A Multi-Agent-Based Simulation Model for the Spreading of Diseases Through Social Interactions During Pandemics

Feres A. Salem · Ubirajara F. Moreno

Received: 14 April 2021 / Revised: 7 March 2022 / Accepted: 19 April 2022 / Published online: 17 May 2022
© Brazilian Society for Automatics–SBA 2022

Abstract
Epidemiological models have a vital and consolidated role in aiding decision-making during crises such as the Coronavirus Disease 2019 (COVID-19) pandemic. However, the influence of social interactions in the spreading of communicable diseases is left aside from the main models in the literature. The main contribution of this work is the introduction of a probabilistic simulation model based on a multi-agent approach that is capable of predicting the spreading of diseases. Our proposal has a simple model for the main source of infections in pandemics of respiratory viruses: social interactions. This simplicity is key for incorporating complex networks topology into the model, which is a more accurate representation for real-world interactions. This flexibility in network structure allows the evaluation of specific phenomena, such as the presence of super-spreaders. We provide the modeling for the dynamical network topology in two different simulation scenarios. Another contribution is the generic microscopic model for infection evolution that enables the evaluation of impact from more specific behaviors and interventions on the overall spreading of the disease. It also enables a more intuitive process for going from data to model parameters. This ease of changing the infection evolution model is key for performing more complete analyses than would be possible in other models from the literature. Further, we give specific parameters for a controlled scenario with quick testing and tracing. We present computational results that illustrate the model utilization for predicting the spreading of COVID-19 in a city. Also, we show the results of applying the model for assessing the risk of resuming on-site activities at a collective use facility.

Keywords Multi-agent-based simulation · Epidemiology · Social interactions · Modeling of complex systems · Agent-based modeling · Complex networks · Pandemics · COVID-19

1 Introduction
The models most widely used to predict epidemics’ evolution date back to the work from Kermack and McKendrick (1927). They belong to the class referred to as compartmental models. In these, the population is divided into compartments, and assumptions are made about the nature and time rate of transfer between compartments (Brauer 2008).

These models are described by differential equations, enabling the ease of utilization for various applications. On the other hand, they are lumped-parameters representations of a complex and distributed system. This means that many simplifications are made to construct such models, which could imply crucial information being left aside and limit the scenarios where the models are accurate. As an example of these limitations, the seminal compartmental models cannot represent a second wave of infections during an epidemic. Roberts et al. (2015) states that many infectious diseases are fundamentally stochastic processes based on the individuals, i.e., they are more naturally represented by stochastic models, and approximating them by deterministic approaches is challenging.

In the ongoing COVID-19 pandemic, compartmental models were broadly used to predict the spreading of the disease (Medrek and Pastuszak 2020), (Mandal et al. 2020), (Scala et al. 2020). Wu et al. (2020) explored metapopulation compartmental models, in which the population is divided into some discrete patches representing spatially dis-
tant populations or classification by age, gender, or any other important factor.

Estrada (2020) list the main approaches to forecast the progression of an epidemic at a global and local scale, being them: deterministic, data-driven (Huang et al. 2020), stochastic (He et al. 2020), agent-based (Cuevas 2020), and their combination. Also, diverse ways of social contact can cause the transmission of a disease, such as social contact networks, transport systems, and metapopulations. Estrada (2020) presents results from combining deterministic compartmental models with contact networks.

Hunter et al. (2020) present a hybrid model, combining agent-based modeling with classical compartmental models to reduce the computational cost of the simulation. Their main assumption is that analyzing at a microscopic level is more relevant when only a few individuals are infected and, after a certain threshold is passed, individual movements should not matter much.

The urgency of the COVID-19 pandemic certainly concerned the worldwide scientific community. A search in some of the most prestigious journals can emphasize these concerns. For instance, searching for papers on epidemiological models for COVID-19 at Nature will lead to a total of 241 works, from which we highlight those by He et al. (2020) and Scala et al. (2020). For more details on the disease, a search in the British Medical Journal will result in more than 3,000 results. However, we emphasize those by Wise (2020) and Mahase (2021).

Recent works present forecasting methods built over the data collected from several countries. Following, we highlight some of these, which use data from the COVID-19 pandemics in Brazil and provide results concerned with predicting the evolution in the country. Medeiros et al. (2021) propose a method based on regression models to forecast the evolution of the pandemic in countries using past data from other countries that have already passed through a similar stage. The work by Ribeiro et al. (2020) brings several approaches from Machine Learning and also presents a comparison between them in terms of accuracy.

Ekinci (2021) models new cases of COVID-19 by conditional variance and asymmetric effects employing variations of Generalized Autoregressive Conditional Heteroscedasticity (GARCH). A comparison between artificial intelligence models is also presented by da Silva et al. (2020), being coupled with pre-processing variational mode decomposition (VMD).

Drews et al. (2022) investigate the difference of forecasts produced by a compartmental model and a statistical time series model. They analyze the sensitivity of the model parameters in data from ten countries. A computational tool is presented by Paiva et al. (2021) to study the pandemic and perform a trend analysis, in which numerical optimization is used to calibrate the model parameters.

