UNIVERSITÀ DI PARMA

ARCHIVIO DELLA RICERCA

University of Parma Research Repository

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

This is the peer reviewd version of the followng article:

Original
Fine-Grained Agent-Based Modeling to Predict Covid-19 Spreading and Effect of Policies in Large-Scale Scenarios / Lombardo, G., Pellegrino, M., Tomaiuolo, M., Cagnoni, S., Mordonini, M., Giacobini, M., Poggi, A.. - In: IEEE JOURNAL OF BIOMEDICAL AND HEALTH INFORMATICS. - ISSN 2168-2194. - 26:(2022), pp. 5.2052-5.2062. [10.1109/JBHI.2022.3160243]

Availability:
This version is available at: 11381/2919970 since: 2023-09-04T08:08:14Z

Publisher:
IEEE

Published
DOI:10.1109/JBHI.2022.3160243

Terms of use:

Anyone can freely access the full text of works made available as "Open Access". Works made available

Publisher copyright

note finali coverpage

(Article begins on next page)

17 October 2026

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

Gianfranco Lombardo, Mattia Pellegrino, Michele Tomaiuolo, Stefano Cagnoni , *Senior Member,IEEE*
Monica Mordonini, Mario Giacobini, Agostino Poggi

***Abstract*— Modeling and forecasting the spread of COVID-19 remains an open problem for several reasons. One of these concerns the difficulty to model a complex system at a high resolution (fine-grained) level at which the spread can be simulated by taking into account individual features. Agent-based modeling usually needs to find an optimal trade-off between the resolution of the simulation and the population size. Indeed, modeling single individuals usually leads to simulations of smaller populations or the use of meta-populations. In this article, we propose a solution to efficiently model the Covid-19 spread in Lombardy, the most populated Italian region with about ten million people. In particular, the model described in this paper is, to the best of our knowledge, the first attempt in literature to model a large population at the single-individual level. To achieve this goal, we propose a framework that implements: i. a scale-free model of the social contacts combining a sociability rate, demographic information, and geographical assumptions; ii. a multi-agent system relying on the actor model and the High-Performance Computing technology to efficiently implement ten million concurrent agents. We simulated the epidemic scenario from January to April 2020 and from August to December 2020, modeling the government's lockdown policies and people's mask-wearing habits. The social modeling approach we propose could be rapidly adapted for modeling future epidemics at their early stage in scenarios where little prior knowledge is available.**

***Index Terms*— Covid-19, Epidemic modeling, Large-scale simulation, Multi-agent simulation, Social modeling, Artificial Intelligence, HPC, ABMS.**

I. Introduction

Since the end of 2019, the number of Covid-19 outbreaks has grown rapidly in many countries. Most governments all over the world have enacted severe measures in full emergency to limit the impact of this new virus upon their communities, in particular lockdown rules and social distancing, to reduce potential transmission events. However, we have witnessed several outbreaks that were difficult to manage and to accurately predict. Most 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 details of SARS-Cov-2 transmission and its dynamics.

In light of this context, the possibility of modeling complex and large scenarios to simulate the spread of Covid-19 becomes a key factor for the prevention and the sustainability of public-health services and the governments' policy-making. Different approaches to this task can be considered, depending on the desired granularity and the level of the available a priori knowledge 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: work sites, transport means, schools, conviviality, free-time activities, and family interactions.

In this research work, our goal is to simulate each person's social behaviors (fine-grained simulation) in a large community of about ten million inhabitants (Lombardy's population in 2019), taking also into account their province of residence. In order to model this system, we used Agent-Based Modeling and Simulation techniques (ABMS). Each agent represents an individual having a certain number of social interactions each day depending on its age, work and sociability level, which we introduced as an individual's feature. We aim to model Covid-19 transmission at a fine-grained resolution, to be able to also 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 organize their behaviors. We implemented this software architecture on the High-Performance Computing facility of the University of Parma, which provided the resources required to scale and distribute the computation workload. 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 Covid-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 according to a power-law degree distribu-

Submitted 20th February 2021

Gianfranco Lombardo, Mattia Pellegrino, Michele Tomaiuolo, Stefano Cagnoni, Monica Mordonini and Agostino Poggi are with the Department of Engineering and Architecture - University of Parma, Parma, Italy

Mario Giacobini is with the Data Analysis and Modeling Unit, Department of Veterinary Sciences, University of Torino, Turin, Italy

tion; (*c*) a model of the lockdown policies and mask-wearing effects; and (*d*) an estimation of the actual impact of Covid-19 in Lombardy during those waves, which is universally considered to have been underestimated.

II. Literature review

Different modeling techniques have been proposed to model and solve real and complex epidemic scenarios. Two widely used techniques are system dynamics (SD) and agent-based modeling (ABM).

System dynamics permits to analyze the modeled system at a high abstraction level 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 in 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 to increase heterogeneity and flexibility [2]–[4].

However, the extensions do not address the main key-issue related with the most relevant parameter, the basic reproduction number (R_0), that is not policy-invariant. Indeed, it depends on the number of s each person's contact and its contact's infection probability. Indeed, as reported in [5], it is hard to translate a policy into the R_0 it induces; the most frequent solution is defining an exponentially decreasing R_0 function or a piecewise function that tries to model the adopted policies to reduce the virus spread. In Riccio et al. [6], Covid-19's spread in Lombardy is simulated by introducing three novel parameters in the SEIR model, taking into account the transmission rate of asymptomatic subjects, the number of unascertained cases in the official data, and, finally, a piecewise exponentially-decreasing R_0 that reflects the effect of the policies adopted in Italy. Another interesting approach is described by Peng et al. [7] where the authors propose a variant of SEIR relying on an exponentially-decreasing R_0 function that depends on the adoption rate of mask-wearing and the number of deaths. In this way the R_0 function takes into account more restrictive policies in the model when the outbreaks become severe. Godio et al. [8] propose an extension of the model proposed in [7] applied to Lombardy, optimizing all the parameters using Computational Swarm Intelligence to model the early stage of the pandemic in Italy. Finally, Chen et al. [9] introduce a SEIAR model that takes into account the asymptomatic patients along with a migration factor that accounts for residents' movements outside a region with an outbreak.

A common approach with SEIR models is assuming that the population distribution is uniformly mixed. However, in a real population, the contacts distribution is mixed heterogeneously and shows some complex network characteristics [10], [11], [12]. Li et al [13] propose a SEIAR model on small-world networks generated using the Watts-Strogatz algorithm [14] but it is computationally infeasible in large scale contexts with millions of people while ignoring the different levels of sociability as a function of age and occupation observed in a real population.

On the other hand, agent-based approaches model the behavior of each agent as well as the interaction between agents. ABMs can be used to study the system at different abstractions levels. In [15] the authors demonstrated that stochastic ABMs can show better performance when several parameters are unknown and there is the need for capturing heterogeneity across individuals and their network of reciprocal interactions. In [16] the authors compare ABMs and meta-population strategies for modeling the epidemic in Italy, concluding that a trade-off between the two methods depends on data availability and suggesting the use of hybrid models. For a further discussion about ABM and its advantages over system dynamics models we refer the reader also to [17] and [18]. However, the main limitation of ABM approaches is the intrinsic trade-off between the desired resolution of the simulation and the population size. Indeed, to model single individuals' characteristics and behaviors it is necessary to reduce the number of individuals to avoid computational issues. This is why system dynamics models or, at least, the use of meta-population approaches are often preferred in large-scale scenarios. This limit has had an impact also on Covid-19 modeling.

