A SIMULATION STUDY OF CORONAVIRUS AS AN EPIDEMIC DISEASE USING AGENT-BASED MODELING

Posted on December 7, 2020 by Matthew

Category: Winter 2021
A Simulation Study of Coronavirus as an Epidemic Disease Using Agent-Based Modeling

By Amal Adel Alzu'bi, PhD; Sanaa Ibrahim Abu Alasal, MS; and Valerie J.M. Watzlaf, PhD, MPH, RHIA, FAHIMA

Abstract
At the end of 2019, the world faced the novel coronavirus, and with it fear of economic collapse and mass fatalities. Simulation systems can be used to monitor the behavior of the virus. Simulation provides an abstract representation of reality by conveying details and characteristics of reality in a simple application. One of the most important ways to simulate is agent-based modeling. The health information professional plays an important role in developing these models. In this research, we simulate the spread of COVID-19 in a region restricted to a population with specific demographic characteristics and social relationships. This study aims to clarify the effects of preventative techniques that suppress the spread of epidemics, such as quarantines, social distancing, and reduced mass transit.

Keywords: COVID-19, coronavirus, agent-based modeling, SIR model, epidemic diseases

Introduction
Coronaviruses are a large family of viruses that typically cause respiratory illnesses. COVID-19 is a zoonotic coronavirus, meaning it can transmit between and among humans and animals. Coronaviruses are often transmitted through the exudates or by contacting the surfaces of infected bodies and is often associated with colds, especially in the winter season.¹

Prior to 2019, two types of coronavirus caused serious epidemics. The first was the 2002-2004 outbreak of severe acute respiratory syndrome (SARS), caused by the SARS-CoV strain, which infected 8,000 and killed 774 in 29 countries. The second major outbreak was caused by Middle East Respiratory Syndrome (MERS-CoV), which was first discovered in Saudi Arabia; in the initial outbreak, 1,841 people were infected and 652 died.¹

In December 2019, cases of acute pneumonia were detected in a wet market in Wuhan, the capital of the Hubei Province, with a population of 11 million. On January 7, 2020, a novel coronavirus—later named COVID-19—was identified, and the first death was recorded.²

The new strain belongs to the extended family of coronaviruses, which includes six strains that target the human respiratory system. The incubation period for this virus is believed to be two to 14 days.¹

COVID-19 mainly threatens people managing chronic respiratory diseases with a weakened immunity, and the elderly. It is transmitted through close contact with the infected person. It is also
transmitted with droplets or aerosol. Symptoms of the disease include shortness of breath, coughing, headache, acute pneumonia, in addition to a fever, which is one of the most common symptoms.\(^3\)

By the middle of May 2020, the world began talking about 4.5 million infections and half a million deaths but to some extent, there are also nearly 2 million recovered cases and the largest number of the infections exist in the USA.\(^2\)

**Figure 1** illustrates the number of cases distributed around the world until May 22, 2020, as taken from the “worldometers.info” website\(^2\), figured using Tableau software.\(^5\)

Computer models can provide comprehensive insight into the behavior of disease outbreaks by simulating the spread of infections over populations with different geographic and demographic characteristics.\(^6\) Computer models can improve the representation and understanding of complex social structures with real-world communication networks that decide the dynamics of transmission, which in turn will be cost-effective as compared with real-time experiments.\(^7\)

One of the most reliable modeling approaches is agent-based modeling, which is a method that can deal with advanced modeling and simulations related to pandemics. Pandemic agent-based modeling re-creates the entire population and its dynamics by the integration of social systems, the heterogeneous fashion of interactions and communities on a single person’s level.\(^8\)

Agent-base modeling simulates a real-time environment (organized system) in an abstract representation, where the main element in this system is the agent (e.g., person, virus, first responders, etc.). The second main element in the model is the factors of each agent, which represent the agent’s characteristics and they are usually taken from real-time experiments (facts about the nature of agents). The third element is the links between agents that work by the values of the selected factors (e.g., a virus infects a person whose respiratory system is compromised by another chronic respiratory condition, such as COPD). And finally, all these elements must be behaving in a predetermined environment.\(^9\)\(^10\) There are many applications for agent-based modeling. However, in this work, we are limited to the epidemic applications.

