

AperTO - Archivio Istituzionale Open Access dell'Università di Torino

**Fine-Grained Agent-Based Modeling to Predict Covid-19 Spreading and Effect of Policies in Large-Scale Scenarios**

**This is a pre print version of the following article:**

*Original Citation:*

*Availability:*

This version is available <http://hdl.handle.net/2318/1872518> since 2022-08-23T07:19:15Z

*Published version:*

DOI:10.1109/JBHI.2022.3160243

*Terms of use:*

Open Access

Anyone can freely access the full text of works made available as "Open Access". Works made available under a Creative Commons license can be used according to the terms and conditions of said license. Use of all other works requires consent of the right holder (author or publisher) if not exempted from copyright protection by the applicable law.

(Article begins on next page)

# Fine-Grained Agent-Based Modeling to Predict Covid-19 Spreading and Effect of Policies in Large-Scale Scenarios

Gianfranco Lombardo<sup>a</sup>, Mattia Pellegrino<sup>a</sup>, Michele Tomaiuolo<sup>a</sup>, Stefano Cagnoni<sup>a</sup>,  
Monica Mordonini<sup>a</sup>, Mario Dante Lucio Giacobini<sup>b</sup> and Agostino Poggi<sup>a</sup>

<sup>a</sup>Department of Engineering and Architecture University of Parma Parma, Italy

<sup>b</sup>DAMU - Data Analysis and Modeling Unit, Department of Veterinary Sciences, University of Torino, Turin, Italy

## ARTICLE INFO

### Keywords:

Covid-19  
Epidemic modeling  
multi-agent simulation  
simulation  
actor model  
ABMS

## ABSTRACT

The Covid-19 pandemic has placed epidemic modeling at the forefront of worldwide public policy making. Nonetheless, modeling and forecasting the spread of COVID-19 remains an open problem. Modeling this kind of complex systems requires the estimation of several parameters and domain assumptions that can faithfully describe the analysis scenario and its dynamics. This task becomes even more challenging for an air-borne disease when a fine grained simulation has to be combined with a large-scale scenario as for Covid-19. In this paper, we modeled Covid-19 spreading in the most populated region of Italy (Lombardy) with about 10 million people. In order to achieve this result we propose a framework that implements: a scale-free modeling of the social contacts combining a sociability rate, demographic information and geographical assumptions; a multi-agent system adopting the actor model and using ActoDeS to efficiently implement 10 million concurrent agents exploiting the High Performance Computing technology. Finally, we simulated the epidemic scenario from January to April 2020 and from August to December 2020, modeling the lockdown policies, to extract valuable information about the transmission rate and probability of Covid-19.

## 1. Introduction

At the end of 2019, a novel coronavirus disease (COVID-19) was declared as a major health hazard by World Health Organization (WHO). This disease has grown rapidly in many countries, and the global number of COVID-19 cases has increased at a rapid rate. Most governments all over the world enacted severe measures in full emergency to limit the impact of this new virus among their communities, in particular lockdown rules and social distancing to reduce social contacts. However, we have witnessed several outbreaks which were difficult to manage and to accurately predict. Most of the difficulties are related to the nature of the virus itself, for example the existence of asymptomatic infectious patients, several days of incubation before the emergence of symptoms, and a lack of knowledge about many aspects of SARS-Cov-2 transmission and its dynamics. In light of this context, the possibility of modeling complex and large scenarios to simulate covid spreading becomes a key factor for prevention, sustainability of public health services and policy making. Different approaches can be considered to achieve this result, depending on the desired granularity and the level of details known a priori about the context. Covid-19 is an air-borne disease whose transmission is facilitated by social contacts in situations where physical distancing is difficult to avoid, for example: job places, transportations, schools, conviviality, free time activities and family interactions. In this research work, our desiderata is to simulate each single person's social behaviors, in a large community of about 10 mil-

lion habitants (Lombardy region population in 2019), taking also into account their province of residence. In order to model this system we exploited Agent-Based Modeling and Simulation techniques (ABMS). Each agent represents an individual who each day has a certain number of social interactions depending on his own age, work and level of sociability, that we introduced as an individual's feature. We aimed to model covid transmission with a fine-grained resolution, also to simulate scenarios with and without social distancing and lockdown policies. However, executing millions of concurrent agents could represent a bottleneck for ABMS. To overcome this issue, we built an efficient framework that implements the agents as distributed and concurrent actors who share asynchronous messages to communicate and to organize their behaviors. Finally, we combined this software architecture with the High Performance Computing facility of the University of Parma to scale and distribute the computational workload on the available resources. Finally, we simulated the Covid-19 outbreaks in Lombardy from January to April 2020 and from August to December 2020, namely the first and second waves of coronavirus disease-19 outbreak. The main contribution of this paper are: (a) a simulation framework for fine-grained and large scale scenarios; (b) a model for social interactions based on sociability rates, demographic based with a power-law degree distribution; (c) modeling for the lockdown policies and mask wearing effects; and (d) an estimation of the real impact of Covid-19 in Lombardy during that waves. The paper is organized as follows: TO DO

\*This research benefits from the HPC (High Performance Computing) facility of the University of Parma

\*Corresponding author

 gianfranco.lombardo@unipr.it (G. Lombardo)

ORCID(S): 0000-0003-1808-4487 (G. Lombardo)

## 2. Literature review

Different modeling techniques have been proposed to modeling and solve real and complex epidemic scenarios. Two

widely used techniques are system dynamics (SD) and agent based modeling (ABM). System dynamics allows analyzing the modeled system at a high level of abstraction where the population is divided into compartments. A common case is the SEIR model (Susceptible Exposed Infective Recovered) [1], where the population can move from one compartment to another according to predefined flow rates. However, the traditional SEIR model is not fine-grained enough to model lockdown policies as for the case of Covid-19 management and control. This limit with the Covid-19 world pandemic has motivated several research works that aim to extend that model in order to allow more heterogeneity and flexibility [2, 3, 4]. However the extensions do not address the main key-issue that is related with the main parameter, the basic reproduction number ( $R_0$ ) that is not policy-invariant. Indeed, it depends on the number of contacts each person has and the infection probability of the contacts. It is hard to translate a real policy into the value of  $R_0$  it will induce [5]. On the other hand, Agent-based approach models the behavior of each individual agent and the interaction between agents. ABM can be used to study the system at different levels of abstractions, ranging from high-level to low level of abstraction that is close to the real world systems. For a discussion about ABM and its advantages over system dynamics models, we refer the reader to [6] and [7].

Covid-19 spreading estimation is challenging and requires to take into account heterogeneous interaction rates among people. These requirements can be easily modeled using ABMS. In [8] and [9] individuals are modeled as moving particles. Covid-19 infections take place when two particles come closer than a certain contact radius. Social distancing for Covid-19 is modeled as changes in the contact radius or momentum equation of the particles but introducing several parameters that are difficult to estimate for large scenarios. In [10], the authors model Covid-19 spreading by replacing the moving particles with contact networks for households, work and random contacts. An extension of this approach is considered also in [11] to simulate the effect of the German lockdown in November 2020. Moreover, the structure of the contact network significantly affects the disease spread as demonstrated also for the case of Covid-19 in [12]. Contact networks are usually modeled with a power-law distribution because the scale-free property ensure to have few people who have contact with many different people while most people interact with a smaller consistent set [13]. However, to the best of our knowledge we found a lack of models to simulate the outbreaks and the lockdown policies with a fine-grained detail in real large scenarios. Most of the examples in literature are suitable to model a limited number of individuals to achieve statistics that enable to project the results in the most general large case.

### 3. ActoDeS

ActoDeS is a software framework that has the goal of both simplifying the development of concurrent and distributed complex systems and guarantying an efficient execution of

applications [14]. In particular, an application is based on a set of actors [15] that perform tasks concurrently and interact with each other by exchanging asynchronous messages. ActoDeS has a layered architecture composed of an application and a runtime layer. The application layer provides the software components that an application developer needs to extend or directly use for implementing the specific actors of an application. The runtime layer provides the software components that implement the middleware infrastructures to support the development of standalone and distributed applications. Depending on the complexity of the application and on the availability of computing and communication resources, one or more actor spaces can manage the actors of the application. An actor space acts as “container” for a set of actors and provides them the services necessary for their execution. In particular, an actor space contains a set of actors (application actors) that perform the specific tasks of the current application and two special actors called executor and service provider. The executor manages the concurrent execution of the actors of the actor space. The service provider enables the actors of an application to perform new kinds of action (e.g., to broadcast a message or to move from an actor space to another one). Each actor has a system-wide unique identifier called reference that allows it to be reached in a location transparent way independently of the location of the sender (i.e., their location can be the same or different). Communication between actors is buffered: incoming messages are stored in a mailbox until the actor is ready to process them; moreover, an actor can set a timeout for waiting for a new message and then can execute some actions if the timeout fires. After its creation, an actor can change several times its behavior until it kills itself. Each behavior has the main duty of processing a set of specific messages through a set of message handlers called cases. Therefore, if an unexpected message arrives, then the actor mailbox maintains it until a next behavior will be able to process it. ActoDeS has been used for modeling and simulation [16], for the development of data analysis tools [17, 18] and their use for the analysis of social networks data [19].

