Individual based Markov model of virus diffusion: comparison with Covid19 incubation period, serial interval and regional time series

Franco Flandoli[1], Eleonora La Fauci[2], Martina Riva[2]

[1] Scuola Normale Superiore, Pisa, Italy

[2] University of Pisa, Italy

January 13, 2021

Abstract

A Markov chain individual based model for virus diffusion is investigated. Both the virus growth within an individual and the complexity of the contagion within a population are taken into account. A careful work of parameter choice is performed. The model captures very well the statistical variability of quantities like the incubation period, the serial interval and the time series of infected people in Tuscany towns.

Keywords: multiscale modeling; Markov Chains; Cellular Automaton; COVID-19; incubation period; serial interval; population graph; social barriers; uncertainty quantification.

AMS Subject Classification 2020: 92C60, 92D30, 60J20

1 Introduction

1.1 Purpose of the work

Covid-19 massive diffusion all over the world motivated the development and investigation of several mathematical models. Many of them are based on differential equations, ordinary or partial, deterministic or stochastic, or based on other stochastic processes and dynamical systems. Among those that use differential equations we cite [41], [23], [44]. All of these papers are based on the fundamental SIR model introduced by Kermack and McKendrick in [28]. Other examples are network models that can be deterministic or stochastic; some examples can be found in [27], [14], [39], [42]; see also [1], [15], [21] for other approaches.

As widely explained in [3] a possible approach in mathematical modeling of an epidemics has to take into account both the complex biology of the virus and the spatial interactions among a population. According to this point of view in our model we stress out the importance to include the different phases of virus growth within an individual and the population graph that allows more realistic contagion between infected and susceptible individuals. Additional comments on this multiscale approach and its relation with the literature are given in Section 1.3 below.

We have developed an individual based model (a particle system) governed by a Markov chain dynamics. The states of individuals and the transitions between them remind those of certain generalizations of SIR model; the interaction between individuals has something in common with network structure; but apart from these conceptual similarities the model is different. This work summarizes our understanding of the performances of this Markov chain individual based model.

We have chosen an individual based model for two reasons:

i) several parameters have a direct interpretation, being associated to individual life

ii) the structure of interaction may include realistic features.

The same reasons may be the source of difficulties:

a) too many parameters to be fitted or chosen

b) too many degrees of freedom and complexity of the interactions, yielding computational troubles.

Let us discuss to which extent we may confirm the advantages (i)-(ii) and which implications they have. Let us also discuss in which measure we have observed the disadvantages (a) and (b) and how much they affect the advantages.

Prediction under new operating conditions is one of the major problems of every government. Usual models like SIR cannot be used for predictions under new conditions; they are excellent to fit existing data and make short term predictions under stationary conditions. More complex models, like individual bases ones, have the drawback of a greater multiplicity of parameters but many of them have a direct interpretation which allows to assign sufficiently robust values to them. Unfortunately, our final conclusion about our model is that there is a parameter, called λ below, which cannot be predicted so easily; we may estimate it by fitting the data, but it is difficult to guess it a priori under a new operating condition (see our analysis below of the predictions after May the 4th, in Tuscany, when the lock-down imposed in Italy in March-April was relaxed). The other parameters however can be chosen a priori and contribute to stabilize most of the model; perhaps after a small number of new regimes one can extract a rule to decide also λ a priori.

Predicting the uncertainty is another fundamental issue, even more difficult sometimes. Our model performs very well from this viewpoint. We do not have to fit uncertainty parameters: a degree of dispersion of the results is intrinsic in the Markov chain dynamics and it is in excellent agreement with real data: we consider the behaviour of Tuscany towns, scaled to 100.000 inhabitants, as a sample, which provide us with an estimate of the true uncertainty. Notice that a prediction of the uncertainty is just absent in most virus spread models or sometimes it is inserted in a rather artificial way by a noise or via additional parameters which have to be fitted by data. In our case it comes for free and it is quite realistic.

Due to its complexity, our model not only allows one to compute usual quantities like the cumulative number of infected people, but also to get more refined information like the statistics of the incubation period and of the serial interval. Similarly to the other remarks above, also here we need information from real data concerning the mean incubation period and mean serial interval in order to fit some parameters, but then the dispersion and more generally the histogram obtained by our simulations, without any other choice of parameters are very close to real data. We are indebted to several studies for these data, see [20], [22], [34],[31], [33], [35], [37], [40], [43], [45], [46].

Concerning the potential drawback (b) above, we had to limit ourselves to environments of the size of a town like Pisa or other Tuscany towns. Completely different computational methods would be needed to treat a full nation, for instance. As a final remark, the model can be used to explore different scenarios for different values of parameters. The detail of realism of the model allows one to vary quite detailed and realistic parameters. An example is given at the end of the paper.

1.2 On the biology of the virus

In the last few months, there have been a huge amount of scientific papers dealing with the complex biology of COVID-19. Due to the incomplete knowledge in the field there are different points of view on many aspects. Therefore, we are aware that our discussion can not be exhaustive.
In order to outline some of the key aspects of the illness, which will be useful to understand the mathematical model we'll present in the following sections, we refer to the report [12], which is bimonthly updated, and to the papers [3] and [13]. A brief description of the biology of the virus is delivered in the following:

- *The virus*: Coronavirus disease 2019 (COVID-19) is initiated by the infection of the SARS-CoV-2 virus, a coronavirus. Coronaviruses are a large family of viruses that cause illness ranging from the common cold which usually occurs in the winter months to more severe diseases such as Middle East Respiratory Syndrome (MERS), Severe Acute Respiratory Syndrome (SARS) and COVID-19.

- *The contagion*: The contagion by the SARS-CoV-2 occurs when breathing respiratory droplets, up to 1 millimetre across, released by an infected person via coughing, sneezing or speaking. Respiratory droplets finally land on various surfaces where the virus maintains its infective capacity for various times. Therefore, SARS-CoV-2 can also be transmitted by contact when a susceptible person touches the mucosa of the mouth, nose or eye after capturing the virus.

- *Entry into human cell*: The large Spike protein (S) forms a sort of crown on the surface of the viral particles. Its receptor-binding domain interacts with high affinity with angiotensin-converting enzyme 2 (ACE2) receptors on the surface of host cells. After the binding, two host cell proteases (Furin and TMPRLRS) cleave spike proteins and their exposed fusion peptides fuse the virus membrane with the membrane of the host cells. The virus RNA enters cells of the upper and lower respiratory tract, and it is translated into viral proteins. The cell dies releasing millions of new viruses that infect other cells and other individuals.

- *Immune system responses*: Cellular and humoral innate immunity represents a first line of resistance which takes care of most encounters with infectious agents. When the virus manages to overcome these barriers, a rapid release of danger signals activates the reaction of the host's immunity. Corona viruses are successful at suppressing various mechanisms in an immune response, but a protective immunity can be however developed.

- *Immunity*: There is evidence that symptomatic COVID-19 elicits immunological memory and resistance to reinfection. Based on SARS, one can expect immunological memory to last 2–3 years.

Moreover, we want to focus on two aspects that we'll be crucial in the following, as they represent important risk factors in the spreading of the illness.

- *Stages of the incubation period*: The incubation period, defined as the time between exposure to the virus and symptom onset, can be divided into two stages. A first one in which the amount of viruses is still small and the individual cannot transmit the virus to others; and a second stage, still of incubation, in which the individual is infectious because, as the virus multiplies, he/she may shed copious amounts of it. Both are pre-symptomatic stages, hence the individual has no sign of being infectious: the immune system prepares its answer but still does not act so strongly. Needless to say, the second stage is very dangerous for other individuals.

- *Unknown patients*: There are several studies, such as [16], [24], [32], that show that a large proportion of the population has SARS-CoV-2 infections that do not result in COVID-19 symptoms, or result in mild symptoms that go unnoticed. Therefore, there is a large proportions of infected individuals who remain unknown to the National Health Systems and who are not isolated. It is clear that this kind of patients are highly dangerous for other individuals, as they may act as they are not infectious.

1.3 Modeling the complexity of COVID-19

As described above, virus pandemic is a complex phenomenon with multiscale features, see also the extensive description in [3]. Inside each individual, the virus performs a certain progression which has strong consequences on the subsequent contagion between individuals, in particular due to the pre-symptomatic period of infectiousness which, in the case of Covid-19, is more pronounced than in other coronaviruses; and also because of the large number of infected people who remain asymptomatic after the incubation period, still infective. The spread through the population is again multiscale, with local (e.g. inside family) and long-distance transmissions. Therefore the full range of scales, from the virus particle dynamics to the individual interactions is active and interconnected.

Mathematics has developed tools with several levels of complexity. The seminal paper of Kermack and McKendrick [28] in 1927, who proposed the so called SIR model, idealizes the phenomenon in few compartments; but still in its simplicity indicates basic logical building blocks which are present in much more refined models. Generalizations and additional complexity due to spatial structure have been introduced later on; let us mention for instance [17], [9], [11], [10], [26], [4], [5], [29] and extensive treatises like [1], [19], [18], [36], [2]. The list of contributions is very large, see the references in [3] and the examples of multiscale models presented there.

Our model fits into these general schemes, simplifies certain aspects but maintains the general multiscale complexity. Among the simplifications, we have not included viral load and the response of the immune system, mastered for instance by [38]; we think that variability of viral load among the population and in different periods of the year may be a key aspect of Covid-19 and could contribute to a more advanced model. We have included, although in a strongly simplified version, the complexity of the evolution of the virus inside each individual by separating the incubation period in a first non-infectious phase and an infection second one, with different (random) length estimated from a careful comparison of data in Section 3. This detail of microscale is not per se but because it is responsible for numerous hidden infections and reflects into the prediction of virus spread. Then we have taken into account the spatial structure of the population, but in the way we think more natural for the diffusion of Covid-19 in the present society under the recent restrictions applied in almost all countries. In particular, we do not use partial differential equations of diffusion type but an individual based model, which propagates the infection but with intrinsic randomness which, at the end, seems to be very close to real measurements. The interaction between the individuals, responsible for the infection spread, includes nearest-neighbor links (what we call family bubbles in Section 2) and long-range ones (what we call pseudo-bubbles and city). Networks, which is another methodology introduced also in mathematical modelling of infectious diseases [42], [30], is certainly inspiring here, but we do not restrict ourselves to specific known schemes and use the ideas of networks in the framework of a Markov chain somewhat similar in structure to a cellular automaton. It is apparent from our modeling proposal that we do not aim to apply a specific methodology but we want to capture the features which we think more important,

4

and thus we mix several ideas and apply simplifications at due places, looking for a reasonable compromise between the great complexity of the true phenomenon and the methodologies which multiscale mathematical analysis may offer. With many differences, the methodology looks similar in spirit to what is done in [3]; see also [2]. Among the peculiarities, let us mention the detailed treatment of the randomness which pervades all steps of our modeling and represents one of main successes in comparison with real data.

2   Individual based Markov chain model

This section presents the basic mathematical model. Precise values of the parameters and other details related to applications are discussed in the next section; here we care only about the structure of the model.

We divide the section in several parts. The first one describes the basic architecture, up to the detailed description of just one but fundamental transition probability, $p_{S,IN}$ defined in (2) below. Then, the next subsections discuss several forms of $p_{S,IN}$, depending on the interactions accounted for in the model.

2.1   Basic structure

Suppose we have N individuals, indexed by the positive integers $i = 1, \ldots, N$. The i-th individual will be in a state $X^i(t)$ at time t. The family of all individuals will be, at time t, in the state

$$X(t) = \left(X^1(t), ..., X^N(t) \right)$$

which will be the *state of the Markov chain*.

The state $X^i(t)$ of each individual may take only a finite number of values: $X^i(t) \in \mathbb{S}_{ind}$, where

$$\mathbb{S}_{ind} = \{S, IN, II, U, K, E\}.$$

The letters refer to stages of the illness or to the status of the individual, and they are abbreviation of:

- S = **Susceptible**
- IN = **Infected but Not infectious**
- II = **Infected and Infectious**
- U = **Unknown**
- K = **Known**
- E = **Exit**.

