, 16, 30, 34, 52, and 66) (February 24, 2020-April 9, 2021) relative to the baseline values (calendar week 10, (March 2-8, 2020)).

Further proof that the reduction in mobility has not occurred in a homogeneous manner and that the structure of the network has changed with respect to the pre-pandemic is shown in Figure 4.11. In this Figure, we represent the mobility change for each location, $\Delta n^{(i)}(t)$ (see Section 3.4.1), which corresponds the number of trips that originate in each province i, for different representative calendar weeks. Results are displayed at a provincial level for clarification. We observe that mobility has been substantially reduced, this decrease being more pronounced during the first weeks of the pandemic. The most important result obtained from this analysis is that, prior to the imposition of the first restrictive measures (week 09), we had a uniform mobility, this structure is maintained in the first reduction in mobility (weeks 12 and 16) and subsequently, until the end of the covered period, mobility is no longer uniform.

4.2 Structural Changes in Mobility Networks for Spain

Once analyzed the changes in the total mobility over the different phases of the pandemic, in this Section we focus on the evolution of the weekly mobility networks over time.

For this purpose, we built weekly mobility networks G_T where edge weights w_{ji} corresponds to the average daily flow along the connection during the week (see Section 3.3.1). Additionally, we compare such networks with the corresponding rescaled weekly mobility networks, $G^R_{T_0}(T)$, where T_0 is the baseline week (calendar week 10). These are built to obtain networks that are structurally

similar to that of the baseline but whose total flow corresponds to that of calendar week T (see Section 3.3.2). For both networks, we apply the same thresholding procedure, discarding those edges whose average daily flow is below $w_{ji} < 200$ trips (see Section 3.3.3).

In Figure 4.12 we show the depiction of our set of representative weekly mobility networks. The thickness of the lines indicates the average daily flow of trips along the connection. For each pair of nodes, we separately represent the flow in both directions in order not to lose information in the case of both flows being significantly different. In agreement with what was observed previously, we see that the networks, after the beginning of the restrictive measures, are considerably less dense. During the weeks corresponding to Phase 1 (weeks 12 and 16), this effect is clearer, observing that in general, all the zones are less connected. During these weeks, the Canary Islands are completely disconnected from the network, dividing it into two weakly connected components. In summer (weeks 30 and 34), as expected, this connectivity increases, obtaining a network more similar to the pre-pandemic one. However, we observe that both at Christmas (week 52) and in the latest week (week 66), the network is less dense than before the lockdown and we have a connectivity more similar to the weeks of lockdown than that before the pandemic.

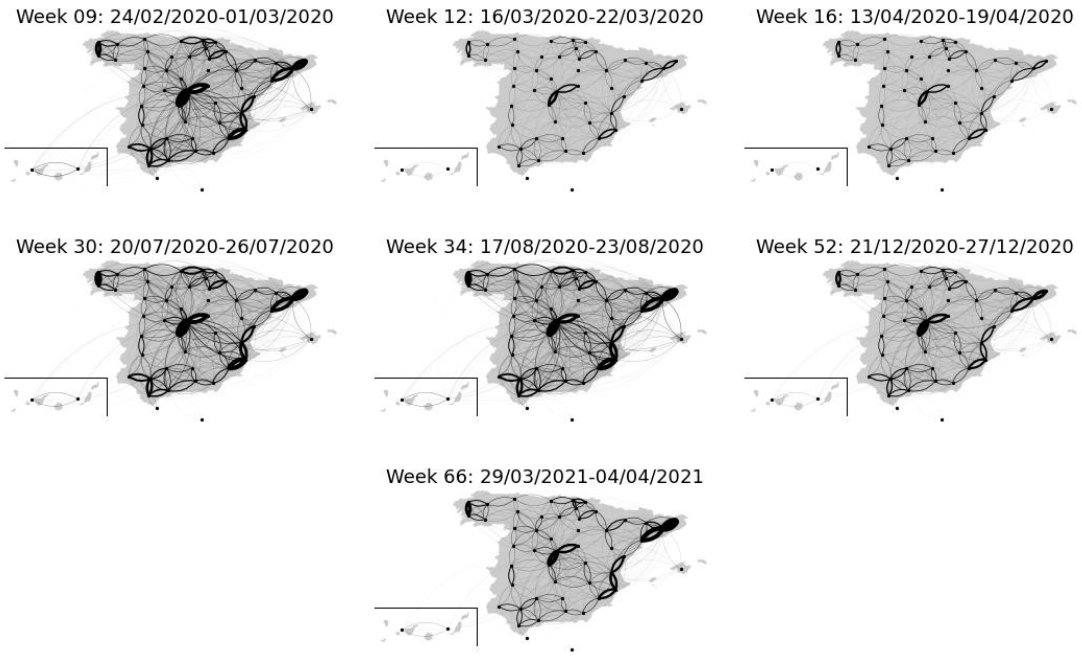
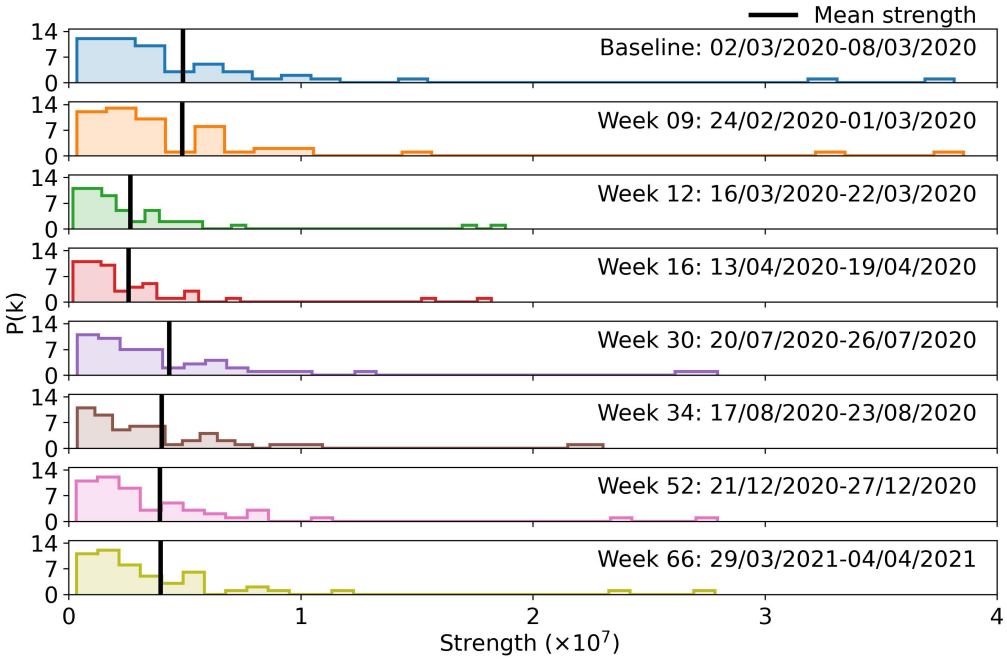


Figure 4.12: Depiction of the networks for the different calendar weekly mobility networks G_T under study (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021). Line widths indicate the average number of daily trips along each connection.

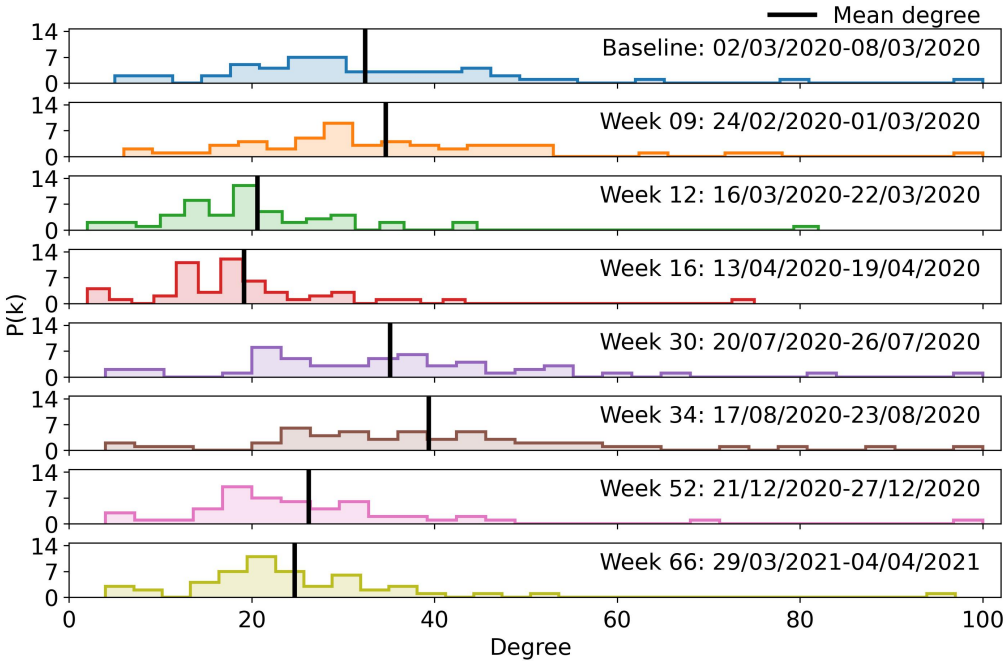
Below, we analyze the evolution of the degree and strength over time. The degree counts the number of links that each node of the network has (see Section 3.4.2), and the strength was calculated in a similar way, but weighting by the flow of each connection (see Section 3.4.2).

The probability distribution of the degree and strength for the different weeks selected in the period studied are shown in Figure 4.13. In what concerns the strength distribution, although there are variations in the different periods studied, a typical distribution is observed throughout the covered period (Figure 4.13(a)). Where we can find with non-negligible probability nodes whose strength is much greater than the mean value ($>2\times10^7$).

When studying the degree distributions, we observe that during the weeks corresponding to the first phase, when the restrictive measures were more severe, the network becomes much less dense (Figure 4.13(b)). In summer, the network connectivity increases, exceeding pre-pandemic values. The important result is that both at Christmas and until the end of the covered period we see that, although the flows have been partially restored, structural changes in the mobility networks are still significant. The network flows have been redistributed in such a way that the degree distribution is shifted towards smaller values. As a result of the presence of a physical constraint, that make less likely to have many connections, we do not observe a power-law behavior.