### 4. Simulator architecture

The number of people that live in Lombardy is about ten millions, therefore a distributed architecture is necessary to achieve the simulation with fine-grained behaviors and details. Due to the large amount of concurrent agents involved in the simulation process, enough memory and processing resources are unavailable on a single computational node. For this reason, we have exploited ActoDeS to design and efficiently implement a distributed epidemic simulator using the actor model []. Every actor represents a generic person with an age, a province of residence, and his own level of sociability. Furthermore, each person has a stochastic behavior depending on some other parameters that will be discussed later [SECTION?]. The simulation process is divided into several “*epochs*”, where every epoch represents a different day. Moreover, people can change their behavior depend-

ing on the current epoch (e.g., normal period or lockdown period). At the end of every epoch, the simulator creates a report file useful to resume the simulation process from a specific point. Individuals are created and divided according to two criteria:

- a. Partitioning according to their belonging province;
- b. Splitting in equal size subsets, depending on the number of actorspaces involved in the simulation.

As a typical distributed framework, the simulator involves a set of computational nodes whose execution is driven by a set of schedulers and managers. In particular, each manager has the duty of creating the subset of agents for its computational node and synchronize the execution of the simulation with the execution of the other computational nodes. Moreover, one of such managers assumes the role of “master”, who has the duties of partitioning the agents involved in the simulation; it sends to the other managers the information they need for creating the agents under their control.

ActoDeS uses passive and active actors to allow large-scale development; this feature can be exploited to optimize and improve the simulation process according to the best suited algorithm for a particular situation.

The simulation process can be described by the following steps:

1. Master manager creates the agents and sends information for creating a subset of agents to each scheduler (including itself).
2. Each manager creates all the actors of its subset.
3. Repeat until the end of the simulation:
  - a. Managers send a synchronization message to the other managers and wait for the corresponding messages from them.
  - b. Schedulers perform an execution step of all their actors.
  - c. Scheduler send a "end step" message to all their actors and managers.

In this architecture, every actorspace acts as a manager. The last generated actorspace plays the master role.

ActoDeS provides a very simple scheduler called "CycleScheduler". This one can be used in a wide variety of applications, more specifically, also in ABMS applications. Furthermore, this scheduler manages the passive actors within its actorspace and cyclically repeats the same actions until the simulation ends:

1. Send a “step” message to all agents and increment the “step” value; this operation triggers the transition from one epoch to the next one.
2. Perform an execution step of all the agents.

In figure 1 there is a diagram that shows the simulator architecture. In the diagram, the population is split in  $N$  actorspaces; the last one acts as a master. However, every actorspace coordinates and manages a subset of the population. Every subset includes, generically, the agents that go

from  $(n - k) \cdot p - 1$  to  $(n - k + 1) \cdot p - 1$ . Where  $n$  is the number of partitions,  $k$  identifies the actual partition and  $p$  represents the constant  $p = \frac{\text{population}}{N}$ .

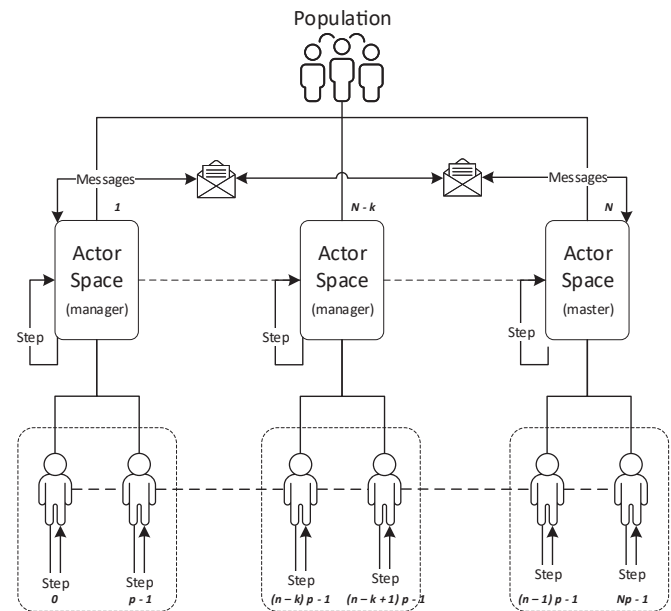


Figure 1: Simulator architecture

#### 4.1. Communication management

The agents need to exchange data and often such interactions involve agents on remote nodes; therefore, the partition of agents on several computational nodes may add a significant communication cost. An important solution to reduce the cost of communication is to reduce the frequency of interactions, merging multiple interactions together. In a conservative distributed simulation system, that synchronizes the simulation step of all the computational nodes involved in the simulation, a solution is to group all the messages directed to the agents of a specific computational node into the message that identifies the subsequent synchronization message.

#### 5. Data collection

In this section we present all of the sources we used to collect data for the simulation of the epidemic scenario in Lombardy and to estimate some of the modeling parameters.

- We retrieved data about COVID-19 spreading in Lombardy from the official institutions [20] and [21]. From these sources we collected data about new daily cases and deaths registered in the Lombardy region over time
- We used also demographic information about population of Lombardy, including its number of inhabitants and their age distribution [22], Table 2
- As a starting point for social interactions modeling, we used data from the National Institute of Health [23]

and [24]. These data reported in Table 1 have been collected by the researchers in order to better understand how respiratory infections might spread. They recruited 7,290 people from eight European countries (Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, The Netherlands, and Poland) to participate in their study. They asked the participants to fill out a diary that documented their physical and nonphysical contacts for a single day. Physical contacts included interactions such as a kiss or a handshake. Nonphysical contacts were situations such as a two-way conversation without skin-to-skin contact. Participants detailed the location and duration of each contact.

- To model Covid-19 compartments we use preliminary data collected by [25], [26] and the age susceptibility to COVID-19 virus [23]
- The modeling of the Italian lockdown required various information about the set of “essential workers” [27]. Moreover, we collected data indicating what percentage of the population used protective devices [28] and their effectiveness [29]
- To have a comparison, in order to prove our results, we finally used the results of the seroprevalence survey conducted during the same period in Lombardy.

Age	Total	Home	School	Work	Transport	Free time	Other activities
0-4	16.54	4.49	5.27	0	0.98	3.06	2.75
5-9	20.49	4.61	8.87	0	1.12	4.53	1.37
10-14	27.38	4.43	11.98	0.2	1.35	5.62	3.8
15-19	29.28	4.59	13.22	0.05	1.74	6.83	2.87
20-24	22.15	3.51	1.17	4.49	0.96	7.23	4.8
25-29	21	3.47	2.23	5.21	1.13	6.3	2.66
30-34	18.03	3.55	0.85	3.92	0.76	5.24	3.72
35-39	21.25	4.38	0.68	7.78	1.05	3.92	3.45
40-44	22.35	3.88	2.53	7	0.67	4.48	3.79
45-49	19.27	2.99	2.61	8.24	0.88	1.93	2.64
50-54	22.3	2.75	5.54	8.05	0.52	2.02	3.41
55-59	18.27	2.88	1.41	4.6	0.68	3.62	5.06
60-64	18.43	3.28	1.07	6.05	0.87	3.53	3.63
65-69	12.74	3.1	0.55	0.48	0.95	3.33	4.33
70+	10.55	3.24	0.06	1.04	0.22	4.22	1.77

Table 1: Average number of interactions by age from [24]

Age	Male	Female	Total	
				%
0-4	205.299	194.557	399.856	4.0%
5-9	239.046	224.761	463.807	4.6%
10-14	254.391	238.914	493.305	4.9%
15-19	247.380	230.340	477.720	4.7%
20-24	253.912	229.596	483.508	4.8%
25-29	262.010	250.130	512.140	5.1%
30-34	282.410	275.681	558.091	5.5%
35-39	313.625	306.736	620.361	6.1%
40-44	369.540	359.909	729.449	7.2%
45-49	420.573	409.422	829.995	8.2%
50-54	425.683	420.177	845.860	8.4%

Age	Male	Female	Total	
				%
55-59	374.754	381.832	756.586	7.5%
60-64	305.783	324.981	630.764	6.2%
65-69	269.084	295.617	564.701	5.6%
70-74	256.056	292.193	548.249	5.4%
75-79	200.480	250.519	450.999	4.5%
80-84	157.532	222.870	380.402	3.8%
85-89	80.444	149.594	230.038	2.3%
90-94	26.783	73.420	100.203	1.0%
95-99	4.737	20.880	25.617	0.3%
100+	248	2.070	2.318	0.0%
<b>Totale</b>	<b>4.949.770</b>	<b>5.154.199</b>	<b>10.103.969</b>	<b>100.0%</b>

Table 2: Age distribution of the Lombardy population [22]

## 6. Social interactions modeling

Modeling physical contacts among people is crucial to simulate an air-borne disease transmission like Covid-19.