One of the most prominent mathematical agent-based models is the Susceptible-Infectious-Recovered (SIR) model, which is based on ordinary differential equations.\(^11\) The model assumes that all the people inside the community are initially equally susceptible to be infected by the virus and that they will have a seasonal immunity after recovery once the infection subsides.\(^11\)

In this study, our main aim is to extend the SIR model to COVID-19 and its factors such as age, gender, smoking status and isolation tendency. Our contribution starts with collecting data about
Coronavirus as it is still a novel virus with unknown behavior since the way of getting infected is still unpredictable and unknown. Additionally, the severity of the disease is heavily dependent on the person’s age and health status. We also aim to study the effect of the controlling procedures to limit the epidemiological effect over closed populations. This paper is structured as follows: the second section lists the most related works about the epidemic disease simulations with agent-based modeling. The third section provides the proposed model, and model validation. Section four talks about the epidemic controlling procedures that might be undertaken. Experiments and results are discussed in section five. And finally, section six provides the conclusion of this work.

It is also important to point out that health informatics and information management professionals should work closely with epidemiologists, computer scientists, and biostatisticians to examine pandemics and epidemics using agent-based modeling. It takes a team with different subject matter expertise to examine this pandemic from all perspectives that involve data, computer systems, and population health.

**Related Works**

One of the most popular agent-based models is the BIOWar, which is a computer model that consists of many computational models for diverse types of environments such as social networks, epidemic disease transmission, weather forecasting, urban areas and the effects of bioterrorist attacks.\(^\text{12}\) BIOWar is also one of the models that adapts the SIR model and uses its main concepts to build its model.

Ge et al., studied H1N1 influenza inside an artificial classroom, in which the agents are students and teachers.\(^\text{13}\) This study considers that there exist some social relations between students and they have some common paths of movement and contact with others. The features they included in their study were age, gender, immunity, and agent activity inside the social network.

Khalil et al., studied H1N1 influenza in 1,000 individuals in an Egyptian society.\(^\text{14}\) They have analyzed the Egypt census data of 2006, where they determined the distribution of the population based on the census records. They also considered the social relations based on the person’s social types (e.g., sibling, child, other families, coworker, etc.). They combined multiple features, including those for the environment, human interactions, and disease characteristics. They adopted the SIR model and customized it to have many more categories for the agents such as (S) Susceptible, (C) in Contact, (E) Exposed, (I) Infectious, (Q) Quarantined, (NQ) Not Quarantined, (D) Dead, (R) Recovered, and (M) Immunized. Unlike Ge, et. al, they selected the contact to be random.\(^\text{13}\)

Another study on the H1N1 was presented by Luo et al.\(^\text{15}\) This study built an artificial society of a town with a population of 1,500 people considered as agents, to study the propagation rules. The model focused on public transportation as being a high-rate infection area with densely connected
agents and movements. They performed experiments with several cases, such as the lack of people’s actions, the closing of major places, and epidemic control surveillance.

The spread of influenza virus infection was simulated in a hospital Emergency Department in Winnipeg, Canada, by adapting the SIR model.\textsuperscript{16} Researchers divided the agent’s types into patients and healthcare workers and inanimate objects (e.g., chairs) that can transmit the virus. The collection of features they used were a group of topography, agent characteristics, behaviors, and interactions. Their method is different than others as the agent moves inside the model in predetermined actions in a circular order controlled by the random arrival of patients.