Covid-19 spread estimation is challenging and needs to consider heterogeneous interaction rates among people. These requirements can be easily modeled using ABMS. In [19] and citecuevas2020agent, 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 introducing a momentum term. However, these parameters are difficult to estimate in large scenarios. In [20], the authors model Covid-19 spread by replacing the moving particles with contact networks for households, work, and random contacts. An extension of this approach is considered also in [21] that simulates the effect of the German lockdown in November 2020. Moreover, the structure of the contact network significantly affects the disease, as demonstrated also for Covid-19 in [22]. Contact networks are usually modeled with a power-law distribution because the scale-free property ensures that only few people interact with many others while most people interact with a smaller consistent set [23]. In [24], the authors modeled Covid-19 spread with single individuals in Piedmont (Italy) using a genetic algorithm to estimate the parameters. Moreover, they modeled the population in groups, considering groups' characteristics such as age, fragility and work conditions to minimize the number of symptomatic people. Another interesting attempt can be found in [25], where the authors propose a high-resolution model using single individuals' features with real data about the Covid-19 outbreak in New Rochelle (NY), where thousands of citizens live. However, to the best of our knowledge, there is a lack of models simulating the outbreaks and the lockdown policies with a fine-grained detail in large real scenarios. Most methods in the literature are suitable for modeling a limited number of individuals and to achieve statistics that are then extended to a larger general case. For this reason, we believe that our

approach can be the first that models and simulates a spreading phenomenon with million of agents at a high resolution for each single individual.

III. Software architecture

About ten million people live in Lombardy, therefore a distributed architecture is necessary to obtain a fine-grained simulation of behaviors and details. To implement our simulator we exploited ActoDemic, a Java framework we developed to model general spreading phenomena [26]. It is based on ActoDes [27], a software framework aimed at both simplifying the development of concurrent and distributed complex systems and ensuring an efficient execution of applications [28], [29]. In particular, an application is based on a set of actors [30] that perform tasks concurrently and interacting with each other by exchanging asynchronous messages. Actors can be passive or active depending on whether they share or not a single thread. An Actor Space (AS) acts as a "container" for a set of actors providing services for their execution by two other special actors: the Executor (EX) and the Service provider (SP). EX manages the concurrent execution of the actors in the actor space; SP enables an application's actors to perform new kinds of action (e.g., to broadcast a message or move from an actor space to another one). Several actor spaces can be defined in a distributed application to exploit the resources of multiple nodes. Each actor has a system-wide unique identifier, called reference, allowing 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). The communication among the actors is buffered and an actor can set a timeout to wait for a new message. After its creation, an actor can change its behavior several times before 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 another behavior is able to process it.

A. Simulator's architecture

In our simulator's architecture, every actor represents a generic subject by with its age, province of residence, and own sociability level. Furthermore, each subject has a stochastic behavior depending on other parameters that will be discussed later in Section VI. The simulation process is divided into several "*epochs*", where every epoch represents a different day. Moreover, people can change their behavior depending on the current epoch (e.g., normal period or lockdown period). Individuals are created and divided according to two criteria:

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

The simulator involves a set of computational nodes whose execution is driven by a set of schedulers and managers. We exploited ActoDeS passive actors to model a large-scale context. In particular, each manager has the duty of creating the subset of actors for its computational node and synchronizing the execution of its node's simulation with the execution of the other computational nodes. Moreover, one such manager assumes the role of "*master*", with the duty of partitioning the agents involved in the simulation; it sends the other managers the information they need for creating the agents under their control.

The simulation process can be described as follows:

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

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

ActoDeS provides the "*CycleScheduler*" that can be used in a wide variety of applications, especially in ABMS applications. Furthermore, this scheduler manages the passive actors within its actor space 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 agents.

Figure 1 shows a diagram that describes the simulator's architecture. In the diagram, the population is split to N actor spaces; the last one acts as the master. However, every actor space coordinates and manages a population subset. 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 $p = \frac{population}{N}$ is a constant.

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 overhead. An important solution that limits the communication cost is reducing the interaction frequency and merging multiple interactions. For this reason, all information is exchanged among the actors at the end of each epoch along with the synchronization messages.

IV. Data collection

We retrieved data about COVID-19 spread in Lombardy from the official institutions [31], [32]. From these sources we collected data about new daily cases and casualties registered in Lombardy over time. We also used demographic information about the population of Lombardy, including the number of inhabitants and their age distribution [33]. As a starting point for modeling social interactions we used data from the Italian National Institute of Health [34] and [35]. These data, reported in Table I, have been collected to better understand how respiratory infections might spread. 7,290 people were

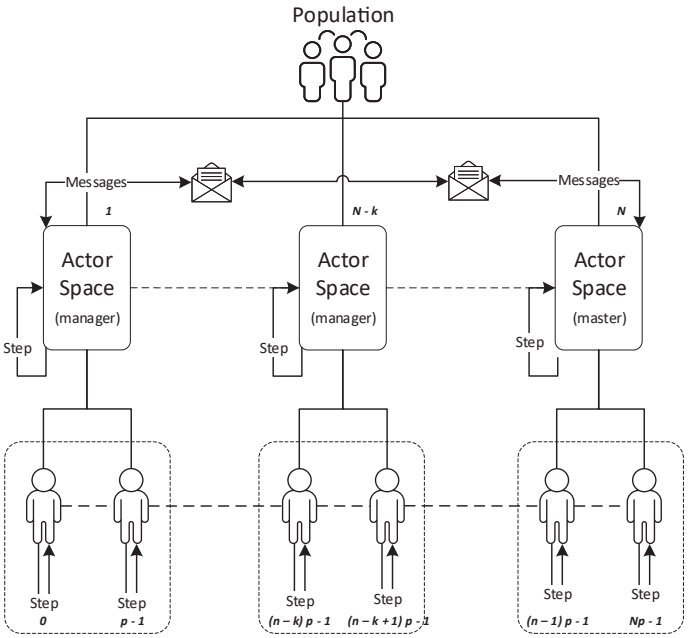


Fig. 1. Simulator architecture

recruited from eight European countries (Belgium, Germany, Finland, Great Britain, Italy, Luxembourg, The Netherlands, and Poland) to participate in the study. Participants were asked to fill out a diary that documented their physical and non-physical contacts in a single day. Physical contacts included interactions such as a kiss or a handshake. Non-physical 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 used preliminary data collected by [36] and [37] and the age susceptibility to COVID-19 virus [34]. Modeling the Italian lockdown required different pieces of information about the set of "*essential workers*" [38]. Moreover, we collected data indicating what percentage of the population used protective devices. [39] and their effectiveness [40]. To evaluate our model, we finally used the results of the seroprevalence survey conducted in that same period in Lombardy.