However, this is still an open problem, when fine-grained details are necessary for the simulation. To tackle this kind of problems, we propose a model of social interactions based on the real daily contacts from [24], that is enriched with a sociability rate that divides people with an high, medium and low level of sociability. These factors are designed in order to increase or decrease the average number of daily contacts of people based on their age. To define these parameters we retrieve a common hypothesis in network science that asserts that social networks have commonly a power-law distribution with an exponent between 2 and 3, also known as the scale-free property [30]. Moreover, the entire population is partitioned in 12 groups, accounting for the provinces of Lombardy. This subdivision ensures that the interactions are not purely casual. Indeed, a generic person meets most of his contacts in his own belonging province. Only a small fraction of interactions will occur with people belonging to another province. The Italian demographic structure and the heterogeneity of social contacts, at different ages, is also taken into account, to estimate the correct average number of interactions that a person should have, daily (See Table 1). More specifically, every agent is characterized by the following attributes:

- Identification number
- Belonging province
- Age
- Social interaction ratio
- Number of contacts
- Current infection phase
- If he/she will use a protective device during the lockdown
- If he/she is an essential worker during the lockdown

Some of these parameters will be further explained in the following sections.

### 6.1. Habitual contacts’ modeling

In the first simulation week, a set of contacts is created for each person. This set will represent, for the entire simulation period, the habitual contacts’ set. More specifically, this set contains all those people that a generic person usually meets during his typical day. For example: family members, friend, co-workers, etc.

It has been assumed that an agent can meet a certain number of people who belong to this set and a certain people number who don’t belong to it. Moreover, it is inferred from table 1, that 65% of the daily meeting of a generic agent is generated by its habitual contacts’ set, while the remaining 35% is made up by new contacts. The formula used to find this percentages is:

$$\sum_{i=0}^K \frac{Home[i]+School[i]+Work[i]}{Total[i]} \quad (1)$$

The values of  $i$  in the above formula represent the age groups which are in table 1.

## 6.2. Sociability rate estimation

The social network generated by the interactions can be theoretically mapped and measured using network science techniques. We assume that a generic agent represents a node in the social graph and the outgoing and incoming edges, from this node, represent respectively: the contacts met by the node and the contacts that meeting the node.

The social interactions' distribution is a determining factor for studying the epidemic spread. For this reason, various hypothesis have been evaluated, but in the end we decided to focus our studies on a *power law* distribution type. A *power law* model fits very well with many real-word use cases and can be used to represent a social interaction model.

In a typical modern society there are different aggregation places: offices, schools, parks, provinces, etc. These situations give place to a particular category called *hub*. A *hub* is a special node that has a greater interaction degree than the others. In a typical *power law* distribution, nodes with relatively high degree are more likely to form. This last assumption makes the *power law* suitable for shaping our social network.

To estimate the sociability rate, the attention is focused on the degree graph distribution, generated by the social interactions. The degree of a node is the number of edges connected to the node. The distribution degree is a function so defined:  $p_k : \mathbb{N} \rightarrow [0, 1]$ , which associates a degree  $k$  with the probability that a node has that degree:

$$p_k = \frac{N_k}{N} \quad (2)$$

where  $N$  is the number of total nodes and  $N_k$  is the number of nodes that have that  $k$  degree.

We have aimed at achieving a *power law scale-free* distribution [30], because it is characterized by the presence of many *hub* nodes that create, in their representation graph, a "long tail". A "*scale-free*" distribution does not change its shape if scaled in dimension. To analyze the distribution, we have referred to the state of the art of curve fitting [31], using the Likelihood ratio test and Kolmogorov-Smirnov distance to determine which probability distribution is compatible with our data looking for a power law distribution. We have exploited the *Complementary cumulative density function* (CCDF), because CCDFs are frequently preferred for visualizing a heavy-tailed distribution. A CCDF measures the probability that a certain value has its degree greater than a reference variable  $x$ , equation 3

$$p_k = \sum_{q=k+1}^{\infty} p_q \quad (3)$$

If  $p_k$  follows a *power law* trend, then the cumulative distribution scales following the law:

$$p_k \sim K^{-\gamma+1} \quad (4)$$

People are modeled according to their own social interaction degree, which can be different from the others. They can be identified according to their *Sociability rate*, that can

have four different values: *high*, *medium*, *low* e *quarantine*. Furthermore, according to this ratio, an agent can meet more or less people in a single simulation day. The interactions occur randomly, generating an interaction graph. People with a *high* degree meet an above average number of people, those with *medium* degree meet exactly the average number, and people with a *low* degree meet a lower than average number of people. The population is made up of 20% of agents with an *high* and *medium* degree and the remaining 80% with a *low* degree. To achieve this distribution, we have introduced three different social interaction multipliers that scale the number of people to meet in a generic simulation day. People with *medium* degree meet exactly the average number of people specified in Table 1, so this particular multiplier is equal to one. To estimate the remaining multipliers we have performed a grid-search over the parameters looking for the optimal combination that returns a power-law distribution with a scale-free property ???. Low and high parameters range respectively between 0.1-0.9 and 1.1 to 1.9. The particular case of *quarantine* will be further detailed later.

We have analyzed the CCDFs obtained with each configuration of the grid-search. In Figure 2, we report some of the results with a comparison with the power-law distribution.

In order to choose the optimal configuration, we have analyzed three different parameters:

- *Likelihood*: The result of the likelihood-ratio test. This tool enables to compare two candidate distributions with the empirical data. The result is positive if the data are more similar to the first distribution and is negative if the data are more similar to the second one. The robustness of this test is also evaluated against the null hypothesis measuring the P-value.
- *KS test*: It is the Kolmogorov-Smirnov distance. This tool tests the distance between a single candidate with the empirical data. In our case it is used to compare a sample with our reference distribution, that is a power law.

Table 3 shows the results obtained according to the used multipliers.

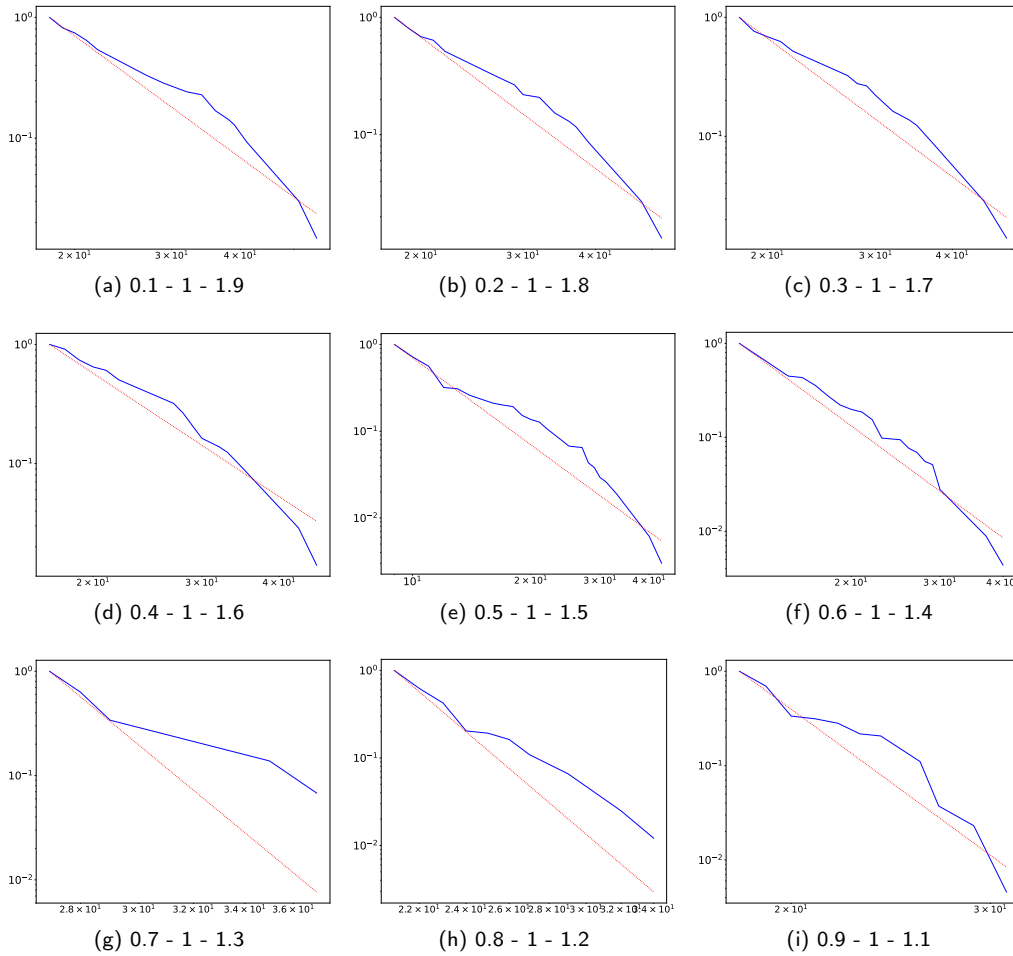
Hence, in light of the results in Table 3, we have decided to use the configuration *0.2 - 1 - 1.8*. This set of multipliers is suitable because it returns the highest likelihood ratio value and in particular a power-law exponent  $\alpha$  equal to 2.64, with an  $x_{min}$  equal to 4. A distribution is *scale-free* if  $\alpha < 3$ .