The logic behind this subdivision lies in the typical history of an individual. Initially, a person is **S**usceptible, namely healthy, never infected by COVID-19. Some individuals, sometime, get infected by others and enter the state **IN**. When a person gets infected, it enters a period of *incubation*, where the virus multiplies in the body and, with some delay, the immune system start developing its answer (which at due time may emerge as symptoms). The incubation period is roughly divided in two stages: a first one in which the virus charge is still small and the individual is not infectious,

cannot transmit the virus to other individuals; this stage is represented by state IN. Then a second stage, still of incubation, when the individual is infectious; it is represented by state **II**. But both are pre-simptomatic stages, hence the individual has no sign of being infectious; the immune system, during states IN and II prepares its answer but still does not act so strongly. Needless to say, state II is very dangerous for the other individuals.

Then the incubation period is over and individuals enter a new stage, where they are infectious and ill. But for many of them the illness is so weak that they do not notice it, or it is below their personal threshold of pain or care. People of this kind are in state **U**, unknown to the national health system because asymptomatic or paucisymptomatic. Again this is a very dangerous state for the community. Other people, when the illness develops, feel symptoms and start medical checks; finally they are tested, recognized as COVID-19 holder and put in isolation or in hospital. This relatively short period, when they are ill, start medical tests but are not yet isolated by the community, has been called **K**, known; they can infect people, hence they contribute to infection. But soon they **E**xit the system, in the sense that they are isolated and do not contribute anymore to infection. Also the individuals in state U will exit, but only when they are not infectious anymore; the period spent in state U is longer than the one spent in K, statistically speaking. The full logical scheme is

$$S \longrightarrow IN \longrightarrow II \quad \begin{array}{c} \nearrow \quad U \quad \searrow \\ \searrow \quad K \quad \nearrow \end{array} \quad E$$
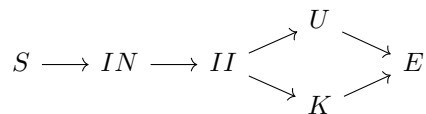
The exit state thus includes individuals of several kinds: isolated, hospitalized, recovered and dead; it is not our purpose to add information on these classes, which are no more of systemic nature (namely they do not have to do with the system of interacting individuals), hence we group all such cases into state E. But it is not difficult to augment the Markov chain with a subdivision of state E, if of interest.

Different would be to discover that recovered individuals may become infected again: such event would cause a link between states E and S which should be understood and carefully modeled. A few instance of this occurrence have been observed but its statistical significance is still poor. In absence of clear proofs that a significant portion of restored individuals may get re-infected, we consider state E as a final state without subseqent transitions.

Another simplification we are adopting, strong with respect to reality, is that we assume that people who are isolated or hospitalized do not contribute any more to infect other people, so we collect them in the same class E as people who are restored or those who die. In reality, we know that the virus spreads also in hospitals, and maybe people asked to stay isolated sometimes break a strict isolation; we have decided to miss this additional complication which, with due modifications, could be taken into account.

The state of the full system $X(t) = \left(X^1(t), ..., X^N(t) \right)$ takes values in

$$\mathbb{S} = \{S, IN, II, U, K, E\}^{\{1,...,N\}}.$$

We assume the Markov chain to be time-homogeneous. Then the dynamic rules are completely prescribed by the transition probabilities

$$p(\eta, \eta') = \mathbb{P}\left(X(1) = \eta' | X(0) = \eta \right)$$

with $\eta, \eta' \in \mathbb{S}$. In the philosophy of **cellular automatons**, we adopt the following rule:

$$p(\eta, \eta') = \prod_{i=1}^{N} \mathbb{P}\left(X^i(1) = \eta'(i)|X(0) = \eta\right). \tag{1}$$

In other words, we impose that, conditionally to the current state, individuals choose their transition independently one each other. For most states $\eta \in \mathbb{S}$ and $A \in \{S, IN, II, U, K, E\}$, we simply have $\mathbb{P}\left(X^i(1) = A|X(0) = \eta\right) = 0$. The exceptions are the transitions of interest for us, that now we list, introducing new names for the corresponding probabilities. Consider scheme (**??**). Most of the transitions, namely all except for $S \to IN$, involve only internal changes of a single individual, not the interaction between different subjects. The corresponding transition probabilities are

$$p_{IN,II} = \mathbb{P}\left(X^i(1) = II|X(0) = \eta\right), \quad \text{if } \eta(i) = IN$$
$$p_{II,U} = \mathbb{P}\left(X^i(1) = U|X(0) = \eta\right), \quad \text{if } \eta(i) = II$$
$$p_{II,K} = \mathbb{P}\left(X^i(1) = K|X(0) = \eta\right), \quad \text{if } \eta(i) = II$$
$$p_{U,E} = \mathbb{P}\left(X^i(1) = E|X(0) = \eta\right), \quad \text{if } \eta(i) = U$$
$$p_{K,E} = \mathbb{P}\left(X^i(1) = E|X(0) = \eta\right), \quad \text{if } \eta(i) = K$$

independently from the values $\eta(j)$ with $j \neq i$. On the contrary, the transition $S \to IN$ depends on the interaction: the value

$$p_{S,IN}(\eta, i) = \mathbb{P}\left(X^i(1) = IN|X(0) = \eta\right), \quad \text{if } \eta(i) = S \tag{2}$$

depends on the whole state η and on the specific individual i. Specifying the value of $p_{S,IN}(\eta, i)$ is a major modeling task, discussed in the next subsections.

In a Markov chain, when we are in state A, the time to exit from A (the number of time steps necessary to exit from A) is geometric distributed; and for a geometric distribution the probability of jump is the inverse average time of the first jump. In our example, this applies to the transitions $IN \to II$, $U \to E$, $K \to E$, and to the exit from state II. Thus, we have

$$p_{IN,II} = \frac{1}{\mathbb{E}[T_{IN,II}]}$$
$$p_{U,E} = \frac{1}{\mathbb{E}[T_{U,E}]}$$
$$p_{K,E} = \frac{1}{\mathbb{E}[T_{K,E}]}$$
$$p_{II,U} + p_{II,K} = \frac{1}{\mathbb{E}[T_{II,\{U,K\}}]}$$

where we have denoted by $T_{IN,II}$ the time to go from IN to II, similarly for $T_{U,E}$ and $T_{K,E}$ and by $T_{II,\{U,K\}}$ the time to go from II to one of $\{U, K\}$. This fact provides the link with quantities which can be interpreted in the examples. For instance, the quantity

$$\mathbb{E}[T_{IN,II}] + \mathbb{E}[T_{II,\{U,K\}}]$$

is the average incubation time.

If the Markov chain is already given, since the jump times require at least one unit, automatically the average times written above are greater or equal to one and thus the corresponding probabilities are smaller than one. However, in the examples, we *choose* a priori the values of the probabilities, guided by the intuition given by the link with average times. Hence we always have to impose the constraint that such probabilities are in $[0,1]$.

2.2 Uniform interaction (only full town)

The first, simplest kind of interaction we discuss is when each individual is exposed to all others, in the community of N individuals. We have in mind, typically, a town of medium size with its surroundings, such that internal interactions are statistically more relevant than the interactions with other towns or regions. Obviously the interactions with the outside are always very relevant, for instance to start the infection in a new town, but we neglect this aspect with could be considered only by a more global model with several large units (towns or regions). We describe only the internal development of a single unit.

Certainly the probability $p_{S,IN}(\eta, i)$ defined by (2) should depend on N and on the number $I(\eta)$ of infectious people:

$$I(\eta) = \sum_{j=1}^{N} 1_{\eta(j) \in \{II, U, K\}}$$

(based on the description of the previous section, we assume that only individuals of class II, U, K are infectious; moreover, for simplicity we assume they have the same potentiality of infection, so that there are no distinguished coefficients in their action). The simplest choice is

$$p_{S,IN}(\eta, i) = \lambda \frac{I(\eta)}{N}. \tag{3}$$

Namely $p_{S,IN}(\eta, i)$ is directly proportional to $I(\eta)$ and inversely proportional to N, with a proportionality constant λ to be chosen. There are heuristic or approximate justifications of this formula which may helps to connect λ to experimental knowledge.

2.2.1 Idea of the code

We now give the idea of the structure of the numerical code we use to simulate this system. The software used is **R** version 3.6.3.

We begin assigning a value to the number of individuals N, the number of days G, and the mean times for the transitions. Then we calculate the transition probabilities, except for $p_{S,IN}$, using the formula we presented in the previous section. After that, we assign values to parameters β, that represents the proportion of unknown sick people, with respect to the known ones, and λ, that represents the probability that an infectious individual infects a susceptible individual in a temporal unit. Finally, we set the number of infectious individuals we have on the first step of the simulation.

We assign a number from 0 to 5 to every state and we create matrices $N \times G$, one for every transition, that will take into account the state of every individual in every time step; and a vector of length N that will be updated at every time step and that contains the states of the individuals that iteration.

Then we use a `for` cycle to iterate over time steps. In its body, we calculate the transition probability

8

$p_{S,IN}$, that can be different at every iteration; we control which are the individuals in every state, and then we decide which individuals will make a transition, according to the suitable transition probability. Finally, we plot the graphs of the number of individuals in every state, using different colours to represent different states.

2.3 Model with family units and the town

A key feature of human communities is that a person does not interact with all the others. This is true in normal conditions, but even more under special restrictions like the so called lockdown related to the COVID-19 emergency. A person has daily contacts with the family, and other contacts with different frequencies and intensities.

Said like this, we could object that the uniform interaction of the previous section is a pure abstraction of no interest. It is not so: it is a building block, as we shall see here. Moreover, fractions of time are spent in potential contact with every subject of the full community: when we travel on public transport or make shopping in certain large market places, or we spend an evening in a downtown local, and so on. Thus the direction of a more realistic model with respect to the previous section is to split the time up so that a fraction of time is spent as above in potential contact with everybody, and other fractions are focused on subgroups of the population.

Subgroups are of several different nature and structure. In this subsection we take into accont only the family. In order to develop a model as simple as possibile, we assume that families are all made of three individuals, which is close to a typical average number. Families are disjoint subsets of the full population, they are a *partition*, opposite to the subdivision discussed in the next subsection. Hence, propagation of virus in a family stops there, unless some member of the family spends time also outside, in contact with members of other families. In principle, if during a lockdown period, a large part of the population is asked to spend all the time within the family; nevertheless some members should continue to work and some to go to buy essential goods for survival, hence full family isolation is not possible and realistic.

In order to unify some aspect of this and the next section, let us introduce a structure here which would be optional: an adjacency matrix. It is a matrix A^{fam} with 0 and 1 entries with the following meaning, in the case of families: for every individual i, there is a set of individuals

$$J_{\text{fam}}(i) \subset \{1, \ldots, N\}$$

that is made up of the individuals that are part of i's family unit. If $j \in J_{\text{fam}}(i)$ and $j \neq i$, we put 1 in entry A_{ij}^{fam}. What we put in A_{ii}^{fam} is unessential, we put zero. If $j \notin J_{\text{fam}}(i)$ then $A_{ij}^{\text{fam}} = 0$. Thus the matrix A, looked at by rows, states who are the individuals that are in the same family unit. This kind of contact between individuals is symmetric:

$$j \in J_{\text{fam}}(i) \Leftrightarrow i \in J_{\text{fam}}(j).$$

Hence the matrix A^{fam} is symmetric. Reordering the indexes of individuals if necessary, the adjacency matrix has the following form:

$$A^{\text{fam}} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$

where we have represented only two families for obvious reasons of space.

We now need to introduce two parameters

$$\alpha_{\text{fam}}, \alpha_{\text{city}} \in [0,1]$$

$$\alpha_{\text{fam}} + \alpha_{\text{city}} = 1$$

that represent the proportion of time spent with people in the family unit with respect to that spent with all the inhabitants of the city. For the time being, these parameters will be the same for everyone; later on we'll introduce categories with particular specifications.

