



# Diffusion as a first model of spread of viral infection

Paulo H. Acioli<sup>a)</sup>

*Department of Physics and Astronomy, Northeastern Illinois University, Chicago, Illinois 60625*

(Received 26 March 2020; accepted 2 June 2020)

The appearance of the coronavirus (COVID-19) in late 2019 has dominated the news in the last few months as it developed into a pandemic. In many mathematics and physics classrooms, instructors are using the time series of the number of cases to show exponential growth of the infection. In this manuscript, we propose a simple diffusion process as the mode of spreading infections. This model is less sophisticated than other models in the literature, but it can capture the exponential growth and it can explain it in terms of mobility (diffusion constant), population density, and probability of transmission. Students can change the parameters and determine the growth rate and predict the total number of cases as a function of time. Students are also given the opportunity to add other factors that are not considered in the simple diffusion model. © 2020

*American Association of Physics Teachers.*

<https://doi.org/10.1119/10.0001464>

## I. INTRODUCTION

The end of 2019 and beginning of 2020 have been dominated by the spread of the coronavirus (COVID-19). The disease started in the Hubei province in China in December 2019 and by early January 2020 started to spread. It grew very rapidly, triggering responses from the Chinese and other governments. On March 11, 2020, the World Health organization declared that COVID-19 was then characterized as a pandemic.<sup>1</sup> The situation triggered different responses from different governments. In the United States, many colleges took the initiative to start their own social distance programs, including sending all students home, extending spring breaks, and ultimately moving all classes to distance learning.<sup>2,3</sup> Many cities and states followed up with shelter at home mandates.<sup>4</sup> The situation in many countries became alarming due to the exponential growth of new cases and deaths.<sup>5,6</sup>

One of the consequences of this pandemic is that instructors at colleges and universities started to monitor and model the data, using this as a teachable moment for students and colleagues. The first step in modeling the data is to plot the number of cases as a function of time and show that it exhibits an exponential growth. For beginners, it is important to introduce them to the logarithmic scale, where the plot becomes a straight line and students can extract the exponent and realize that the fit to the US data on March 20, 2020 shows that the number of cases doubled every 2.4 days. Fits to the initial data for Chicago and New York City show that New York has a higher growth rate than Chicago (Fig. 1), and the total number of cases in Chicago is substantially smaller. One can speculate that this might be due to difference in population densities in these cities. After the first 20 days and after mitigation and social distance measures were imposed, it is clear that the rate of infection slowed down.

In this work, we propose to look at some of the factors affecting the spread of viruses using a simple diffusion model in which each individual in a population is treated as a Brownian<sup>9</sup> particle with diffusion constant  $D$ . Also added to this model is the incubation period of the virus and a probability of transmission of the virus if individuals are closer than a certain distance. This model is to be used as a project in a computational physics course and verify if it adequately

predicts the exponential growth of the number of cases, as well as if the population density, mobility, and probability of transmission play roles in the percentage of the population that will be infected as a function of time. Students will be asked to modify a code, analyze the output for different sets of parameters, and write a critical analysis of its predictive effectiveness.

## II. COMPUTATIONAL PROJECT DETAILS

In this section, we propose a project to be implemented in computational physics courses to study the spread of infectious diseases as a simple diffusion of individuals. The benefits of this project are that its implementation is simple, but it can lead to a qualitative understanding of how diseases are spread, and it can also allow the student to understand factors that can affect it. The main ingredients of the model are: Individuals are considered particles that obey a Brownian diffusion process; each individual will have three possible states, healthy, sick (contagious), and cured; a healthy individual has a probability of getting infected if its distance to a sick individual is smaller than a certain threshold; the incubation and sickness periods are the same; once an individual gets cured, it cannot be infected or contagious again. VPython (or GlowScript)<sup>10</sup> is the language of choice as it allows for a real time visualization of the infection spread and it will allow for fast simulations of small populations on a laptop. In the assignment, students will implement the code, analyze the data generated, critique the initial assumptions, and propose improvements for more realistic simulations. Below we describe the algorithm starting with the standard diffusion equation.