## 7. Modeling methodology

In this section, we present the details about the epidemic diffusion model used for the simulations, the main hypothesis we adopted to model the Covid-19, and the methodology we propose to model the lockdown phase and the wearing of protective masks among the population.

### 7.1. Epidemic diffusion model

We have implemented a Covid-19 diffusion model starting from the compartments of the mathematical model SEIR (Susceptible-Exposed-Infective-Recovered)[1]. This model



**Figure 2:** Complementary cumulative density function according to interaction multipliers - Red line (PowerLaw Fit) , Blue line (Empirical Data)

Multiplier			Likelihood		P-value	KS test
Low	Medium	High	1° Distributon Powerlaw	2° Distribution Lognormal		
0.1	1	1.9	1.99		$\approx 0$	0.11
0.2	1	1.8	364		$\approx 0$	0.21
0.3	1	1.7	-1.46		0.145	0.116
0.4	1	1.6	-150		$\approx 0$	0.122
0.5	1	1.5	-403		$\approx 0$	0.093
0.6	1	1.4	-465		$\approx 0$	0.113
0.7	1	1.3	-205		$\approx 0$	0.12
0.8	1	1.2	-301		$\approx 0$	0.09
0.9	1	1.1	-803		$\approx 0$	0.127

**Table 3:** Grid-search results of the sociability rates considering the most promising distribution candidates

represents one of the most widely adopted mathematical models to characterize epidemic dynamics and to predict possible contagion scenarios. It is based on a series of dynamic ordinary differential equations that consider the amount of the population subject to contagion, the trend over time of individuals who recover after infection, and the individuals who unfortunately die. However, a limit of the SEIR model

is its coarse grain nature with respect to individuals behaviors. In light of this, we have used the same compartments provided by SEIR, adding an intermediate one specifically for the Covid-19 context between Infective and Recovered, named Positive. This additional compartments is required to distinguish people who are officially recognized as positive to the virus using throat swabs in the real data. Moreover,

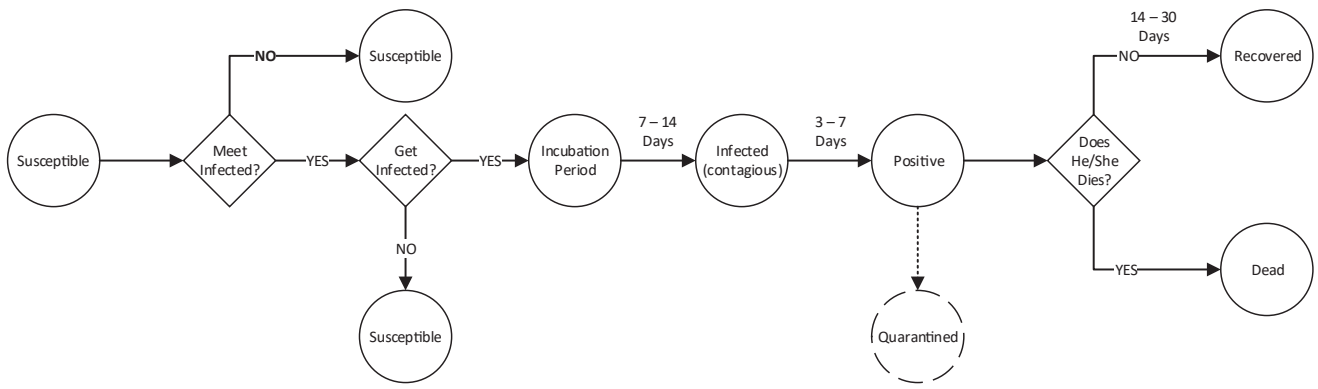


Figure 3: Infection cycle

this distinction has an impact on the behavior definition: it is supposed that a generic positive will be quarantined, to prevent the infection spread, while an infected patient can theoretically be asymptomatic and unaware of his condition, leading a normal life with a normal number of social interactions. Finally, this difference between Infected and Positive is fundamental to analyze the epidemic dynamics, especially during the first wave, when the possibility of testing large amount of people was limited. This has probably introduced significant estimation errors in the official data, about the real number of infected people. Since our work is agent-based, we have not used the differential equations model but we have defined the individual behaviors of the actors. To describe the pathogen spread dynamics, we have assumed that a generic person can transition through various phases when he/she contracts the disease. Such phases represent the virus life cycle inside the human body. At the beginning, all persons are in a susceptibility stage. In this stage, every person can be infected by another one, who is contagious. A person who is infected, moves from a susceptibility phase to an incubation phase and remains in this stage for a certain time period and then moves into an infection stage. A person in this condition can infect other people. When this phase ends, the person becomes positive. After a certain time amount, a positive can heal or die. There is no death probability, but deaths follow the real death curve trend, in Lombardy. When a person heals, he/she cannot be infected any more. In particular, the incubation phase lasts from 7 to 14 days, that infectious one from 3 to 7 days, and the positive one from 14 to 30 [25] [26]. In Figure 3 there is a diagram that represents a generic infection cycle.

Moreover, people susceptibility is different according to their age [23], as shown in Table 4.

Age	Susceptibility
0-14	31%
15-64	47%
65+	100%

Table 4: Susceptibility by age

## 7.2. Lockdown modeling

On 8 March 2020, the Italian government decided to apply some containment measures to the whole national territory to stop the COVID-19 spread. We will refer to this condition as the “Lockdown”. Additionally, these measures included:

- non-essential activities closure,
- crowd prohibition,
- travel prohibition, if not justified by proven needs.

In this phase there was a significant decrease in the overall social interactions. For most people, the reduction was homogeneous, because they could meet only family members or cohabitants, except for the few times they need to buy groceries. To model the lockdown condition, we have exploited the information shown in Table 1. We compute the average number of interactions allowed for each age range using formula 5, assuming that these interactions are habitual contacts.

$$\frac{Home[age]}{Total[age]} \cdot 100 \quad (5)$$

However, a small percentage of people in working age were still allowed to go to their job place because of the essential nature of their job 5. Those essential workers experienced a different decrease in interactions. Indeed, to model this different condition we compute a different average number of interactions for the essential workers that takes into account the possibility of meeting people at home, at work and on public transports. This hypothesis collapses into formula 6, where the fields always refer to Table 1:

$$\frac{Home[age] + Work[age] + Transports[age]}{Total[age]} \cdot 100 \quad (6)$$



Age	Essential activities
20-29	14.6 %
30-39	25.4 %
40-49	28.7 %
50-59	22.7 %
60+	8.5 %

Table 5: Percentage of the Italian population divided by age, that was allowed to work in presence during the lockdown phase because of the essential nature of their job [27]

### 7.3. Protective devices

With the Lockdown, people started to wear protective devices (e.g., surgical masks). According to the study of [28], about 83.81% of Italian population used a protective mask during the Lockdown period, with a 2.23% approximation error. When the simulation starts, it is defined which people will use a mask during the Lockdown and which ones not. The masks' effectiveness is based on the study performed by [29]. Three different masks types were considered: cloth ones, surgical ones, and N95 ones. A mask's inward efficiency could range from 20 to 80% for a cloth one, 70-90% for a typical surgical mask and >95% for a N95 masks. Moreover, outward efficiency could range from 0 to 80% for cloth mask, while surgical masks and N95 ones are likely 50-90% and 70-100% outwardly protective, respectively. The effectiveness of a generic mask is gotten from the average effectiveness of the three previous types. The effectiveness reduces the virus transmission probability and protects inwardly from 62 to 90% and outwardly from 40 to 90%

### 7.4. Covid-19 transmission probability and contagion modeling

The transmission probability, in this model, shows what is the probability to be in a condition that support the virus spread. In theory what has been asserted is that the spreading is favored when people are in a closed place, slightly ventilated and less than one meter away from an infected person for more than 15 minutes. Modeling this situation is not trivial, thus we decided to model the contagion with a transmission probability that is related and specific for the virus, named *Covid-19 Transmission Probability (CTV)*. This choice is common in several modeling approaches, however currently in literature there are no referral to this parameter. In light of this, we decided to estimate this parameter empirically with a data-driven approach, that will be further described in section 8.2. The contagion mechanism is stochastic and is based on all of the probabilities mentioned in section 7. Each actor is described by a transmission probability ( $TP$ ) that represents the ability of transmit his state to another actor. Practically, the only actors that can transfer their state are the one that are in the infectious and positive stages. This probability is described in Figure 4 where the general transmission probability of Covid-19 and the outward mask protection probability of the actor ( $P_{outward}$ ) is considered. The Contagion probability ( $CP$ ) for a susceptible actor

A who meets an infectious actor B is presented also in Figure 4, where:  $TP_A$  is the transmission probability of actor A and  $P_{inward}$  is the inward mask protection probability of actor A. Finally the contagion happens by randomly sampling from a uniform distribution considering the Contagion probability and the susceptibility of actor A.