Age	Total	Home	School	Work	Transport	Free time
0-4	16.54	4.49	5.27	0	0.98	5.81
5-9	20.49	4.61	8.87	0	1.12	5.9
10-14	27.38	4.43	11.98	0.2	1.35	9.42
15-19	29.28	4.59	13.22	0.05	1.74	9.7
20-24	22.15	3.51	1.17	4.49	0.96	12.03
25-29	21	3.47	2.23	5.21	1.13	8.96
30-34	18.03	3.55	0.85	3.92	0.76	8.962
35-39	21.25	4.38	0.68	7.78	1.05	7.37
40-44	22.35	3.88	2.53	7	0.67	8.27
45-49	19.27	2.99	2.61	8.24	0.88	4.57
50-54	22.3	2.75	5.54	8.05	0.52	5.43
55-59	18.27	2.88	1.41	4.6	0.68	8.68
60-64	18.43	3.28	1.07	6.05	0.87	7.16
65-69	12.74	3.1	0.55	0.48	0.95	7.66
70+	10.55	3.24	0.06	1.04	0.22	5.99

TABLE I
AVERAGE NUMBER OF INTERACTIONS BY AGE (FROM [35])

V. SOCIAL INTERACTIONS MODELING

Modeling physical contacts among people is essential to simulate an air-borne disease transmission like Covid-19. However, this is still an open problem when fine-grained details are necessary to the simulation. To tackle this kind of problems, we propose a social interaction model based on the actual daily contacts from [35], enriched with a sociability rate that divides people with a high, medium, and low sociability level. These have been introduced to increase or decrease the average number of the subjects' daily contacts with other people, based on their age. To define these parameters, we hypothesize, as is common in network science, that social networks commonly have a power-law distribution with an exponent between 2 and 3, also known as the scale-free property [12] or power-law graph [41]. Moreover, the entire population is partitioned into 12 groups, accounting for Lombardy's 12 provinces. This subdivision ensures that the interactions are not purely casual. Indeed, a generic subject meets most of his contacts in his own province. Only a small fraction of interactions 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 each subject's average daily number of interactions (See Table I). More specifically, every agent is characterized by the following attributes: ID, province of residence, age group, social interaction ratio, number of contacts, current infection phase. A bit defines if the subject uses a protective device during the lockdown, another bit if the subject is an essential worker during the lockdown. Some of these parameters will be further described in the following sections.

A. Modeling regular contacts

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

We assumed that an agent can meet a certain number of people belonging to this set and a certain number who do not belong to it. Moreover, Table I shows that 65% of a generic agent's daily contacts occur with its regular contacts set, while the remaining 35% is with new contacts. We considered as regular contacts the ones in the Home, School and Work fields of Table I for each age group.

B. Sociability rate estimation

The social network generated by the interactions among people can be theoretically mapped and measured using network science techniques. We assume that a generic agent represents a node in the social graph and its outgoing and incoming edges represent the contacts had by the node and the contacts of other nodes with it, respectively.

The social interactions distribution is a crucial factor for studying the epidemic spread. For this reason, we evaluated

various hypotheses, but, in the end, we decided to focus our studies on a *power-law* distribution. A *power-law* model fits very well many real-world use cases and can be used to model social interactions.

In a typical modern society there are different aggregation sites: offices, schools, parks. These contexts favor large gatherings of people and lead to the creation of particular nodes in a contact network called *hubs* (or *super-spreaders*). A *hub* is a node whose interaction degree is greater than the others'. The degree of a node is the number of edges connected to it. The distribution degree is a function $p_k : \mathbb{N} \rightarrow [0, 1]$, that associates a degree k with the probability for a node to have that degree:

$$p_k = \frac{N_k}{N} \tag{1}$$

where N is the number of total nodes and N_k is the number of nodes having k as their degree.

In a typical *power-law* distribution, nodes with a 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 distribution degree of the graph generated by the social interactions.

We aimed at achieving a *power-law scale-free* distribution [12], because it is characterized by the presence of many *hubs* that create a "*long tail*", in their representation graph. 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 [42], using the Likelihood Ratio test and the Kolmgorov-Smirnov distance to determine which probability distribution is compatible with our data, looking for a power law distribution. We have chosen the *Complementary Cumulative Density Function* (CCDF), because CCDFs are often preferable for visualizing a heavy-tailed distribution. A CCDF measures the probability that a certain node's degree is greater than a reference variable x (see Equation 2).

$$p_k = \sum_{q=k+1}^{\infty} p_q \tag{2}$$

If p_k follows a *power-law* trend, then the cumulative distribution scales according to the law:

$$p_k \sim K^{-\gamma+1} \tag{3}$$

People are modeled according to their social interaction degree, which can be different from the others'. They can be identified by their *Sociability Rate* (SR), that admits four different values: *high, medium, low and quarantine*. Furthermore, according to this ratio, an person can meet more or fewer other people in a single simulation day. The interactions occur randomly, generating an interaction graph. People with a *high* degree meet a number of people above average, those with *medium* degree meet exactly the average number, and people with a *low* degree meet a number of people below average. The population includes 20% of agents having a *high* and a *medium* degree, as opposed to the remaining 80% of agents that have a *low* degree. To achieve this distribution, we have introduced three different social interaction multipliers

Multipliers			LR test	LR p-value	KS test
0.1	1	1.9	1.99	$\simeq 0$	0.11
0.2	1	1.8	364	$\simeq 0$	0.21
0.3	1	1.7	-1.46	0.145	0.116
0.4	1	1.6	-150	$\simeq 0$	0.122
0.5	1	1.5	-403	$\simeq 0$	0.093
0.6	1	1.4	-465	$\simeq 0$	0.113
0.7	1	1.3	-205	$\simeq 0$	0.12
0.8	1	1.2	-301	$\simeq 0$	0.09
0.9	1	1.1	-803	$\simeq 0$	0.127

TABLE II
GRID-SEARCH RESULTS OF THE SOCIABILITY RATES CONSIDERING THE MOST PROMISING DISTRIBUTION CANDIDATES. WHEN COMPUTING THE LIKELIHOOD, THE FIRST CANDIDATE IS ALWAYS THE POWER-LAW AND THE SECOND THE LOG-NORMAL DISTRIBUTION.

that scale the number of people met in a generic simulation day. The multiplier for people having a *medium* degree is equal to 1. To estimate the remaining multipliers we have performed a grid search over the parameters looking for the optimal combination, i.e., one that returns a power-law distribution with a scale-free property II. The multiplier's range for SR=*low* is [0.1-0.9] while, for SR=*high*, it is [1.1-1.9]. The particular case of *quarantine* will be detailed later.

We have analyzed the CCDFs obtained with each configuration considered by the grid search. In Figure 2, we report some of the results along with a comparison with a power-law distribution. In order to choose the optimal configuration, we have analyzed three different parameters:

- *LR (Likelihood-Ratio) test*: This test compares two candidate distributions. 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 significance degree of the result (p-value) refers to the null hypothesis that the two distributions are the same.
- *KS (Kolmogorov-Smirnov) test*. This tool measures the distance between a candidate distribution and the empirical data. In our case, it is used to compare a sample with our reference distribution, that is a power law.

Table II shows the results obtained for the multipliers we selected.