Also, Hackl et al.,\textsuperscript{17} studied the spread of influenza within mass transit systems in Zurich, Switzerland. Their main agent included individuals who move inside the transit system. They used basic socio-demographic data, housing locations, and the activity schedule for each of these people. In contrast to other research, they did not consider the effects of the vaccination existence, immunity, virus incubation periods, age distribution, and gender.

Chickenpox (varicella-zoster virus), also considered an epidemic virus, was studied by Rafferty et al.\textsuperscript{18} This study was conducted in an urban center and rural regions in Alberta, Canada. The authors used a wide range of properties for the model, including population size, mortality and fertility rates, initial cell-mediated immunity, force of reactivation, duration, probability of infection, connection range, and others.

EpiSims is an agent-based simulation tool that studies the population movability based on the census of people using a set of parameters to simulate the progress of the transmission of an epidemic disease between the infected and susceptible people.\textsuperscript{19} It has been built based on the Transportation Analysis and Simulation System (TRAN-SIMS).\textsuperscript{20}

TRANSIMS is a model built to predict the social networks of the population by assuming that the transportation infrastructure controls people’s intentions regarding where or when to carry out their activities. EpiSims was applied to Portland, Oregon, where people are carrying out their activities such as, studying, working, or shopping and they are moving between several locations and thus exposing themselves to various viral infections.

Proposed Model

In this study, we extend the SIR model and propose an agent-based model to simulate the spread of COVID-19 in an urban area. The environment that we are studying is two urban neighborhoods separated by crossings (A and B), where there are many options for movement and communication between people in both regions. Also, there are two groups of people, represented in either squares or circles, and are geographically divided by a border. These individuals are randomly moving around their environment.
The individual’s health status is represented by the colors of agents. We used four colors: white is uninfected, red is infected, green is recovered, and blue are immune. When a person is recovered, it is permanently immune to the virus. The other type of agent we have is the health official or ambulance worker, symbolized by yellow, who patrols the environment for people who are ill. Once a person comes in contact with an infected agent, the ambulance will immediately deliver the infected person to the hospital inside his region of residence. An illustration is provided in Figure 2.

SIR Model Extended

We used the SIR model as a base model to develop our proposed agent-based model, by extending extra states of the agents to be more realistic and related to the characteristics of COVID-19. Also, we used random distribution for the movements between agents and then the infection would be random. The agents traverse a series of classes showing the states of the disease over the infection duration. In the declaration of the state’s classes, we followed the model presented by Khalil et al.\textsuperscript{14}, where they also adapted the SIR model. The first class is Susceptible (S), which is the state of the person that is not in contact with other people in the environment and each person is initialized to this class with a 0.1 probability to be infected.

The second class is Contact (C), where people are in contact with each other, randomly or in predetermined links. Since the infection is controllable with a percentage, we will assume that the next state of the agent after the contact will remain Susceptible (S) or will become Infected (I) based on that infection chance, which basically depends on the immunity level of the person.

After that, if the infected individuals are still moving around and contacting the other susceptible people, they also will be exposed to the danger of being infected. Hence, there must be a controlling procedure by quarantining the infected individuals driven by a probability of isolation tendency.

In this state, the types of class would be Quarantine (Q), which is divided into two types (at the hospital or at home). And in case the person is not quarantined, his class would be Non-Quarantine (NQ). If people are not quarantined, they would not have adequate health care, so they would be exposed to the danger of death and their classification would be (D). People who have been quarantined in hospitals are assumed to have perfect healthcare, and they are expected to recover from the disease five times faster than people who are quarantined in their homes.

If people have finished the quarantine period with a high recovery chance, they will move to the Recovered (R) class. This recovery chance depends on the demographic properties assigned randomly to the individuals. At the end of the infection period, the person will recover from the disease and become a cured person with a state class named Cured (Cu). Figure 3 presents the chart of our proposed model, and Figure 4 illustrates the sequence flow chart.

Our model reflects the behavior of a group of people with various characteristics, such as gender,
age, immune ability, and smoking, who live in a closed area.