With the previous definitions and quantities, arguing similarly to (3), we choose

$$p_{S,IN}(\eta, i) = \alpha_{\text{city}} \lambda_{\text{city}} \frac{I_{\text{city}}(\eta)}{N} + \alpha_{\text{fam}} \lambda_{\text{fam}} \frac{I_{\text{fam}}(\eta, i)}{N_{\text{fam}}}$$

where $N_{\text{fam}} = 2$, $I_{\text{city}}(\eta)$ and $I_{\text{fam}}(\eta, i)$ are the number of infectious individuals in the city and in the family, respectively,

$$I_{\text{city}}(\eta) = \sum_{j=1}^{N} 1_{\eta(j) \in \{II, U, K\}}$$

$$I_{\text{fam}}(\eta, i) = \sum_{j \in J_{\text{fam}}(i)} 1_{\eta(j) \in \{II, U, K\}}$$

and $\lambda_{\text{city}}, \lambda_{\text{fam}}$ are parameters with analogous meaning to the parameter λ of the previous section, possibly different for the city and the family (for instance, during pandemic emergency, in the town people wear masks and keep the safety distance, but not at home, and thus the propensity to infect is different).

2.4 Model with pseudo-bubbles

Families are a partition of the population. Other groups which limit the interactions exist, related to work, sport, friendship etc. but they are not (or not always) a partition: a person may have two friends who do not know each other. Transitivity is not always true. We model such groups by a family of sets

$$J_{\text{ps}}(i) \subset \{1, \dots, N\}$$

one for each subject $i \in \{1, \dots, N\}$. We still impose the property

$$j \in J_{\text{ps}}(i) \Leftrightarrow i \in J_{\text{ps}}(j)$$

which implies that the matrix A is symmetric. As in the case of families, we impose $i \in J_{\text{ps}}(i)$, although it is irrelevant. But we do not impose transitivity. Associated to these sets we write an adjacency matrix. It is a matrix A^{ps} with 0 and 1 entries as in the case of A^{fam}: if $j \in J_{\text{ps}}(i)$ and $j \neq i$, we put 1 in entry A_{ij}^{ps}. This adjacency matrix has, for instance, the following form:

$$A^{\text{ps}} = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

where we have represented the case of two people, the first and the last one, who have all other people in their pseudo-bubble; and the other people (different from the first and last subject) who have only the first and the last one in their pseudo-bubbles.

In the full model we consider city, family and pseudo-bubbles at the same time. Hence we need to introduce three parameters

$$\alpha_{\text{fam}}, \alpha_{\text{city}}, \alpha_{\text{ps}} \in [0,1]$$

$$\alpha_{\text{fam}} + \alpha_{\text{city}} + \alpha_{\text{ps}} = 1$$

that represent the proportion of time spent with people in the family unit with respect to the city or the pseudo-bubble. For the time being, this parameters will be the same for everyone; later on we'll introduce categories with particular specifications.

With the previous definitions and quantities, arguing similarly to (3), we choose

$$p_{S,IN}(\eta, i) = \alpha_{\text{city}} \lambda_{\text{city}} \frac{I_{\text{city}}(\eta)}{N} + \alpha_{\text{fam}} \lambda_{\text{fam}} \frac{I_{\text{fam}}(\eta, i)}{N_{\text{fam}}} + \alpha_{\text{ps}} \lambda_{\text{ps}} \frac{I_{\text{ps}}(\eta, i)}{N_{\text{ps}}(i)}$$

where the old quantities are defined as in the previous section and

$$I_{\text{ps}}(\eta, i) = \sum_{j \in J_{\text{ps}}(i)} 1_{\eta(j) \in \{II, U, K\}}$$

$$N_{\text{ps}}(i) = |J_{\text{ps}}(i)| - 1$$

where $|J_{\text{ps}}(i)|$ is the cardinality of $J_{\text{ps}}(i)$.

2.5   Model with layers

The last scenario we considered is the most complicated because we stratified the society in various groups and each of them has its own parameters due to the fact that individuals in the same group have similar lifestyle.

With *layers* we mean a partition of the set $\{1, \ldots, N\}$ that are the indexes wherewith we enumerate the individuals.

We decide to split out into two sets the individuals, one for workers and the other for students, senior citizen and smart workers. This choice is justified by the fact that in general workers have different lifestyle in comparison to students, seniors or smart workers. This difference emerged in an outstanding way during the lockdown; indeed many workers continued to go to work, maybe with reduced hours, but they have spent much more time away from home compared to other categories. To model this partition we take a sample of a certain percentage of the population that will represent the individuals in the group of students, seniors and smart workers; the individuals left form the second group. Hence this scenario contains all the modifications we progressively did, plus this partition. This means that every group of people carries his set of parameters, that are $\alpha_{city}^1, \alpha_{fam}^1, \alpha_{ps}^1$ for *smart* citizens and $\alpha_{city}^2, \alpha_{fam}^2, \alpha_{ps}^2$ for workers. Hence, the probability to get infected is the same as the third scenario but has to be calculated with different parameters within the same group of people, so for the j-th group the probability is

$$p_{S,IN}(i, \eta) = \lambda \cdot \left(\alpha_{city}^j \cdot \frac{I_{city}(\eta)}{N} + \alpha_{fam}^j \cdot \frac{I_{fam}(\eta)}{|J(i)| - 1} + \alpha_{ps}^j \cdot \frac{I_{ps}(\eta)}{|J'(i)| - 1} \right).$$

The setting of this scenario will be useful to model situations in which the difference of the parameters $\alpha^i_{(\cdot)}$ is considerable. Indeed, this scenario has the same properties of the third one if we want to model a normal situation, but allow us more realism if we want to model a period of time that includes both a period before the lockdown and the lockdown.

2.6 The delay caused by family and pseudo-bubbles

Restricting part of the life in families or pseudo-bubbles is very useful as a limitation to virus spread. In this section we show this fact by a numerical simulation. Since the aim of this section (opposite to the next ones) is only speculative and not of comparison with real data, we attribute simple values to the model parameters: we use the same parameters $\mathbb{E}[T_{IN,II}]$, $\mathbb{E}[T_{II,\{U,K\}}]$, $\mathbb{E}[T_{U,E}]$, $\mathbb{E}[T_{K,E}]$, λ, β for all the three versions we compare. For the model with family units, we set $\alpha_{fam} = \alpha_{city} = 0.5$, with the meaning that, on average, an individual spends half of a day at home and the other half outside. For the model with family units and pseudo-bubbles, we set $\alpha_{fam} = 0.5$, $\alpha_{city} = 0.1$, $\alpha_{ps} = 0.4$, with the meaning that, on average, an individual spends half of a day at home and the other half outside, but only for a proportion of time equal to 0.1 he can be in contact with all the inhabitants of the city, for example on public transportation, while for the rest of the time he can be in contact only with the people that form his pseudo-bubble. We can observe that, as we mentioned, adding social structures, such as family units and pseudo-bubbles, to the population results in a slowdown of the spreading of the infection. Indeed, it is possible to see that the peaks are delayed in the second and third figures, with respect to the first one. We divide the Exit state into two states, E_1 and E_2, in order to see on the graphs how many individuals have been in state U or K, respectively.
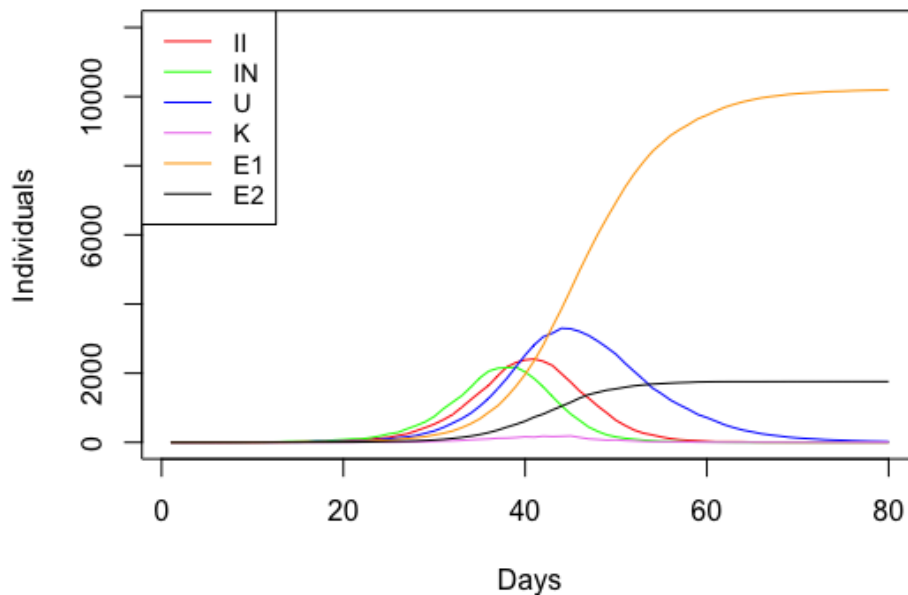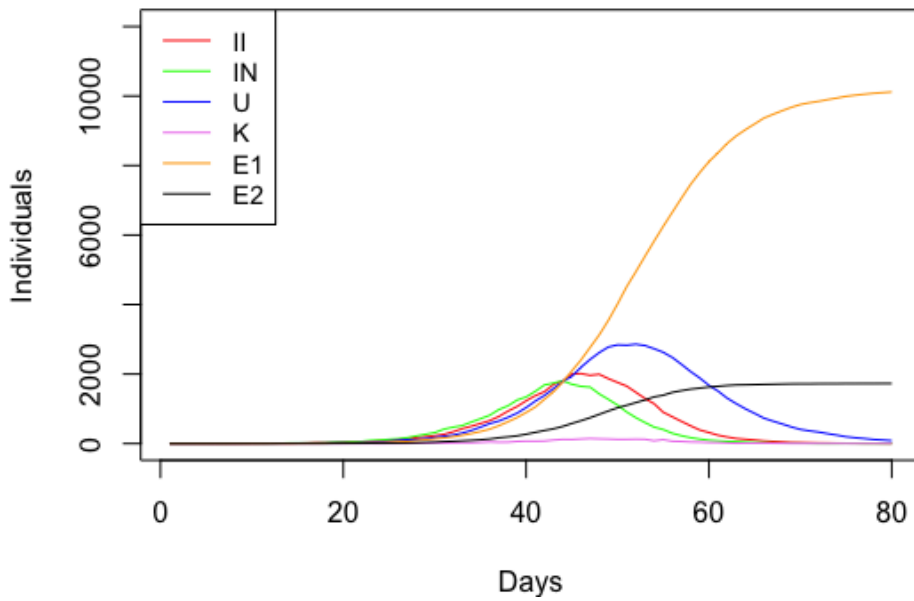

Figure 1: Uniform interaction

12

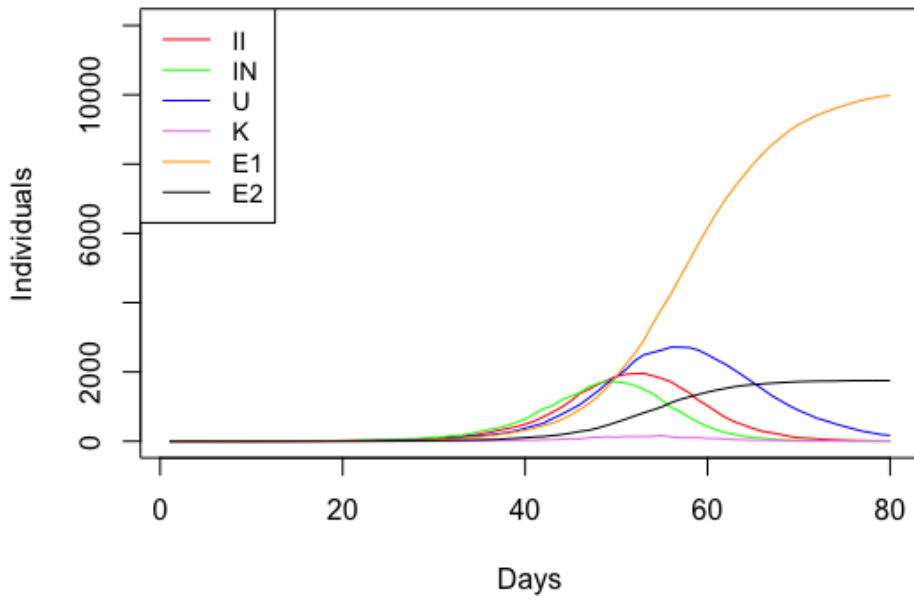Figure 2: Model with town and family units

Figure 3: Model with town, family units and pseudo-bubbles

13

3   Characterization of the model parameters

Let us list the model parameters emerged until now. A first group is represented by are those appearing in the transition probabilities different from $p_{S,IN}(\eta, i)$: $p_{IN,II}$, $p_{II,U}$, $p_{II,K}$, $p_{U,E}$, $p_{K,E}$.