The automation and control community in Brazil is also concerned with the pandemic situation, counting with contributions of many researchers in the modeling and control of epidemics. Santos et al. (2022) present an equation-based model within a sliding modes control strategy. The work of Batista et al. (2020) focuses on regression using the available data.

Batistela et al. (2020) also present an equation-based model, adding an immune compartment to enable vaccination effects analysis. The methodology brought by Tosin et al. (2020) is based on global sensitivity analysis, and they use it to compare variations of a six-compartments model. Almeida et al. (2020) apply a Markov chain-based stochastic compartmental model to regionally predict the pandemic’s evolution.

The computational model proposed by Gomes and Serra (2021) aims in analyzing the spread of COVID-19 by using type-2 fuzzy systems and Kalman filtering over the experimental data. Oro et al. (2020) use autoregressive distributed lag model to forecast the number of new daily cases of the disease in a specific population.

Vasconcelos et al. (2022) analyze the epidemic situation through a computational tool that fits logistic curves of growth to the observed data. The paper by Costa et al. (2022) presents a probabilistic dynamic model of the SIR class that describes the temporal behavior of epidemics in discrete time.

The model presented by Dias et al. (2022b) combines deterministic and stochastic features and allows the population stratification into subgroups. Pazos et al. (2020) apply proportional control to non-pharmaceutical interventions (NPIs) aiming at the prevention of COVID-19 spreading over a six-compartments model.

Purposing to avoid the health system overload, Dias et al. (2022a) implements a proportional-integral controller on a compartmental model. Pataro et al. (2022) suggest using model predictive control to plan social distancing policies to mitigate the epidemic’s spread.

Despite the variety of modeling views presented by the community, there is still a gap concerning microscopic approaches. According to Pastor-Satorras et al. (2015), we are currently in a second golden age in epidemic modeling, led by the improvement in the real-world accuracy of the models. This is being achieved by the combination of available data in large amounts and the explicit simulation of entire populations down to a microscopic scale.

A microscopic model, which considers the network’s structure for social interactions, could be useful for generating better predictions related to the spreading of diseases. This is due to the possibility of adding crucial factors in the analysis, such as isolation level, hygiene habits, number of hospital beds needed, among others (Manzo 2020).

A more recent search on the literature led us to papers that reinforce the significance and contributions of our work.
because they are concerned with nuances of the problem that are addressed in our proposal. Parino et al. (2021) introduce a meta-population model to evaluate more sophisticated NPIs. Ramos et al. (2021) modify a compartmental model for including control measures, such as social distancing and contact tracing. A partial differential equations model is presented by Viguerie et al. (2021) for capturing dynamics based on human habits.

Gumel et al. (2021) present a brief overview of models used to study COVID-19, which includes agent-based, network, and statistical models. They also illustrate how to incorporate new features to seminal epidemic models, such as vaccination and NPIs. In their work, Sturniolo et al. (2021) state that effects of testing, contact tracing, and isolation at individual levels can be approximated to aggregate measurements on the population level.

The work from Nielsen et al. (2021) warns about the evidence mounting that the COVID-19 pandemic is characterized by “superspreading.” They also state that to capture this phenomenon an individual-scale model is needed. Mo et al. (2021) consider passenger contact in public transit to be a crucial factor in the spreading of infectious diseases. Hence, they introduce a model for this network of contacts that is time-varying. They also refer to the need of identifying “influential passengers,” which is the same as the concern with super-spreaders raised by Nielsen et al. (2021).

After an extensive literature review, we could identify many different approaches to the forecasting of epidemic spreading. However, the complex networks governing social interactions are yet to be investigated in this context. Hence, the main contribution of this paper is the proposal of a simulation model based on a multi-agent network to assess the spread of infectious diseases in the agents’ population that is flexible enough to be fitted for different diseases and network structures.

Social interaction is the main source of infections in pandemics such as the one caused by the Severe Acute Respiratory Syndrome CoronaVirus-2 (SARS-CoV-2). So, in our work, we propose a simple model for the social interactions, giving space to use complex networks for connecting the individuals, since they are a more accurate representation of real-world interactions. Flexibility in the network structure of social interactions allows the analysis of more refined non-pharmaceutical interventions (NPIs), such as restricting the social interactions only over long-distance contacts. It also enables the assessment of scenarios with super-spreaders (that would be represented as highly connected nodes in the network) and how much they impact the spreading. As minor contributions, we can highlight the modeling of the dynamical network topology for two different scenarios: the social interactions considering an entire city, and considering a single building from an educational institution.

Another contribution is the generic microscopic (i.e., at individual level) infection evolution model that we propose. We believe that this is the first time that a model considers non-deterministic transitions between stages of a disease, in such a way that the probability of occurring a transition is time-dependent on the last transition. Also, as another minor contribution, we provide the definition of the transition probabilities for the scenario addressing the resumption of on-site teaching activities. These probabilities are established considering a disease modeled with 4 stages (Susceptible–Exposed–Infected–Removed), and a fast screening and isolation of infected individuals.