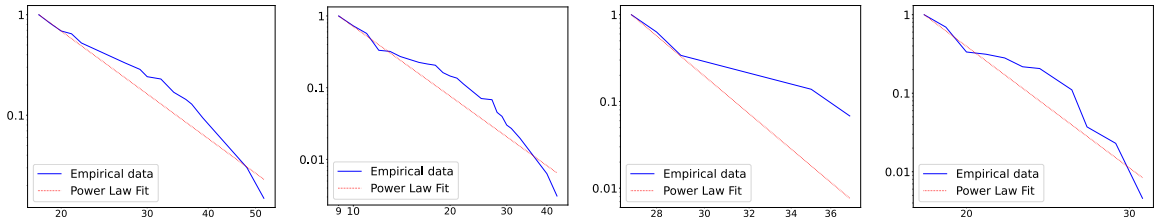
Hence, in light of the results in Table II, we have decided to use the configuration *0.2 - 1 - 1.8*. This set of multipliers is suitable because it returns the highest likelihood ratio and, in particular, a power-law exponent α equal to 2.64, with x_{min} equal to 4. A distribution is called *scale-free*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

VI. MODELING METHODOLOGY

In this section, we present the details of the epidemic diffusion model used in the simulations, the main hypotheses we made to model Covid-19, and the methodology we propose to model the lockdown phase and the mask-wearing protective effects on people.

A. Epidemic diffusion model

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



(a) 0.2 - 1 - 1.8 (b) 0.5 - 1 - 1.5 (c) 0.7 - 1 - 1.3 (d) 0.9 - 1 - 1.1

Fig. 2. Complementary cumulative density function according to interaction multipliers: power-law fit in red, empirical data in blue.

represents one of the most widely adopted mathematical models for characterizing epidemic dynamics and predicting 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 the number of individuals who recover after infection and of the casualties. A limit of the SEIR model is its coarse-grain nature with respect to modeling individual behaviors. In light of this, we have used the same compartments provided by SEIR, extended by the addition of an intermediate one between Infective and Recovered, specific of the Covid-19 context, named Positive. In the real-world data (actual data), this compartment identifies people officially recognized as positive to the virus after undergoing a throat-swab test.

This distinction has an impact on the definition of behaviors: we suppose that generic positive will be quarantined to prevent the infection spread, while an infected patient may be asymptomatic and unaware of her/his condition and lead a normal life with its typical number of social interactions. Finally, the difference between Infected and Positive is fundamental for analyzing the epidemic dynamics, especially during the first wave, when the possibility of testing large amounts of people was limited. This has probably introduced relevant estimation errors in the actual number of infected people into the official data. Since our work is agent-based, we have not used a differential-equation model but we have defined the individual behaviors of the actors. To describe the pathogen spread dynamics, we have assumed that a subject can move through various phases when he/she contracts the disease. Such phases represent the virus' life cycle inside the human body. At the beginning, all subjects are in a susceptibility stage. In this stage, every subject can be infected by another one who is contagious. An individual who is infected moves from a susceptibility phase to an incubation phase and remains in this stage for a certain time period, then moves into an infection stage. A subject in this condition can infect other people. When this phase ends, the person becomes positive. After a certain time, a positive subject will either heal or die. There is no death probability, but deaths follow the real death curve trend in Lombardy. When an individual heals, it cannot be infected any more. In particular, the incubation phase lasts from 7 to 14 days, the infectious phase from 3 to 7 days, and the positive phase from 14 to 30 days [36] [37]. Figure 3 shows a diagram representing a generic infection cycle.

Moreover, people's susceptibility is different according to their age [34]: For the age range (0-14) it is 31%, for age range (15-64) it is 47%, and for age range (65+) it is 100%.

B. 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 "*Lockdown*". Additionally, these measures included:

- closing non-essential activities,
- forbidding crowding,
- forbidding travel, unless 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 needed to buy essential goods, e.g., food. To model the lockdown condition, we have relied upon the information shown in Table I. We computed the average number of interactions allowed for each age range, assuming the Home interactions to be the ones with habitual contacts.

However, a subset of people in working age were still allowed to go to their job site as a requirement of their job (see Table III). Those essential workers experienced a different decrease in interactions. Indeed, to model this different condition, we computed a different average number of interactions for the essential workers that takes into account the probability of meeting people at home, at work, and on public transports, as reported in Table I.

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

TABLE III

PERCENTAGE OF THE ITALIAN POPULATION, BY AGE GROUPS, THAT WAS ALLOWED TO WORK AT THEIR USUAL WORK SITE DURING THE LOCKDOWN PHASE AS A REQUIREMENT OF THEIR JOB [38]

.

C. Protective devices

During Lockdown, people started to wear protective devices (e.g., surgical masks). According to the study of [39], about 83.81% of the Italian population used a protective mask during the Lockdown period, with a 2.23% approximation error.

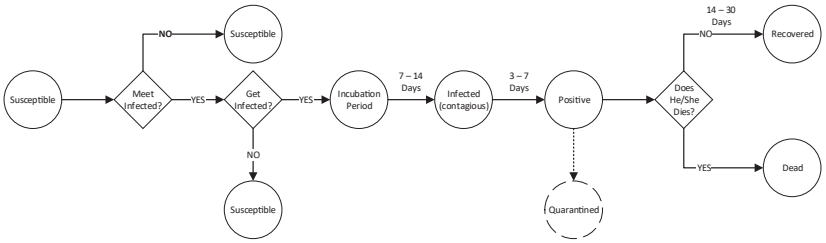


Fig. 3. Infection cycle

In our model, when the simulation starts, we define which subjects will use a mask during the Lockdown and which ones will not. The masks' effectiveness data are based on the study performed by [40]. Three different masks types were considered: cloth, surgical, and N95 masks. A mask's inward efficiency could vary from 20% to 80% for cloth, 70%-90% for surgical and above 95% for N95 masks. Moreover, outward efficiency could range from 0 to 80% for cloth mask, while surgical and N95 masks are 50-90% and 70-100% outwardly protective, respectively. The effectiveness of a generic mask is assumed to be equal to the average effectiveness of the three previous types. The effectiveness reduces the virus transmission probability by 62 to 90% inwardly and by 40% to 90% outwardly.

D. Transmission probability and contagion modeling

The transmission probability, in this model, represents the probability to be in a condition that supports the virus spread. In theory, what has been asserted is that the spread is favored when people are in a poorly ventilated closed place 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 to the virus, named *Covid-19 Transmission Probability (CTV)*. This choice is common in several modeling approaches. However, currently, in the literature there are no references to this parameter. In light of this, we decided to estimate this parameter empirically through a data-driven approach that will be further described in Section VII-B. The simulation of the contagion mechanism is stochastic and based on all the probabilities mentioned in Section V. Each actor is described by a transmission probability *(TP)* that represents its ability of transmitting its state to another actor. Practically, the only actors that can transfer their state are the ones that are in the infectious or positive stages. The computation of this probability is described in Figure 4 where the general transmission probability of Covid-19 and the actor's outward mask protection probability ($P_{outward}$) are considered. The Contagion probability *(CP)* computation for a susceptible actor A who meets an infectious actor B is detailed also in Figure 4, where: TP_A is actor A's transmission probability and P_{inward} is actor A's inward mask protection probability. Finally, the contagion happens by randomly sampling from a uniform distribution considering the Contagion probability and actor A's susceptibility.

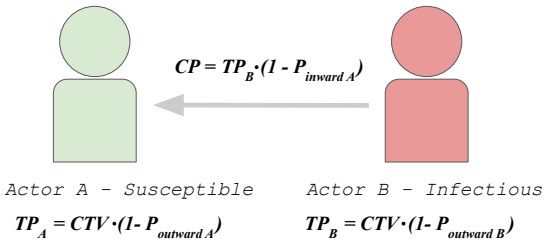


Fig. 4. Covid-19 transmission model among the agents

VII. Experimentation for the Lombardy region

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

A. Initial conditions

When the simulation starts, all agents are in the susceptibility status. However, this way no one can start transmitting the 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 such that they respect the real positives' number between 20 and 29 February in Lombardy on a provincial basis.