The diffusion equation in 2D

$$\frac{\partial f(x, y, t)}{\partial t} = D \left[ \frac{\partial^2 f(x, y, t)}{\partial x^2} + \frac{\partial^2 f(x, y, t)}{\partial y^2} \right] \quad (1)$$

is used to study many phenomena<sup>11–20</sup> from diffusion inside the nucleus to population dynamics to solving the Schrödinger equation. The function  $f$  will have different meanings depending on the application, from temperature to density. In this work, we will use it for the diffusion of individuals, treated as particles, over closed boundaries subject

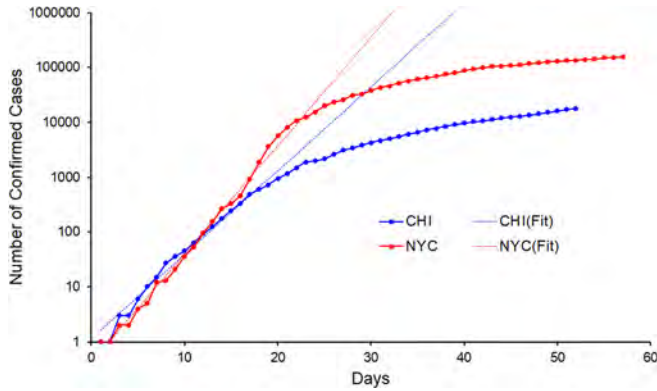


Fig. 1. COVID-19 cases in Chicago (CHI, dark gray) (Ref. 7) and New York City (Ref. 8) (NYC, light gray). The dashed lines are the best fit to the data after the initial spike in cases. The  $x$ -axis is the number of days from the date that the first case in each city was detected.

to contamination of a viral infection. The normalized solution to Eq. (1) for a single particle is given by

$$f(x, y, t) = f_0 \frac{1}{\sqrt{4\pi Dt}} e^{-(x^2+y^2/4Dt)}. \quad (2)$$

Therefore, one can simulate the diffusion of a particle from its previous position by generating a Gaussian distribution of zero mean and variance  $\sqrt{2Dt}$ . For a system of  $N$  non-interacting particles with the same diffusion constant, we use Eq. (2) for each particle at each simulation time step.

The next ingredients in the simulation will be the population density  $\rho$ , the number of inhabitants (particles in the simulation cell)  $N_{\text{pop}}$ , the diffusion constant  $D$ , the number of simulation steps  $N_{\text{step}}$ , the time step  $dt$ , the incubation period  $t_{\text{inc}}$ , the transmission radius  $r_{\text{transm}}$ , and the probability of transmission from an infected to a healthy individual,  $prob$ . All of these variables are set at the beginning of the simulation. We preset the total simulation time to be 90 days and that each time step is 0.01 days, therefore each simulation takes 9000 steps. The algorithm is described below:

- (1) Input  $\rho$ ,  $N_{\text{pop}}$ ,  $N_{\text{step}}$ ,  $D$ ,  $dt$ ,  $t_{\text{inc}}$ ,  $prob$ ,  $r_{\text{transm}}$ .
- (2) Calculate the size of the square cell as  $L = \sqrt{N_{\text{pop}}/\rho}$ .
- (3) Initialize the population.
- (4) Choose a fraction of the initial population to be infected, and set the timer for the sickness,  $t_{\text{sick}}$ .
- (5) Loop over  $N_{\text{step}}$ .
- (6) Move all individuals according to the Gaussian distribution, Eq. (2).
- (7) Compute the distance between each healthy and infected individuals. If the distance is less than  $r_{\text{transm}}$ , the healthy individual becomes sick with probability  $prob$ .
- (8) Subtract the sickness timer by  $dt$ .
- (9) If  $t_{\text{sick}} < 0$  the sick individual gets cured.