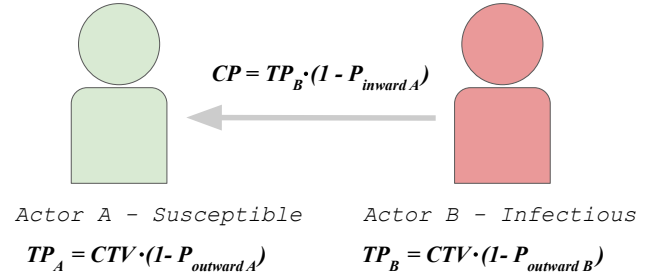


Figure 4: Covid-19 transmission

## 8. Experimentation for the Lombardy region

In order to evaluate our simulator and the social interactions model, we considered as use-case the Covid-19 outbreaks in Lombardy (Italy) during 2020. In particular, we are mainly interested into modeling the first wave from January to April 2020. Secondary, the second wave between August and December has been taken in consideration to validate the model and its parameters.

### 8.1. Initial conditions

When the simulation starts, every agent are born in the susceptibility stage. However, this way no one can start an hypothetical contagion. Thus, at the beginning of the simulation, more specifically when the population is created, some actors are randomly chosen and start directly from the incubation phase. Moreover, these people are selected in such a way to respect the real positives' number between 20 and 29 February in Lombardy on a provincial basis. The data used are shown in Table 6.

Date	Positive people by province											
	BG	BS	CO	CR	LC	LO	MN	MI	MB	SO	PA	VA
2020-02-20	1	0	0	0	0	2	0	0	0	0	0	0
2020-02-21	1	0	0	3	0	24	0	1	1	0	1	0
2020-02-22	3	0	0	9	0	19	0	0	0	0	1	0
2020-02-23	3	1	0	18	0	13	0	1	2	1	8	0
2020-02-24	10	2	0	25	1	10	1	6	1	0	5	0
2020-02-25	23	5	1	21	1	26	0	7	0	2	5	2
2020-02-26	36	11	0	26	0	17	0	7	1	0	16	1
2020-02-27	43	11	1	30	1	32	0	11	1	0	17	1
2020-02-28	76	25	0	36	1	37	3	10	2	0	12	1
2020-02-29	75	32	2	20	1	78	6	14	3	0	13	1

Table 6: Positive people between 20 and 29 February in Lombardy

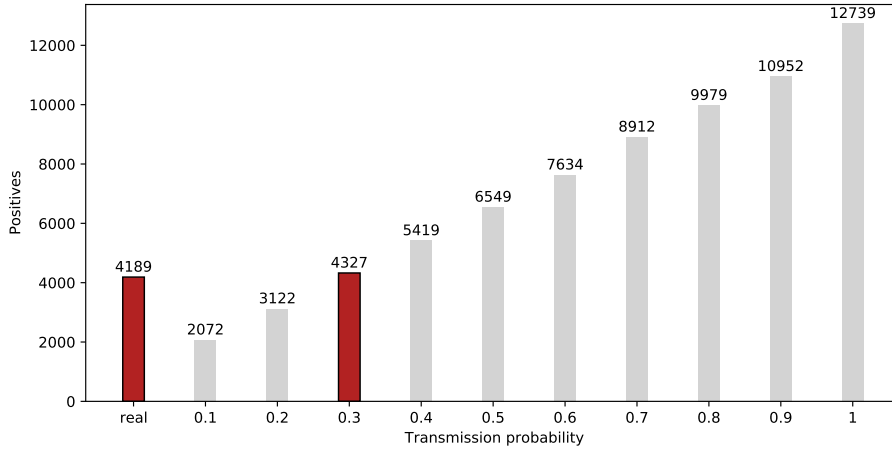


Figure 5: Positive people on 8 March with different transmission rates

## 8.2. Covid-19 Transmission Probability estimation

We estimated this probability with a random search over the probability space with a data-driven procedure. We estimated this value in two different scenarios. The first one, considering the early stage of the pandemic in the period before the national lockdown and when people were not using protective devices. In particular, we considered a key date, the 8th March, that represents the last day before the lockdown. The transmission probability parameter is a positive real number that ranges between 0 and 1. Our goal was to find a value that in the simulation without lockdown and without the usage of protective devices could return a positive people number close to that identified in real data until the 8th March 2020. The value satisfying these hypotheses in the first scenario is: 0.3. The simulation result is a random process, therefore an average of five tests were made for every single tested value. A summary histogram is shown in figure 5.

On the other hand, the second scenario we took into account is the second outbreak occurred after the summer (second wave) that is deeply analyzed in section 9.2 in the results. This second estimation was necessary because real-data in that period are more robust and significant due to the higher number of Covid-19 tests performed each day in Italy among people.

## 9. Results and discussion

In this section we present the results we obtained simulating different scenarios and considering each time an average of 10 different runs since the entire simulation process is stochastic in most of its steps. For each case we measured the simulation goodness using the Pearson correlation and the Root-mean-square error (RMSE) between simulation data and real data from the Italian government [21]. The Pearson correlation expresses any linear relationship between two statistical variables. This value ranges from  $-1$  to  $1$ , where  $1$

corresponds to a strong linear correlation,  $0$  corresponds to the absence of a correlation and  $-1$  corresponds to a negative linear correlation. In our case, this particular correlation will explain how much the simulated contagion curve has a similar trend to the real one, see equation 7.

$$r = \frac{\sum (x - m_x)(y - m_y)}{\sqrt{\sum (x - m_x)^2 \sum (y - m_y)^2}} \quad (7)$$

where  $m_x$  is the vector  $x$  mean value and  $m_y$  is the vector  $y$  mean value. The Root-mean-square error instead, shows the difference between the predicted values and those observed in the real data, Equation 8

$$\sqrt{\frac{1}{n} \sum_{i=1}^n (f_i - o_i)^2} \quad (8)$$

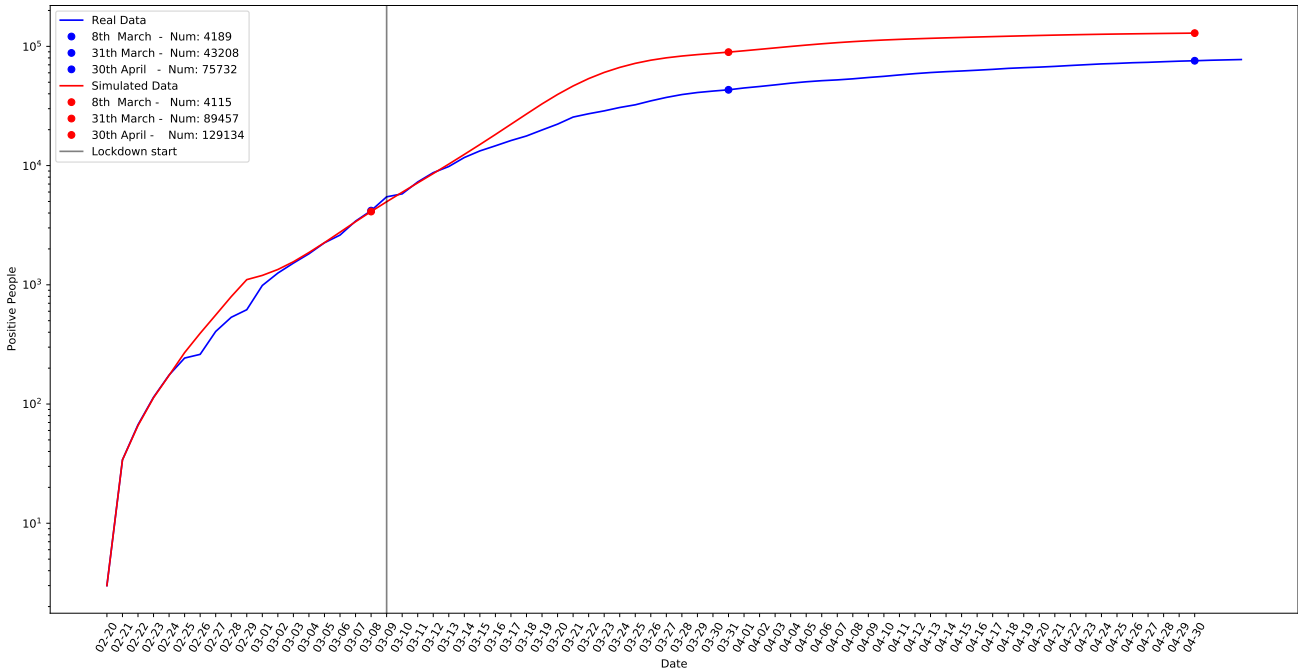
where  $n$  is the samples number,  $f$  are the predictions and  $o$  are the observed real data. Figure 6 shows the results about the simulation of the early-stage of the pandemic between January and April 2020 with the Covid-10 Transmission probability (CTP) equal to 0.3. The blue curve represents the real contagion data, while the red one represents the simulated ones. Table 7 shows the values regarding the Pearson correlation and the RMSE. We analyzed the Pearson correlation and the RMSE until April 30th. The number of positives obtained, in Figure 6, thanks to the use of the simulator, exceeds by about 53,000 units the number of real positives on that date.