Given any state of a Markov chain, the time T of transition outside it, equal to 1,2,3,$\ldots$, is geometrically distributed. Hence the probability of transition outside in a time step is equal to the inverse of the average time of transition outside the state. Therefore we have

$$p_{IN,II} = \frac{1}{\mathbb{E}\left[T_{IN,II}\right]}$$

$$p_{II,U} + p_{II,K} = \frac{1}{\mathbb{E}\left[T_{II,\{U,K\}}\right]}$$

$$p_{U,E} = \frac{1}{\mathbb{E}\left[T_{U,E}\right]}$$

$$p_{K,E} = \frac{1}{\mathbb{E}\left[T_{K,E}\right]}.$$

where $p_{i,j}$ is the transition probability from state i to state j, $T_{i,j}$ is the random time needed for that transition, and $\mathbb{E}[T_{i,j}]$ is its mean value. Concerning $p_{II,U}$ and $p_{II,K}$ there is a proportionality parameter β.

such that

$$p_{II,U} = \beta \frac{1}{\mathbb{E}\left[T_{II,\{U,K\}}\right]}, \qquad p_{II,K} = (1 - \beta) \frac{1}{\mathbb{E}\left[T_{II,\{U,K\}}\right]}.$$

The parameter β represents the probability of becoming U instead of K for an individual in state II, and it is one of the parameters of the model.

These relations are very important for the characterization of these transition probabilities because they link them to quantities for which we have information from the literature. More precisely, among the several information present in the literature, we have decided to consider sufficiently accurate the following ones.

If we denote by $\hat{t}_{i.p.}$ the estimate of the mean duration of the incubation period as found in the literature, it follows that

$$\hat{t}_{i.p.} = \mathbb{E}[T_{IN,II}] + \mathbb{E}[T_{II,\{UK\}}]. \tag{4}$$

The overall range for the duration of the incubation period, as found in many studies like [33], [45], [35], is 1.8–9.0 days. However, we decide to rely on the World Health Organization's website[1], that reports that the incubation period of COVID-19 is, on average, $5 - 6$ days, therefore we restrict our research to articles that present a mean value for the incubation period between 5 and 6. We decide to fix

$$\hat{t}_{i.p.} = 5.5,$$

as this value respects the requirement to be between 5 and 6, and we can find on two articles that are based on a reasonable number of real cases, [31], [34], values really close to this one (5.5 in the first one and 5.6 in the second one).

---

[1]https://www.who.int/news-room/commentaries/detail/transmission-of-sars-cov-2-implications-for-infection-prevention-precautions

We now need to choose a value either for $\mathbb{E}[T_{IN,II}]$ or $\mathbb{E}[T_{II,\{UK\}}]$. This is not a simple task, as it is not clear when an individual enters the infectious stage of the incubation period. We refer to the paper [25] that analyzes the temporal dynamics of the transmissibility of COVID-19 and we decide to make the infectiousness of an individual begin, on average, 3 days before the exit from state II, which means that we choose

$$\mathbb{E}[T_{II,\{UK\}}] = 3.$$

Now, we can derive the mean duration of the other stage of the incubation period:

$$\mathbb{E}[T_{IN,II}] = 5.5 - 3 = 2.5.$$

From the duration of infectiousness we can also obtain information about $\mathbb{E}[T_{U,E}]$. Indeed, when an individual exits state II and enters state U, he remains unknown to the National Health System, and, therefore, he is not isolated and may continue to infect other people for the whole time he is infectious. From the figures that represent the inferred infectiousness profile in [25], we can observe that, after an individual exits state II, the infectiousness profile decreases monotonically, and it becomes quite low after 6 days from that moment. Thus, we decide to make the infectiousness of an individual end, on average, 6 days after the exit from state II, which means that we choose

$$\mathbb{E}[T_{U,E}] = 6.$$

This means that the infectiousness lasts on average 9 days. It's an estimate not far away from other papers that we found later on; for example in [43] they say that the infectious period lasts on average 10.9 days. Things are different for individuals that enter state K after the exit from state II. The intuition is that the time spent in state K is really short, in the range of one to two days, the time needed to notice to have some symptoms, contact the doctor and be put in isolation. Thus, we decide to set

$$\mathbb{E}[T_{K,E}] = 1.5.$$

We now come to β, the proportion of unknown cases with respect to known cases. To set this parameter, we refer to the recent research carried out by Istat about the seroprevalence of COVID-19. There it is possible to read that the individuals that have been in contact with the virus are 6 times more than the total cases officially intercepted during the pandemic. Therefore we choose

$$\beta = 0.85.$$

It is remarkable to note that this number is not the same for the whole duration of the epidemic: in the first stages, only people with evident symptoms were tested and this led to a greater proportion of unknown cases. After a while they started with contact tracing and could intercept also cases that were not evidently ill; causing a reduction of the proportion of unknown cases.

Next, let us discuss the transition probability $p_{S,IN}(\eta, i)$. In the simplified model of full city interaction, the only parameter needed to compute $p_{S,IN}(\eta, i)$ is λ.

The heuristic meaning of λ is, in our opinion, not sufficient to characterize its value in examples. It is necessary to fit it. We shall describe in the next section a quite elaborate strategy to fit this value on the basis of auxiliary quantities related to experiments.

When the model with family and pseudo-bubbles is used (or just with family), it is necessary to characterize several additional parameters:

- $\lambda_{\text{city}}, \lambda_{\text{fam}}, \lambda_{\text{ps}}$

- $\alpha_{\text{city}}, \alpha_{\text{fam}}, \alpha_{\text{ps}}$

- $N_{\text{ps}}(i)$ for each i in the population.

In addition, the fact that we have chosen families of three people is also a hidden parameter fit, which could be changed.

Our choice about λ_{fam} is simply

$$\lambda_{\text{fam}} = \lambda$$

where, by λ, we mean the general parameter fitted according to the next section. It corresponds to a sort of biological basic parameter applicable to individuals that do not know about the existence of the virus or do not care about it. This is, with some degree of simplification, the behavior of people when they spend time with their family.

However is very difficult to fit the values of λ_{city} and λ_{ps} from data. We use different values of λ_{city} and λ_{ps} when considering the phase prior to lockdown or the lockdown itself. During the phase prior to lockdown, we have decided to apply a simplification, namely to assume that

$$\lambda_{\text{city}} = \lambda, \quad \lambda_{\text{ps}} = \lambda.$$

Differently, during the lockdown and post-lockdown phase we set $\lambda_{\text{city}}, \lambda_{\text{ps}}$ as fractions of λ, to model the effect of social distancing and the introduction of masks as a personal defence against the virus. We decided that for the pseudo-bubble the value of λ is higher to model a smaller perception of risk when an individual spends time with people he knows. Obviously, the parameter λ_{fam} is the same as before because people at home do not follow the same rules. The coefficients $\alpha_{\text{city}}, \alpha_{\text{fam}}$ and α_{ps} represent the percentage of time spent in each group. We have chosen different values of these parameters for pre-lockdown and lockdown period.

Note that if we use layers we have to introduce other coefficients $\alpha^i_{city}, \alpha^i_{fam}, \alpha^i_{ps}$ but we will explain it later. Finally, $N_{\text{ps}}(i)$ is obviously too difficult to be known in detail and thus we have made an average choice - as it is the choice that a family is made of three people - namely, we fix a parameter ξ, and for every individual i, we generate a number $N_{ps}(i)$ from a Poisson distribution of parameter ξ, which is the size of the set $J_{ps}(i)$. Then we identify $N_{ps}(i)$ random positions in the string $\{1, \ldots, N\}$, which will be the entries equal to 1 in the $i - th$ row of the adjacency matrix. Finally, we make the matrix symmetric.

3.1 The fit of parameter λ

In order to fix a value for the parameter λ, we need to introduce the concept of serial interval. The serial interval is defined as the time duration between a primary case (infector) having symptom onset and a secondary case (infectee) having symptom onset. The serial interval can present different features, depending on when the infection occurs: it can happen when the individual who infects the other is still in the pre-symptomatic stage, or when he is already in the symptomatic one. In the event of pre-symptomatic transmission, it may happen that the serial interval is negative, if the infectee develops symptoms before the infector. Therefore, there are three possible scenarios for the considered pair:
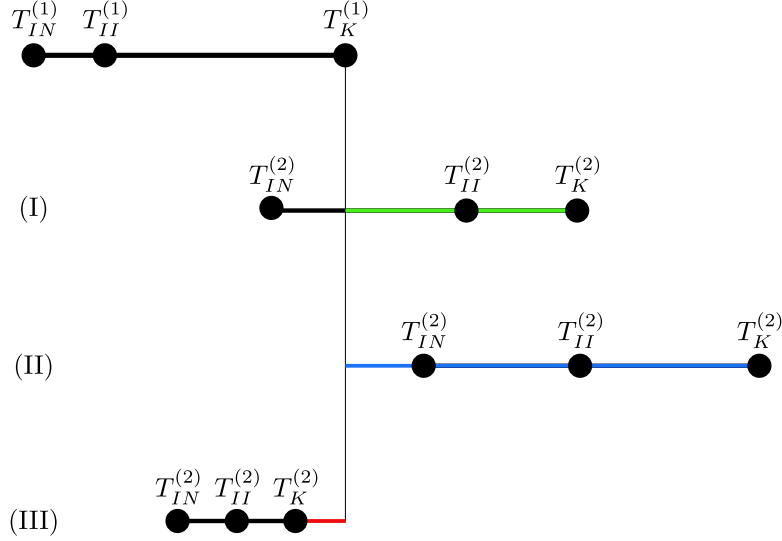
Figure 4: Three different scenarios for serial interval

We denote with the index 1 times referred to the infector and with the index 2 times referred to the infectee. Since we are not interested in what happened when the infector was not yet infectious, we consider the system formed by the infector-infectee pair from the moment the first individual enters state II. Looking at Figure 4, we can derive that

$$\text{Serial interval} = T_{IN_2,II_2} + T_{II_2,UK_2} + T_I - T_{II_1,UK_1};$$

where we denote by T_I the time necessary to the first individual to infect the second one. But, recalling that

$$\text{Incubation period} = T_{IN,II} + T_{II,UK},$$

we have

$$\text{Serial interval} = I.P._2 + T_I - T_{II_1,UK_1}.$$

For what concerns the expected values, we have

$$
\begin{aligned}
\mathbb{E}[S.I.] &= \mathbb{E}[I.P._2] - \mathbb{E}[T_{II_1,UK_1}] + \mathbb{E}[T_I] \\
&= \mathbb{E}[T_{IN_2,II_2} + T_{II_2,UK_2}] - \mathbb{E}[T_{II_2,UK_2}] + \mathbb{E}[T_I] \\
&= \mathbb{E}[T_{IN_2,II_2}] + \mathbb{E}[T_{II_2,UK_2}] - \mathbb{E}[T_{II_1,UK_1}] + \mathbb{E}[T_I];
\end{aligned}
$$

but $\mathbb{E}[T_{II_2,UK_2}] = \mathbb{E}[T_{II_1,UK_1}]$, therefore

$$\mathbb{E}[S.I.] = \mathbb{E}[T_{IN,II}] + \mathbb{E}[T_I]. \tag{5}$$

Notice that $\mathbb{E}[T_I]$ is not a parameter of our model. Different values of λ affect only $\mathbb{E}[T_I]$; indeed, greater values of λ imply a bigger probability of infection and, therefore, a smaller value of T_I while smaller values of λ imply a smaller probability of infection and a greater value of T_I. Different studies estimated the mean serial interval in the range of 2.86–7.5 days, [33], [45], [40],