In order to generate more accurate statistics, we suggest the students run simulations with the same initial parameters multiple times, between 20 and 100, depending on the size of the system and the speed of the student's computer. We suggest that the smallest population contains 100 individuals, as even one sick individual corresponds to an initial infected population of 1%. Depending on how much time the students have to complete the analysis of the project, they can use a population of 1000, being aware that each individual 9000 steps simulation can take up to 30 min. They can speed up

the process by using larger time steps, however, they must test if the results of an individual simulation with the same diffusion constant and different time steps lead to similar outcomes.

### III. AN EXAMPLE OF A PROJECT

In this section, we provide an example of a project that students could pursue and also validate the method by comparing the results of the simulation with the data presented in Fig. 1. In this project, students will study the differences of the infection proliferation in New York City (NYC) and in Chicago. New York City was chosen because it is experiencing a very rapid growth in the number of cases and preventive measures such as shelter-at-home were taken by the city and state governments at very early stages of the infection. Chicago was chosen as a local connection to Northeastern Illinois University (NEIU) students, and is also experiencing an exponential growth in the number of cases and has also been affected by shelter-at-home mandates. We must be mindful that the current model will not be able to study the effects of the measures that the government is taking, but it might be able to justify their need.

The population density of New York city is 10,194 people/km<sup>2</sup>,<sup>21</sup> while Chicago's is 4,665 people/km<sup>2</sup>.<sup>22</sup> It is natural to offer the students the hypothesis that if all other variables are the same, the spread of infections in New York City will be faster than in Chicago and that the number of cases will be much larger for the same period of time after the first case. With this hypothesis alone, students should be able to generate enough data for the project. They can also discuss if the mobility (i.e., the diffusion constant) should be the same in both cases and study the effect of mobility in the spread of the disease. In the proposed model, the period of incubation is the same as the period of sickness, but students can be offered the option to modify this assumption. The model also assumes that once cured, an individual will attain immunity, and will not be able to spread the disease. They can discuss modifications to the model to incorporate relapse. It is clear that with this very simple model and this very limited two-city project, they can perform a very thorough study that can give us some qualitative understanding of infectious disease spread.

In order to validate our model, we will try to recreate the same initial exponential growth as experience by Chicago and New York City. The exponential fit to the data in Fig. 1 yields exponents of 0.352 and 0.455 for CHI and NYC, respectively. Although we envision that we can see such growth in a 100 individual population, this growth is unattainable for the same time frame as the real data, as 100% of the population would be infected in days 13 and 11 if we start with one individual infected in day 0. Deviations from the exponential behavior will be seen even earlier as there is not enough susceptible population available to be infected. In order to get similar exponential growth, we start with simulations of populations of 100 individuals and 1 sick individual chosen at random. The initial diffusion constant was chosen as  $D = 100 \text{ m}^2/\text{day}$  with a time step of  $dt = 0.01 \text{ day}$ . Therefore, the variance of the Gaussian distribution is  $2 \text{ m}^2$ . In this case, the number of simulation steps is 9000 for each individual simulation, corresponding to 27 s on a laptop with an Intel(R) Core(TM) i5-8250U CPU @ 1.60 GHz. Therefore, 50 simulations will take about 22.5 min of computation. We assume that the transmission radius is 2 m, and