B. Covid-19 Transmission Probability estimation

We estimated Covid-19 transmission 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 corresponds to the early stage of the pandemic in the period before the national lockdown, when people were not using protective devices. In particular, we considered a key date, March 8^{th}, 2020, 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 such that, in the simulation without lockdown and without the usage of protective devices, the number of positives would be close to the real data until March 8^{th}, 2020. The value satisfying these hypotheses in the first scenario was

0.3. The simulation result is a random process, therefore an average over five runs was computed for every value we tested. Results are summarized in figure 5.

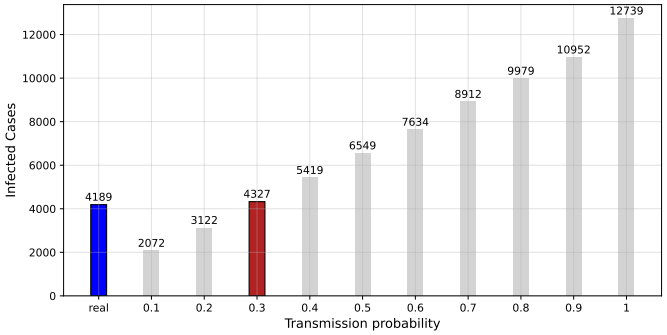


Fig. 5. Positive people on March 8^{th}, **2020** vs. transmission rates

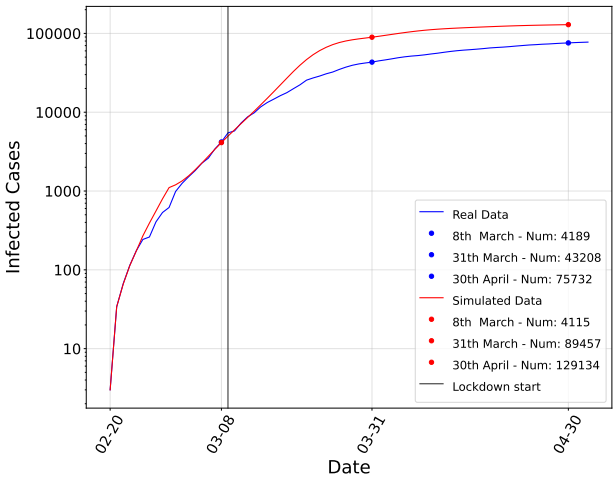


Fig. 6. Simulation results compared to real data with starting assumptions (CTP=0.3) - Positives' incremental representation

The second scenario we took into account is the second outbreak, occurred at the end of summer (second wave), that is deeply analyzed in section VIII-B. This second estimation was necessary because real-data in that period are more robust and significant because of the higher number of Covid-19 tests performed each day in Italy.

VIII. Results and Discussion

In this section, we present the results we obtained simulating different scenarios and considering, for each of them, results averaged over ten different runs since the entire simulation process is stochastic in most of its steps. For each case, we assessed the simulation quality using the Pearson correlation coefficient and the root mean square error (RMSE) between simulated data and real data from the Italian government [32]. The Pearson correlation coefficient expresses a linear relationship between two statistical variables. This value ranges from -1 to 1, where 1 corresponds to a strong positive linear correlation and -1 corresponds to a strong negative linear correlation. In our case, it explains how much the simulated contagion curve trend is similar to the real one. The RMSE, instead, is the average over all samples of the squared error between the predicted values and those observed in the real data. Figure 6 shows the results of the simulation in 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 curve represents the simulated ones. The values of the Pearson correlation coefficient and RMSE referred to Figure 6 until April 30th are 0.992 for Pearson correlation and 38818 for the RMSE. The total number of positives obtained using the simulator on that date, exceeds by about 53,000 units the number of actual positives.

A. Comparison with national screening activities

The results previously obtained seem to support our thesis and hypotheses. However, the comparison with the serological investigations made by *ISTAT* [43] revealed a different scenario. The study about the seroprevalence showed that, on July 15^th, the actual number of cases in Lombardy was about eight time as large as the data from Covid-19 tests, more specifically 7.92 times larger. This study, through the national screening campaign, showed that about 7.5% of the Lombard population had developed antibodies for the novel Corona virus. Lombardy's population amounts to about 10,060,000 people. 7.5% of the population amounts to about 754,500 people. On the other hand, the number of positive cases identified by throat swab on July 15^{th} was only 95,236. Hence, the numerical ratio is 7.92. 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 positives can be obtained. The resulting value can be used as a comparison for the model.

B. Second wave: parameters fine tuning

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 manage. Moreover, there was no contagion tracking system nor was a number of throat swabs suitable for correctly tracking infections available. Instead, when the second wave hit Italy, there was a larger availability of tools needed to face and analyze the epidemiological data more correctly.

In a first analysis, we used the model previously obtained and validated it, to verify if the simulated data properly followed the actual data. The time window used covered a temporal period from September 15^{th}, 2020 to October 30^{th}2020. Obviously, the initial conditions were changed to make the simulator compatible with the autumn data.

The values obtained greatly underestimated the actual data. Figure 7 compares the simulated data (red curve) and the actual data (blue curve). In this case, the Pearson correlation coefficient was 0.988 and the RMSE 27415.

Due to the large estimation error, 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 (Figure 8). In

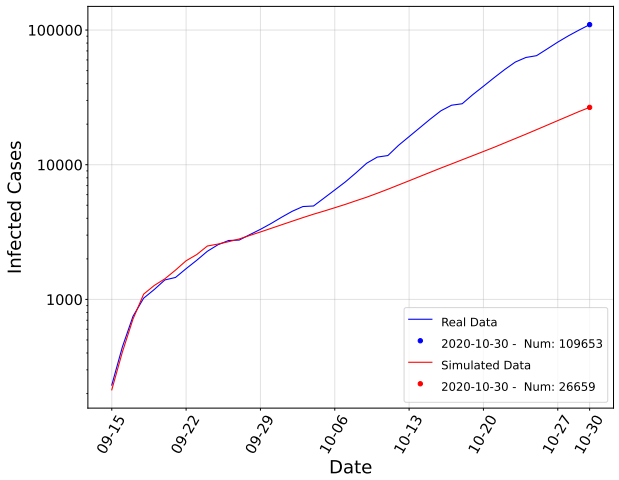


Fig. 7. Simulation results with real data with starting assumptions (CTP=0.3) - Autumn case - Positives' incremental representation

this final case, the Pearson correlation coefficient was 0.996 while the RMSE was 6405.