There are two types of agents. The first is ordinary people. The other is the paramedic. The characteristics mentioned above are randomly distributed to all people, while the paramedics have no specific characteristics since they are considered immune assuming they wear personal protective equipment (PPE) that prevents them from contracting the virus. In our model, we study the behavior of normal people. So, we excluded the paramedics from our model and we just added them to accelerate patients’ transmission to hospitals and to prevent people’s contact while transmission to hospitals. Therefore, we expect that the paramedics are not going to be infected, as we are studying the behavior of the people not the medical staff. The environment contains a certain number of randomly distributed residents who move in a period of time, measured in days, to suit the average period of the disease.

The ordinary individuals in this model could fall into many different categories listed as follows:

- **Infection chance**: probability that the disease would be transmitted from one individual to another.
- **Recovery chance**: probability of an individual to recover once the infection period ends.
- **Average recovery time**: the time needed (on average) for an individual to recover. The actual individual’s recovery time is pulled from a normal distribution centered around the average recovery time at its mean, with a standard deviation of a quarter of the average recovery time in days.
- **Average isolation tendency**: the average tendency of individuals to isolate themselves and to not spread the infection.
- **Average hospital tendency**: the average tendency of a person to go to a hospital when sick. If the infected individual is predefined as a “hospital goer”, then he will be isolated in the hospital, and the recovery time is half the time of the normal average recovery period because he will get better medication and
- **Cured**: probability of the person to get better and be healthy, and hence be immune from the virus infection again.
- **The number of males**
- **Smoking percentage**
- **Mortality rates**: from Meters et al.\(^2\) and being normalized into proportions consistent with the population on which our model was applied.

In addition, there are two conditions that might be used to control the behavior of the model, the first one is the "LINKs", and when this option is ON, there will be random links to represent the social network of individuals. In this case, the disease will spread twice as fast. However, when the option is OFF, then the disease will spread with equal probability between people who are moving around. The other condition is the "TRAVEL", when this option is on, then the people can travel between the two separate regions and the population will be mixed. There is also the Travel Tendency feature,
which will be active when the Travel option is turned on. This feature indicates 1 percent of individuals to travel per each unit of time.

**Model Validation**

Often, because of the lack of reliable field data and the lack of the real position of the individual cases, epidemiological modeling is very difficult to verify. Hence, we adopt the SIR model to develop and validate our proposed model because it has solid mathematical assumptions and has been validated for many real epidemic scenarios.

We have validated our proposed model by experimenting with the same parameters on both the SIR model and our model. Then we aligned the plot that resulted from both of them to check the behavior for each of the disease types: Susceptible, Infected, and Recovered.

Khalil et al.\(^ {14} \) used Mathematica\(^ {21} \), which is a computer program widely used in the fields of mathematics, engineering, and various sciences, to simulate the SIR model with the following variables: duration of infection=15 days; initial immune=0; and initial infection chance=0.01.

Results are shown in [Figure 5](#). The original SIR model\(^ {11} \) showed that same behavior in [Figure 6](#). The same values of the previously mentioned parameters were used in our proposed model, assuming that there is no quarantining. Results are shown in [Figure 7](#). We can notice that the three behaviors are not perfectly matched, but at the end, our model matches the general behavior of the SIR model.

We can hypothesize that the difference is related to (a) the different allocation of community demography information such as sex, age, and smoking and (b) the random variable allocation for the infection time which is the infection chance where the SIR model uses deterministic values. We can also notice that in our model, the number of non-infected people starts to increase after a while because we are adding them back to the (not infected counter).

**Epidemic Controlling Procedures**

Suppression procedures are often applied to control epidemic diseases, but only after the disease reaches the highest level of outbreak. This would bring chaos among individuals, especially after a large number of deaths and this is what we are seeing across the world today with COVID-19.

Therefore, a course of events should be predicted in epidemiological situations to ensure the control of the situation before reaching a high peak of the outbreak. One of the most important methods of this control is to quarantine infected patients inside their homes or in private hospitals.