Pearson correlation	Root Mean Squared Error
0.992	38818

Table 7: Pearson correlation and RMSE referred to Figure 6

### 9.1. Comparison with national screening activities

The result previously obtained would support the thesis and the hypothesis. However, the comparison with the sero-



**Figure 6:** Simulation results with real data with starting assumptions (CTP=0.3) - Incremental positives representation

logical investigations made by *ISTAT* [32] revealed a different scenario. The study about the seroprevalence showed that, on July 15, the real cases number in Lombardy was about eight time greater than the data reported from Covid-19 tests, more specifically 7.92 times greater. This study, through the national screening campaign, showed that about the 7.5% of the Lombard population had developed antibodies for the novel Corona virus. The Lombardy's population is made up of about 10.060.000 people. The 7.5% of the population is therefore equivalent to about 754.500. On the other hand, the positive people number identified by throat swab on July 15 was only 95.236. Hence, the numerical ratio is obtained by doing:

$$\frac{754.500}{95.236} = 7.92 \quad (9)$$

This means that, most likely, the data gathered in the spring were underestimated. Assuming that this ratio is constant over time, we can retro-project this data and observe how many positive people can be obtained. The resulting value can be used as a comparison for the model.

## 9.2. Second wave: parameters fine tuning

Thanks to the lockdown period, in March and April, the contagion curve has slowly decreased. The curve lowering led to a grip loosening of the restrictive measures on the population. The second wave of contagions produced new epidemiological data that can be analyzed to improve the model proposed in this paper. We have already highlighted that, most likely, the data gathered in the first epidemic wave were inaccurate, because the country found itself in a situation that it was not able to deal with. Moreover, there was

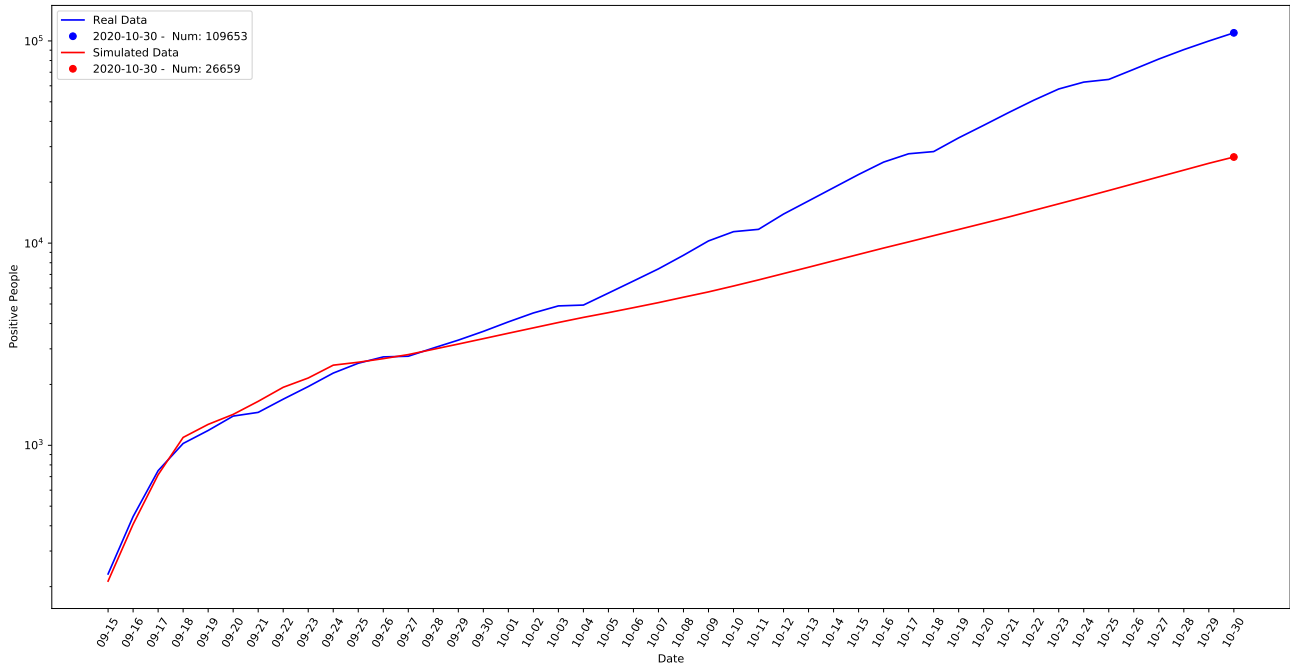
not any working contagion tracking system and not even the number of throat swab to carry out a correct infections tracking. Instead, when the second wave hit Italy, there were the necessary tools to be able to face and analyze the epidemiological data in a more correct way.

A first approach was to use the model previously obtained and validate, to verify if the simulated data followed the real data in a suitable way. The used time window covered a temporal period from 15 September 2020 to 30 October 2020. Obviously, the initial conditions have been changed to make the simulator compatible with the autumn data, table 8.

Date	Positive people by province											
	BG	BS	CO	CR	LC	LO	MN	MI	MB	SO	PA	VA
2020-09-15	25	28	11	9	0	1	5	89	34	0	11	0
2020-09-16	0	37	0	7	0	8	2	80	21	0	12	27
2020-09-17	44	36	0	0	28	0	32	144	0	0	10	14
2020-09-18	28	36	0	18	0	0	29	185	56	0	6	21
2020-09-19	0	28	44	0	0	0	7	86	0	8	0	0
2020-09-20	20	14	0	11	0	0	0	47	40	0	2	12
2020-09-21	0	30	0	4	17	9	8	100	25	0	24	0
2020-09-22	21	24	16	7	3	5	13	88	14	5	16	53
2020-09-23	11	17	7	7	0	4	0	91	0	0	14	26
2020-09-24	15	41	6	4	12	3	28	78	53	0	21	23

Table 8: Positive people between 15 and 24 September in Lombardy

The values obtained greatly underestimated the real data. In Figure 7 is possible to observe a comparison between the two curves: simulated data (red one) and real data (blue one).



**Figure 7:** Simulation results with real data with starting assumptions (CTP=0.3) - Autumn case - Incremental positives representation

Pearson correlation	Root Mean Squared Error
0.988	27145

Table 9: Pearson correlation and RMSE referred to Figure 7

Due to the estimate imprecision, we decided to estimate again the transmission probability parameter, using the autumn data. We searched for a value that correctly followed the contagion curve in the previously specified time period. The best value that satisfied our hypothesis was: 0.53. See Figure 8

Pearson correlation	Root Mean Squared Error
0.996	6405

Table 10: Pearson correlation and RMSE referred to Figure 8

### 9.3. Final Projections

In the light of the considerations and the results in the previous section, we decided to simulate the spring case again with the new transmission probability value. Moreover, we considered also the data obtained from the comparison with the national screening activity. The whole process is represented graphically in figure 9. The blue curve represents the real data, the green one represent the serological data projection on the real data and the red one represents the simulated data with transmission probability equal to 0.53. The second estimation of the Covid-19 Transmission Probability using the autumn data is confirmed as a better choice to validate our model. Indeed, the results we obtained simulating the early-stage are in line with the scientific observations