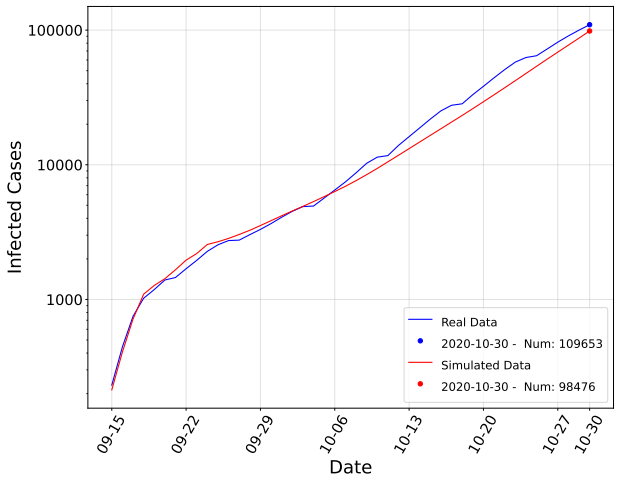


Fig. 8. Simulation results with real data with transmission probability equal to 0.53 - Autumn case - Positives' incremental representation

C. Final Projections

In light of the considerations and the results reported in the previous section, we decided to simulate the spring case again using the new transmission probability value. We also considered 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 actual data, the green curve represents the serological data projection on the real data, and the red curve represents the simulated data with transmission probability equal to 0.53. The second estimate of the Covid-19 Transmission Probability using the autumn data was confirmed to be 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 April 30^{th}, corresponding to the last simulation day, is only 82,369 units. Seroprevalence analysis is much more reliable than the data collected during the months of March and April, because it also takes into account asymptomatic people, which is a very crucial factor. We believe that the matching data confirm the validity of our hypothesis and of our simulation model. Considering the cumulative curves in Figure 9, for the comparison between the simulated and real data, the Pearson correlation coeffcient is 0.996 and the RMSE is 249529. In the comparison between the simulated and serological data the Pearson correlation coefficient is again 0.996, while the RMSE is 56009.

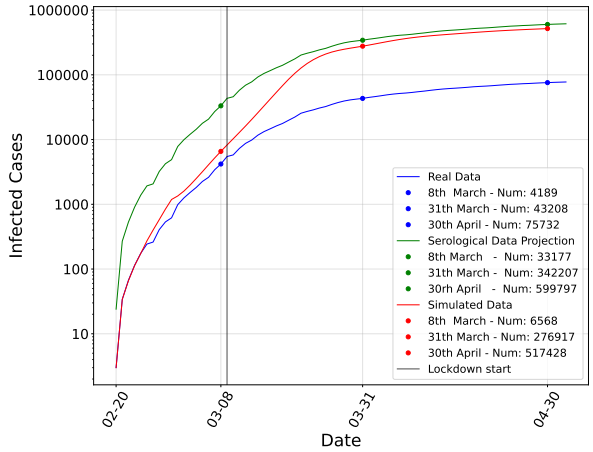


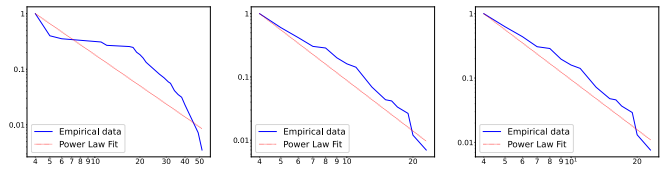
Fig. 9. Simulation results with real data with transmission probability equal to 0.53 - Spring case - Positives' incremental representation

D. 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 as 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 distribution pairs: power-law and log-normal, log-normal and exponential. Figure 10 shows the CCDF's trends during the Lockdown. We measured the likelihood ratio among the empirical data and the candidate distributions. With a p-value always virtually equal to 0, we found that, during the lockdown period, the social interactions distribution follows mostly a log-normal trend. This modification of the degree distribution clearly describes the effects of lockdown measures on the social contact pattern, where the reduction of the number of subjects' interactions and gathering drastically reduces the presence of super-spreader actors in the populations. This modification has a clear impact on the epidemic process.

IX. Evaluation of the Proposed Social Modeling Approach

Modeling social interactions according to a Power law represents one of the key contributions of our approach.



(a) March 8^{th} (b) March 31^{st} (c) April 30^{th}

Fig. 10. Complementary cumulative density function during the lockdown period - Power-law fit in red, empirical data in blue

As previously reported, we modeled the contact network with a power-law distribution because of its property reflected in the resulting heterogeneity of people interactions. In order to assess the validity of our choice, we tested our simulator using a random network instead of the power-law for social modeling. In this way, we built a Null model that matches all other features of our model except for the contact network, replaced by the random one. With such a model, in the simulation process the agents do not interact with a different number of contacts according to their social habits, but all of them always meet the average number of contacts obtained as the weighted average of the data in Table I that is equal to 19. We then simulated the spring case again with different TP values: 0.1, 0.3, 0.53 (equal to our model) to ensure a fair search of the TP parameter. Every outcome was averaged over five tests. Considering the cumulative curves in Figure 11 and the data reported in Table IV, we can assert that our power-law based approach to social modeling gives an important contribution to model the Covid-19 spread since it is more reliable both in terms of RMSE and Pearson correlation coefficient considering the Serological real data.

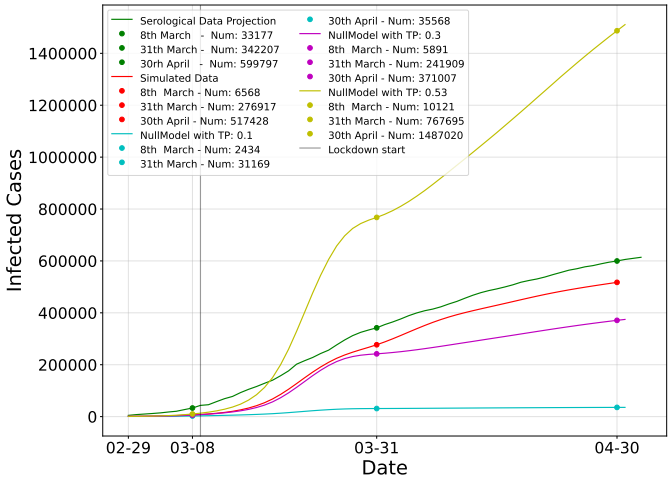


Fig. 11. Simulation results without power-law based social modeling with different transmission probability (0.1-0.3- 0.53). Spring case - Positives' incremental representation

X. Comparison with other state-of-art models

In order to validate our Covid-19 model we compared with different SEIR models we already presented in Section II. The comparison is presented in Figure 12. For each SEIR model we used the same starting days as for our model (February $20^{th} - 29^{th}$, 2020). In particular, we used a traditional SEIR Model which does not model the lockdown, mask-wearing and restrictions measures in general. We also tested two more complex SEIR models: a. the one from Riccio et al. [6] that has been used for the Lombardy context taking into account the transmission rate of asymptomatic subjects, a piece-wise R0 to model the lockdown in Lombardy, and another parameter weighting the effect of the policies adopted in Italy; b. the model from Godio et al. [8] that introduces a SEIR model optimized using a Swarm Intelligence algorithm and based on an exponentially decreasing R0 function that takes into account the adoption-rate of masks, the lockdown, and the number of deaths. The results for March-April 2021 show that all the SEIR models we considered produce less accurate prediction of Covid-19 spread in Lombardy (Serological Data) in terms of RMSE and Pearson Correlation (Table V).