The individuals have their behavior defined according to their current state and they transition between states by two distinct processes: social interaction (for transitioning from a Susceptible state) and infection evolution (for transitioning from any state after the infection). The model allows to define individual and time-varying behaviors for the agents and to describe the pattern of social interactions between them. These concerns to both the influence of social interactions and the individual behaviors, in a microscopic point of view, are distinct features from the proposed model.

The way we have modeled the infection evolution eases the customization for other diseases based on available data about them. This process will be more intuitive than in other models since it is a matter of defining the relevant states of the disease and building a histogram of people transitioning between the states on each given day. This ease of adding and removing states for the infection evolution also makes more complete analyses possible. For instance, a state representing people in need of a scarce resource, such as mechanical ventilators, could enable adjusting to the forecasted demand. Other examples would be the possibility to set different behaviors for people in isolation or after getting vaccinated to see how it affects the overall spreading.

Our proposal takes into account the complexity of the social network’s structure connecting the agents, as well as their particular behaviors. By these means, the model can be easily adapted to depict situations in which social interaction occurs in a particular way. For instance, it can be used to assess the spread of diseases in spaces of public social interaction, such as schools, industries, and shopping centers. Considering the schools’ scenario, for example, social interaction occurs more often between classmates than otherwise. For this case, the network could be defined by clusters representing classrooms and some random links between them.

Besides that, we can set distinctive behaviors to only a fraction of the population and analyze their impact on the population as a whole. To list a few examples of distinctive behaviors, we can mention: considering part of the population as “careless” about social distancing and hygiene measures; evaluating different levels of isolation in the population; analyzing the effect of a partial and progressive
immunization through vaccination plans. Further, since the model allows us to create additional states for the agents, we can gather more detailed information about the epidemic when necessary.

The results presented in this article comprise the impact evaluation of resuming on-site teaching activities at the campus of our University. We analyzed different scenarios, varying the number of students per classroom and the number of interactions between them per day. Our findings were useful for the University’s technical committee to decide whether it was safe or not to resume the activities (Salem et al. 2020).

Additionally, the proposed simulation model accepts time-dependent parameters. This allows representing situations in which social distancing is relaxed at a given time, for instance. This also enables to model changes in the etiologic agent, like an increase in the probability of infection as seen in new strains of the SARS-CoV-2 (Wise 2020). Also, these variants are thought to allow the virus to evade the immune response and some vaccines were found to be less effective (Mahase 2021).

This article is organized as follows: In Sect. 2, we introduce the simulation model, giving details on how the social interactions and the infection evolution occur. In Sect. 3, we present an illustrative example of usage for the model, applying it to a COVID-19 epidemic in a hypothetical city; In Sect. 4, we show a specific analysis scenario, in which the model can be applied. Finally, in Sect. 5, we summarize the main conclusions drawn from this work and outline some research directions to follow in the future.

2 Presentation of the Model

The analysis of interconnected complex systems usually requires some simplifications to be feasible (Strogatz 2001). To evaluate the effect of impacts related to the network topology, the agents’ dynamics are relaxed (e.g., nodes with single integrator dynamics). On the other hand, when analyzing the effect of each node’s behavior, the network is simplified (e.g., considering a full-connected network). This work suggests a model that combines both a complex structure for the network and agents that evolve their behavior over time.

Of course, many aspects will still be neglected to permit the carrying of simulations. However, we believe that combining the complexity of these two points in a model can bring benefits to analyzing such systems. Also, its complexity is parsimonious, considering the current computational power available.

We built on the idea of random pairwise encounters constrained by the network structure. However, we gave more complexity to the agents’ behavior, since it could change according to the disease stage, policies of social distancing, etc. We propose a model in which the agents are defined as a finite set of states and the transitions between them can be defined by a time-dependent probability function.

2.1 Background

Most epidemic models divide the population into compartments, aggregating individuals that share the same status with respect to the disease of interest (Earn 2008). The number of compartments and what they represent changes according to the disease being modeled. A common choice of compartments for modeling the COVID-19 is called SEIR (Susceptible–Exposed–Infected–Removed).

This abbreviation stands for Susceptible, Exposed, Infected, and Removed. Its difference for the seminal SIR (Susceptible–Infected–Removed) model is the insertion of a latent stage (Exposed) to represent diseases that do not make persons infectious right after they get in touch with the etiologic agent. This happens to be the case for a SARS-CoV-2 infection.

As previously mentioned, compartmental models are governed by differential equations. The number of persons in each compartment should be an integer. However, for a large enough population, we can assume these quantities as real numbers representing the fraction of the population in each compartment. Therefore, the SEIR model is given by the following set of equations

\[
\begin{align*}
\frac{dS(t)}{dt} &= -\beta S(t)I(t) \quad (1a) \\
\frac{dE(t)}{dt} &= \beta S(t)I(t) - \gamma E(t) \quad (1b) \\
\frac{dI(t)}{dt} &= \gamma E(t) - (\lambda + \mu)I(t) \quad (1c) \\
\frac{dR(t)}{dt} &= (\lambda + \mu)I(t) \quad (1d)
\end{align*}
\]

where \(S(t), E(t), I(t), \) and \(R(t)\) are the population’s fractions in each compartment in time; \(\beta, \gamma, \lambda, \) and \(\mu\) are rates defining the flow of persons from one compartment to another; \(\lambda\) and \(\mu\) are, respectively, the recovery and fatality rates, and they could be simplified to a single parameter.