17

[37], [22], [46], We set $\mathbb{E}[S.I.] = 3.96 \simeq 4$, as we can find in [20], and we calculate an estimate of $\mathbb{E}[T_I]$ with a Monte Carlo method. Notice that we used the uniform scenario to calculate the probability to get infected. In this way, we determine a value of λ that produces an empirical value of $\mathbb{E}[T_I]$ close to 1.5, as we want to obtain

$$\mathbb{E}[T_I] = \mathbb{E}[S.I.] - \mathbb{E}[T_{IN,II}] \simeq 1.5.$$

This is obtained with

$$\lambda = 0.75.$$

4 Results

4.1 Distribution of the incubation period

To determine the probability distribution of the incubation period we write a routine that returns the incubation period of an individual. With Monte Carlo method we extract a large sample of incubation periods and we use the results to make an histogram and to fit the *I.P.* density distribution. The function takes as argument the number of hours, the mean value of the random jumps from state to state. At the beginning we have a random individual in state IN. The function returns how many days this individual stays in states IN or II. Despite the fact that incubation period needs individuals known to the sanitary system that develops symptoms we do not get rid of the values of the function obtained from individuals that go in state U after they exit state II. This is not a mistake because the time a known individual and an unknown one take to exit from state II is the same, so this choice doesn't affect our results. It is noteworthy that to calculate the incubation period it is sufficient to use the first scenario in which the way an individual gets infected is uniform, every other individual can infect the others. Another remark is about the temporal unit: in this function we use hours instead of days as temporal units. The reason is simple: the incubation period is given by the sum of two exit times, thus if we use days as temporal units, its value will be greater than 2. Choosing hours in the histogram we will also have values smaller than two. Once the transition probabilities have been chosen in the case in which the temporal unit is a day, we have only to divide by 24 the probabilities, to obtain the case with hours as temporal units. This is true for those transitions whose probability is equal to the inverse of the mean of the exit times. For the first transition we don't have this property but we are not interested in the way the individual gets infected, therefore we can divide by 24 every probability and it doesn't affect our results. Calling the function 5000 times we can plot an histogram with relative frequencies and try to fit its distribution with different densities. As shown in Figure 5, we fitted a Gamma distribution using the function *fitdistr* of the *library(MASS)*, here plotted with a black line. In [25] the incubation period density is estimated to be like in Figure 6. The authors obtained mean equal to 5.2 days hence for this reason we can make a comparison.

Both estimates are similar: they peaked at the third day even though the peak of the second estimate is greater than the one we obtained. Moreover, in both cases after 12th day the density decreases rapidly becoming negligible at the 20th day. The parameters obtained with this method are mean equal to 5.5, *shape* = 1.99 and *rate* = 0.37.

In Figure 5 there are also different fits with Lognormal and Weibull densities. It is clear that in our case the best fit is given by the Gamma distribution, whereas in [25] the authors found out that the Lognormal best fitted their data. If we calculate the mean and sd of a Lognormal fit we

find the parameters: $meanlog = 1.43$, $sdlog = 0.78$ while their parameters are $meanlog = 1.62$, $sdlog = 0.42$.
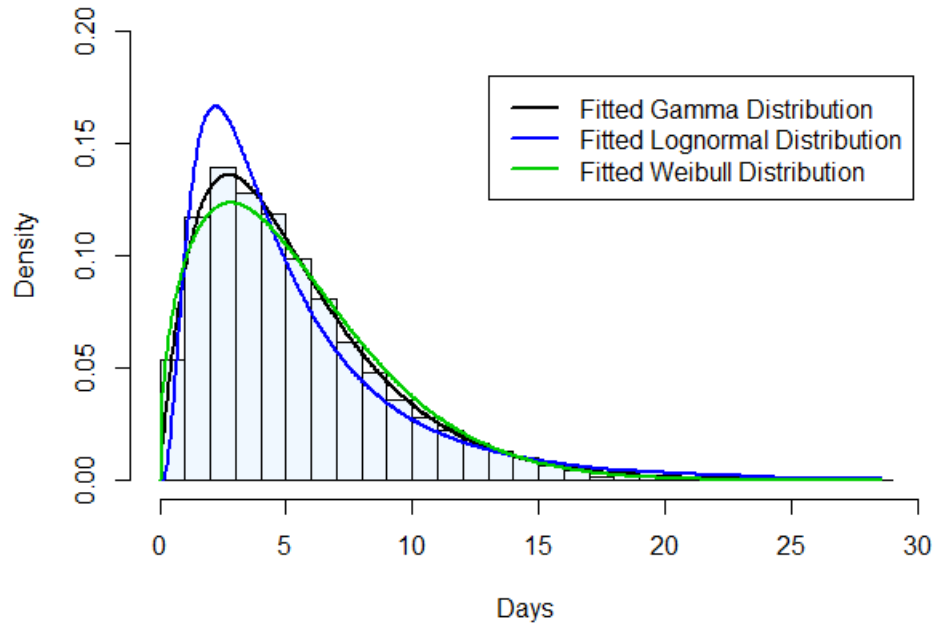

Figure 5: Histogram obtained calling the function that calculates the incubation period 5000 times.
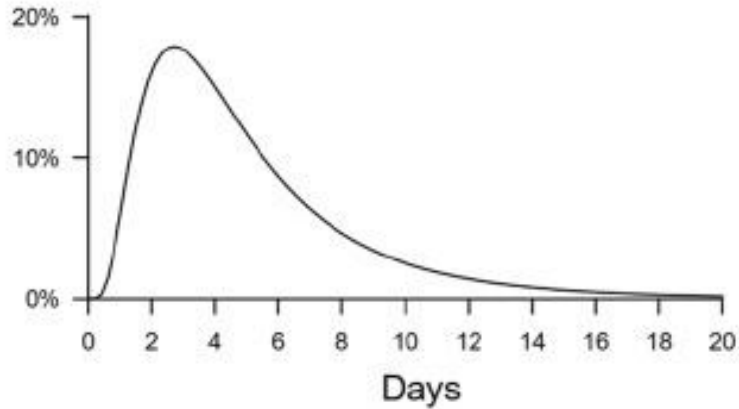
Figure 6: An example of incubation period density from literature.

The empirical cumulative function in our case can be compared with results in [31]. In Figure 8 they showed levels corresponding to the 2.5th, 50th and 97.5th percentiles respectively. The cumulative function in our case is the empirical cumulative function while the one in [31] is obtained by a supposed Lognormal distribution with parameters obtained using the observations. Their lognormal parameters are 1.65 and 0.42 (see Table 2 in [31]) The main difference between Figure 8 and Figure 7 is in the value of the cumulative function for the first 2.5 days. Indeed, in our case the cumulative function increases more rapidly than the other cumulative function, and this means that there is a not negligible proportion of symptomatic individuals within the first 2.5 days.
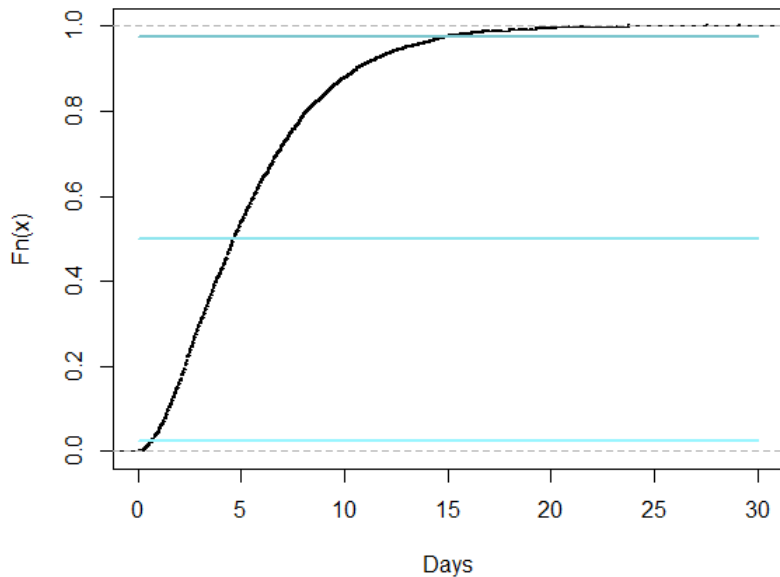
Figure 7: Cumulative function obtained with the function that calculates the incubation period.
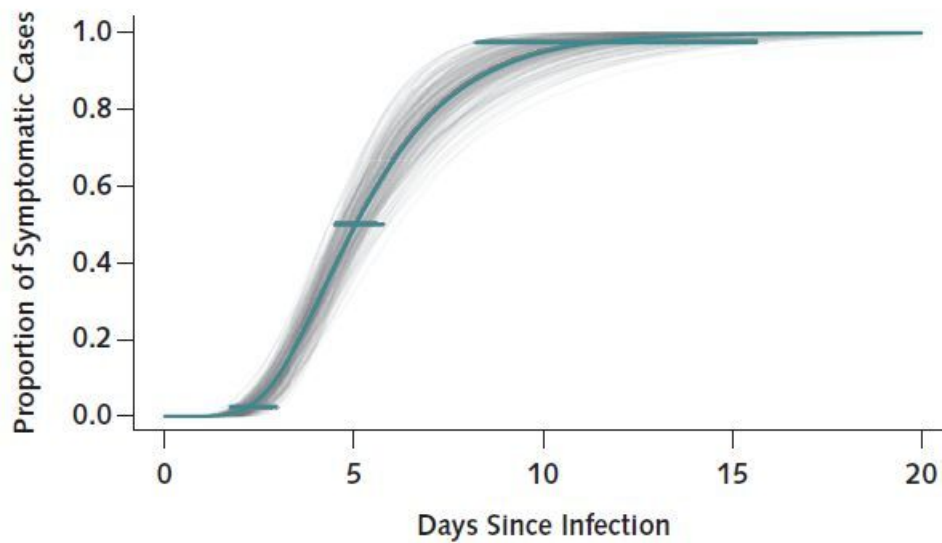
Figure 8: Cumulative function for incubation period from literature.

4.2   Distribution of serial interval

To obtain information about the serial interval probability distribution we write a routine that returns the serial interval for a random pair of infector-infectee, in which we choose the individual indexed by one as the infector. This is because, as for the incubation period, we use the code for the uniform scenario to calculate the serial interval, hence it is not important which individual we choose as infector. Despite the fact that, differently from the incubation period, the serial interval depends on the way an individual gets infected, with scenarios that includes families the probability that one of the relatives of the infector gets infected is high, so this will result in a shorter serial interval because in no time he will be infected. This function returns how many days there are between the day the infector becomes II and the day the infectee becomes II. The infectee is chosen in the set of the first individual infected by the infector; in particular we choose the infectee with the smaller index. Differently from the case of the incubation period we have to exclude the infectors that become U because one individual of that state can remain infectious for many days and this can affect the results.