(a)

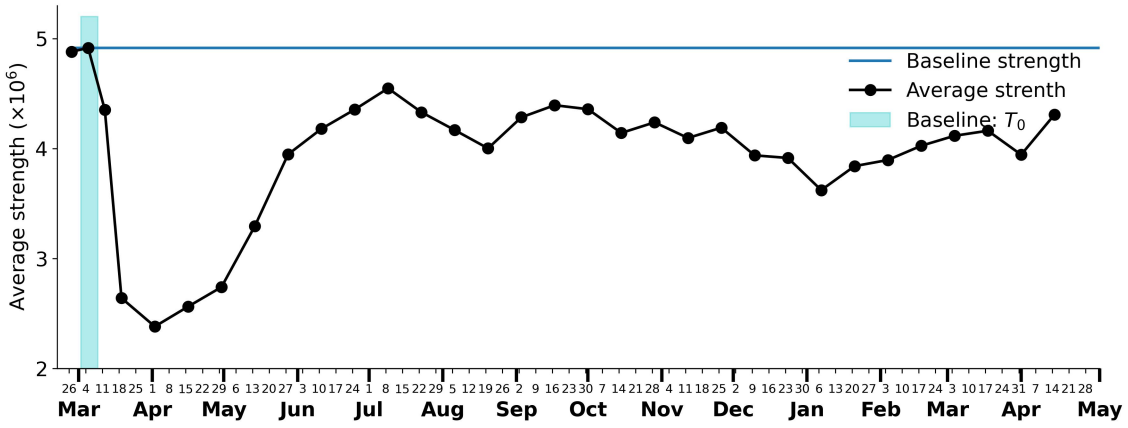


(b)

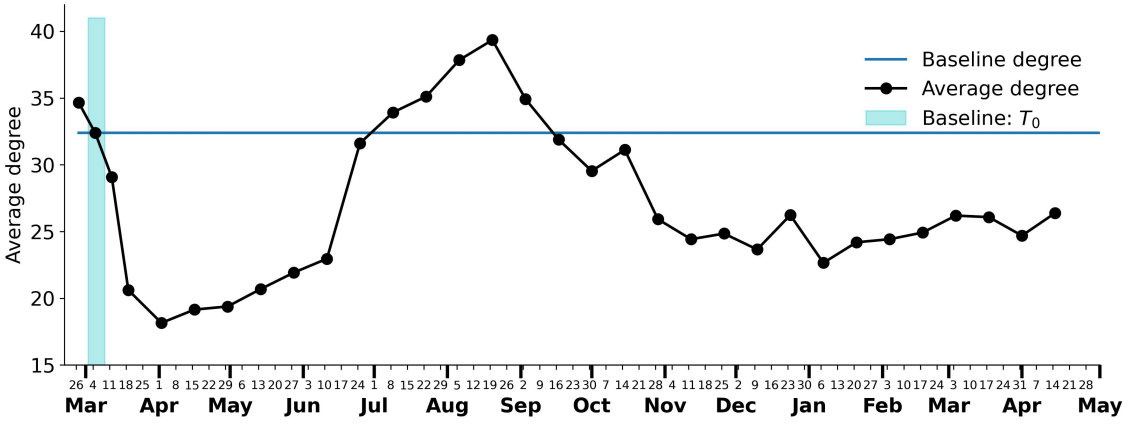
Figure 4.13: Distribution of the node strength 4.13(a) and node degree 4.13(b) for the different weekly mobility networks G_T over time (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021). Vertical lines mark the mean value.

Moreover, we analyzed the temporal evolution of the average node strength and degree. What is observed in the figures is in agreement with what observed before. The global trend followed by the average strength shown in Figure 4.14(a) is what can be expected by looking at the relative total mobility in Figure 4.4. We observe a significant reduction in the first phase of the epidemic,

that is followed by an increase (mid-April) when the global flow is partially restored but always maintaining values below the reference values. It is clear that the mobility peaks observed in Figure 4.4, which we place at the origin of the different waves, are also observed in the dynamics of the mean flow. On the other hand, the evolution of the average degree shown in Figure 4.14(b) follows a similar behavior to the long-distance trips shown in Figure 4.5. The average connectivity of the network increases and recovers similar values to those of the pre-pandemic, when a peak in long-distance trips is observed. In the summer months 2020 we observe that long-distance trips exceeded the values of the reference period. This behavior is also observed in the evolution of the average degree. This is because the network connectivity increases when long-distance connections are recovered.



(a)



(b)

Figure 4.14: Evolution of the average strength 4.14(a) and average degree 4.14(b) for weekly mobility networks G_T over time (February, 2020-April, 2021), relative to the values in the baseline week $T_0 = 10$ (March 2-8, 2020). Blue vertical lines mark the baseline period, $T_0 = 10$, and horizontal lines the corresponding baseline value.

In addition, and to complement the results presented till now, we have also studied how the structural changes in the mobility networks, affect the ability of the networks to support diffusion.

The temporal evolution of relaxation and mixing time is shown in Figure 4.15. The relaxation time measures the time scale in which an edge-centered random walk approaches the uniform distribution at equilibrium. Analogously the mixing time is defined for a node-centered random walk (see Section 3.4.2). We can see that t_{rlx} increases significantly during the isolation period

in relation to the baseline. The mixing time also shows a relative increase of similar order at this time. Both magnitudes show slight variations in the rest of the period, showing lower relative values at mobility peaks, such as in August or during Christmas, coinciding with the increase in long-distance trips that favors the mixing of people. After these periods, the containment measures were reinforced and consequently both observables show higher relative values. During the second half of June and the first weeks of July we find abnormal values of the t_{mix}. This effect is also observed, to a lesser extent, in the relaxation time. Despite this abnormality, the minimum of both observables, relative to the baseline values, is recorded in mid-August, coinciding with the relative maximum of long-distance trips.

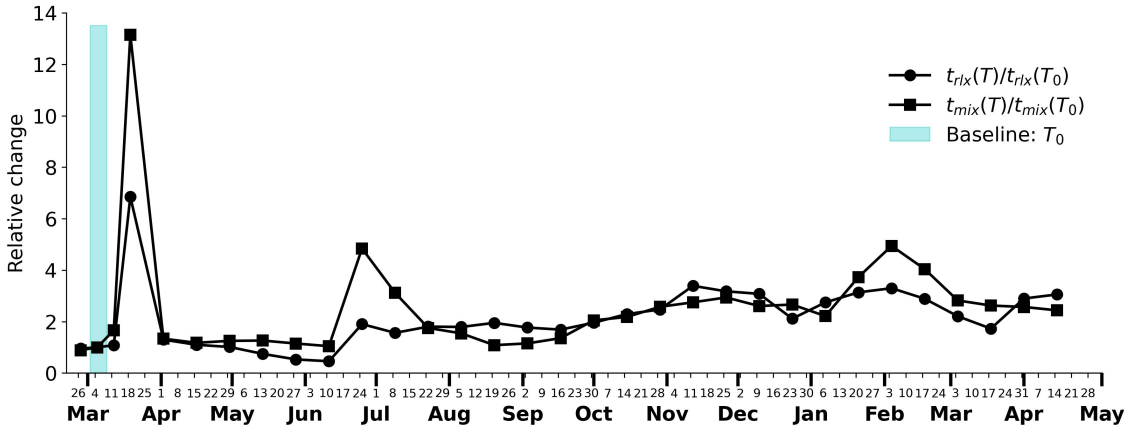


Figure 4.15: Evolution of the relaxation (t_{rlx}) and mixing (t_{mix}) time throughout the considered period (February, 2020-April, 2021), for weekly mobility networks G_T over time, relative to the values in the baseline week $T_0 = 10$ (blue bar) (March 2-8, 2020).

The temporal evolution, relative to the baseline values, of the average shortest path length and the average clustering coefficient has been analyzed. Additionally, we have studied the temporal evolution of the average shortest path length for the rescaled networks. In this way, we are able to discern whether the changes shown by the observables can be explained purely by changes in the total flow or, structural changes in the network must also be considered. Results can be seen in Figure 4.16.

Once again, we observe substantial changes in the structural properties of the networks. As expected, the behavior of the average shortest path length is similar to the relaxation time. It shows high relative values during the first phase of the epidemic, then recovers very similar values to those of the pre-pandemic in summer. After the reinforcement of the containment policies and the consequent reduction in mobility, it increases again, a phenomenon that is observed again after the Christmas drop. $L(T)$ owes its increase to the absence of long-distance connections during these periods and, therefore, several steps must be included when making long-distance trips. Regarding the behavior of the average clustering coefficient, we observe a sharp increase during the first epidemic wave, caused, once again, by the general loss of connectivity in the network. After this period, it remains relatively constant and with values similar to those of the baseline. During this period, the network looses long-distance connections, as a result the clustering coefficient presents only small fluctuations since short-distance connections are the main contributors to the clustering coefficient of the network.

A fundamental aspect to determine the importance of the structural changes and to isolate the effects caused by a simple global reduction in mobility, is to study how the average shortest path length varies in the rescaled weekly networks, $G^R_{T_0}(T)$. We see that part of the increase in these observables cannot be explained by a global reduction in the flux and that, on the contrary, it is

necessary to take into account the topological contributions in order to explain their totality. For example, we observe high values of the average shortest path length that are due, almost entirely, to structural changes that remain present in the network even after the global flow has been practically restored.

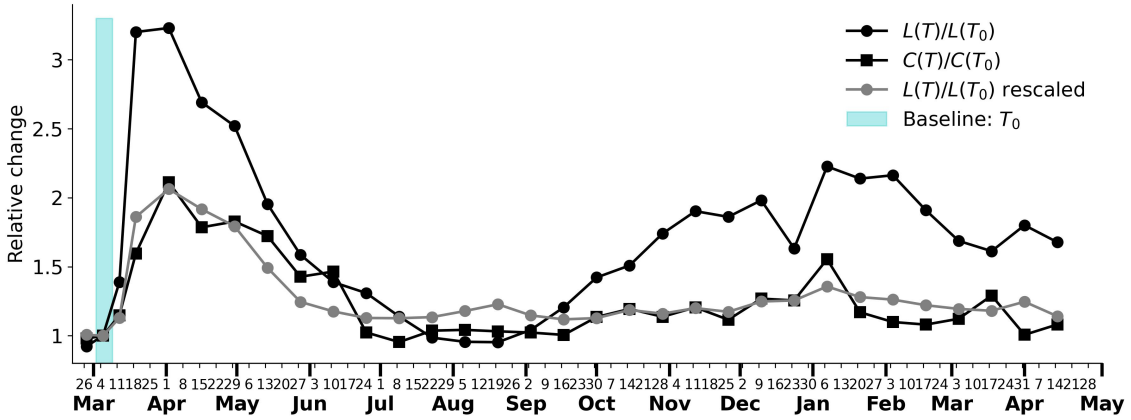


Figure 4.16: The average shortest path length $L(T)$ and the average clustering coefficient $C(T)$ for weekly mobility networks G_T over time (February, 2020-April, 2021), relative to the values in the baseline week $T_0 = 10$ (blue bar) (March 2-8, 2020). In grey, the evolution of the $L(T)$ for the corresponding rescaled mobility networks, $G^R_{T_0}(T)$, is plotted in order to show how much of this observables behavior can be explained by a global reduction of mobility.