The meaning of each compartment will be further explained when we present the illustrative example in Sect. 3.

2.2 Overview

Each agent in the simulation is modeled as a finite set of states and the transitions between them can occur in two main ways. The first one is given by a social interaction process that can cause a susceptible agent to become infected. Once infected, a second process starts, which we called infection evolution. In this second stage, the states and transitions are defined
Social Interactions Infection Evolution

Fig. 1 Overview of the agents’ evolution, where SA means “Susceptible Agent” and IA means “Infected Agent”

Regular Small-world Random

ρ = 0 ρ = 1 Increasing randomness

Fig. 2 Random rewiring procedure, from a regular ring lattice to a random network. Source: Watts and Strogatz (1998)

The social interactions cause a susceptible agent to be infected by the disease. We based the social interactions on the asynchronous gossip algorithm (Boyd et al. 2006), such that they occur pairwise and at random. The probability of two agents interacting is given by the weight attributed to the link between them in the adjacency matrix that describes the network. Hence, the configuration of the network plays a fundamental role in this model.

Watts and Strogatz (1998) state that the topology from networks of coupled dynamical systems is usually assumed to be completely regular or completely random. Hence, they introduce a model of networks that can be tuned in the middle ground between these two extremes. They name these networks as “small-world” due to the small-world phenomenon, also known as six degrees of separation.

The procedure for creating a small-world network is dependent on a single parameter. It starts from a regular ring lattice, where every node is connected to its closest neighbors only. Then, every edge can be randomly “rewired” with probability ρ, avoiding duplicates. Figure 2 depicts the outputs of this process as the value of ρ changes.

According to Strogatz (2001), the slightest bit of rewiring transforms the network into a “small-world,” with short paths between any two nodes (like in a completely random graph) and maintaining the network highly clustered (as in a regular lattice). These few “long-distance” connections are responsible to increase the speed of spreading (of diseases, information, etc.) in this type of network (Watts 2004). These links are the “weak ties” that Granovetter (1973) addressed in his work.

Also, the social interaction pattern can be time-dependent to represent periods of quarantine and relaxation in the same simulation. There is room also to define diverse classes of agents, representing people that are more careful or others that have more links than average, for instance.

The infection evolution model describes the states and transitions one agent can experience after being infected. Persons react in different ways to the same diseases. Hence, we defined probabilistic transitions between states. The probability of transitioning is time-varying and depends on how many days have passed since the last transition. We illustrate that concept in Fig. 3.

These probabilities can be inferred from real-world data, e.g., accounting for how many people display symptoms each day after being infected by an etiologic agent. Also, the probabilities are cumulative, in the sense that even if few people transition on the 8th day, the probability of transitioning on this day is higher, once given that the transition did not occur before. This guarantees that the agents do not get caught in a deadlock.

Further, if a state can lead to multiple other states, the probability functions should be multiplied by factors that add up to one. These factors represent the percentage of people that transition to each state and can also be inferred from data. For instance, in the COVID-19 pandemic, the infected agents can either present Mild Symptoms, Severe Symptoms, or Recover without showing any symptoms. These concepts will become clearer in the results presented in Sect. 4.
3 Illustrative Example

In this section, we present simulations set up to depict the disease spreading in a city. We link the agents’ possible states to the compartments considered at the SEIR model presented in Eq. 1. The network is defined by a set of clusters, representing the neighborhoods, in which most of the interactions occur. However, we consider that some random inter-clusters contacts also occur on any given day. In Fig. 4, we present a flowchart summarizing every step of our simulation model.

The initial step is the scenario setup, consisting of configuring:

- The Infection Evolution Model;
- The Social Interaction Model.

To define the infection evolution model, we need to describe which states an agent can go through and how (or if) one state can lead to another. This definition should be reasonable following the disease of interest. Also, in this step, the probabilities distributions of state transitioning are set up.

In the social interaction model definition, we should set parameters related to the population and the network connecting its individuals. We must define the number of agents and their initial state, which can be drawn by randomly selecting agents based on given percentages of the population that start at each state. We can also, for instance, select part of the individuals to represent careless people, i.e., people with a higher probability of infecting/getting infected during social interaction.

Regarding the network governing social interactions, we should define: if it is static or dynamic; the number of clusters; the edges, representing possible social interactions, both inside and outside the clusters.

Following, we give more detailed information on the scenario of this illustrative example, along with the simulations’ results.

3.1 Infection Evolution Model

In essence, each agent is defined as a finite state machine, i.e., they can be at only one state per time, and they can perform transitions from one state to another if a given pre-established condition is fulfilled. Figure 5 depicts the state machine chosen to represent the agents in the present analysis.

These states are equivalents to the compartments considered in the SEIR epidemiological model, which is widely used in the literature. However, it is valid to describe their meaning to clear any question related to the simulations shown here.