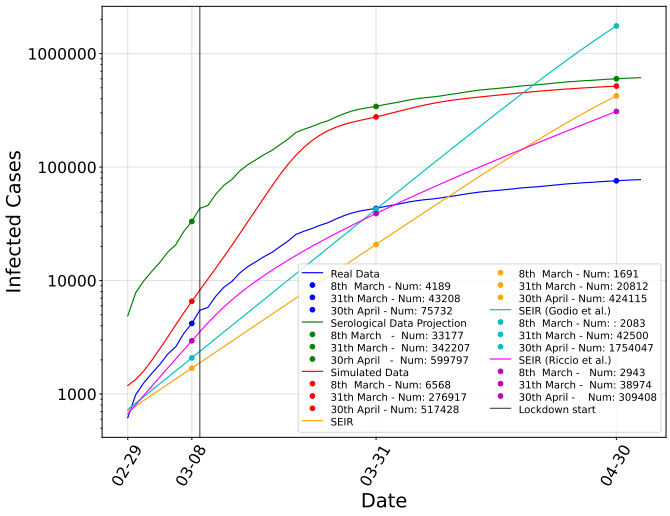


Fig. 12. Comparison among our model, the Real data, the Serological data and the SEIR models. Spring case - Incremental positives representation

	RMSE	Pearson Correlation
Our model	56009	0.996
Null model with TP = 0.1	346040	0.956
Null model with TP = 0.3	139227	0.990
Null model with TP = 0.53	481493	0.994

TABLE IV
COMPARISON OF OUR MODEL TO THE NULL MODELS IN TERMS OF ROOT MEAN SQUARED ERROR (RMSE) AND PEARSON CORRELATION

	RMSE	Pearson
Our model - Serological	56009	0.996
Riccio et al. - Serological	270060	0.859
Godio et al. - Serological	318034	0.713
SEIR - Serological	269197	0.769

TABLE V
COMPARISON AMONG OUR MODEL AND THE SEIR MODELS IN TERMS OF ROOT MEAN SQUARED ERROR (RMSE) AND PEARSON CORRELATION

XI. Conclusion

We presented a framework that aims to combine a large-scale epidemic scenario with a fine-grained simulation level that considers millions of single individuals. This result is achieved by an efficient multi-agent approach implemented on the HPC facility at the University of Parma and a novel method of individuals' social modeling. Moreover, we analyzed the outbreaks of Covid-19 in Lombardy (Italy) during 2020 as a use case.

We also proposed a model of Covid-19 spread that involves demographic data, lockdown policies, protective devices, and a social interactions network model. We validated our simulation architecture using real data, making a comparison with the most recent results of the seroprevalence study in Italy. Finally, we contributed to estimate a transmission probability for Covid-19 that can be useful for future work in this field. The results have highlighted the quality of our model that is able to predict a number of infectious people that is close to the projection of the seroprevalence on the population of Lombardy. Moreover, they confirm that our fine-grained simulation and hypotheses could be used in a real context to simulate the epidemic scenario and to support decisions about lockdown policies. Finally, as expected, our results are in line with the widespread opinion that, during the Spring 2020 outbreak, the number of real cases in Northern Italy were probably underestimated due to the problems with performing massive tests and the presence of asymptomatic patients. Future work will be aimed at improving this model, considering different scenarios, scaling to larger scenarios, and fine-tuning it for social modeling when more results are available.

Acknowledgment

The work of G. Lombardo is partially financed by the Supercomputing Unified Platform Emilia-Romagna - SUPER project POR FSE 2014/2020. The work of M. Pellegrino is partially financed by the Programme "FIL-Quota Incentivante" of University of Parma and co-sponsored by Fondazione Cariparma. Our experiments were run on the High-Performance Computing (HPC) facility of the University of Parma.

References

[1] H. W. Hethcote, "The mathematics of infectious diseases," *SIAM review*, vol. 42, no. 4, pp. 599–653, 2000.

[2] D. Acemoglu, V. Chernozhukov, I. Werning, and M. D. Whinston, *Optimal targeted lockdowns in a multi-group SIR model*. National Bureau of Economic Research, 2020, vol. 27102.

[3] J. R. Donsimoni, R. Glawion, B. Plachter, and K. Wälde, "Projecting the spread of covid19 for germany," 2020.

[4] V. Grimm, F. Mengel, and M. Schmidt, "Extensions of the seir model for the analysis of tailored social distancing and tracing approaches to cope with covid-19," *medRxiv*, 2020.

[5] C. Avery, W. Bossert, A. Clark, G. Ellison, and S. F. Ellison, "An economist's guide to epidemiology models of infectious disease," *Journal of Economic Perspectives*, vol. 34, no. 4, pp. 79–104, 2020.

[6] A. Riccio, "Analysis of the sars-cov-2 epidemic in lombardy (italy) in its early phase. are we going in the right direction?" *medRxiv*, 2020.

[7] L. Peng, W. Yang, D. Zhang, C. Zhuge, and L. Hong, "Epidemic analysis of covid-19 in china by dynamical modeling," *arXiv preprint arXiv:2002.06563*, 2020.

[8] A. Godio, F. Pace, and A. Vergnano, "Seir modeling of the italian epidemic of sars-cov-2 using computational swarm intelligence," *International Journal of Environmental Research and Public Health*, vol. 17, no. 10, p. 3535, 2020.

[9] M. Chen, M. Li, Y. Hao, Z. Liu, L. Hu, and L. Wang, "The introduction of population migration to seiar for covid-19 epidemic modeling with an efficient intervention strategy," *Information Fusion*, vol. 64, pp. 252–258, 2020.

[10] C. Cattuto, W. Van den Broeck, A. Barrat, V. Colizza, J.-F. Pinton, and A. Vespignani, "Dynamics of person-to-person interactions from distributed rfid sensor networks," *PloS one*, vol. 5, no. 7, p. e11596, 2010.

[11] G. Lombardo, A. Ferrari, P. Fornacciari, M. Mordonini, L. Sani, and M. Tomaiuolo, "Dynamics of emotions and relations in a facebook group of patients with hidradenitis suppurativa," in *International Conference on Smart Objects and Technologies for Social Good*. Springer, 2017, pp. 269–278.

[12] A.-L. Barabási, *Network Science*. Cambridge University Press, 2016, ch. 4. The scale-free property and 10. Spreading Phenomena.

[13] J. Li, J. Zhong, Y.-M. Ji, and F. Yang, "A new seiar model on small-world networks to assess the intervention measures in the covid-19 pandemics," *Results in Physics*, vol. 25, p. 104283, 2021.

[14] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[15] H. Rahmandad and J. Sterman, "Heterogeneity and network structure in the dynamics of diffusion: Comparing agent-based and differential equation models," *Management science*, vol. 54, no. 5, pp. 998–1014, 2008.

[16] M. Ajelli, B. Gonçalves, D. Balcan, V. Colizza, H. Hu, J. J. Ramasco, S. Merler, and A. Vespignani, "Comparing large-scale computational approaches to epidemic modeling: agent-based versus structured metapopulation models," *BMC infectious diseases*, vol. 10, no. 1, pp. 1–13, 2010.

[17] G. P. Figueredo, P.-O. Siebers, M. R. Owen, J. Reps, and U. Aickelin, "Comparing stochastic differential equations and agent-based modelling and simulation for early-stage cancer," *PloS one*, vol. 9, no. 4, p. e95150, 2014.

[18] H. V. D. Parunak, R. Savit, and R. 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*. Springer, 1998, pp. 10–25.

[19] P. C. Silva, P. V. Batista, H. S. Lima, M. A. Alves, F. G. Guimarães, and R. C. Silva, "Covid-abs: An agent-based model of covid-19 epidemic to simulate health and economic effects of social distancing interventions," *Chaos, Solitons & Fractals*, vol. 139, p. 110088, 2020.