Finally, regarding the study of the evolution of the Small-World characteristic of the networks, we have studied the dependence on the geographical distance of the average shortest length, as shown in Figure 4.17. In Small-World spatial networks, the expected shortest path grows with geographic distance, until it levels out at a critical distance where it remains nearly constant regardless of distance. Taking into account our lack of statistics for long-distances (few provinces are more than 470km from each other), we observe this behavior for the majority of weeks, but not for those of Christmas, the latest one in time, and the two representative weeks of the first phase. For the last case we observe stronger dependency with geographical distance. This property is typical of lattice-like networks. When we study the dependence of $L(T)$ on geographic distance for rescaled networks, we see that this behavior is only partially explained by a global flow reduction. The results obtained are not very conclusive but, even so, they help us to observe changes in geographic dependence with the SPL. This may be due, for example, to the lack of statistics for the largest distances or to the fact that the shortest paths were included, regardless of the corresponding distance traveled.

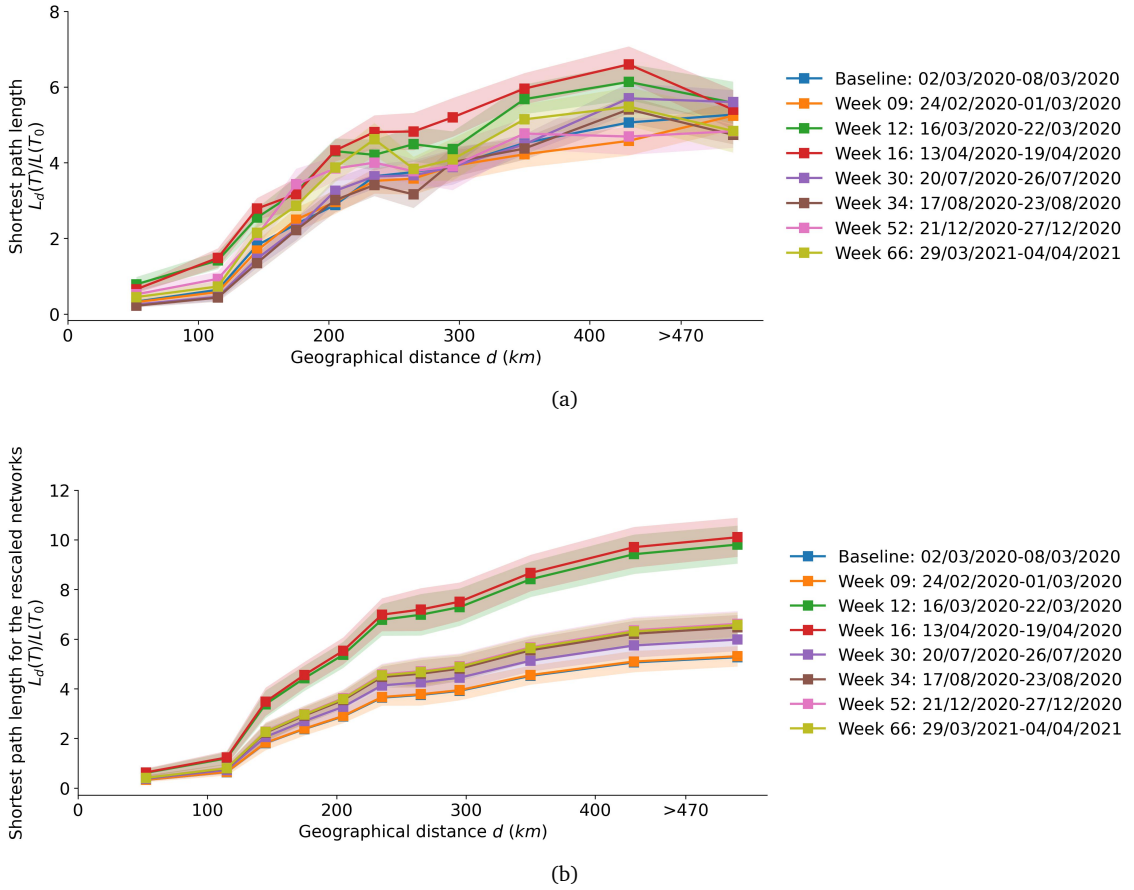


Figure 4.17: The shortest path length $L_d(T)$ at a distance d for different weekly mobility networks (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021), relative to the values in the baseline week $T_0 = 10$ (blue bar) (March 2-8, 2020). For the weekly mobility networks (G_T) 4.17(a) and the corresponding rescaled ($G^R_{10}(T)$) ones 4.17(b). Shaded area represents the standard error.

Additionally, we have focused on community structures to study how the internal organization of the network has evolved over time. For this we used OSLOM, a freely accessible community detection program [62]. The map of the distribution of the communities, where each province is represented with colors according to the community to which it belongs, is shown in Figure 4.18. In some cases, we find that the same province belongs to three different communities. In order to be able to visualize these results, we have built three different levels of color maps where each one assigns a color to the province corresponding to the three different levels we encounter. In addition, in Figures 4.19 we show the evolution of the number of communities, their average size average (in number of provinces) and the average population per community.

We find that before the pandemic the network is heavily connected, consisting of only 3 communities each with a large number of provinces and population (as can be see by the deviations of both magnitudes in Figures 4.19(b) and 4.19(c)). One of them covers the entire south and the islands of the country. During the first phase of the quarantine, connectivity is lost and consequently the number of communities increases. Then we have 5 communities of similar sizes. During the summer months, connectivity is partially restored, structurally resembling what was observed before the pandemic. The coast is connected to the more inland regions due to vacation travel. On the other hand, both at Christmas and in April, 2021, we recovered, and even exceeded, the number of communities during the first phase. The network becomes less connected, and communities are smaller in size and contain considerably less population than before the pandemic.

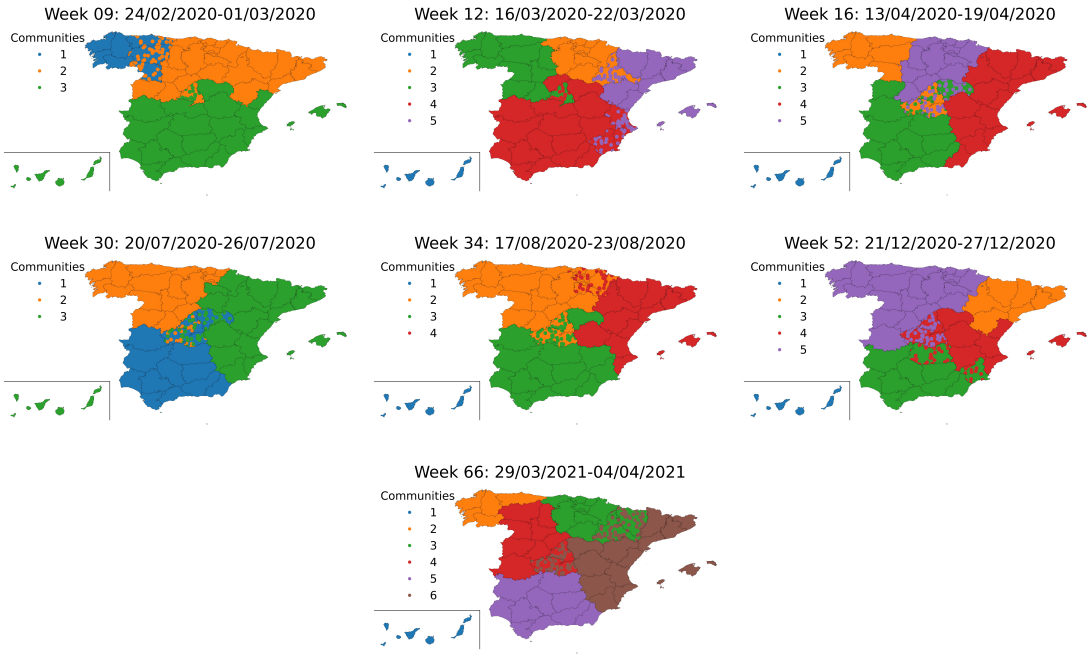
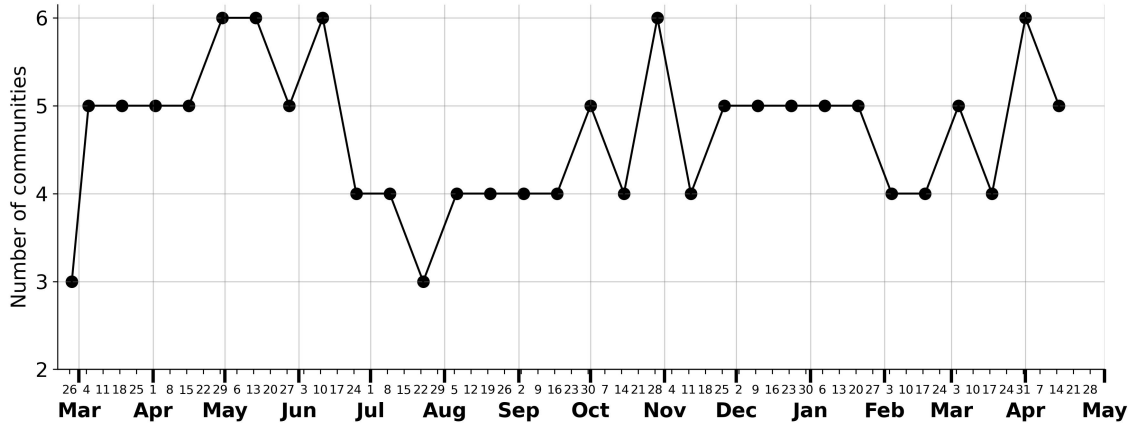
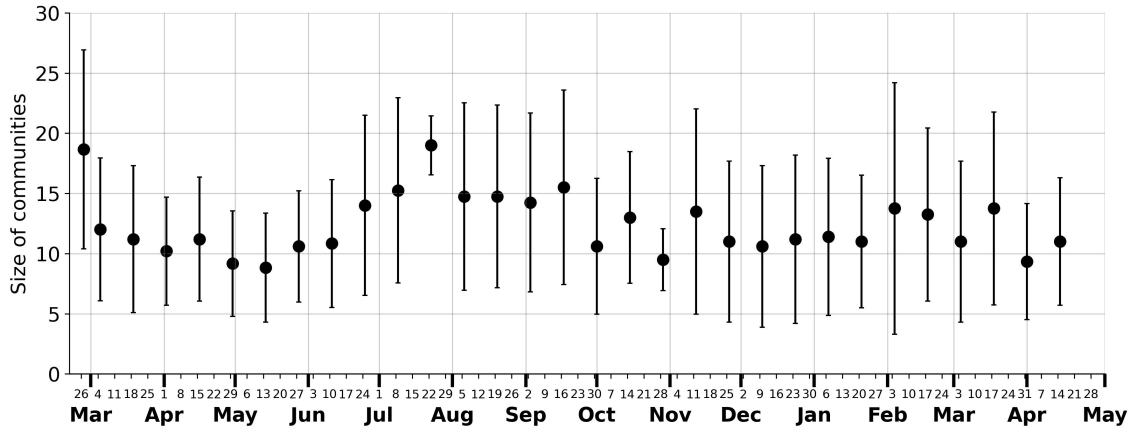