Fig. 4 Step-by-step of the simulation process applied to the illustrative example

Fig. 5 The state machine for the individuals, where: $S \rightarrow$ Susceptible; $E \rightarrow$ Exposed; $I \rightarrow$ Infected; $R \rightarrow$ Removed
• [S] Susceptible: A person that was never infected by the virus, therefore, he/she can pass to the Exposed state when in contact with an Infected person;
• [E] Exposed: A person that has been in contact with an Infected and caught the virus, but his/hers viral load is insufficient to transmit it for others;
• [I] Infected: A person that caught the virus some time ago and can infect Susceptible persons, if social contact occurs;
• [R] Removed: A person that caught the virus and already passed the period in which he/she could infect other persons.

The state [R] may require further explanation. The term Removed was chosen because this state could represent many things, such as a person that kept in quarantine after symptoms onset; a person that was hospitalized due to more severe symptoms; a person that recovered fully from the infection; or a person that died from disease complications. In short, it represents any causes that lead to the person not being able to infect others anymore.

It is also valid to emphasize that in the present simulations, people in [R] state are not factually removed from the system. This means that as the number of agents in [R] state raises, the probability of an agent in [S] state getting infected drops. This is due to the fact that “removed” nodes keep on being randomly picked for interacting. However, when this occurs, they do not interfere in the state of others. If the simulation was not carried out in this way, it would be more reasonable if the number of interactions per day dropped as the number of people to interact with also shrinks.

The transition triggers defined for the state machine in Fig. 5 are non-deterministic events. The first transition, from [S] to [E], occurs when the agent in [S] randomly chooses to interact with another that is in [I]. And, given that this contact occurred, there is still a probability of the agent in [S] not being infected.

The other two transitions, from [E] to [I] and from [I] to [R], happen after a certain number of days have passed since the last transition. This means that there is a transition probability for each given day, counting from the day the agent entered the current state. Figures 6 and 7 depict that concept. It is worth emphasizing that these probabilities accumulate. For instance, given that, on the fourth day since the last transition, the agent still not passed from [E] to [I] the probability of that occurring on that day is 1 (that is, it will always occur in these circumstances).

The probability distributions presented in Figs. 6 and 7 are based on the incubation time (Guan et al. 2020) and serial interval (He et al. 2020; Du et al. 2020) estimates found in the specialized literature. Although infected persons shed the virus for more than three days, the chosen values agree with the assumption that most people will self-isolate shortly after acknowledging their symptoms. Also, these distributions can be fitted to data of occurrence for each day if available.

3.2 Social Interaction Model

In the simulations presented here, the network topology was defined as clusters representing the neighborhoods, in which a Watts–Strogatz network defines how people are related. Besides that, a certain number of random temporary connections between clusters are created each day.

Figure 8 depicts the initial graph for the network in the simulation. A total population of 5000 agents is considered, equally divided into 50 clusters (neighborhoods). At the beginning of each day, 50 random links (1% of the population) are created between different clusters. At the end of each day, these links are removed.

To initialize the simulation, 74 agents are randomly drawn from the entire population to start in [I] state (represented by red nodes in the graph from Fig. 8)). We considered that no agent starts in [E] or [R] states, i.e., the remaining 4926 agents will begin as susceptible ([S] state, represented by blue nodes in the graph).
3.3 Results

We run 100 simulations to reduce the effects of randomness on the results. Each simulation is carried out until the epidemic can be considered as ended. The end of disease spreading is recognized when no person is in [E] or [I] states. The spreading evolution averaged over these simulations is depicted in Figs. 9 and 10. In Fig. 9, we isolate the data concerning only the [E] and [I] states to give a better visualization of how the infections evolved in time. Also, from this plot, we can extract some insights regarding the epidemic’s extent.

No simulation lasted longer than 210 days. On average, we can state that the epidemic would span over approximately 160 days for the scenario specified here. It is worth mentioning that we consider the city as a closed world, not taking into account infected persons coming from other cities.

We present a stacked bars plot in Fig. 10. The height of each bar represents the average population’s fraction in each state on a given day. This representation is useful for analyzing the entire spreading evolution.

Even though the peak of simultaneous infections stayed close to 10% of the population, approximately 70% of the population have been infected at the end of the epidemic. This overall infection ratio is directly proportional to the number of lives lost and public expenditure in the health system.

4 Assessing a Specific Scenario

The model can adapt to a variety of specific scenarios and analyses. We present an experiment that was used to support the decision of keeping the on-site teaching activities suspended at the University, in the outbreak of the COVID-19 crisis.

The simulations carried objective was to analyze the impact of resuming presental teaching activities at educational facilities for young adults. For this purpose, a non-deterministic simulation model was used, grounded on the interaction of connected agents in a network. Each connection represents a possible (direct or indirect) contact between two persons, which may lead to the infection by SARS-CoV-2.