[20] R. Hinch, W. J. Probert, A. Nurtay, M. Kendall, C. Wymatt, M. Hall, K. Lythgoe, A. B. Cruz, L. Zhao, A. Stewart *et al.*, "Openabm-covid19-an agent-based model for non-pharmaceutical interventions against covid-19 including contact tracing," *medRxiv*, 2020.

[21] J. Gabler, T. Raabe, and K. Röhrl, "People meet people: A microlevel approach to predicting the effect of policies on the spread of covid-19," 2020.

[22] C. Wolfram, "An agent-based model of covid-19," *Complex Systems*, vol. 29, no. 1, 2020.

[23] S. Eubank, H. Guclu, V. A. Kumar, M. V. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," *Nature*, vol. 429, no. 6988, pp. 180–184, 2004.

[24] G. Pescarmona, P. Terna, A. Acquadro, P. Pescarmona, G. Russo, E. Sulis, and S. Terna, "An agent-based model of covid-19 diffusion to plan and evaluate intervention policies," *arXiv preprint arXiv:2108.08885*, 2021.

[25] A. Truszkowska, B. Behring, J. Hasanyan, L. Zino, S. Butail, E. Caroppo, Z.-P. Jiang, A. Rizzo, and M. Porfiri, "High-resolution agent-based modeling of covid-19 spreading in a small town," *Advanced theory and simulations*, vol. 4, no. 3, p. 2000277, 2021.

[26] M. Pellegrino, G. Lombardo, M. Mordonini, M. Tomaiuolo, S. Cagnoni, and A. Poggi, "Actodemic: A distributed framework for fine-grained spreading modeling and simulation in large scale scenarios," 2021.

[27] F. Bergenti, E. Iotti, A. Poggi, and M. Tomaiuolo, "Concurrent and distributed applications with actodes," in *MATEC Web of Conferences*, vol. 76. EDP Sciences, 2016, p. 04043.

[28] G. Angiani, P. Fornacciari, G. Lombardo, A. Poggi, and M. Tomaiuolo, "Actors based agent modelling and simulation," in *International Conference on Practical Applications of Agents and Multi-Agent Systems*. Springer, 2018, pp. 443–455.

[29] G. Lombardo and A. Poggi, "Actornode2vec: An actor-based solution for node embedding over large networks," *Intelligenza Artificiale*, vol. 14, no. 1, pp. 77–88, 2020.

[30] G. A. Agha, "Actors: A model of concurrent computation in distributed systems." Massachusetts Inst of Tech Cambridge Artificial Intelligence Lab, Tech. Rep., 1985.

[31] "Report real-time regione lombardia," 2020.

[32] "Github protezione civile," https://github.com/pcm-dpc/COVID-19, Real Time.
[33] "Popolazione per età, sesso e stato civile 2020," 2020.
[34] C. T. Scientifico, "Valutazione di politiche di riapertura utilizzando contatti sociali e rischio di esposizione professionale," 2020.
[35] J. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. S. Tomba, J. Wallinga *et al.*, "Social contacts and mixing patterns relevant to the spread of infectious diseases," *PLoS Med*, vol. 5, no. 3, p. e74, 2008.
[36] "Covid-19: nuovo studio sull'incubazione del sars-cov-2," 2020.
[37] "Epidemia covid-19-epicentro iss," 2020.
[38] "Transition steps to stop covid-19 without killing the world economy," 2020.
[39] D. Perrotta, A. Grow, F. Rampazzo, J. Cimentada, E. D. Fava, S. Gil-Clavel, and E. Zagheni, "Behaviours and attitudes in response to the COVID-19 pandemic: Insights from a cross-national facebook survey," May 2020. [Online]. Available: https://doi.org/10.1101/2020.05.09.20096388
[40] S. E. Eikenberry, M. Mancuso, E. Iboi, T. Phan, K. Eikenberry, Y. Kuang, E. Kostelich, and A. 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*, vol. 5, pp. 293–308, 2020. [Online]. Available: https://doi.org/10.1016/j.idm.2020.04.001
[41] G. Lombardo, A. Poggi, and M. Tomaiuolo, "Continual representation learning for node classification in power-law graphs," *Future Generation Computer Systems*, vol. 128, pp. 420–428, 2022.
[42] "Correction: powerlaw: A python package for analysis of heavy-tailed distributions," *PLoS ONE*, vol. 9, no. 4, p. e95816, Apr. 2014. [Online]. Available: https://doi.org/10.1371/journal.pone.0095816
[43] "Studio della sieroprevalenza in italia," http://www.salute.gov.it/portale/news/p3$_{2111}$.$jsp?lingua = italianomenu = notiziep = dalministeroid = 4998$, 2020.

Gianfranco Lombardo Ph.D., University of Parma, 2021. He is a Postdoctoral Researcher at the Department of Engineering and Architecture at the University of Parma. His research activity is focused on Machine Learning, Network Science and Distributed systems. During Fall 2019 he has been Visiting Researcher at Center for Applied Optimization (Herbert Wertheim College of Engineering, University of Florida).

Mattia Pellegrino received a M.Eng. in Computer Engineering at the University of Parma in 2020. His thesis concerned the coronavirus spread using an actor model. Currently he is a Ph. D. student at the University of Parma. His branch of interest concerns epidemiological simulation using distributed computing and the development of IoT technologies.

Michele Tomaiuolo received a M.Eng. in Computer Engineering and a PhD in Information Technologies from the University of Parma. Currently he is an assistant professor at the Department of Engineering and Architecture, University of Parma. He participated in various research projects, including the EU funded @lis TechNet, Agentcities, Collaborator, Comma, and the national project Anemone. His current research activity is focused on social media analysis, trust management, multi-agent systems.

Stefano Cagnoni Stefano Cagnoni (M 85, SM 04) graduated in Electronic Engineering at the University of Florence in 1988 where he has been a PhD student and a post-doc until 1997. In 1994 he was a visiting scientist at the Whitaker College Biomedical Imaging and Computation Laboratory at the Massachusetts Institute of Technology. Since 1997 he has been with the University of Parma, where he has been Associate Professor since 2004. His research interests regard Artificial Intelligence, with particular regard to Evolutionary Computation and its applications to image analysis and Pattern Recognition He has published more than 150 papers in international journals and conference proceedings.

Monica Mordonini received the degree in electronics engineering and the Ph.D. degree in information technologies from the University of Parma, Italy. She is an Aggregate Professor at the Department of Engineering and Architecture, University of Parma since 1999. Her current research interests include artificial intelligence. Recently, the research has focused on issues of information and knowledge extraction from complex systems, applications of semantic Web, analysis of social networks and opinion mining.

Mario Giacobini is Associate Professor in Data Science at the Department of Veterinary Sciences - Univerisity of Torino, where he leads the Data Analysis and Modeling Unit. His research interests concentrate on Artificial Life, both as bio-inspired computational techniques - mainly Machine Learning and Evolutionary Algorithms and their application to biodata mining - and as modeling of biological phenomena - ranging from Epidemiology to Molecular Biology - mainly using concepts and instruments of Network Science. Since November 2019 Mario Giacobini is the Rector's Delegate for European Relations.

Agostino Poggi is Full Professor of Computer Engineering at the Department of Engineering and Architecture of the University of Parma. His research focuses on agent, Web and object-oriented technologies and their use to develop distributed and complex systems. In these research areas, he has been and is involved both in national and international projects. Moreover, he was one of the authors of JADE, an "open source" software development framework for realizing multi-agent systems that is become the international reference software to develop multi-agent applications.