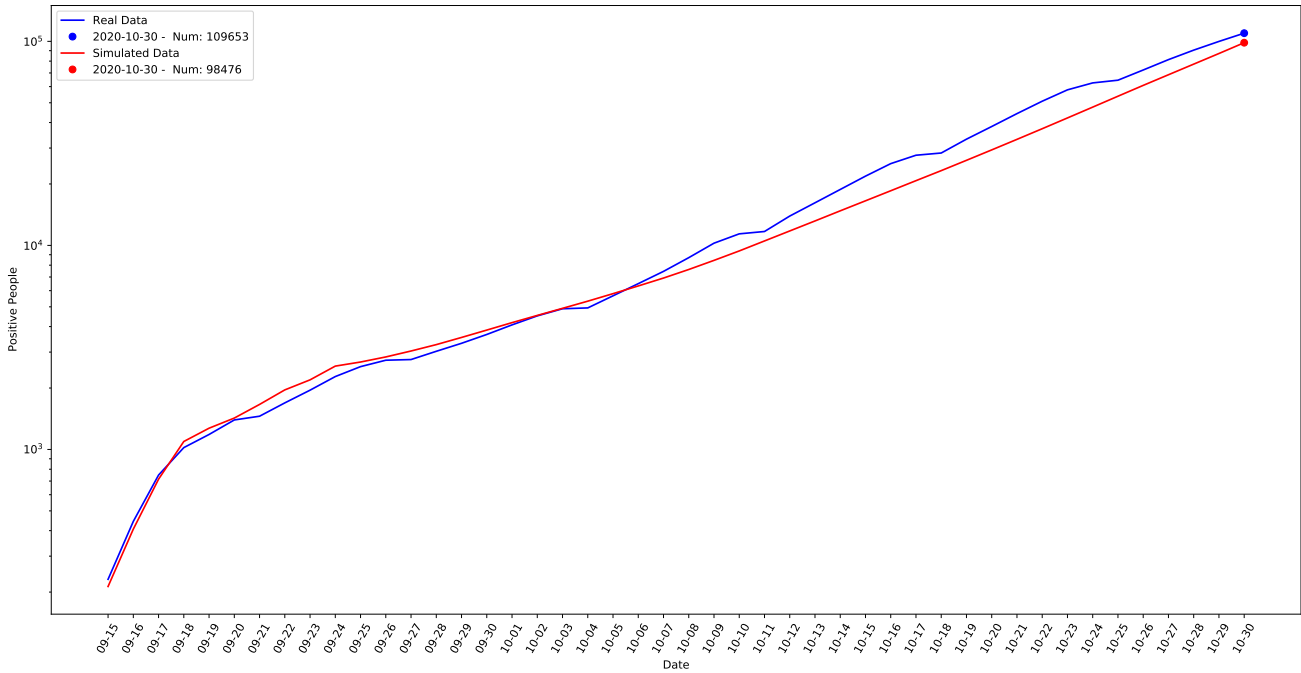
based on the seroprevalence study. The difference with the serological data projection on 30Th April, that corresponds to the last simulation day, is only 83.369 units. Seroprevalence analysis is much more reliable than the data collected during the months of March and April, because it also takes into account the asymptomatic people, a very crucial factor. We believe that the data accordance confirms the validity of our hypothesis and of the simulation model.

### 9.4. Social interactions changes due to the Lockdown

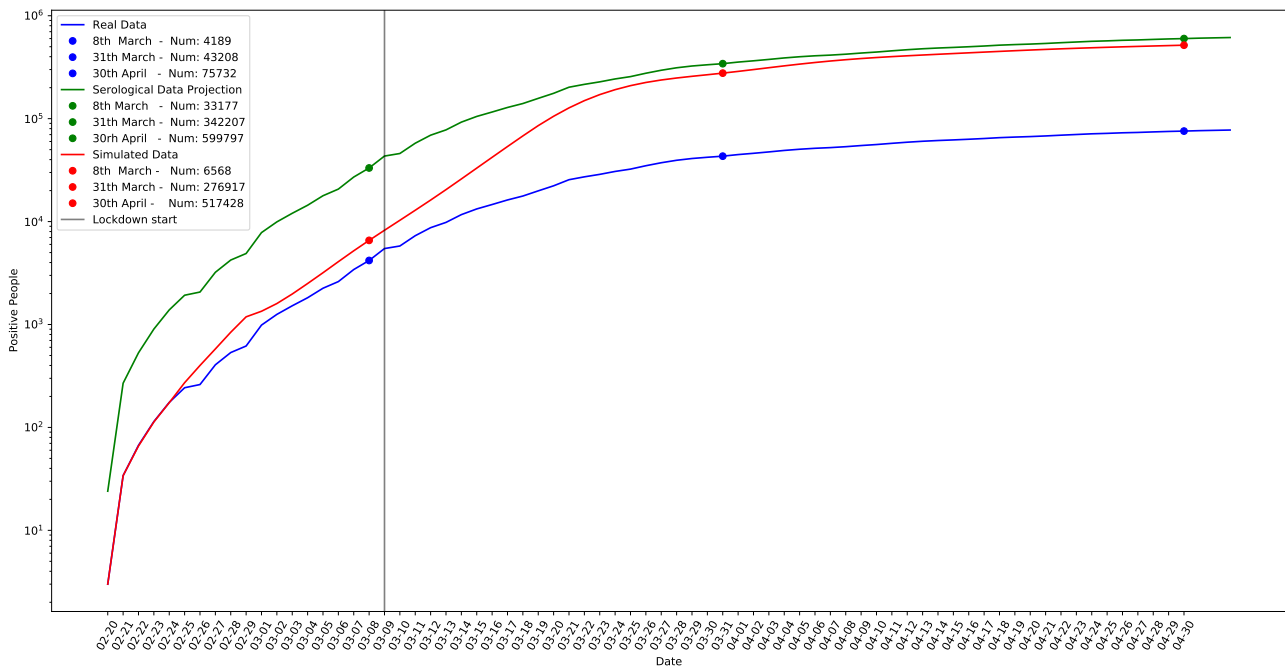
As reported, modeling the lockdown phase involves a general reduction of social interactions although some people were allowed to work because essential workers. We analyzed the dynamic changes of social contacts during the lockdown, in particular we analyzed the likelihood ratio of the degree distribution in different time snapshots to understand what kind of distribution could better describe this condition. In particular, we considered two candidate distributions pairs: power-law and log-normal, log-normal and exponential, table 12. In figure 10 it is possible to observe the CCDF's trends during the lockdown. Observing the results it is possible to infer that, during the lockdown period, the social interactions distribution follows mostly a log-normal trend. (SE MARIO IPOTIZZA ALTRI COMMENTI IN MERITO SONO I BENVENUTI)

## 10. Conclusions

In this paper we presented a framework that aims to combine large-scale epidemic scenario with a fine-grained level of simulation. This result is achieved by exploiting an effi-



**Figure 8:** Simulation results with real data with transmission probability equal to 0.53 - Autumn case - Incremental positives representation



**Figure 9:** Simulation results with real data with transmission probability equal to 0.53 - Spring case - Incremental positives representation

cient multi-agent systems approach combined with the HPC facility. Moreover, we analyzed as use case the outbreaks of Covid-19 in Lombardy (Italy) during 2020. We propose also a modeling for Covid-19 that involves demographic data, lockdown policies, protective devices and a social interactions network modeling. We validated our simulation architecture using real-data and making comparison with the most

recent results of the seroprevalence study in Italy. Finally we contribute to estimate a transmission probability for Covid-19 that can be useful for future works in this field. The results prove the goodness of our model that is able to predict a number of infectious people that is close to the projection of the seroprevalence among the inhabitants of Lombardy. Moreover they confirm that our fine-grained simulation and

Data 1 - Data 2	Pearson correlation	Root Mean Squared Error
Simulated data - Real Data	0.996	249529
Simulated data - Serological data projection	0.996	56009

Table 11: Pearson correlation and RMSE referred to Figure 9

Epoch	Likelihood #1		P-value	Likelihood #2		P-value
	1° Distribution Power-law	2° Distribution Log-normal		1° Distribution Log-normal	2° Distribution Exponential	
40	-340		$\approx 0$	2114		$\approx 0$
61	-267		$\approx 0$	337		$\approx 0$
92	-360		$\approx 0$	349		$\approx 0$

Table 12: Social interactions distribution analysis during the lockdown period

hypothesis could be used in real context to simulate epidemic scenario and to support decision about lockdown policies. Finally, as expected, our results are in line with the diffused hypothesis that during Spring 2020 the number of real cases in North of Italy were probably underestimated due to the difficulties into performing massive tests and the presence of asymptomatic patients. Future works are related to improvements of this model considering different scenario, scaling to larger scenarios and fine-tuning for the social modeling when more results will be available.

### CRedit authorship contribution statement

**Gianfranco Lombardo:** Conceptualization of this study, Methodology, Software, Writing. **Mattia Pellegrino:** Software, Methodology, Writing. **Michele Tomaiuolo:** Methodology, Writing, Supervision. **Stefano Cagnoni:** Methodology, Writing, Supervision. **Monica Mordonini:** Methodology, Writing, Supervision. **Mario Dante Lucio Giacobini:** Methodology, Writing, Supervision. **Agostino Poggi:** Conceptualization of this study, Methodology, Software, Writing.