The scenario considered was a single building, assuming that this could be scaled to the total number of students without significant losses. Besides that, it was also assumed that part of the students would be resistant or careless to prevention measures and, for that reason, they would have a greater probability of being infected. The effect of limiting the number of students per classroom was analyzed, taking into account that this would also reduce the number of interactions per day for each agent. Finally, due to the probabilistic nature of the model, 100 simulations were run in each scenario to obtain more meaningful information from the results.

The explanation for the simulation model was divided into two parts: the model for the agents, which defines how each person acts “internally,” i.e., how they pass from one state to the other; and the model for the network, which indicates what are the possible interactions between different agents in the simulation. Following, both components would be detailed.

4.1 Infection Evolution Model

Here, we use the same SEIR-based infection evolution model presented for the simulations from the previous section. The state machine is shown in Fig. 5. Figures 6 and 7 present the probabilities for transition between states as a function of the time past since entering the current state.

4.2 Social Interaction Model

In these simulations, a single building was considered, being composed of 52 classrooms. The number of students per classroom will be given when presenting the simulation scenarios. Hence, it is assumed that each classroom is a full-connected network, i.e., a contact may occur between any members of this cluster. An instance of this setting is shown in Fig. 11, with 20 students in each classroom.

The number of classrooms and students was chosen based on the structure of buildings present at our University. The assumption of a complete graph for the contacts inside the classroom is based on the fact that all students will be sharing the same closed space, which allows interactions, as well as infections, to occur between any pair of students.

Besides the contacts in the classroom, random encounters outside the cluster were considered. These latter contacts
Fig. 9  Average number of persons in [E] and [I] states

Fig. 10  Stacked bars plot for the average number of persons in each state

Fig. 11  Graph representing the isolated clusters (classrooms)

are referred to in the text as “hallway contacts” for simplicity. However, they may occur in bathrooms or any other environments shared between clusters. For these hallway contacts, at each day a Watts–Strogatz graph is generated for all the students in the building, as exemplified in Fig. 12. The Watts–Strogatz graphs were generated with a mean-degree of 6 and a re-wire probability of 0.25.

Therefore, each day, a given number of interactions happens inside the classroom and another number of contacts occurs in the “hallways.” The interleaving of these two kinds of interaction is given in a random manner.

4.3 Results

Three different scenarios were considered for the simulations. In all of them, it was set that, given that a potentially infectious contact happened (that is, between a person in [S]
with another in [I]), the infection probability is 0.03 (3%). In case the person is “careless,” this probability rises to 0.15 (15%).

Also, in all scenarios, 5% of the total population is randomly picked to be “careless,” rounding down the result to the nearer integer. The number of individuals initially in [I] state is 0.5% of the whole population, rounding up the result to the nearer integer. Individuals that do not start in [I] state will be in [S] state, i.e., no individual starts in [E] state at the present simulations.

Hence, what differentiates the three scenarios is the variation of the following parameters:

- Number of persons per cluster;
- Number of contacts inside the cluster (classroom) per day;
- Number of hallway contacts per day.

The name chosen to represent each scenario indicates the value set for each of these parameters in the same order that they were presented. Since the number of classrooms (clusters) is fixed, varying the number of persons per cluster will directly affect the size of the population in each scenario. Therefore, the first scenario, called 20-10-6, will have 1040 agents in total. The second scenario, 15-7-4, will have 780 agents. And the third one, 10-5-4, has a total population of 520. The characterization of each scenario is summarized in Table 1.

As previously mentioned, 100 replications of the simulation were carried out for each scenario. According to Law (2015), terminating simulations are those having an event that specifies the length of each run. Since our goal in this computational experiment is to evaluate the spreading after a month of resuming on-site teaching activities, our natural event to terminate the replication is the arrival of the 31st day of the simulation. The data referring to the 31st day is presented to depict the situation at the end of the 30th day. That is the same as saying, the final state of each day is the initial state to the following day.

All random numbers used in the replications follow a uniform distribution. That includes the generations of each day’s Watts–Strogatz graph for hallway encounters and the selection of edges for interactions both in classrooms and hallways. We decided to keep the initial conditions (i.e., which agents are initially infected or exposed) as a constant between every run of the simulation on the same scenario.

The numbers of persons in each state at each time instance are separated random variables, which we define as the outputs of the simulation experiment. For instance, we call $S_i(k)$ the number of Susceptible individuals at the $k$-th day of the $i$-th replication. And the same applies to $E_i(k)$, $I_i(k)$, and $R_i(k)$ for holding the number of Exposed, Infected, and Removed individuals respectively.

For simplicity, we will introduce the concepts for the random variable $S_i$ but they are also extendable for the others simulation outputs. For a fixed day $k$, let $S_i(k)$ be a random variable defined on the $i$-th replication, for $i = 1, 2, \ldots, 100$; assume $S_i(k)$’s comparable for different replications, then $S_i(k)$’s are independent and identically distributed (IID) random variables (Law 2015). Therefore, we use the mean and the standard deviation of these random variables to depict the consistency of the simulation model, demonstrating that the variations in their values are only caused by the non-deterministic nature of the simulations.

We present the results for scenario 20-10-6 in Figs. 13 and 14. For each day $k$, a different random variable is considered.