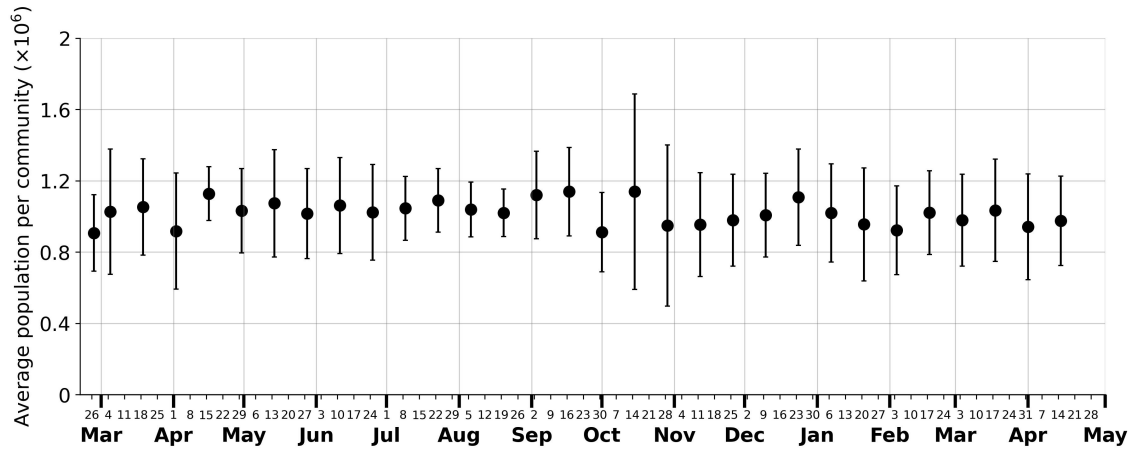
Figure 4.18: Communities evolution for different calendar weeks under study (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021). Provinces are filled in different colors when they are found to belong to multiple communities.



(a)



(b)



(c)

Figure 4.19: Evolution of the communities over the course of the pandemic (February, 2020-April, 2021). 4.19(a) shows the temporal evolution of the number of communities, 4.19(b) the evolution of their average size (in number of provinces), and 4.19(c) the evolution of their average population per community. Vertical bars represent the standard deviation.

4.3 Effect on Spreading Processes in Spain

This Section aims at studying to what extent the different changes in mobility observed throughout the COVID-19 epidemic in Spain, affect the spread of an infectious disease. For this, we implemented the SIR metapopulation model, introduced in Section 3.5.3.

Along with the changes in mobility, mediated by the different mobility networks, the model also includes the effect of different non-pharmaceutical interventions, allowing for the study of different scenarios. The first one, called **quarantine distancing**, considers that the reduction in total mobility leads to a proportional reduction in the average number of contacts. The second scenario, **quarantine isolation**, accounts for a more drastic reduction, in which an equivalent percentage of the population is assumed to be completely isolated at home, while the rest of the population does not alter their behavior. Finally, the last scenario, **no quarantine**, does not contemplate a reduction in total mobility, although the mobility network of the week under study continues to mediate the epidemic.

Once again, we emphasize that, despite using epidemiological parameters similar to those of COVID-19 in Spain ($\mathcal{R}_0$=3 and μ=1/8 d), we do not intend to reproduce the spread of the epidemic in Spain, but rather study the effects of changes in mobility in the spread of epidemics in general. In addition, the model takes into account the drastic reduction in mobility, as well as its restructuring, which is known to have substantial effects on the spread of epidemics. However, other pharmaceutical or non-pharmaceutical interventions implemented throughout the study period have important effects in mitigating the epidemic and are not reflected.

Due to the high computational demand of the simulations, we have implemented them for the set of mobility networks built from the trips of those calendar weeks selected as representative of the period studied (see Section 4.1.2). We chose to perform the simulations on this reduced set of mobility networks, which allows us to average over a greater number of realizations, thus obtaining more accurate results (the results shown are averaged over a hundred realizations unless otherwise indicated). Table 4.1 summarizes the period covered by each selected week, indicating its start and end.

Calendar week	Initial date	Final date
09	24/02/2020	01/03/2020
12	16/03/2020	22/03/2020
16	13/04/2020	19/04/2020
30	20/07/2020	26/07/2020
34	17/08/2020	23/08/2020
52	21/12/2020	27/12/2020
66	29/03/2021	04/04/2021

Table 4.1: Period covered by the different weeks for which the epidemiological model has been implemented.

4.3.1 Study on the dependence of the model with the outbreak source

We have studied the dependence of the model on the outbreak origin location. Initially, a set of I_0=100 infected individuals are introduced into a single province.

With this objective, we carried out the simulation for the different selected weeks and the different containment scenarios, but choosing different outbreak locations. In each case, we repeated the simulations for three outbreak locations (Madrid, Ceuta and Granada). These three provinces correspond to those with the highest (Madrid) and lowest (Ceuta) census , and Granada with an average census. In each case, we let the epidemic evolve until we are sure that it has ended and that, therefore, the percentage of the infected population at the end is zero.

We first focused on studying the effect on the final size of the epidemic. To do this, we measured the final percentage of the population that is in the recovered compartment. This quantity tells us which percentage of the total population has passed the infection, and therefore, it is an indicator of its size. We take into account that, in the quarantine isolation scenario, we must remove from this compartment the population that is initially considered to be recovered without having passed the infection since is effectively isolated at home without the possibility of contracting the disease. The results obtained are shown in Figure 4.20.

Moreover, we studied how the origin effects the speed at which the epidemics spreads. To do this, we looked at the time at which the peak of the incidence has been reached and compare it with its size. We show the results obtained for a particular week (week 12) in Figure 4.21. In this way, we can see if the epidemic has spread more quickly, when the maximum is at earlier times or, on the contrary, it has been slower when the peak is shifted to later times.

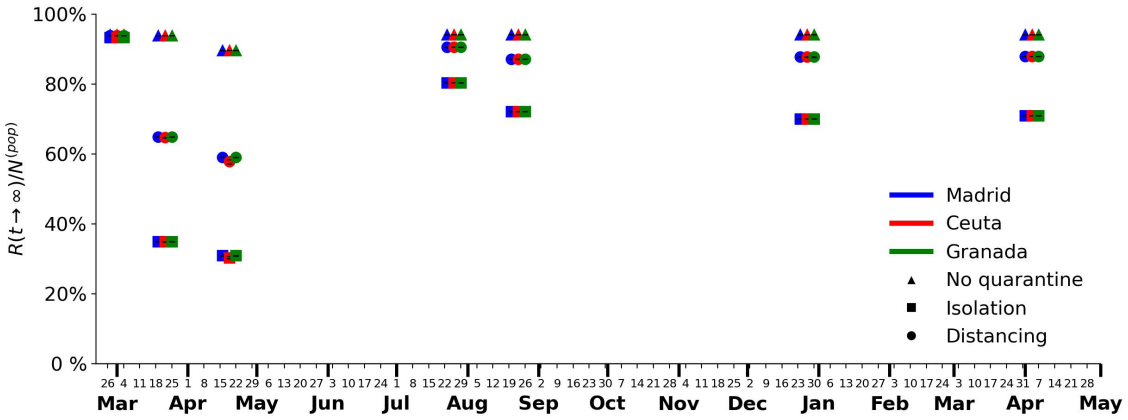
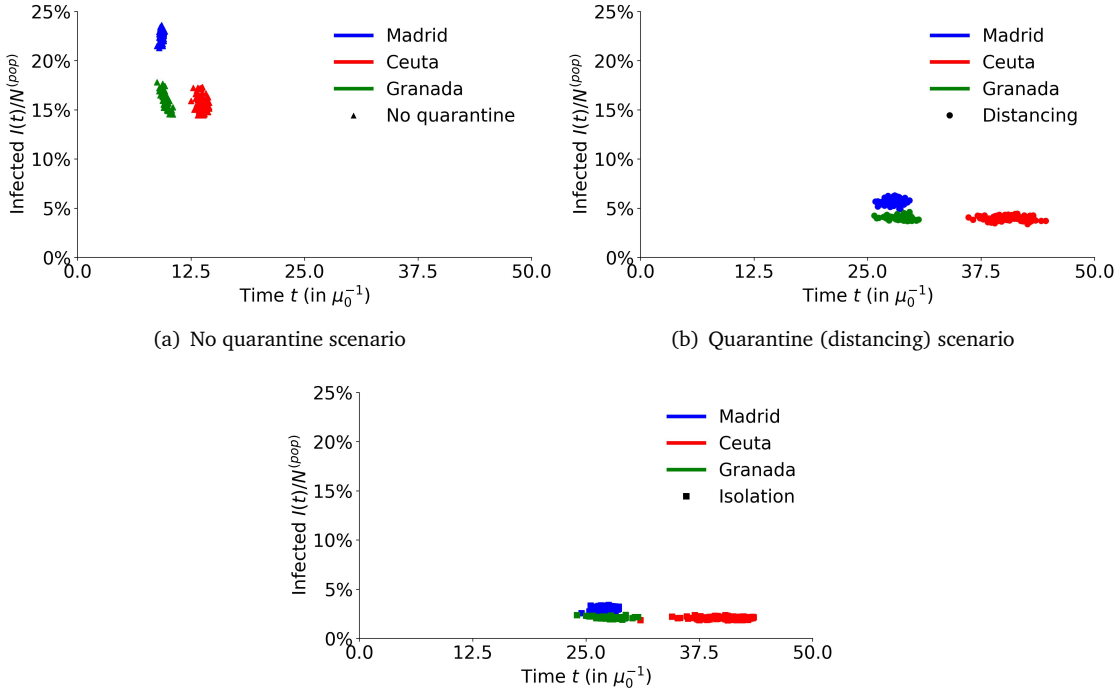


Figure 4.20: Evolution of the accumulated incidence of the epidemic $R(t \to \infty)/N^{pop}$ in the two quarantine scenarios and in the no quarantine scenario, for each weekly mobility networks G_T (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021) originating the epidemic at Madrid, Ceuta and Granada, the provinces with the maximum, minimum and intermediate census, respectively. Results are averaged over a hundred realizations. Black bars represent the standard deviation. Parameters employed in the simulation: $\mathcal{R}_0$=3, μ=1/8 d and I_0=100.