As in the case of incubation period, with a Monte Carlo method we extract a large sample of serial intervals to make an histogram and then we compared our result with density estimates from the literature. In Figure 9 we fitted our histogram with a normal distribution of mean equal to 4.03 and standard deviation equal to 4.06 and also with a t-Student density with mean, sd and degrees of freedom respectively equal to 3.9, 3.11 and 5. From [20] we took Figure 10, even in this case they fitted with a normal distribution with mean interval 3.96 days and sd 4.75 days. In both cases, the density is zero before -10 and after 20th day. The relative frequencies in our case are increasing until the values from 4 to 5 and then decreasing, whereas this behaviour is not found in their histogram. Despite this, they both peaked around 0.09. In [20] authors reported 12.6% of negative serial intervals whereas in our case this value is around the 10%. From Figure 9 it is clear that t-Student density fits better our histogram. Indeed, it follows better the peaks compared to the normal density, and it seems to be closer to our data. Confirming what it was suggested by the plots, with a Shapiro-Wilk test we obtain a p-value smaller than 10^{-16} and this means we have to reject the normality hypothesis. Moreover, the value of skewness is positive and around 0.22 and this confirms the fact that the peaks in the histogram are slightly concentrated on the left. The value of kurtosis is around 2, hence definitely we can't say that serial interval is normally distributed.
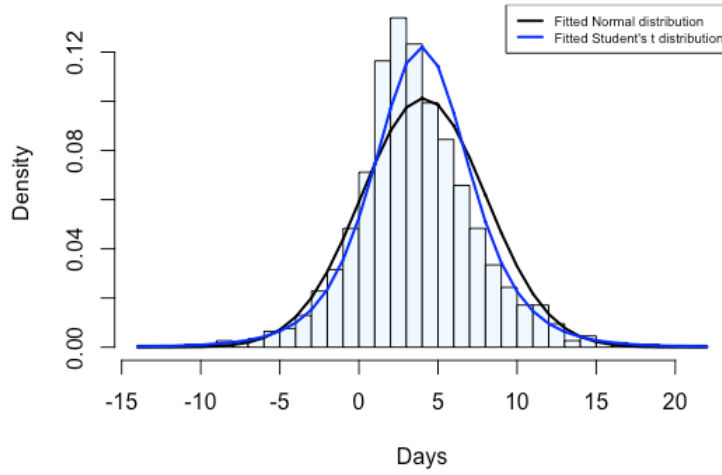
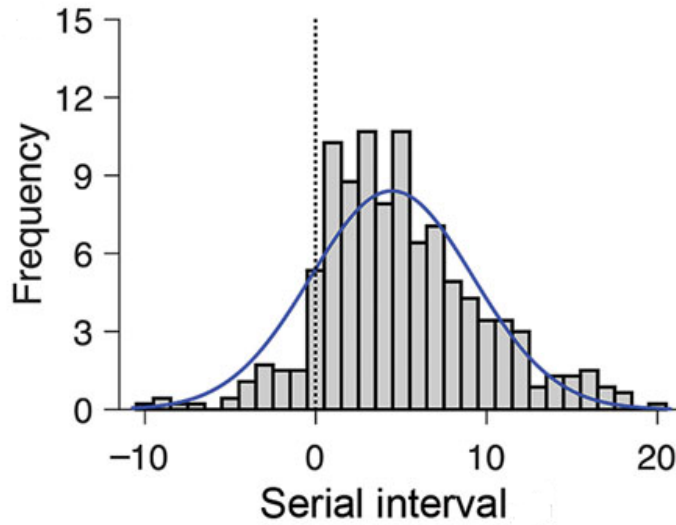Figure 9: Histogram obtained calling the function *serial* 5000 times.

Figure 10: Fit with normal distribution from paper [20].

4.3   Time series of Tuscany

In this section we deal with real cases that are the infections in Tuscany. We consider the total cases for every of ten provinces and we try to fit the mean of these cases. For computational reasons we run our routine over 9999 individual and approximate the cases over 99990 inhabitants multiplying

the result by ten. Indeed we standardize the cases over 99990 to make the routine execution faster. In Figure 11 there is the plot of the total cases over this standard number. Real data are taken from the page of Franco Mossotto [2].
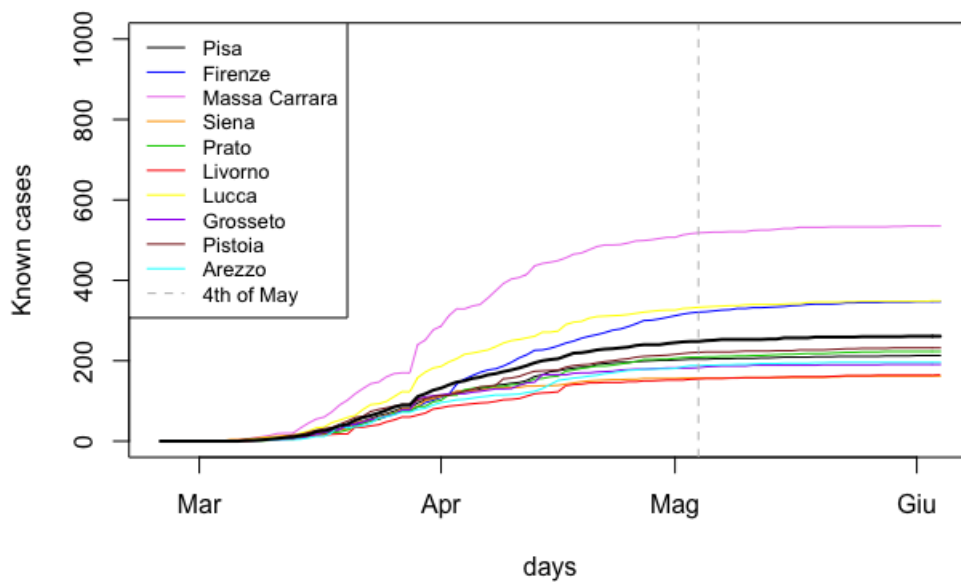

Figure 11: Total cases over 99990 inhabitants.

The variability of these real cases goes about from 100 to 500. Hence we expect that a good fit of the mean is given by trials with variability in this range. Notice that the mean is the black thick line and the majority of the other lines stay under the mean. In Figure 12 we try to fit the mean using only the families as social barriers.

---

[2]https://datastudio.google.com/u/0/reporting/91350339-2c97-49b5-92b8-965996530f00/page/RdlHB
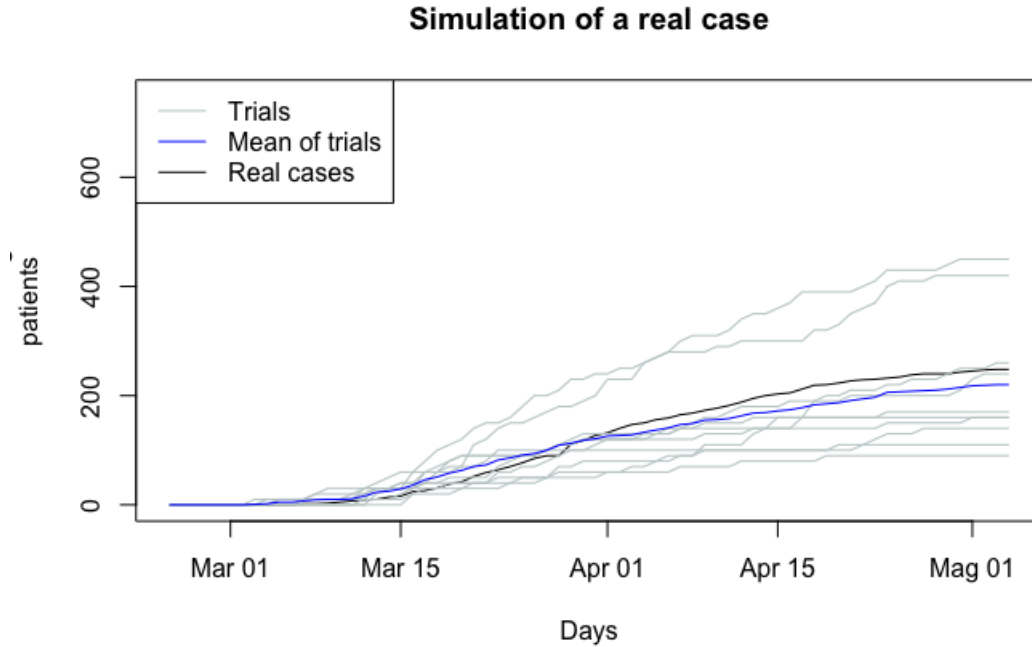
Figure 12: Fit of mean using only families.

The coefficients we used are:

- $\lambda = 0.75$ during the pre-lockdown, $\lambda/4$ during the lockdown for bubbles and pseudo-bubble respectively

- α:
  pre-lockdown: $\alpha_{fam} = 0.5$, $\alpha_{city} = 0.5$
  lockdown:

  - workers: $\alpha_{fam} = 0.65$, $\alpha_{city} = 0.35$
  - students: $\alpha_{fam} = 0.975$, $\alpha_{city} = 0.025$

- $\beta_1 = 0.85$ for the first 18 days and $\beta_2 = 0.7$ for the remaining days.

We manage to fit the total case of Tuscany mean also adding the pseudo-bubbles ad social barriers. In Figure 13 there is the output obtained using the following coefficients:

- $\lambda = 0.75$ during the prelockdown, $\lambda/4$ and $\lambda/3.5$ during the lockdown for bubbles and pseudo-bubble respectively

- α :
  pre-lockdown:

  - $\alpha_{fam} = 0.5$

- $\alpha_{city} = 0.1$
- $\alpha_{ps} = 0.4$

lockdown:

- workers: $\alpha_{fam} = 0.7$, $\alpha_{city} = 0.05$, $\alpha_{ps} = 0.25$
- students: $\alpha_{fam} = 0.975$, $\alpha_{city} = 0.025$
- $\lambda_{fam} = \frac{\lambda}{4}$, $\lambda_{ps} = \frac{\lambda}{3.5}$

- $\beta_1 = 0.85$ for the first 18 days and $\beta_2 = 0.7$ for the remaining days.

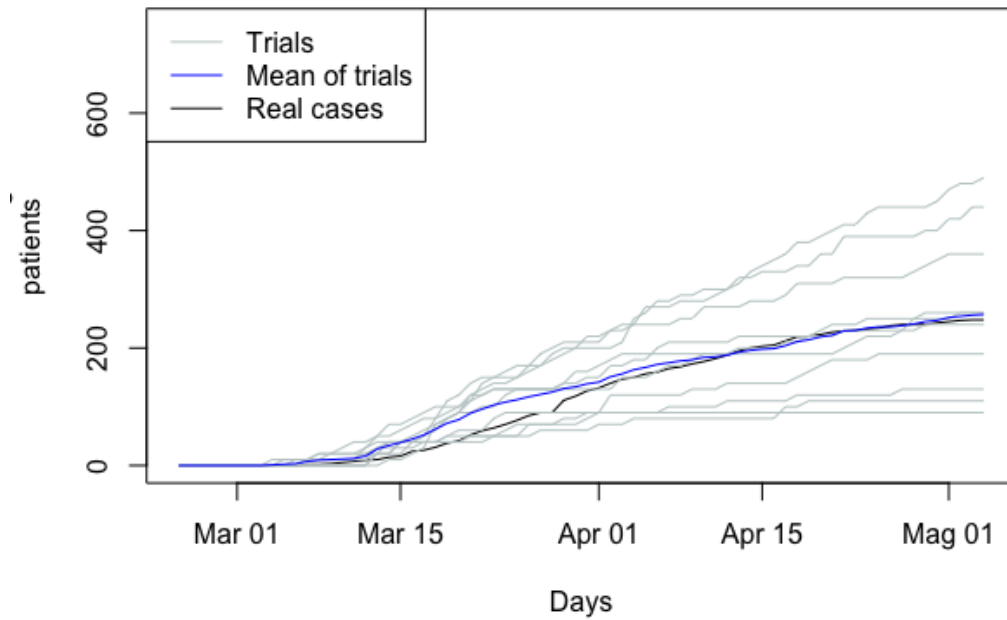Figure 13: Fit of the mean using also pseudo-bubbles.

We have noticed which is the range of variability of the real cases in Tuscany and now we want to make some considerations upon the variability of our model. Making one-hundred trials of our code we show an histogram of the total cases in the last day, that is the 4th of May, both for the model only with bubbles and for the model with pseudo-bubbles too.