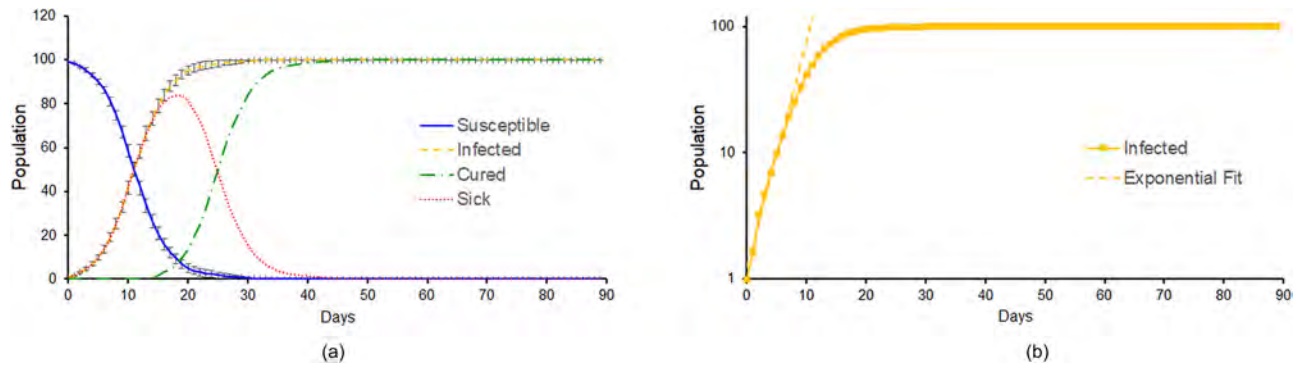


Fig. 2. (a) Average of 50 90-day simulations of the spread of a virus in a population of 100 individuals in a square cell with the population density of NYC ( $\rho = 0.012$  people/m<sup>2</sup>).  $D = 100$  m<sup>2</sup>/day,  $prob = 0.2$ ,  $dt = 0.01$  day. (b) Graph of the number of infected individuals (solid line) in a logarithmic scale and the exponential fit for the initial 10 days of the simulation (dashed line).

that the probability of contamination is 20% per time step. The incubation period is taken as 14 days. Below are some of the results of these simulations.

In Fig. 2(a), we present the results of the simulation for New York City with the set of parameters in the previous paragraph. As one can see, under this assumption 50% of the population is sick on day 11 and 99% of the population will be infected after 28 days. One can see that after day 18 the number of sick people starts to decline. However, if the death rate is similar to what has been observed for the COVID-19 pandemic, about 5.5% of the population of NYC would perish, and the numbers would be even worse, since no major city in the world would be able to have hospital beds for 83.8% of its population at the peak of the infection. In the figure, we also observe that the rate of cure follows the number of infected individuals with a lag time of 14 days, which is the incubation/sickness period. An exponential fit, shown in Fig. 2(b), yields an exponent of 0.41 which is a bit lower than what was observed in the real data. Students will need to search for the set of parameters that is able to fully match the initial growth rate data that they choose to model.

In Fig. 3(a), we report the results for the city of Chicago. Most of the parameters are the same, with the exception of the density that will be changed to  $\rho = 0.0047$  people/m<sup>2</sup>. For this set of parameters, 50% of the population will be infected after 26 days reaching a maximum of 87.8% after day 86. The peak of the number of sick people is 42.3%, and it happens about 32 days after the first 1% of the population is infected. In Fig. 3(b), we present a logarithmic scale plot of the total number of infections together with an exponential fit for the initial stages of the infection. Similar to what

happened in the simulation for NYC, it is clear that the number of infections deviates from the exponential growth after 13 days. The exponent in the fit is 0.2 which is also smaller than the value for the data for Chicago. A closer fit is likely to take place by increasing the probability of infection.

These two examples do support our initial hypothesis that under the same conditions, one would expect a slower growth rate in a less dense population. This leads to the conclusion that keeping all parameters the same, the percentage of the population infected is correlated with the population density. In order to determine how mobility affects the rate of infection, one would change the diffusion constant and repeat the simulations. The expectation is that the lower the mobility, the lower the total infection rate. In addition, even with the limitations of a small population simulation we were able to find growth rates of infection that are similar to what was observed for the cities of Chicago and New York, thus validating this approach to simulate the spread of infectious diseases.