The place of onset of the epidemic is not a determining factor of the final proportion of the affected population, as can be seen in Figure 4.20, where the dependence of the final size of the epidemic in all three scenarios and province of origin for the weeks included in the selected sample, are shown. Comparing the results obtained for the different weeks, we see that those with lower mobility flows originate a smaller final size. We also observe that the final size of the epidemic is reduced based on containment scenarios that assume a reduction in traffic. This reduction being greater in the more restrictive scenario. Regarding the change in the evolution over the weeks, firstly and in accordance with what was observed previously, we see that the mobility network in the first week considered does not produce changes when the contention scenarios are imposed. This is due to the similarity between that network and the baseline, both topologically and at the flow level. Confinement scenarios are most effective in the weeks of lower mobility that temporally coincide with the most restrictive measures at the national level. Consistent with what was observed previously, this effect

is mostly lost in the summer months and is later partially recovered at Christmas and at the end of the studied period.



(a) No quarantine scenario

(b) Quarantine (distancing) scenario

(c) Quarantine (isolation) scenario

Figure 4.21: Maximum of the infected curve for one hundred simulations for the three outbreak sources (Madrid, Ceuta and Granada) obtained from the epidemiological model when using the weekly mobility networks G_{12} (March 16-22, 2020) in the no quarantine scenario 4.21(a), the distancing scenario 4.21(b) and the isolation scenario 4.21(c). Results shown for a hundred realizations. Parameters employed in the simulation: $\mathcal{R}_0$=3, μ=1/8 d and I_0=100.

Taking into account the time series of active infected for the different scenarios mediated by the mobility network of week 12 (shown in Figure 4.21), we can extract information on both the influence in the selected province of origin and the containment model used. The reason for choosing this week is that, it is the one in which the effect of the scenario used is most evident. In the first place, the peak of the outbreak is greater when the province of origin is Madrid, in the three scenarios. This maximum is reached in shorter times for both Madrid and Granada. If we look at the differences between scenarios, we can see that a reduction in the overall traffic means that the peak is delayed to later times and the curve flattens. Once again, this effect is greater in the isolation scenario.

We see that, whether the disease originates in a particular province is not as significant as the choice of containment method. However, it does affect its development by altering the times and the maximum pressure on the health system by having a higher or lower peak in the $I(t)$ curve.

4.3.2 Simulations for random outbreak source

In this Section, we performed the same analysis as in the previous Section, but taking, for each realization, a randomly chosen outbreak source. In the first place (Figure 4.22) we obtained a result consistent with what is observed in Figure 4.20, based on the different mobility networks and

using the three contention scenarios. The main difference is found in the weeks of March and April, 2020, where there is a dispersion that we did not appreciate in the previous Section. To understand this effect, we studied the analysis of the maxima of $I(t)$ shown in Figure 4.23 for the different scenarios and weeks. As we can see, for weeks 12 and 16 results present a greater dispersion, and in some cases there is no outbreak in the epidemic, which is explained by the lower connectivity of the networks in these weeks.

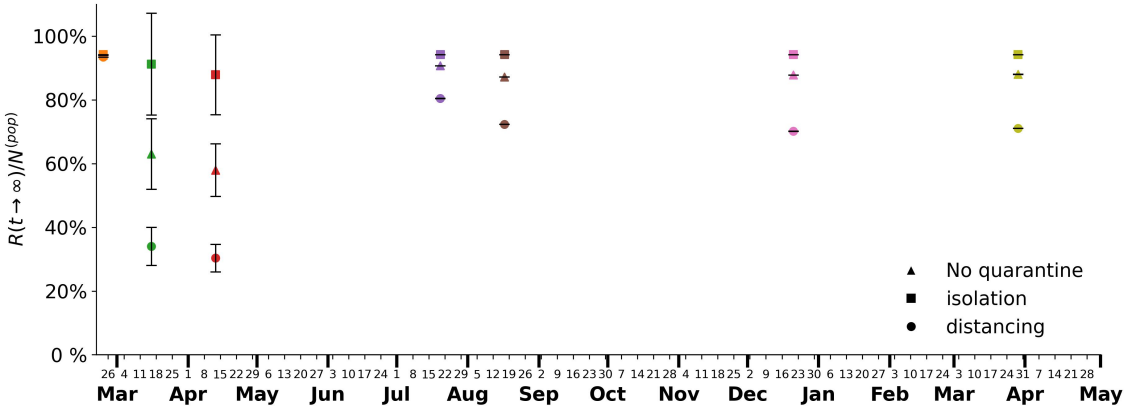


Figure 4.22: Evolution of the accumulated incidence of the epidemic $R(t \to \infty)/N^{pop}$ in the two quarantine scenarios and with no quarantine scenario, setting as random, the outbreak source for each weekly mobility networks G_T (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021). Results are averaged over a hundred realizations with a randomly selected outbreak source. Black bars represent the standard deviation. Parameters employed in the simulation: $\mathcal{R}_0$=3, μ=1/8 d and I_0=100.

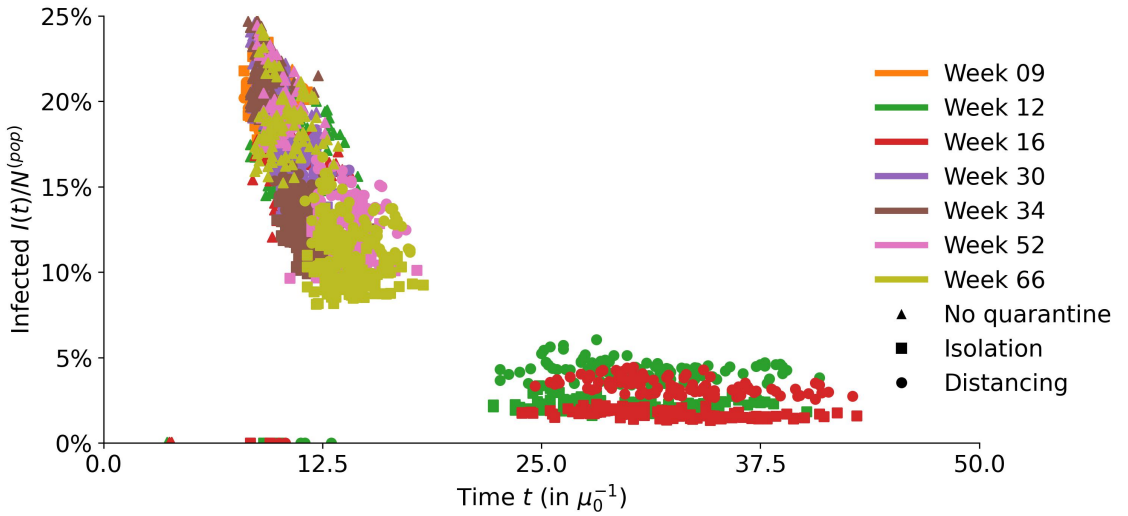


Figure 4.23: Maximum of the infected curve $I(t)$ obtained from the epidemiological model when using the weekly mobility networks G_{12} (March 16-22, 2020) in the no quarantine scenario, the distancing scenario and the isolation scenario for each of the one hundred realizations. Parameters employed in the simulation: $\mathcal{R}_0$=3, μ=1/8 d and I_0=100.

We have also studied the impact on the final size of the epidemic, sampling different values of the parameter $\mathcal{R}_0$ in the range $(0, 5]$, left panels of Figure 4.24. The results obtained from the respective rescaled networks (these networks are topologically the same as the weekly mobility network of the baseline T_0 but the edge weights have been rescaled in such a way that the average

total flow observed corresponds to that of each week T, see 3.3.2) are also shown (right panels of Figure 4.24). Results are averaged over 50 realizations. The epidemiological threshold shifts upward when containment scenarios are taken into account. In both distancing and isolation scenarios, the threshold appears at the same values of $\mathcal{R}_0$, but differences in the final size of the epidemic are observed for higher values of the threshold, in agreement with what was observed previously. We observe the same behavior when the networks are rescaled, showing that this effect is explained only by a global reduction of mobility, and not by the structural changes in the networks.