![Fig. 12 Example of a graph representing the small-world network governing the random hallway interactions between all nodes](image)

**Table 1** Parameters of the simulation in each scenario

| Scenario name | Persons per cluster | Daily classroom contacts | Daily hallway contacts | Total population | Careless persons | Initially infected persons |
|---------------|---------------------|-------------------------|------------------------|------------------|------------------|--------------------------|
| 20-10-6       | 20                  | 10                      | 6                      | 1040             | 52               | 6                        |
| 15-7-4        | 15                  | 7                       | 4                      | 780              | 39               | 4                        |
| 10-5-4        | 10                  | 5                       | 4                      | 520              | 26               | 3                        |
Hence, a mean and a standard deviation may be computed separately for each $k$.

Figure 13 shows the mean values for the 100 simulations, along with the corresponding standard deviation. Figure 14 is a stacked bar plot depicting only the average values. In this, we omit the number of persons in [S] state for better visualizing the total of persons that caught the virus in this scenario. Besides that, the plot in Fig. 14 present the normalized values over the total population for the scenario.

Since presenting the statistical analysis for every time step would be too extensive, we selected two days to illustrate the behavior of these random variables. We elected a point in the middle ($k = 15$) and at the end ($k = 31$) the simulations. In Table 2, we present the analysis for scenario 20-10-6, where $\mu$ is the mean, CI (95%) is the 95\% confidence interval for $\mu$, $\sigma$ is the standard deviation, and $\sigma_n = \sigma / \mu$ is the normalized standard deviation.

From the values in Table 2, we can observe that the data variance grows as the simulation days increase. This behavior is expected since the random variables always start from the same initial conditions and any deviation between two different runs of the simulation will accumulate over time.

We have run a maximum likelihood estimator (MLE) for assessing if the simulations’ outputs could be fitted by a normal distribution. In Fig. 15, we present the results of this process for the random variable $R_i(k)$ on the same days of the previous analyses ($k = \{15, 31\}$), still for the 20-10-6 scenario. The purpose of this analysis is to illustrate that the outputs approximate a normal curve. Hence, we will not extend this representation for every random variable, on each day, and at each scenario.

Furthermore, we show the results for scenario 15-7-4 in Figs. 16 and 17. They are presented in the same order and meaning as the previous scenario.
We present the statistical analysis for the 15-7-4 scenario in Table 3. The interpretation of these measurements is similar to the previous case. However, we can notice that there was an increase in the dispersion of the results, which is clearer when analyzing the normalized standard deviation, $\sigma_n$.

In Figs. 18 and 19, we show the results for scenario 10-5-4. They are presented in the same order and meaning as the previous scenario.

Furthermore, in Table 4, we summarize the data from the statistical analysis over the 10-5-4 scenario. The interpretation of these measurements is similar to the previous scenarios. Once again, we can observe an increase in the standard deviation of most of the random variables. This can indicate a general behavior brought to sight by our simulations, which is the increase in results variance in inverse proportion to the population size.

Further conclusions drawn from the results are that even considering a low infection probability, resuming presential activities in these conditions could lead to a new wave of infections in the city as a whole. However, the simulations indicate that a restriction in the number of students per classroom could be effective in counteracting contagion. In relative terms, classrooms with 20 students would leave to an infection of about 19% of the whole population at the end of 30 days. In classrooms with only 10 students, the infection would drop to little more than 8% of the total population. It is worthy to remember that both scenarios start with 0.5% of infected agents.

Even though this restriction has a positive impact, it is not sufficient for definitively holding the virus spread on the campus. It would be necessary to take some measures for reducing the time a person is actively infecting others (state [I]). However, testing all students every day, or even every two days, is unfeasible.

Hence, a high risk was verified in the possibility of resuming fully presential activities in a near future. The shrinking of this risk is associated with restricting the number of contacts, either by reducing the number of students per classroom or the time that students remain in the University. It is important to emphasize that this study did not take into account the contamination in the displacement towards the University, nor the interaction with models of propagation in the city as a whole. Both factors would impact significantly in worsening the results presented here.

### 5 Conclusions and Directions

The main contribution of this paper is merging social interactions with the modeling of epidemics. This contribution allows us to simulate a variety of different situations. Further, our proposal enables us to evaluate counteracting measures that rely on the behaviors of each individual.
The results until this point are promising but there are a few directions that we envision for future research works. Following, we list a few of these possibilities: assessing the efficiency of diverse immunization strategies, such as vaccination programs; evaluating the emergence of new strains of the etiologic agent during the epidemic and adapting the model to handle outbreaks of more than one disease at the same time, including the possibility of the interplay between different diseases.

Furthermore, we point out the need for an identification process for the most common measures taken by public administrators. Regardless of the model used to make predictions, if we want control inputs to be translated into actual political measures, we need to know how much prohibiting a given activity impacts in truth to the social isolation level.