In order to show the limitations of small population simulations, we run a simulation for the city of Chicago with a probability of infection of 30% per time step but increased the number of individuals to 10 000. The results are shown in Fig. 4(a). With this set of parameters, we obtain the same exponential growth as observed in Fig. 1 for a period of 15 days. Starting with 0.01% of the population infected, we see that at the end of 90 days about 40% of the population will have been infected. One can also see that even a small diffusion constant of 100 m<sup>2</sup>/day and a probability of transmission of 30%, the virus will spread to a large percentage of the population. It is clear that a diffusion model does a

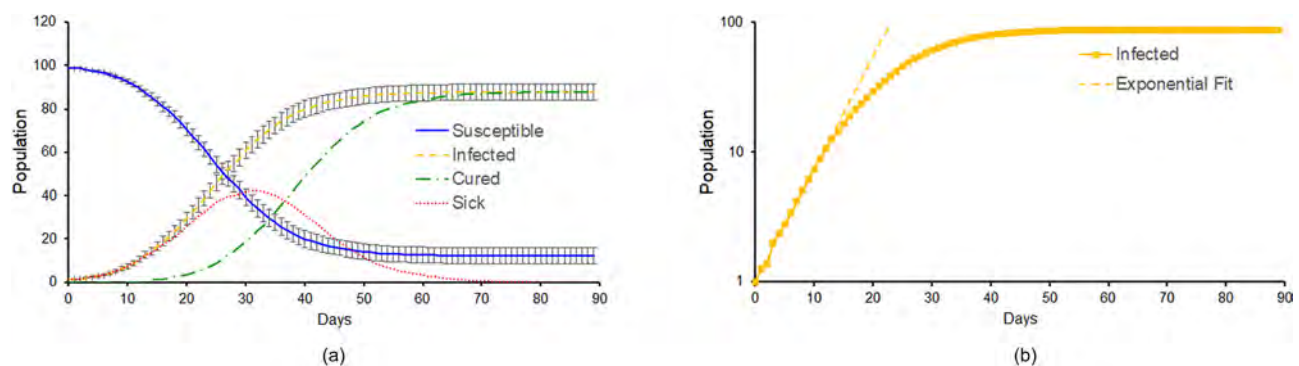


Fig. 3. Average of 50 90-day simulations of the spread of a virus in a population of 100 individuals in a square cell with the population density of the city of Chicago ( $\rho = 0.0047$  people/m<sup>2</sup>).  $D = 100$  m<sup>2</sup>/day,  $prob = 0.2$ ,  $dt = 0.01$  day. (b) Graph of the number of infected individuals (solid line) in a logarithmic scale and the exponential fit for the initial 10 days of the simulation (dashed line).

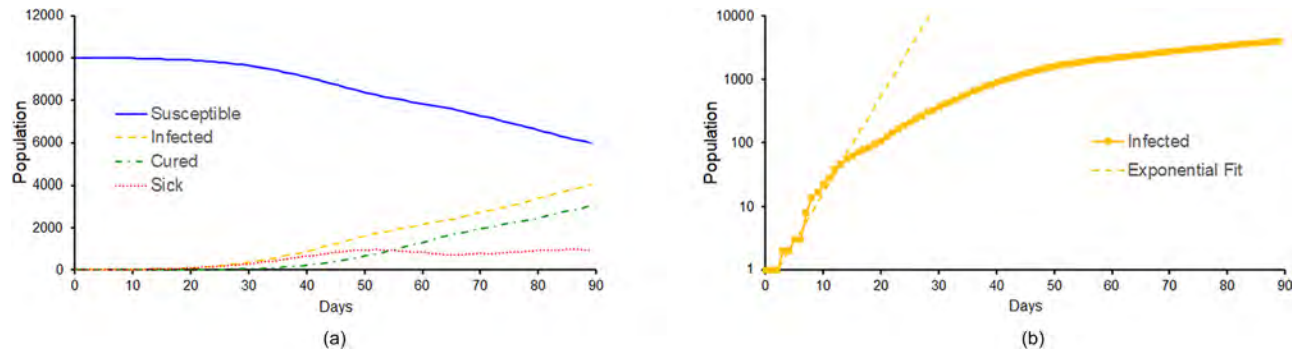


Fig. 4. One 90-day simulations of the spread of a virus in a population of 10000 individuals in a square cell with the population density of the city of Chicago ( $\rho = 0.0047$  people/m<sup>2</sup>).  $D = 100$  m<sup>2</sup>/day,  $prob = 0.3$ ,  $dt = 0.01$  day. (b) Graph of the number of infected individuals (solid line) in a logarithmic scale and the exponential fit for the initial 15 days of the simulation (dashed line).