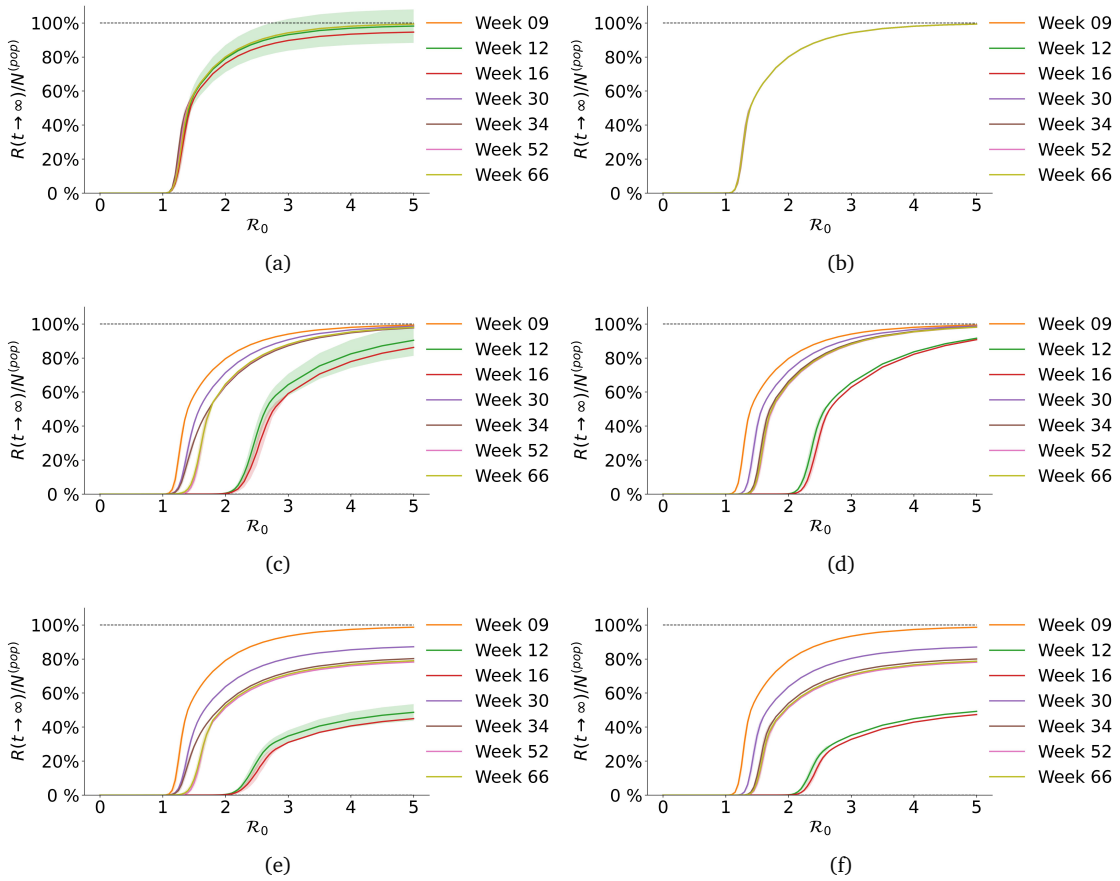
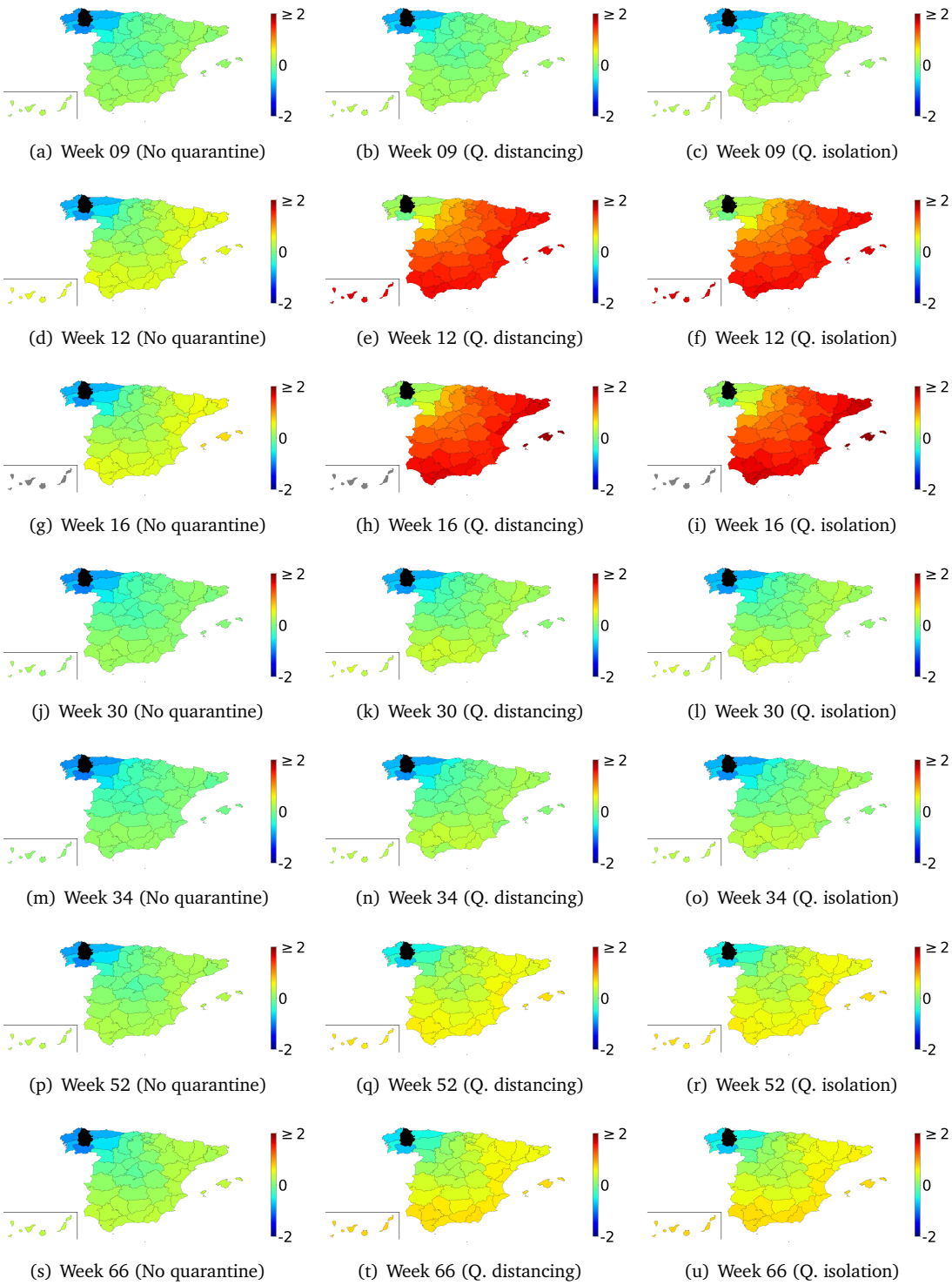


Figure 4.24: Final incidence of the epidemic $R(t\to\infty)/N^{pop}$ as a function of $\mathcal{R}_0$ for different weekly mobility networks G_T (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021) and the corresponding rescaled mobility networks $G^R_{10}(T)$. In the no quarantine 4.24(a), no quarantine rescaled 4.24(b), quarantine distancing 4.24(c), quarantine distancing rescaled 4.24(d), quarantine isolation 4.24(e) and quarantine isolation rescaled 4.24(f) scenarios. Shaded area represents the standard deviation. Each point is averaged over 50 realizations for a randomly selected outbreak source. Parameters employed in the simulation: μ=1/8 d and I_0=100.

4.3.3 Geographical spreading of the epidemic. A single random outbreak source

Finally, we studied the geographical spread of the epidemic. To do this, we repeated the previous simulations but selecting a single random outbreak origin and record how long it took for each province to become infected. This magnitude is measured with the arrival time, t^*, defined as the first time in which those infected in a location exceed the threshold of 0.1% of the population in the location.

Before the start of the restrictions (week 09), and during the summer (weeks 30 and 34), we find that the arrival times are very similar to those obtained with the baseline network. However, both in the weeks of March and April, 2020 (weeks 12 and 16), at Christmas (week 52), and in the last one considered week (week 66), these times increased substantially with geographical distance. Moreover, distancing and isolation scenarios exacerbate this effect. These results can be seen in Figure 4.25, where the geographical spreading in the country, representing the arrival time relative to the corresponding one in the baseline week in the no-quarantine scenario, is shown.



(a) Week 09 (No quarantine) (b) Week 09 (Q. distancing) (c) Week 09 (Q. isolation)

(d) Week 12 (No quarantine) (e) Week 12 (Q. distancing) (f) Week 12 (Q. isolation)

(g) Week 16 (No quarantine) (h) Week 16 (Q. distancing) (i) Week 16 (Q. isolation)

(j) Week 30 (No quarantine) (k) Week 30 (Q. distancing) (l) Week 30 (Q. isolation)

(m) Week 34 (No quarantine) (n) Week 34 (Q. distancing) (o) Week 34 (Q. isolation)

(p) Week 52 (No quarantine) (q) Week 52 (Q. distancing) (r) Week 52 (Q. isolation)

(s) Week 66 (No quarantine) (t) Week 66 (Q. distancing) (u) Week 66 (Q. isolation)

Figure 4.25: Logarithm of arrival times t^* relative to the values in the no quarantine scenario of G_{10} (March 2-8, 2020), $\log\left(t^*/t^*(0)\right)$ for a single random outbreak origin (black) and averaged over a hundred realizations for different weekly mobility networks (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021). Provinces in grey indicate that the epidemic does not arrive at a finite time due to the lack of connections. Parameters employed in the simulation: $\mathcal{R}_0$=3, μ=1/8 d and I_0=100.

In the second and last place, we studied the arrival times varies as a function of the distance to the

outbreak source origin (Figure 4.26). The arrival times show a strong and increasing dependence on distance in the weeks with lower mobility (12 and 16). During the weeks prior to the pandemic (09), as well as in the summer months (30 and 34), arrival times increase slowly with geographical distance and saturate at approximately 400km. Both at Christmas and at the end of the considered period (weeks 52 and 66, respectively), the dependence saturates from 400km, but is considerably stronger than last weeks commented. Considering the results obtained for the rescaled networks (Figures 4.26(b), 4.26(d) and 4.26(f)), we see that the changed scaling relation of the arrival times is not replicated in the weeks of lower mobility and in the last two considered. This indicates that it is not caused by a reduction in the total flow. In addition, if we compare these results with those shown in Figure 4.17, we see that the geographical spread of the epidemic shows a similar functional dependence with distance as the shortest path. This suggests that the structural changes observed during these weeks, have a significant impact on geographic spreading of the disease.

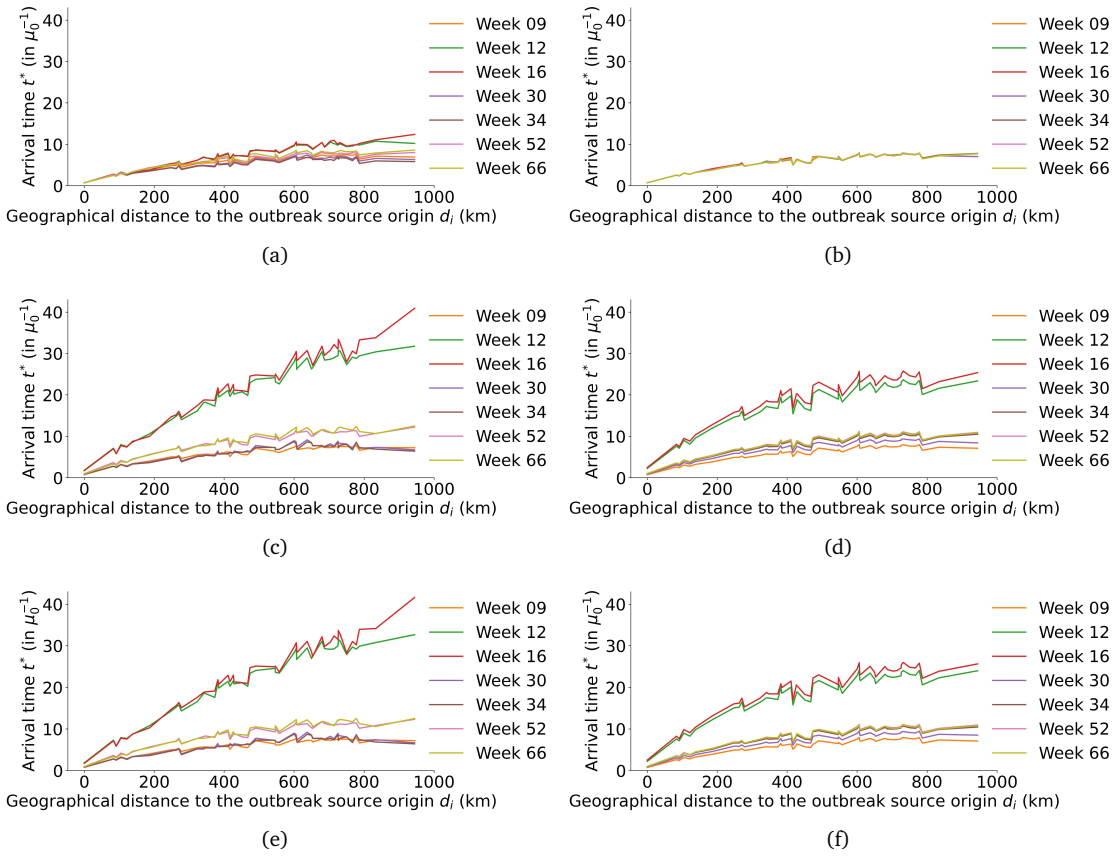


Figure 4.26: The average arrival time t^* in provinces as a function of the geographic distance d_{i^*} from the outbreak origin i^* for different weekly mobility networks G_T (weeks 09, 12, 16, 30, 34, 52, and 66) (February, 2020-April, 2021) and the corresponding rescaled mobility networks $G_{10}^R(T)$. Results are shown for the no quarantine 4.26(a), no quarantine rescaled 4.26(b), quarantine distancing 4.26(c), quarantine distancing rescaled 4.26(d), quarantine isolation 4.26(e) and quarantine isolation rescaled 4.26(f) scenarios. Shaded area represents the standard error. Results for a single random outbreak origin. Parameters employed in the simulation: $\mathcal{R}_0$=3, μ=1/8 d and I_0=100.