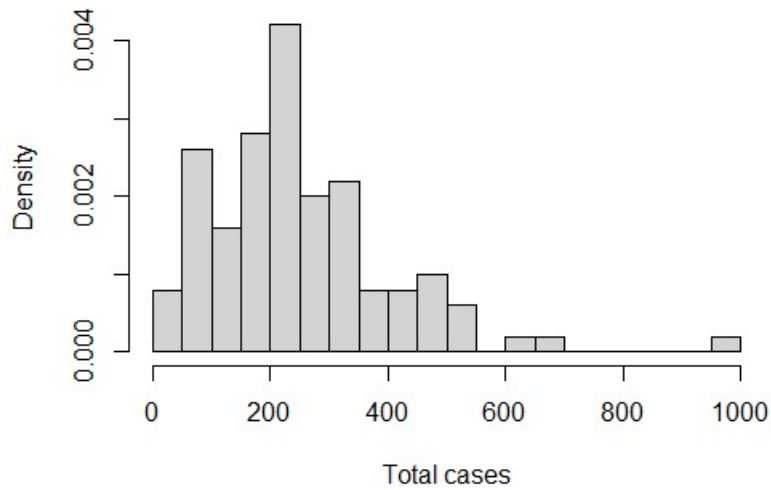For the model without pseudo-bubbles we obtain the following figure.

Figure 14: Total cases on the 4th of May in 100 trials. Model without pseudo-bubbles.

Looking at Figure 14 it is clear that the distribution of the results is asymmetric with a long tail indeed we calculate a skewness value of 1.34. The most frequent values are those between 200 and 250. The mean of the whole trials is 256 and the standard deviation is 156.

We now show the same histogram as obtained using the model with pseudo-bubbles.
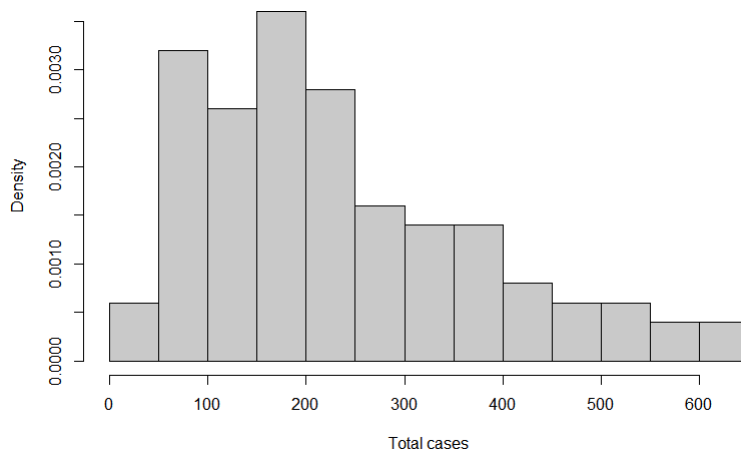

Figure 15: Total cases on the 4th of May in 100 trials. Model with pseudo-bubbles. The histogram is obtained with seed equal to 30.

As we expected, the variability is smaller than in the previous case. The distribution is asymmetric, with a long tail and this is confirmed by a skewness value equal to 0.93. The mean is equal to 237.7, while the standard deviation is equal to 143.9.

27

In both cases it is noteworthy that the variability is a little bit wider if compared to the real cases, especially for the the lower extreme of the range. Using the uniform scenario it is not easy to fit the mean because the introduction of the families slows down a lot the epidemics. Differently using the pseudo-bubble we managed to fit the mean, introducing more realism. However more realism goes together with more parameters, thus for the period under exam is sufficient to use the second scenario to model the real cases.

What if the government would have allowed each individual to meet his friend, colleagues or other acquaintance, during the lockdown? To answer this question we try to launch our code and the result of the simulations is showed in Figure 16.
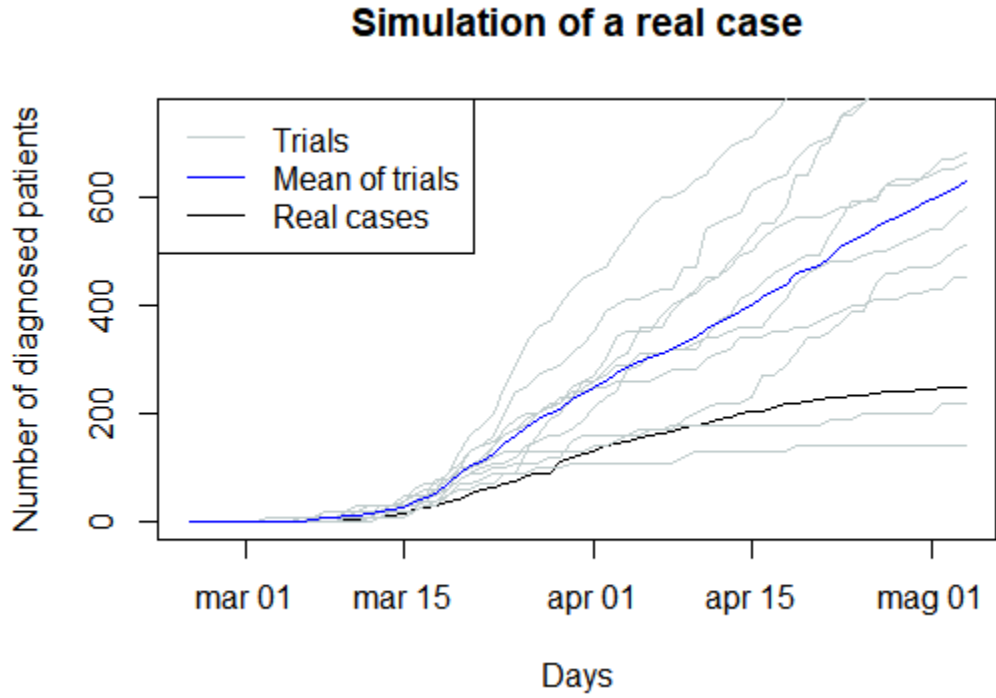
Simulation of a real case

Figure 16: Case in which each individual is allowed to spend a few hours a day with is pseudo-bubble.

We found out that the pseudo-bubbles are useful to describe the number of diagnosed patients also after the lockdown period. In Figure 17 and Figure 18 we manage to fit the total cases beyond the 4th of May, until 3th of June, using two different scenarios.
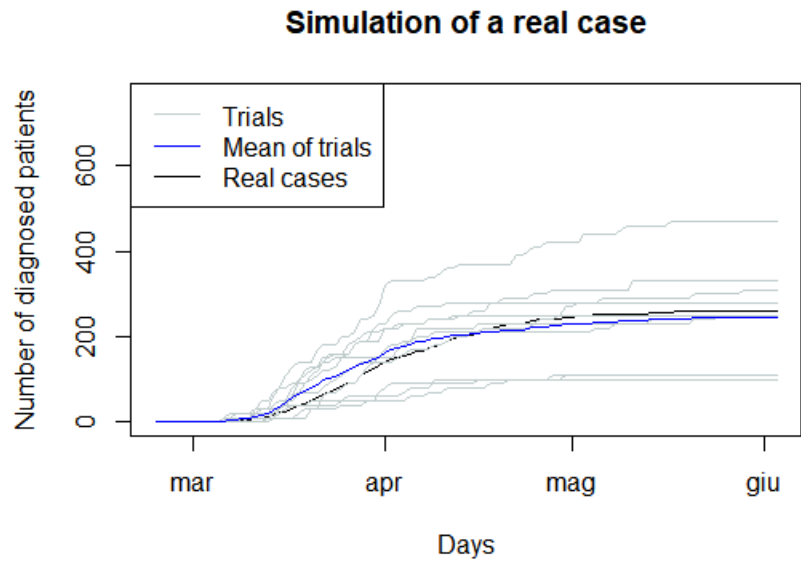
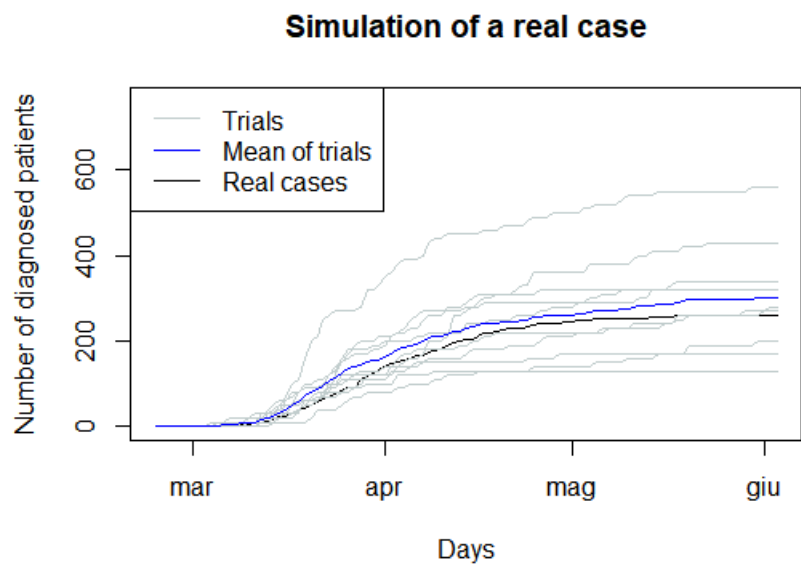Figure 17: Fit post-lockdown after the 4th of May using only bubbles.

Figure 18: Fit of total cases beyond the 4th of May.

To obtain this results we use values of α similar to those of the pre-lockdown but different values of λ. In the short period of time after the end of the lockdown people were very careful and continued to spend their time with their family mostly.

Moreover in May or June individuals stayed outside and as observed this has as a result that being infected was more difficult. For this reasons we used smaller values of λ. In Figure 17 we set $\lambda/9$ for the city. We have already noticed that the pseudo-bubbles slow down the epidemics, indeed if we use the same values of λ as we did in the previous case, the result is that the mean of the trials is smaller than that of the real cases. Hence, we can use smaller values of λ such as $\lambda/7$ in the city and $\lambda/6.5$ for the pseudo-bubbles. The variability is a little bit wider in the case with the pseudo-bubbles but it can be due with the particular simulation.

5   Conclusions

We have developed a Markovian model following natural intuitions about the behavior of a population of individuals, properly simplified in structure and interactions; e.g., during the March-April 2020 lock-down phase, the model takes into account elements like how much time part of the population had contact with other people out of their family, even though with masks or social distancing.

The model requires the specification of several parameters related to biological or sociological issues (e.g. the average time between infection and emergence of symptoms, or the time spent in contact with different subsets of the community). We have identified reasonable values of most of these parameters, either by common sense intuition or looking at the biomedical literature depending on the nature of the parameter; sometimes choosing the value in a straightforward way, sometimes else by more refined logical arguments. However, one parameter remains uncertain, very difficult to be determined a priori, before the time series of infected people of a certain period and regime is known: it is the parameter λ (with its different versions) described in the previous sections, a sort of rate of infectiousness of individuals, which takes into account the level of protection (distances masks, open air etc.) of a specific period.

The value of λ at home, inside the family where individuals spend most of their time and do not use special devices of protections, is taken from data fitted before lock-down phases, when we suppose that the same level of lack of protection was the rule in the society. Such baseline value of λ was used for family interactions in every period of our simulations (pre-lock-down, lock-down, post-lock-down). But the value of λ, during lock-down, in the contacts external to the family was very different and we have deduced it by fitting data; it is an average value of infectiousness under protection strategies which was impossible to predict, due to the novelty of the regime.

However, a natural conjecture was that the same value of λ would hold in the post-lock-down period, concerning the interactions external to the family. The difference between lock-down and post-lock-down should be only due to the different times of exposure to external interactions. Our simulations did not confirm this simple hypothesis: if the same value of λ is used, with the new exposal times, the simulations would produce a much worse behaviour than reality, with an increase of infection after a very short period, opposite to what happend (a long period of rest until September 2020). The only way we have to fit the data during the post-lock-down is to reduce the value of λ (the order of reduction is a division by two). In spite of the same protection rules (distances and masks), starting from May 2020 the infectiousness decreased, a fact that has many potential explanations; the difficulty to conjecture this reduction of λ is the main source of limitation in the hope to use the model of this paper for future predictions.