very good job of simulating the spread of the COVID-19 virus and could be used in a classroom setting, and with the modifications suggested below one can even use it for more realistic predictions when simulating larger populations. The main limitation will be access to better computational resources such as a parallelized code and a computer cluster.

To conclude, we show snapshots of the simulation cell and its population in Fig. 5. White spheres represent the susceptible people, dark gray represent sick, and light gray are those individuals who recovered. This window is very useful for a quick analysis of what is happening and can show how

the infection spreads over time. We found it useful to change the radius of the individuals so that they are visible as the population increases, and the cell size on the screen remains the same. Because we are displaying the results for 10000 individuals it is interesting to see that the infection has a clear origin and the diffusion process spreads the disease outwards. In these simulations, we chose closed boundaries and as a result there will be a limitation on the growth as there will be less susceptible neighbors to be infected when the disease reaches a boundary. To avoid this limitation, one can use periodic boundary conditions.

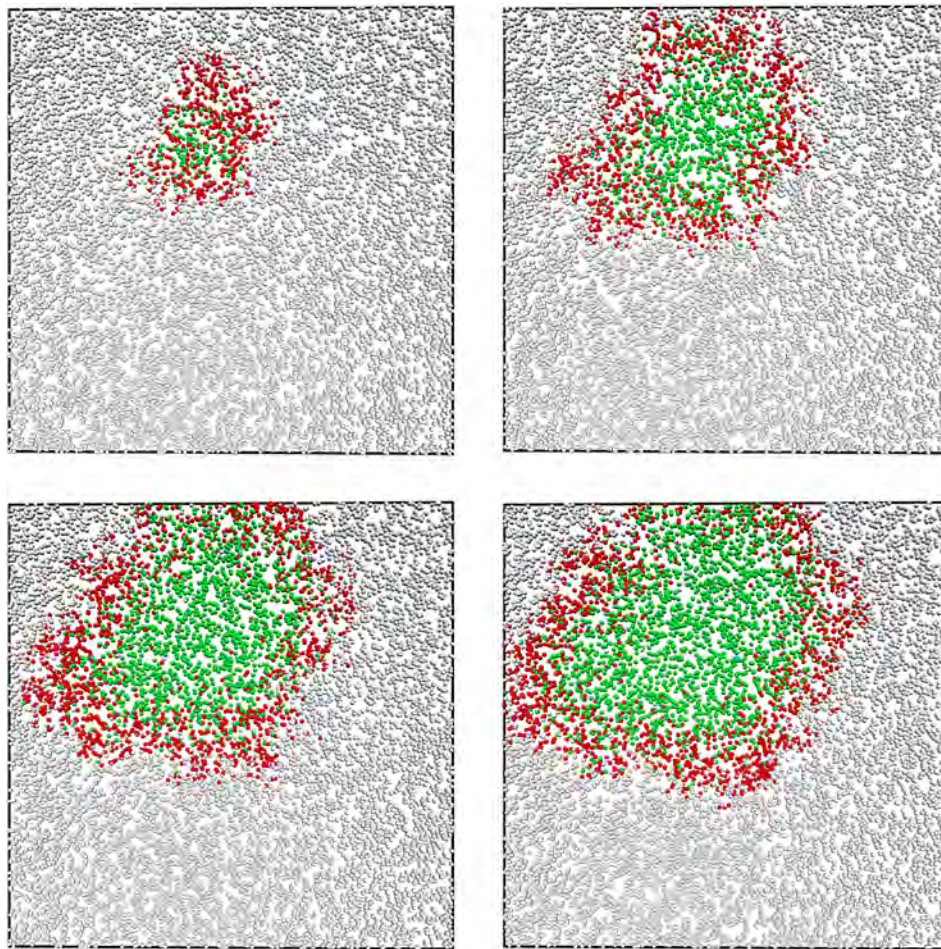


Fig. 5. Snapshots of one of the 90-day simulations of the spread of a virus in a population of 10000 individuals in a square cell with the population density of CHI ( $\rho = 0.0047$  people/m<sup>2</sup>).  $D = 100$  m<sup>2</sup>/day,  $prob = 0.3$ ,  $dt = 0.01$  day. White spheres represent the susceptible people, dark gray represent sick, and light gray are those individuals who recovered.

The main limitation of this model is that it assumes the same mobility for all individuals and a constant density in the simulation cell, and no travel between cities. In addition, the average distance traveled by Brownian particles is proportional to the square of the simulation time. Thus, there is a limit on how far they will travel and how many people an individual can infect. However, depending on the level of the students, instructors can use the same code and create sub-regions in the simulation cell that would have different population densities and subsets of the populations can be confined to these regions. The students could allow individuals to go from one region to another with a given probability. This would allow the possibility of an infected individual to move to a region that otherwise would have no infections. This modification would allow for the spread of the infection across borders. One could also include quarantine effects, by creating small cells where a single infected individual will be confined for a period of time and no other individuals are allowed in. Social distancing can be implemented by a small repulsive potential attributed to some individuals. These modifications will bring this simple model closer to more realistic simulations as those of Refs. 23 and 24. To help with the implementation of these projects, the original code can be downloaded from the author's website.<sup>25</sup>