Chapter 5

Conclusions

The onset of COVID-19, at the end of 2019, had important and varied impacts on our lives. Among them, of course, the great loss of human lives. It has also affected the way people interact, with an impact on the way we commute and travel and therefore in many other socioeconomic aspects. Since the beginning of this global crisis, there has been strong evidence pointing to airborne transmission as the main form of contagion for COVID-19, with individual proximity and closed environments as the major infection routes. As a result, many countries have implemented containment policies aimed at controlling the spread of the disease at all scales. From limiting long-range international travels to reducing urban mobility and finally lockdowns. Even if, the different restrictions proved, in many cases, to be effective in mitigating the spread of the disease they implied disruptions at the social and economic levels rarely seen in human history. For these reasons, there is a need to understand their effects on our society in detail and to better face similar future scenarios.

In this Master's Thesis, we study how containment measures and other changes in people's behavior have affected mobility in Spain and Germany. Knowledge on dynamic changes in human mobility, and spatial interaction models is critical for understanding and predicting the dynamics of COVID-19 infection. For the case of Germany, we focused on the first months (till June, 2020) of the epidemics, whereas in Spain, we extended the period covered to over a year later (till April, 2021). To this end, we analyzed changes in the total mobility flows in both countries in relation to a reference period with "normal" mobility (March 2 to 8, 2020 for Spain and March, 2019 for Germany). We also studied this change in relation to the distance travelled for assessing whether it happened homogeneously. Changes in mobility flows influence mobility networks. Hence, in the case of Spain, we studied whether mobility networks are simply less dense, or on the contrary, their structure have been affected. For this purpose, we used the tools provided by Network Science. Finally, using a numerical epidemiological model, we analyzed the impact of these changes on the spread of an epidemics in Spain, which allows us to understand more in depth what is the role of human mobility in an epidemic.

We used mobility flows, gathered from mobile phone data that account for the number of trips between each pair of locations in Germany and Spain. For Germany, we focused on flows between the 401 counties that compose the country. For Spain, the data is available at the municipal level (there are 8,131 municipalities). Moreover, also for Spain, we extended the period covered beyond the first wave of the epidemic. This extension has allowed to study the differences between the first four waves of the COVID-19 pandemic. In order to study how these changes affect the structure of mobility networks, we built weekly mobility networks, where nodes represent all possible locations, and edge weights the average flow along the connection. We performed this analysis at the level of the 52 provinces of Spain. Additionally, for sake of comparison, we built weekly rescaled mobility networks, that are structurally similar to those of the reference period, but have the same total amount of trips as the corresponding week. This allowed us to isolate the effects of reduced mobility and structural changes in the networks. To characterize the networks, we used various measures

from Network Theory, that allowed us to quantify changes over time in their ability to support diffusion and Small-World behavior. Finally, we have implemented a SIR metapopulation model with containment, that takes into account the drastic reduction in total mobility. It assumes that reduced mobility results in a lower rate at which contacts between infected and susceptible individuals leads to infection. Two possibilities to reduce mobility are considered, a social distancing scenario and a more drastic isolation mimicking lockdown measures.

Studying mobility trends, we found that traffic was effectively reduced during the COVID-19 pandemic both countries. The largest reduction took place in the middle of March, 2020, coinciding with the impositions of the vast majority of restrictive measures. In Germany, pre-pandemic values were reached at the beginning of June, 2020, while in Spain, pre-pandemic values are never reached, in the 7-d moving average, throughout the considered period. This effect is stronger the greater the distance traveled. In Spain, over a year later, long-distance flows have not been fully restored and present a greater reduction compared to short-distance flows, with the exception of summer 2020. During this period, with the easing of containment policies and summer vacations, long-haul travel increased significantly, exceeding the reference values. Moreover, when comparing flows of long-distance trips with the number of individuals recorded as infected in Spain, we found that long-distance travels play an important role in the development of the epidemic at a national scale, during its early stages. Other key factors, such as easing containment policies, are also determinant in the evolution of the epidemic. For that, we have considered periods when a linear correlation between the two amounts is a valid first approximation. We found an optimum delay of 47 days between the long-distance trips and the number of infected individuals. In addition, the Pearson correlation remained above 0.7 for time lags greater than 14 days.

To identify key structural changes over time, we analyzed the mobility networks, for various calendar weeks throughout the covered period. By studying the distributions of the degree and strength, as well as their changes over time, we find that, even if flows are practically restored, many connections are still lost. In agreement with the above mentioned, the mobility network have fewer long-distance connections compared to the pre-pandemic values. The observed structural changes in mobility have an impact on the network's properties, generally associated with their Small-World nature. We observed that paths between locations are generally longer, since more local were needed to be included when traveling. Consequently, the average shortest path length increased, being this effect more important during the first epidemic wave. This is also confirmed by the study of relaxing and mixing times. Abnormal values of the average shortest path length and relaxation and mixing time are measured throughout the considered period. When we looked at these quantities for the corresponding rescaled mobility networks, we find that they cannot be explained a uniform reduction of mobility alone, which underlines that the observed mobility changes are mainly of topological nature. We also examined how the meso-scale organization of the networks evolved over time. Results confirmed that the networks were more fragmented, resulting in a greater number of communities. Once again, this loss of connectivity remains until the end of the studied period, where we found that the network is divided into twice as many communities as there were before the pandemic. We conclude that, since the implementation of the containment measures began, there are mainly local connections and fewer long-haul connections. Hence, evidences of profound structural changes, that remain present in the long-term, are found. The resulting mobility networks are less dense, more clustered and local and hence more homogeneous.

To study the effect that the different changes have on the spread of epidemics, we have implemented an epidemiological metapopulation model. The model, based on the SIR compartmental model, takes into account different contention scenarios that account for a drastic reduction in mobility flows. Specifically, we consider three different scenarios, a normal one, without containment, and two containment scenarios, implementing social distancing and a stricter isolation. We found that, lockdown measures have a significant impact on epidemic spreading. A reduction in mobility results an abatement in the total impact of the epidemic. The epidemic curve is flattened and the peak is shifted to later times. This applies to both containment scenarios and, as expected, the effect is greater in the more stringent situation. Additionally, confinement scenarios are more effective

during those periods of lower mobility. Finally, the geographic spread of the outbreak was analyzed. We found that a reduction in long-distance travels reduces the geographic spread of the epidemic. This effect is mainly due to the structural changes present in the mobility networks. We therefore conclude that the different changes in the mobility networks have a direct and significant impact on dynamic processes such as the spread of epidemics.

The research conducted as part of this Master's Thesis, is a first approach towards understanding how human mobility responds to restrictive measures. We also investigate how these changes persist over time. We hope that this preliminary work will facilitate further investigations in this area. An interesting approach would be to look at these changes at a lower spatial scale. In Spain, confinement measures are not applied or withdrawn at the national level, provincial authorities have considerable room for action. Therefore, one could study how the different measures affect the epidemic locally and how this ultimately affects both the structure of the mobility network and the epidemic at the national level. In the same way, and given the importance of long-distance travel, in the spread of epidemics, it would be interesting to extend the study to the scale of countries or continents. A clear reflection of this importance was advanced at the introduction of this work, where it is stated that the first cases of COVID-19 registered in Spain were imported from other countries. Furthermore, the period covered by our study runs until April, 2021. However, the COVID-19 epidemic persists over one year later. We observe that the changes in the network are still present at the end of the covered period. At the time of writing these conclusions, six waves of the COVID-19 have been registered in Spain, and we are very likely at the gates of the seventh. Extending the period covered by this study would allow to know whether the network has finally returned, both at a flows and topological level, to the pre-pandemic situation. This may provide a long-term perspective on the impact of containment measures on human mobility.

From the late 19th century to the late 20th century, five major pandemics threatened mankind. However, in the first 20 years of the 21st century, we have already experienced four large pandemics. This increase in the frequency of epidemics highlights the importance of understanding the effects of restrictive policies on human mobility. These results would be of great help to face epidemics that are very likely to emerge in the coming years. From our analysis, we find strong evidences of long-lasting, disease-mitigating structural changes in mobility networks. Results are consistent with that of [15], though on a smaller scale, as the study performed in Germany only covers the first wave of the epidemic, where the greatest reduction of mobility took place. All of this suggests that specific mobility restrictions, that target long-distance connections, should be used to limit the spread of a disease. This may help us alleviate the current COVID-19 pandemic, prepare for other similar scenarios and prevent us from experiencing another massive lockdown. "*What is not defined cannot be measured. What is not measured, cannot be improved.*", William T. Kelvin.

Bibliography

[1] Ricard Ferrer. "Pandemia por COVID-19: el mayor reto de la historia del intensivismo". *Medicina Intensiva* **44** (2020). DOI: 10.1016/j.medin.2020.04.002.

[2] *Centers for Disease Control and Prevention (CDC): Basics of Covid-19*. https://www.cdc.gov/coronavirus/2019-ncov/your-health/about-covid-19/basics-covid-19.html. (Last accessed: December 31, 2021). 2021.

[3] *World Health organization*. https://www.who.int/en/activities/tracking-SARS-CoV-2-variants/. (Last accessed: December 31, 2021).

[4] *Ministerio de Sanidad: Información Científica-Técnica. Parámetros epidemiológicos*. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/20210810_EPIDEMIOLOGIA.pdf. (Last accessed: January 1, 2022). 2021.

[5] Jian Wu et al. "Household Transmission of SARS-CoV-2, Zhuhai, China, 2020". *Clinical Infectious Diseases* **71** (2020), pages 2099–2108. DOI: 10.1093/cid/ciaa557.

[6] Qun Li et al. "Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia". *New England Journal of Medicine* **382** (2020). PMID: 31995857, pages 1199–1207. DOI: 10.1056/NEJMoa2001316.

[7] Katherine Randall et al. "How Did We Get Here: What Are Droplets and Aerosols and How Far Do They Go? A Historical Perspective on the Transmission of Respiratory Infectious Diseases". *A Historical Perspective on the Transmission of Respiratory Infectious Diseases (April 15, 2021)* (2021).

[8] Trisha Greenhalgh et al. "Ten scientific reasons in support of airborne transmission of SARS-CoV-2". *The lancet* **397** (2021), pages 1603–1605.