Additionally, it also is preferable to reduce the movements resulting from the behavior in the practical life or the social life network where uninfected persons are required to limit their relationships and movements to ensure that no infection is transmitted to them. In the worst cases,
people of a particular city or neighborhood are required to remain within the range of their city and not move to other cities and preventing other individuals from other cities to enter. Furthermore, people who are at risk of infection (susceptible) are advised to exercise healthy habits that increase pathological resistance (immunity) such as adequate sleep, exercise, and smoking cessation.

**Experiments and Results**

In this study, we focus on the controlling strategies that could be taken during epidemic situations. We have built multiple scenarios in our model using NetLogo software to simulate the effect of each strategy. The population we used is 1,000 persons and the infection chance is 25 percent (this means each person has a chance less than or equal to 25 percent of contracting the virus based on their immunity), recovery chance is 25 percent, average recovery time is 15 days, intra-mobility (movements inside the city range) is 60 percent, the ages were randomly distributed, and smoking was also randomly distributed.

The controlling strategies include:

- Applying the model without any controlling procedures and without considering any relationships between people or traveling between the two regions
- Applying the model with average isolation tendency = 10 percent
- Applying the model with an average hospital going tendency = 10 percent
- Applying the model with adding a cured chance = 20 percent
- Applying the model with employing the paramedics = two persons
- Applying the model with relationships between people (a) without applying controlling procedures, (b) with applying controlling procedures
- Applying the model with allowing the travel between the two regions with travel tendency = 60 percent, (a) with links, (b) without links, (c) without controlling procedures

The results of the scenarios are plotted and listed in Figures 8-14. We notice from the plots that the peak of the infection changes based on the application of the controlling strategies. Also, as controlling strategies are implemented, the epidemic duration is decreased.

For Scenario 1 ([Figure 8](#)), the peak of the epidemic started early from day number seven. Since the values of the parameters are random, we conducted the experiment five times and noticed that the values are very close. The number of deaths was between 100 and 150, where we can predict that poor controlling of such a disease can yield high mortality rates.

In the second scenario, shown in [Figure 9](#), we did the home isolation option by 10 percent; the peak of the infection was on the tenth day, which means that it has reduced the prevalence of infection with comparison to the first scenario. The number of deaths did not diminish too much, but it ranged between 100 and 120 people.
The third scenario (Figure 10) was about adding both home and hospital isolation with 10% each. The peak of the infection was also on the tenth day since we did not alter the basic parameters of the model. This experiment was to check the effect of hospital isolation, where we suppose that the healthcare provided in the hospital is better than home isolation, as it limits the person's communication with his surroundings completely, to reduce the spread of infection. The number of deaths was reduced to the range of 90 to 100.

For the fourth experimental scenario (Figure 11), we assumed that people have been cured and they are safe from being infected again with a chance of 20 percent. We noticed that the peak of infection was on day five, but with a lower number of infections. The number of deaths was also reduced to 75 to 95. We tried to increase the immunity for 25 percent, 30 percent and so forth, and we found that the number of infections and deaths decreased.

In the fifth scenario (Figure 12), we added the possibility of ambulance personnel transporting infected people to hospitals. This will accelerate the process of transferring patients to the hospital, in addition to reducing the infection rate, as ambulance personnel are designed in this model, so that they would be isolated with PPE from infection. The number of deaths was decreased at a tiny level but still in the same range of scenario four.

The sixth scenario was to apply the LINKs option, which indicates the relationships among the community. First, we applied these links without applying controlling procedures (Figure 13 a), then with controlling procedures in place (Figure 13 b).

The first try resulted in epidemic, where the rise in infection occurred on the sixth day, as the number of infected people at that time was more than half of the community, and the number of deaths rose to 150.