---

**Table 3** Statistical analysis for scenario 15-7-4

| Scenario | \( \mu \) | CI (95%) | \( \sigma \) | \( \sigma \_n \) |
|----------|---------|----------|---------|---------|
| \( S_i(15) \) | 742.9 | ± 2.5 | 12.7 | 0.02 |
| \( S_i(31) \) | 681.6 | ± 8.3 | 41.6 | 0.06 |
| \( E_i(15) \) | 8.3 | ± 0.9 | 4.6 | 0.55 |
| \( E_i(31) \) | 13.12 | ± 1.4 | 7.3 | 0.56 |
| \( I_i(15) \) | 7.0 | ± 0.8 | 4.0 | 0.57 |
| \( I_i(31) \) | 12.8 | ± 1.5 | 7.4 | 0.58 |
| \( R_i(15) \) | 21.8 | ± 1.2 | 5.9 | 0.27 |
| \( R_i(31) \) | 72.5 | ± 5.8 | 29.2 | 0.40 |

It can be used as support to decision-making involving policies for counteracting epidemics, such as restrictions to social interactions by preventing commercial activities and events, testing strategies, ICU beds management, etc.
Fig. 18 Mean and standard deviation for the spreading in the 10-5-4 scenario

Fig. 19 Stacked bars plot for the average percentage of people that caught the virus, for the 10-5-4 scenario

Table 4 Statistical analysis for scenario 10-5-4

|        | \( \mu \)  | CI (95%) | \( \sigma \) | \( \sigma_n \) |
|--------|------------|----------|-------------|---------------|
| \( S_i \) (15) | 498.9      | \( \pm 1.5 \) | 7.7         | 0.01          |
| \( S_i \) (31) | 477.6      | \( \pm 4.2 \) | 21.2        | 0.04          |
| \( E_i \) (15) | 4.2        | \( \pm 0.5 \) | 2.8         | 0.66          |
| \( E_i \) (31) | 3.9        | \( \pm 0.7 \) | 3.7         | 0.96          |
| \( I_i \) (15) | 3.3        | \( \pm 0.5 \) | 2.3         | 0.69          |
| \( I_i \) (31) | 3.8        | \( \pm 0.7 \) | 3.5         | 0.92          |
| \( R_i \) (15) | 13.6       | \( \pm 0.9 \) | 4.5         | 0.33          |
| \( R_i \) (31) | 34.7       | \( \pm 3.1 \) | 15.8        | 0.45          |

Declarations

Conflicts of interest The authors declare no conflict of interest.

References

Almeida, M.L., Morgado, I., Arenas, Z.G. 2020. Modelo SIR Estocástico na Transmissão da COVID-19 no Rio de Janeiro - Abordagem preliminar Modelo SIR estocástico na transmissão da COVID-19 no Rio de Janeiro - Abordagem preliminar. Anais da Sociedade Brasileira de Automática, 2(1). https://doi.org/10.48011/asba.v2i1.983

F.C., & Cunha, A., Jr. 2020. Inferência dos Estágios Iniciais da COVID-19 em Portugal. Anais da Sociedade Brasileira de Automática, 2(1). https://doi.org/10.48011/asba.v2i1.978

C.M., Cabrera, M.A., et al. 2020. COVID-19: Estudo da imunização usando modelo SIR. Anais da Sociedade Brasileira de Automática, 2(1). https://doi.org/10.48011/asba.v2i1.979

Funding The authors acknowledge the financial support of Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil) under Grant 142396/2020-6 and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES, Brazil) under Grant 88887.198893/2018-00.
Pazos, F. A., & Felicioni, F. 2020. A control approach to the COVID-19 disease using a SEIHRD dynamical model. Anais da Sociedade Brasileira de Automática, 2(1). https://doi.org/10.48011/asba.v2i1.1002

Ramos, A. M., Ferrández, M. R., Vela-Pérez, M., Kubik, A. B., & Ivorra, B. (2021). A simple but complex enough θ-sir type model to be used with COVID-19 real data. Application to the case of Italy. Physica D: Nonlinear Phenomena, 421, 132839. https://doi.org/10.1016/j.physd.2020.132839.

Ribeiro, M. H. D. M., da Silva, R. G., Mariani, V. C., & dos Santos Coelho, L. (2020). Short-term forecasting covid-19 cumulative confirmed cases: Perspectives for brazil. Chaos, Solitons Fractals, 135, 109853. https://doi.org/10.1016/j.chaos.2020.109853.

Roberts, M., Andreasen, V., Lloyd, A., & Pellis, L. (2015). Nine challenges for deterministic epidemic models. Epidemics, 10, 49–53. https://doi.org/10.1016/j.epidem.2014.09.006.

Salem, F. A., & Moreno, U. F. 2020. Simulação estocástica. Retrieved June 26, 2020, from https://covid19sc.github.io/simulacoes_probabilisticas.html

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of ‘small-world’ networks. Nature, 393(6684), 440–442. https://doi.org/10.1038/30918.

Wise, J. (2020). COVID-19: New coronavirus variant is identified in UK. British Medical Journal, 371, 4857. https://doi.org/10.1136/bmj.m4857.

Wu, J. T., Leung, K., & Leung, G. M. (2020). Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: A modelling study. The Lancet, 395(10225), 689–697. https://doi.org/10.1016/S0140-6736(20)30260-9.

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.