[9] *Ministerio de Sanidad: Evaluación del Riesgo de la Transmisión de SARS-CoV-2 mediante Aerosoles. Medidas de prevención y recomendaciones, documento Técnico*. https://www.sanidad.gob.es/profesionales/saludPublica/ccayes/alertasActual/nCov/documentos/COVID19_Aerosoles.pdf. (Last accessed: January 1, 2022). 2021.

[10] Byung Uk Lee. "Minimum Sizes of Respiratory Particles Carrying SARS-CoV-2 and the Possibility of Aerosol Generation". *International Journal of Environmental Research and Public Health* **17** (2020). DOI: 10.3390/ijerph17196960.

[11] Joseph G Allen and Andrew M Ibrahim. "Indoor Air Changes and Potential Implications for SARS-CoV-2 Transmission". *JAMA* **325** (2021), pages 2112–2113.

[12] R.M. Anderson and R.M. May. *Infectious Diseases of Humans: Dynamics and Control*. Infectious Diseases of Humans: Dynamics and Control. OUP Oxford, 1992.

[13] Paolo Bajardi et al. "Human Mobility Networks, Travel Restrictions, and the Global Spread of 2009 H1N1 Pandemic". *PLOS ONE* **6** (2011), pages 1–8. DOI: 10.1371/journal.pone.0016591.

[14] Steven Riley. "Large-Scale Spatial-Transmission Models of Infectious Disease". *Science* **316** (2007), pages 1298–1301. DOI: 10.1126/science.1134695.

[15] Frank Schlosser et al. "COVID-19 lockdown induces disease-mitigating structural changes in mobility networks". *Proceedings of the National Academy of Sciences* **117** (2020), pages 32883–32890. DOI: 10.1073/pnas.2012326117.

[16] Vitaly Belik, Theo Geisel, and D. Brockmann. "Natural Human Mobility Patterns and Spatial Spread of Infectious Diseases". *Physical Review X* **1** (2011). DOI: 10.1103/PhysRevX.1.011001.

[17] Carlos Castillo-Chavez, Derdei Bichara, and Benjamin R. Morin. "Perspectives on the role of mobility, behavior, and time scales in the spread of diseases". *Proceedings of the National Academy of Sciences* **113** (2016), pages 14582–14588. DOI: 10.1073/pnas.1604994113.

[18] Alex Arenas et al. "A mathematical model for the spatiotemporal epidemic spreading of COVID19". *medRxiv* (2020). DOI: 10.1101/2020.03.21.20040022.

[19] Dilip Kumar Bagal et al. "Estimating the parameters of susceptible-infected-recovered model of COVID-19 cases in India during lockdown periods". *Chaos, Solitons & Fractals* **140** (2020), page 110154. DOI: https://doi.org/10.1016/j.chaos.2020.110154.

[20] Matteo Chinazzi et al. "The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak". *Science* **368** (2020), pages 395–400.

[21] *Equipo COVID-19-RENAVE-CNE-CNM (ISCIII): Informe no 109 Situación de COVID-19 en España a 15 de diciembre de 2021*. https://www.isciii.es/QueHacemos/Servicios/VigilanciaSaludPublicaRENAVE/EnfermedadesTransmisibles/Paginas/InformesCOVID-19.aspx. (Last accessed: January 8, 2022). 2021.

[22] Oxford University. *Our World in Data*. https://ourworldindata.org/covid-vaccinations?country=OWID_WRL. (Last accessed: May 9, 2022).

[23] *Ministerio de Sanidad: Covid en España*. https://covid19.isciii.es/. (Last accessed: December 31, 2021). 2021.

[24] Lijing Wang et al. "Using Mobility Data to Understand and Forecast COVID19 Dynamics". *medRxiv* (2020). DOI: 10.1101/2020.12.13.20248129.

[25] William Ogilvy Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics". *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* **115** (1927), pages 700–721.

[26] Mark EJ Newman. "Spread of epidemic disease on networks". *Physical review E* **66** (2002), page 016128.

[27] Nuria Oliver et al. *Mobile phone data for informing public health actions across the COVID-19 pandemic life cycle*. 2020.

[28] Benjamin F Maier and Dirk Brockmann. "Effective containment explains subexponential growth in recent confirmed COVID-19 cases in China". *Science* **368** (2020), pages 742–746.

[29] Joseph Chadi Lemaitre et al. "Assessing the impact of non-pharmaceutical interventions on SARS-CoV-2 transmission in Switzerland". *medRxiv* (2020).

[30] Giulia Pullano et al. "Population mobility reductions during COVID-19 epidemic in France under lockdown". *MedRxiv* **29** (2020), page 2020.

[31] Jayson S Jia et al. "Population flow drives spatio-temporal distribution of COVID-19 in China". *Nature* **582** (2020), pages 389–394.

[32] Lei Zhang et al. "Human Mobility Trends during the COVID-19 Pandemic in the United States" (2020).

[33] Emanuele Pepe et al. "COVID-19 outbreak response: a first assessment of mobility changes in Italy following national lockdown". *MedRxiv* (2020).

[34] Alessandro Galeazzi et al. "Human mobility in response to COVID-19 in France, Italy and UK". *Scientific Reports* **11** (2021), pages 1–10.

[35] Jonas Dehning et al. "Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions". *Science* **369** (2020).

[36] Seth Flaxman et al. "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe". *Nature* **584** (2020), pages 257–261.

[37] Romualdo Pastor-Satorras et al. "Epidemic processes in complex networks". *Rev. Mod. Phys.* **87** (3 2015), pages 925–979. DOI: 10.1103/RevModPhys.87.925.

[38] Song Gao et al. “Mapping county-level mobility pattern changes in the United States in response to COVID-19”. *SIGSpatial Special* **12** (2020), pages 16–26.

[39] Alex Smolyak et al. “Effects of mobility restrictions during COVID19 in Italy”. *Scientific reports* **11** (2021), pages 1–15.

[40] Frank Schlosser et al. *Covid-19 Mobility Germany*. https://osf.io/n53cz/. (Last accessed: May 9, 2021). 2021. DOI: 10.17605/OSF.IO/N53CZ.

[41] *Ministerio de Transportes, Movilidad y Agenda Urbana*. https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad. (Last accessed: May 9, 2021).

[42] *Opendatasoft*. https://public.opendatasoft.com/explore/dataset/provincias-espanolas/table/?sort=provincia. (Last accessed: May 9, 2021).

[43] *Instituto de Salud Carlos III*. https://www.isciii.es/Paginas/Inicio.aspx. (Last accessed: May 9, 2021).

[44] *Instituto Nacional de Estadística*. https://www.ine.es/. (Last accessed: May 9, 2021).

[45] Rezania Azdy and Febriyanti Darnis. “Use of Haversine Formula in Finding Distance Between Temporary Shelter and Waste End Processing Sites”. *Journal of Physics: Conference Series* **1500** (2020), page 012104. DOI: 10.1088/1742-6596/1500/1/012104.

[46] Hamada Badr et al. “Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study”. *The Lancet Infectious Diseases* **20** (2020). DOI: 10.1016/S1473-3099(20)30553-3.

[47] Jacob Benesty et al. “Pearson correlation coefficient”. In: *Noise reduction in speech processing*. Springer, 2009, pages 1–4.

[48] Ernesto Estrada and Philip A Knight. *A first course in network theory*. Oxford University Press, USA, 2015.

[49] E. Estrada. *The Structure of Complex Networks: Theory and Applications*. EBSCO ebook academic collection. OUP Oxford, 2012.

[50] A.L. Barabási and M. Pósfai. *Network Science*. Cambridge University Press, 2016.

[51] M. Newman. *Networks: An Introduction*. OUP Oxford, 2010.

[52] A. Barrat, M. Barthélemy, and A. Vespignani. *Dynamical Processes on Complex Networks*. Cambridge University Press, 2008.

[53] Shi-Ze GUO et al. “Strength-Strength and Strength-Degree Correlation Measures for Directed Weighted Complex Network Analysis”. *IEICE Transactions on Information and Systems* **E94.D** (2011), pages 2284–2287. DOI: 10.1587/transinf.E94.D.2284.

[54] Guido Caldarelli. “Complex Networks: Principles, Methods and Applications by Vito Latora, Vincenzo Nicosia and Giovanni Russo”. *Journal of Complex Networks* **6** (2018), pages 830–830. DOI: 10.1093/comnet/cnx062.

[55] Edsger W. Dijkstra. “A note on two problems in connexion with graphs”. *Numerische Mathematik* **1** (1959), pages 269–271.

[56] M. Newman. *Networks*. OUP Oxford, 2018.

[57] Bojan Mohar. “Some Applications of Laplace Eigenvalues of Graphs”. *Graph Symmetry: Algebraic Methods and Applicaions. NATO ASI Ser. C* **197** (2001). DOI: 10.1007/978-94-015-8937-6_6.

[58] Naoki Masuda, Mason A. Porter, and Renaud Lambiotte. “Random walks and diffusion on networks”. *Physics Reports* **716-717** (2017). Random walks and diffusion on networks, pages 1–58. DOI: https://doi.org/10.1016/j.physrep.2017.07.007.

[59] Giorgio Fagiolo. “Clustering in Complex Directed Networks”. *Physical review. E, Statistical, nonlinear, and soft matter physics* **76** (2007), page 026107. DOI: 10.1103/PhysRevE.76.026107.

[60] Jari Saramäki et al. “Generalizations of the clustering coefficient to weighted complex networks”. *Physical Review E* **75** (2007), page 027105. DOI: 10.1103/PhysRevE.75.027105.

[61] Jukka-Pekka Onnela et al. “Intensity and coherence of motifs in weighted complex networks”. *Phys. Rev. E* **71** (6 2005), page 065103. DOI: 10.1103/PhysRevE.71.065103.

[62] Andrea Lancichinetti et al. “Finding Statistically Significant Communities in Networks”. *PLOS ONE* **6** (2011), pages 1–18. DOI: 10.1371/journal.pone.0018961.

[63] Fred Brauer. “Mathematical epidemiology: Past, present, and future”. *Infectious Disease Modelling* **2** (2017), pages 113–127.

[64] Maia Martcheva. *An introduction to mathematical epidemiology*. Volume 61. Springer, 2015.

[65] J.D. Murray. *Mathematical Biology: I. An Introduction*. Interdisciplinary Applied Mathematics. Springer New York, 2013.

[66] Fred Brauer. “Compartmental models in epidemiology”. In: *Mathematical epidemiology*. Springer, 2008, pages 19–79.

[67] Daniel T. Citron et al. “Comparing metapopulation dynamics of infectious diseases under different models of human movement”. *Proceedings of the National Academy of Sciences* **118** (2021). DOI: 10.1073/pnas.2007488118.

[68] Frank Schlosser and Benjamin F Maier. *EpiCommute*. https://github.com/franksh/EpiCommute. 2020.

[69] Michele Tizzoni et al. “On the use of human mobility proxies for modeling epidemics”. *PLoS computational biology* **10** (2014), e1003716.