However, when also the parameter λ is chosen, the statistical variability of all quantities emerges automatically from the random structure of the model and turns out to be very close to reality: both

30

for quantities like the incubation period and serial interval, and for time series like the number of detected infected people in a region like Tuscany, relatively homogeneous with respect to different towns. The model is therefore realistic in its random performances and may be used to assess the uncertainty of predictions and to explore different scenarios: as an example, we have simulated what we should have expected if a government, during the March-April 2020 lock-down, would have allowed each individual to meet a small prescribed number of friend and colleagues, see Figure 18.

Let us finally mention that models of the form of the one of this paper can easily incorporate further developments and improvements like creating more than two layers, creating a subgroup of individuals that can be re-infected after a certain period of time or giving a different infectiousness to some individuals because of their high viral load. We have analyzed the performances of a basic case to identify strengths and weaknesses. Concerning possible applications of a refined system like this one, let us mention the opportunity to simulate details of the past given the present (this requires much more work than presented here), which could be of help in contact tracing strategies.

Acknowledgements

References

[1] BAILEY, N. T., ET AL. *The mathematical theory of infectious diseases and its applications.* Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[2] BELLOMO, N., BELLOUQUID, A., GIBELLI, L., AND OUTADA, N. *A quest towards a mathematical theory of living systems.* Springer, 2017.

[3] BELLOMO, N., BINGHAM, R., CHAPLAIN, M. A., DOSI, G., FORNI, G., KNOPOFF, D. A., LOWENGRUB, J., TWAROCK, R., AND VIRGILLITO, M. E. A multi-scale model of virus pandemic: Heterogeneous interactive entities in a globally connected world. *Math. Models Methods and Appl. Sci 30*, 8 (2020), 1591–1651.

[4] BELLOMO, N., GIBELLI, L., AND OUTADA, N. On the interplay between behavioral dynamics and social interactions in human crowds. *arXiv preprint arXiv:1709.06603* (2017).

[5] BELLOMO, N., PAINTER, K. J., TAO, Y., AND WINKLER, M. Occurrence vs. absence of taxis-driven instabilities in a may–nowak model for virus infection. *SIAM Journal on Applied Mathematics 79*, 5 (2019), 1990–2010.

[6] BIGGERSTAFF, M., COWLING, B. J., CUCUNUBÁ, Z. M., DINH, L., FERGUSON, N. M., GAO, H., HILL, V., IMAI, N., JOHANSSON, M. A., KADA, S., ET AL. Early insights from statistical and mathematical modeling of key epidemiologic parameters of COVID-19. *Emerging infectious diseases 26*, 11 (2020).

[7] BRITTON, T. Stochastic epidemic models: a survey. *Mathematical biosciences 225*, 1 (2010), 24–35.

[8] BRITTON, T., HOUSE, T., LLOYD, A. L., MOLLISON, D., RILEY, S., AND TRAPMAN, P. Five challenges for stochastic epidemic models involving global transmission. *Epidemics 10* (2015), 54–57.

[9] CAPASSO, V., GROSSO, E., AND SERIO, G. Mathematical models in epidemiological studies. i. application to the epidemic of cholera verified in bari in 1973. *Annali Sclavo; Rivista di Microbiologia e di Immunologia 19*, 2 (1977), 193–208.

[10] CAPASSO, V., AND MADDALENA, L. Convergence to equilibrium states for a reaction-diffusion system modelling the spatial spread of a class of bacterial and viral diseases. *Journal of Mathematical Biology 13*, 2 (1981), 173–184.

[11] CAPASSO, V., AND PAVERI-FONTANA, S. A mathematical model for the 1973 cholera epidemic in the european mediterranean region. *Revue d'épidémiologie et de Santé Publiqué 27*, 2 (1979), 121–132.

[12] CECCONI, M., FORNI, G., AND MANTOVANI, A. Covid-19: Summer 2020 report. *Accademia Nazionale dei Lincei, Commissione Salute* (2020).

[13] CECCONI, M., FORNI, G., AND MANTOVANI, A. Ten things we learned about covid-19. *Intensive Care Medicine 46*, 8 (2020), 1590–1593.

[14] CRAIG, B. R., PHELAN, T., SIEDLAREK, J.-P., AND STEINBERG, J. Improving epidemic modeling with networks. *Economic Commentary*, 2020-23 (2020).

[15] DALEY, D. J., AND GANI, J. *Epidemic modelling: an introduction*, vol. 15. Cambridge University Press, 2001.

[16] DAY, M. Covid-19: four fifths of cases are asymptomatic, china figures indicate, 2020.

[17] DIEKMANN, O. Limiting behaviour in an epidemic model. *Nonlinear Analysis: Theory, Methods & Applications 1*, 5 (1977), 459–470.

[18] DIEKMANN, O., HEESTERBEEK, H., AND BRITTON, T. *Mathematical tools for understanding infectious disease dynamics*, vol. 7. Princeton University Press, 2012.

[19] DIEKMANN, O., AND HEESTERBEEK, J. Mathematical epidemiology of infectious diseases. *Model Building, Analysis* (1989).

[20] DU, Z., XU, X., WU, Y., WANG, L., COWLING, B. J., AND MEYERS, L. A. Serial interval of COVID-19 among publicly reported confirmed cases. *Emerging infectious diseases 26*, 6 (2020), 1341.

[21] FLAXMAN, S., MISHRA, S., GANDY, A., UNWIN, H. J. T., MELLAN, T. A., COUPLAND, H., WHITTAKER, C., ZHU, H., BERAH, T., EATON, J. W., ET AL. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature 584*, 7820 (2020), 257–261.

[22] GANYANI, T., KREMER, C., CHEN, D., TORNERI, A., FAES, C., WALLINGA, J., AND HENS, N. Estimating the generation interval for coronavirus disease (COVID-19) based on symptom onset data, March 2020. *Eurosurveillance 25*, 17 (2020), 2000257.

[23] GIORDANO, G., BLANCHINI, F., BRUNO, R., COLANERI, P., DI FILIPPO, A., DI MATTEO, A., AND COLANERI, M. Modelling the COVID-19 epidemic and implementation of population-wide interventions in Italy. *Nature Medicine* (2020), 1–6.

[24] GUDBJARTSSON, D. F., HELGASON, A., JONSSON, H., MAGNUSSON, O. T., MELSTED, P., NORDDAHL, G. L., SAEMUNDSDOTTIR, J., SIGURDSSON, A., SULEM, P., AGUSTSDOTTIR, A. B., ET AL. Spread of sars-cov-2 in the icelandic population. *New England Journal of Medicine* (2020).

[25] HE, X., LAU, E. H., WU, P., DENG, X., WANG, J., HAO, X., LAU, Y. C., WONG, J. Y., GUAN, Y., TAN, X., ET AL. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine 26*, 5 (2020), 672–675.

[26] KÄLLEN, A., ARCURI, P., AND MURRAY, J. A simple model for the spatial spread and control of rabies. *Journal of theoretical biology 116*, 3 (1985), 377–393.

[27] KEELING, M. J., AND EAMES, K. T. Networks and epidemic models. *Journal of the Royal Society Interface 2*, 4 (2005), 295–307.

[28] KERMACK, W. O., AND MCKENDRICK, A. G. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of London. Series A, Containing papers of a mathematical and physical character 115*, 772 (1927), 700–721.

[29] KIM, D., AND QUAINI, A. Coupling kinetic theory approaches for pedestrian dynamics and disease contagion in a confined environment. *arXiv preprint arXiv:2003.08357* (2020).

[30] KNOPOFF, D. A., AND TRUCCO, F. A. A compartmental model for antibiotic resistant bacterial infections over networks. *International Journal of Biomathematics 13*, 01 (2020), 2050001.

[31] LAUER, S. A., GRANTZ, K. H., BI, Q., JONES, F. K., ZHENG, Q., MEREDITH, H. R., AZMAN, A. S., REICH, N. G., AND LESSLER, J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine 172*, 9 (2020), 577–582.

[32] LAVEZZO, E., FRANCHIN, E., CIAVARELLA, C., CUOMO-DANNENBURG, G., BARZON, L., DEL VECCHIO, C., ROSSI, L., MANGANELLI, R., LOREGIAN, A., NAVARIN, N., ET AL. Suppression of a sars-cov-2 outbreak in the italian municipality of vo'. *Nature 584*, 7821 (2020), 425–429.

[33] LI, Q., GUAN, X., WU, P., WANG, X., ZHOU, L., TONG, Y., REN, R., LEUNG, K. S., LAU, E. H., WONG, J. Y., ET AL. Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia. *New England Journal of Medicine* (2020).

[34] LINTON, N. M., KOBAYASHI, T., YANG, Y., HAYASHI, K., AKHMETZHANOV, A. R., JUNG, S.-M., YUAN, B., KINOSHITA, R., AND NISHIURA, H. Incubation period and other epidemiological characteristics of 2019 novel coronavirus infections with right truncation: a statistical analysis of publicly available case data. *Journal of clinical medicine 9*, 2 (2020), 538.

[35] LU, H., AI, J., SHEN, Y., LI, Y., LI, T., ZHOU, X., ZHANG, H., ZHANG, Q., LING, Y., WANG, S., ET AL. A descriptive study of the impact of diseases control and prevention on the epidemics dynamics and clinical features of SARS-CoV-2 outbreak in Shanghai, lessons learned for metropolis epidemics prevention. *Medrxiv* (2020).

[36] MURRAY, J. *Mathematical biology II: spatial models and biomedical applications*, vol. 3. Springer-Verlag, 2001.

[37] NISHIURA, H., LINTON, N. M., AND AKHMETZHANOV, A. R. Serial interval of novel coronavirus (COVID-19) infections. *International journal of infectious diseases* (2020).

[38] PERELSON, A. S. Modelling viral and immune system dynamics. *Nature Reviews Immunology 2*, 1 (2002), 28–36.

[39] PORCO, T. C., HOLBROOK, K. A., FERNYAK, S. E., PORTNOY, D. L., REITER, R., AND ARAGÓN, T. J. Logistics of community smallpox control through contact tracing and ring vaccination: a stochastic network model. *BMC public health 4*, 1 (2004), 34.

[40] TINDALE, L., COOMBE, M., STOCKDALE, J. E., GARLOCK, E., LAU, W. Y. V., SARASWAT, M., LEE, Y.-H. B., ZHANG, L., CHEN, D., WALLINGA, J., ET AL. Transmission interval estimates suggest pre-symptomatic spread of COVID-19. *MedRxiv* (2020).

[41] TUCKWELL, H. C., AND WILLIAMS, R. J. Some properties of a simple stochastic epidemic model of SIR type. *Mathematical biosciences 208*, 1 (2007), 76–97.

[42] VESPIGNANI, A. Modelling dynamical processes in complex socio-technical systems. *Nature physics 8*, 1 (2012), 32–39.

[43] YOU, C., DENG, Y., HU, W., SUN, J., LIN, Q., ZHOU, F., PANG, C. H., ZHANG, Y., CHEN, Z., AND ZHOU, X.-H. Estimation of the time-varying reproduction number of COVID-19 outbreak in China. *International Journal of Hygiene and Environmental Health* (2020), 113555.

[44] ZEB, A., ALZAHRANI, E., ERTURK, V. S., AND ZAMAN, G. Mathematical model for coronavirus disease 2019 (COVID-19) containing isolation class. *BioMed research international 2020* (2020).

[45] ZHANG, J., LITVINOVA, M., WANG, W., WANG, Y., DENG, X., CHEN, X., LI, M., ZHENG, W., YI, L., CHEN, X., ET AL. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: a descriptive and modelling study. *The Lancet Infectious Diseases* (2020).

[46] ZHAO, S., GAO, D., ZHUANG, Z., CHONG, M. K., CAI, Y., RAN, J., CAO, P., WANG, K., LOU, Y., WANG, W., ET AL. Estimating the serial interval of the novel coronavirus disease (COVID-19): A statistical analysis using the public data in Hong Kong from January 16 to February 15, 2020.