#### IV. CONCLUSION

This project is simple enough that it can be implemented and analyzed in a first course on computational physics. However, it is also rich enough that it can lead to a large amount of data that can be used to qualitatively and quantitatively analyze the spread of viral infection. In addition, it can help students understand the phenomenon of particle diffusion and Brownian motion. In the example presented, the simulation was limited to the same probability of infection, the same time step, the size of the population was fixed, and the length of the disease was the same as the incubation period. But students are also able to change these parameters to see if they can reproduce data readily available in the news. In addition, they will have the opportunity to determine if the initial growth rate in the number of cases follows an exponential trend, and if once the growth rate reaches a certain threshold it will follow the expected logistic behavior.

In conclusion, we hope to have convinced the reader that a simple diffusion model can be used to qualitatively explain the spread of disease and even in some cases quantify it. It is our judgment that it will allow students to work on a problem that is directly affecting them, and that these simulations will help them offer valuable insight on the issue.

#### ACKNOWLEDGMENTS

The author would like to thank Dr. Greg Anderson for sparking the discussion of using the COVID-19 pandemic as a teachable moment and Dr. Orin Harris for inspiring the idea of comparing the data for different populations.

<sup>a)</sup>Electronic mail: p-acioli@neu.edu

<sup>1</sup>World Health Organization Director-General's March 11, 2020 COVID-19 Media Briefing <<https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>>

<sup>2</sup>The Wall Street Journal, "Coronavirus prompts colleges to send students home," <<https://www.wsj.com/articles/coronavirus-prompts-colleges-to-send-students-home-11583862936>>.

<sup>3</sup>USA Today, "Coronavirus prompts colleges to send students home," <<https://www.usatoday.com/videos/news/2020/03/14/coronavirus-colleges-spring-break/5046501002/>>.

<sup>4</sup>USA Today, "'Stay home, stay healthy': These states have ordered residents to avoid nonessential travel amid coronavirus," <<https://www.usatoday.com/story/news/nation/2020/03/21/coronavirus-lockdown-orders-shelter-place-stay-home-state-list/2891193001/>>.

<sup>5</sup>Johns Hopkins: Coronavirus Resource Center, <<https://coronavirus.jhu.edu/map.html>>.