### References

- [1] Herbert W Hethcote. The mathematics of infectious diseases. *SIAM review*, 42(4):599–653, 2000.
- [2] Daron Acemoglu, Victor Chernozhukov, Iván Werning, and Michael D Whinston. *Optimal targeted lockdowns in a multi-group SIR model*, volume 27102. National Bureau of Economic Research, 2020.
- [3] Jean Roch Donsimoni, René Glawion, Bodo Plachter, and Klaus Wälde. Projecting the spread of covid19 for germany. 2020.
- [4] Veronika Grimm, Friederike Mengel, and Martin Schmidt. Extensions of the seir model for the analysis of tailored social distancing and tracing approaches to cope with covid-19. *medRxiv*, 2020.
- [5] Christopher Avery, William Bossert, Adam Clark, Glenn Ellison, and Sara Fisher Ellison. An economist’s guide to epidemiology models of infectious disease. *Journal of Economic Perspectives*, 34(4):79–104, 2020.
- [6] Graziela P Figueredo, Peer-Olaf Siebers, Markus R Owen, Jenna Reps, and Uwe Aickelin. Comparing stochastic differential equations and agent-based modelling and simulation for early-stage cancer. *PLoS one*, 9(4):e95150, 2014.
- [7] H Van Dyke Parunak, Robert Savit, and Rick L Riolo. Agent-based modeling vs. equation-based modeling: A case study and users’ guide. In *International Workshop on Multi-Agent Systems and Agent-Based Simulation*, pages 10–25. Springer, 1998.
- [8] Petrônio CL Silva, Paulo VC Batista, Hélder S Lima, Marcos A Alves, Frederico G Guimarães, and Rodrigo CP Silva. Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions. *Chaos, Solitons & Fractals*,

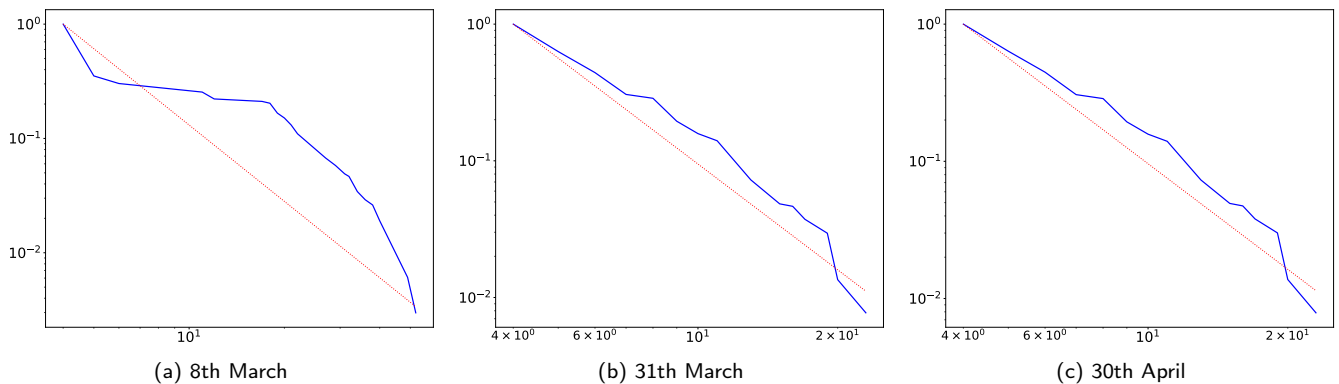


Figure 10: Complementary cumulative density function during the lockdown period - Red Line (PowerLaw Fit) , Blue Line (Empirical Data)

- 139:110088, 2020.
- [9] Erik Cuevas. An agent-based model to evaluate the covid-19 transmission risks in facilities. *Computers in biology and medicine*, 121:103827, 2020.
- [10] Robert Hinch, William JM Probert, Anel Nurtay, Michelle Kendall, Chris Wymatt, Matthew Hall, Katrina Lythgoe, Ana Bulas Cruz, Lele Zhao, Andrea Stewart, et al. Openabm-covid19-an agent-based model for non-pharmaceutical interventions against covid-19 including contact tracing. *medRxiv*, 2020.
- [11] Janos Gabler, Tobias Raabe, and Klara Röhr. People meet people: A microlevel approach to predicting the effect of policies on the spread of covid-19. 2020.
- [12] Christopher Wolfram. An agent-based model of covid-19. *Complex Systems*, 29(1), 2020.
- [13] Stephen Eubank, Hasan Guclu, VS Anil Kumar, Madhav V Marathe, Aravind Srinivasan, Zoltan Toroczkai, and Nan Wang. Modelling disease outbreaks in realistic urban social networks. *Nature*, 429(6988):180–184, 2004.
- [14] Federico Bergenti, Eleonora Iotti, Agostino Poggi, and Michele Tomaiuolo. Concurrent and distributed applications with actodes. In *MATEC Web of Conferences*, volume 76, page 04043. EDP Sciences, 2016.
- [15] Gul A Agha. Actors: A model of concurrent computation in distributed systems. Technical report, Massachusetts Inst of Tech Cambridge Artificial Intelligence Lab, 1985.
- [16] Giulio Angiani, Paolo Fornacciari, Gianfranco Lombardo, Agostino Poggi, and Michele Tomaiuolo. Actors based agent modelling and simulation. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 443–455. Springer, 2018.
- [17] Gianfranco Lombardo, Paolo Fornacciari, Monica Mordonini, Michele Tomaiuolo, and Agostino Poggi. A multi-agent architecture for data analysis. *Future Internet*, 11(2):49, 2019.
- [18] Gianfranco Lombardo and Agostino Poggi. Actornode2vec: An actor-based solution for node embedding over large networks. *Intelligenza Artificiale*, 14(1):77–88, 2020.
- [19] Michele Tomaiuolo, Gianfranco Lombardo, Monica Mordonini, Stefano Cagnoni, and Agostino Poggi. A survey on troll detection. *Future internet*, 12(2):31, 2020.
- [20] Report real-time regione lombardia. <https://experience.arcgis.com/experience/0a5dfcc103d0468bbb6b14e713ec1e30/>, 2020.
- [21] Github protezione civile. <https://github.com/pcm-dpc/COVID-19, Real Time>.
- [22] Popolazione per età, sesso e stato civile 2020. <https://www.tuttitalia.it/lombardia/statistiche/popolazione-eta-sesso-stato-civile-2020/>, 2020.
- [23] Comitato Tecnico Scientifico. Valutazione di politiche di riapertura utilizzando contatti sociali e rischio di esposizione professionale. 2020.
- [24] Joël Mossong, Niel Hens, Mark Jit, Philippe Beutels, Kari Auranen, Rafael Mikolajczyk, Marco Massari, Stefania Salmaso, Gianpaolo Scalia Tomba, Jacco Wallinga, et al. Social contacts and mixing patterns relevant to the spread of infectious diseases. *PLoS Med*, 5(3):e74, 2008.
- [25] Covid-19: nuovo studio sull’incubazione del sars-cov-2. <https://www.osservatoriomalattiere.it/news/ricerca-scientifica/15771-covid-19-nuovo-studio-sull-incubazione-del-sars-cov-2>, 2020.
- [26] Epidemia covid-19. [https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19\\_12-marzo-2020.pdf](https://www.epicentro.iss.it/coronavirus/bollettino/Bollettino-sorveglianza-integrata-COVID-19_12-marzo-2020.pdf), 2020.
- [27] Coronavirus, la demografia del lockdown: ecco l’identikit di chi lavora e di chi sta a casa. <https://www.ilfattoquotidiano.it/2020/04/13/coronavirus-la-demografia-del-lockdown-ecco-lidentikit-di-chi-lavora-e-di-chi-sta-a-casa/5763773/>, 2020.
- [28] Daniela Perrotta, André Grow, Francesco Rampazzo, Jorge Cimentada, Emanuele Del Fava, Sofia Gil-Clavel, and Emilio Zagheni. Behaviours and attitudes in response to the COVID-19 pandemic: Insights from a cross-national facebook survey. May 2020.
- [29] Steffen E. Eikenberry, Marina Mancuso, Enahoro Iboi, Tin Phan, Keenan Eikenberry, Yang Kuang, Eric Kostelich, and Abba B. Gumel. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. *Infectious Disease Modelling*, 5:293–308, 2020.
- [30] Albert-László Barabási. *Network Science*, chapter 4. The scale-free property and 10. Spreading Phenomena. Cambridge University Press, 2016.
- [31] Correction: powerlaw: A python package for analysis of heavy-tailed distributions. *PLoS ONE*, 9(4):e95816, April 2014.
- [32] Studio della sieroprevalenza in italia. [http://www.salute.gov.it/portale/news/p3\\_2\\_1\\_1\\_1.jsp?lingua=italiano&menu=notizie&p=dalministro&id=4998](http://www.salute.gov.it/portale/news/p3_2_1_1_1.jsp?lingua=italiano&menu=notizie&p=dalministro&id=4998), 2020.