This situation looks like the first scenario with more rapid infection spreading related to social network interactions. The second part of this experiment was much less dangerous than the first, because the necessary precautions—isolations and the presence of ambulance personnel—are in place. The number of infected people also reached its peak on the sixth day, but with fewer infections, so that the largest number of infections was up to 423, and the number of deaths decreased to 85 to 95.

The final scenario allowed travel between both regions and applied controlling procedures on people traveling. The first experiment was to allow the traveling between two regions with applying the LINKs and with the existence of the controlling procedures. This situation means that people from region A will contact people from region B whom they linked with only. This assumption can let us predict that the mobility would not be chaos, but it will be directed by the social network between the two communities. In this experiment the peak of the infection was on day seven, the number of infections was about 450 cases only 75 to 95 out of them have died (Figure 14a). The second experiment was to allow the mobility but without links limits, which means that the people
will move between the two regions in a total random chaos but with controlling procedures. The peak of infection was on day six, the number of deaths was around 70 to 90 people (Figure 14b). The third experiment was to study the community with traveling but without links or controlling procedures. The situation was totally epidemic, where the number of infections was more than 50 percent and the number of deaths was around 200 (Figure 14c).

**Discussion**

We have discussed the mechanisms for controlling epidemiological conditions. Through the results that we obtained for multiple scenarios and after repeating every scenario more than five times, we found that the implementation of isolation policies can result in significant reductions in death rates. In addition, the process of separating people and limiting their social relations can lead to a clear effect of reducing the infection rates in society. Mobility ratios also have an impact on the spread of infection, whether it is going only according to social proximity or a random way. And we have noticed that the failure to take preventive measures can cause a rapid increase in the spread of the infection, as we are seeing in some countries across the world.

Countries are now heading to apply difficult coercive measures to cities and urban residents, at the peak time of the spread of the disease. This behavior is correct, but it must be taken in the necessary time so that the peak of spread can be controlled before reaching the stage of the outbreak and the high death numbers, which is the goal of simulation models.

Health informatics and information management professionals can aid in the early detection of such outbreaks since they are the ones dealing with health data every day. As health information professionals scour the electronic health record, they can establish trends in clinical documentation, coded data, and quality measures that can relate to signs and common symptoms of the virus. All of this information is used in building the agent-based modeling that was previously described. Health information professionals, working with other members of the healthcare team can be at the epicenter of pandemics since they have data at their fingertips that is crucial to building simulation models that can be used to slow the progress of a devastating illness such as COVID-19.

**Conclusion**

In this work, we studied COVID-19 on a closed community using agent-based modeling. We matched the details of the model with the current status of the disease, considering that there is some ambiguity in the details of this disease as the infection rate varies from one country to another and from one community to another. We did an extension of a mathematically proven basic model (SIR) so that we added other scenarios to the model to represent the stages of disease transmission, infection, and recovery. The main contribution of this work lies in reviewing the main data of the disease and in representing the results of the controlling measures in cases of the spread of epidemic diseases, such as COVID-19. We also found that health information professionals working with epidemiologists, computer scientists, and statisticians can aid in the development of the
simulation models by providing accurate data to assist in the development of such models. Finally, we found that staying home and hospital isolation policies, in addition to preventing travel between cities, will reduce the prevalence and thus reduce the deaths.

Authors

Amal Adel Alzu’bi, PhD. (aazoubi9@just.edu.jo) is Assistant Professor, Department of Computer Information Systems, Jordan University of Science and Technology, in Irbid, Jordan.

Sanaa Ibrahim Abu Alasal, MS. (sanaasal11@gmail.com), is Research Assistant, Department of Computer Science, Jordan University of Science and Technology, in Irbid, Jordan.

Valerie J.M. Watzlaf, PhD, MPH, RHIA, FAHIMA, (valgeo@pitt.edu) is Vice Chair of Education and Associate Professor, Department of Health Information Management, School of Health and Rehabilitation Science, University of Pittsburgh, in Pennsylvania.

References and Notes


There are no comments yet.