<sup>6</sup>World Health Organization, Coronavirus disease (COVID-2019) situation reports, <<https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports>>.

<sup>7</sup>Chicago: Coronavirus Response Center, <<https://www.chicago.gov/city/en/sites/covid-19/home.html>>.

<sup>8</sup>Spectrum News NY1, "Coronavirus in New York," <<https://www.ny1.com/nyc/all-boroughs/news/health-and-medicine/coronavirus-blog>>.

<sup>9</sup>R. Feynman, "The Brownian movement," Feynman Lect. Phys. I, 41-1-41-18 (1964).

<sup>10</sup>VPython: 3D Programming for Ordinary Mortals, <<https://vpython.org/>>.

<sup>11</sup>F. A. Williams, "Elementary derivation of the multicomponent diffusion equation," *Am. J. Phys.* **26**, 467-469 (1958).

<sup>12</sup>D. C. Kelly, "Diffusion: A relativistic appraisal," *Am. J. Phys.* **36**, 585-591 (1968).

<sup>13</sup>Karmeshu and L. S. Kothari, "Neutron diffusion as a random walk problem," *Am. J. Phys.* **40**, 1264-1269 (1972).

<sup>14</sup>C. Domb and E. L. Offenbacher, "Random walks and diffusion," *Am. J. Phys.* **46**, 49-56 (1978).

<sup>15</sup>J. DArruda and E. W. Larsen, "Simple derivation of the diffusion equation from the Fokker-Planck equation using perturbation methods," *Am. J. Phys.* **46**, 392-393 (1978).

<sup>16</sup>A. M. Albano, N. B. Abraham, D. E. Chyba, and M. Martelli, "Bifurcations, propagating solutions, and phase transitions in a nonlinear chemical reaction with diffusion," *Am. J. Phys.* **52**, 161-167 (1984).

<sup>17</sup>S. Redner and P. L. Krapivsky, "Capture of the lamb: Diffusing predators seeking a diffusing prey," *Am. J. Phys.* **67**, 1277-1283 (1999).

<sup>18</sup>K. Ghosh, K. A. Dill, M. M. Inamdar, E. Seitaridou, and R. Phillips, "Teaching the principles of statistical dynamics," *Am. J. Phys.* **74**, 123-133 (2006).

<sup>19</sup>D. R. Spiegel and S. Tuli, "Transient diffraction grating measurements of molecular diffusion in the undergraduate laboratory," *Am. J. Phys.* **79**, 747-751 (2011).

<sup>20</sup>T. Pang, "Diffusion Monte Carlo: A powerful tool for studying quantum many-body systems," *Am. J. Phys.* **82**, 980-988 (2014).

<sup>21</sup>"NYC population: Current and projected populations," <[www1.nyc.gov](http://www1.nyc.gov)>. Accessed on May 2, 2019.

<sup>22</sup>Open Data Network Demographics Data, <[https://www.opendatane트워크.com/entity/1600000US1714000/Chicago\\_IL/geographic.population.density?year=2018](https://www.opendatane트워크.com/entity/1600000US1714000/Chicago_IL/geographic.population.density?year=2018)>.

<sup>23</sup>G. Rossetti, L. Milli, S. Rinzivillo, A. Sirbu, D. Pedreschi, and F. Giannotti, "NDlib: A Python library to model and analyze diffusion processes over complex networks," *Int. J. Data Sci. Anal.* **5**, 61-79 (2018).

<sup>24</sup>"Towards data science: Modeling coronavirus with Python," <<https://towardsdatascience.com/modelling-the-coronavirus-epidemic-spreading-in-a-city-with-python-babd14d82fa2>>.

<sup>25</sup>The code can be downloaded from <[http://physics.neu.edu/~acioli/research/viraldiffusion/Viral\\_Diffusion.py](http://physics.neu.edu/~acioli/research/viraldiffusion/Viral_Diffusion.py)>. Instructors and students using the code should cite the present manuscript in any presentation or publication that uses this code as downloaded or modified.