Implementing Cellular Automata in Epidemiology *

Arav Bhardwaj
NCSSM Online
North Carolina School of Science and Mathematics
Durham, North Carolina

December 7, 2022

Abstract: Epidemiology is the study and analysis of the disease spread within a given population. Without proper measures, epidemics can easily lead to pandemic situations, like the COVID-19 pandemic that shook our world. In this paper, I demonstrate how cellular automata can be used to simulate the spread of epidemics and how they can be more informative than traditional models that rely on partial-differential equations. More informative models can lead to better evacuation and safety measures, which can help to prevent future pandemics. It is important to note that the models examined in this paper are not based on real-world data; they are wholly theoretical and rely on contrived parameters for proof of concept.

Key words: SIR model, compartmental model, cellular automata, partial-differential equations, Moore neighborhoods

Introduction

The study of epidemiology is the cornerstone of public health. Over the years, epidemiologists have developed methodology that have been used in clinical research, public health studies, and more. The study of epidemiology daties back to ancient times, to the father of medicine, Hippocrates. Mathematical methods were first introduced to the field in the early 20th century, by epidemiologists such as William O. Kermack and A.G. Kendrick, who developed the first SIR model for modeling disease spread. Since then, technology and computation has played a vital role in the study of epidemics. Epidemiologists can now collect, study, and analyze data to better understand how diseases spread. This has led to more effective preventative measures, which has significantly decreased the risk of outbreak for many diseases. Until date, epidemiology models have traditionally been restricted to the realm of mathematical equations, specifically partial-differential equations. However, modeling epidemic spread requires careful consideration of not only the temporal but spacial aspect of dis-

*Correspondence to: bhardwaj24a@ncssm.edu

ease spread, which is where traditional methods struggle. In this paper, I propose an alternative type of model, one based on cellular automata, first formulated by John Von Neumann in the 1950s.[1] Cellular automata are most easily defined as a set of individual units that have simple rules which determine their behavior. One of the simplest rules, Rule 22, is depicted below (Figure 1). In 2002, Stephen Wolfram published "A New Kind of Science," where he extensively studied cellular automata and their behaviors. He explained two important principles: the principle of computational equivalence, and the principle of computational irreducibilty [2]. The first asserts that nearly all processes can be considered computations of equivalent sophistication, and the second that many computational systems in the universe cannot be accurately predicted ahead of time with mathematics – the only approach is to run a computation to simulate the behavior of the system.

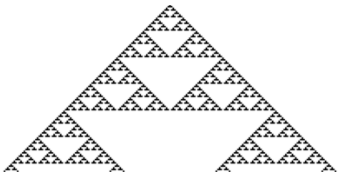


Figure 1: Rule 22

Many scientists were skeptical about the paradigm shift that Wolfram's book called for, since important scientific models were traditionally built on mathematical equations. However, there has been a trend where many dissipative models – such as urban growth and ecosystem models – have being converted to cellular automata-based models. Cellular automata thrive the in theoretical studies, where they are easier to understand and can work as well as mathematical methods with less consideration. This paper focuses on the application of cellular automata in the field of epidemiology, to prove that existing mathematical methods in the field can not only be matched but outperformed by cellular automata-based models.

Computational Approach

The SIR model framework is a classic for disease spread, and will serve as the basis for the models explored in this paper. It divides a fixed a population of individuals into three groups as a function of time:

- ***S(t)***, corresponding to the number of susceptible individuals,
- ***I(t)***, corresponding to the number of infected individuals, and
- ***R(t)***, corresponding to the number of recovered individuals.

The SIR model is known as a compartmental model, as it each individual of the given population falls into one of the three categories defines above. [3] The SIR model works as follows:

1. Susceptibles encounter an infected individual and may become infected
2. Infected individuals recover after a certain time period and are moved to the Recovered group
3. Recovered individuals are assumed to be immune to the disease and cannot become susceptible again

The illustration below describes this methodology (Figure 2).

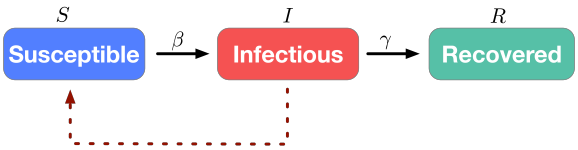


Figure 2: Illustration of SIR model

The SIR model additionally requires two parameters: the contact rate of the disease – the infection rate – and the recovery rate. [6] The most basic form of the SIR model – and the one that will be used in this paper – assumes that the population count remains fixed. Therefore, demographic factors such as births, deaths, and immigration and emigration that change the population size are not considered.

The first model will be a traditional implementation of the SIR model, and will serve as the baseline for which we can examine other models. It is founded upon the following group of partial-differential equations, where N represents the total population, β represents the contact rate, and γ represents the recovery rate (measured in 1/days). 1

$$\begin{aligned} \frac{dS}{dt} &= \frac{-\beta SI}{N}, \\ \frac{dI}{dt} &= \frac{\beta SI}{N} - \gamma I, \\ \frac{dR}{dt} &= \gamma I \end{aligned} \tag{1}$$

The second model is a probabilistic cellular automata model. Specifically, cellular automata are discrete computational systems consisting of a predefined two-dimensional space made up of rows and columns. Each individual unit in this grid is called a "cell," which will be be considered to represent a person for the purposes of this paper. For example, a 100×100 grid would contain 10,000 cells, or a population of 10,000 individuals. In accordance with the SIR methodology defined above, each cell in the model can have one of three possible states: susceptible, infected, or recovered. The state of each cell in the model is modified according to a rule applied to the cells in its Moore neighborhood, which is designated by the cell itself and the eight cells directly surrounding it in the north, south, east, west, northeast, southeast, northwest, and southwest directions. The modification rule in this case corresponds to probabilities for the susceptibility, contact rate, and recovery rate of a particular individual. The characteristics of the cellular automata model are more explicitly defined below:

- The population remains constant, i.e. there are no deaths, no births, etc. Therefore, ***S(t)*** + ***I(t)*** + ***S(t)*** is always equal to ***N***.
- A susceptible individual may become in-

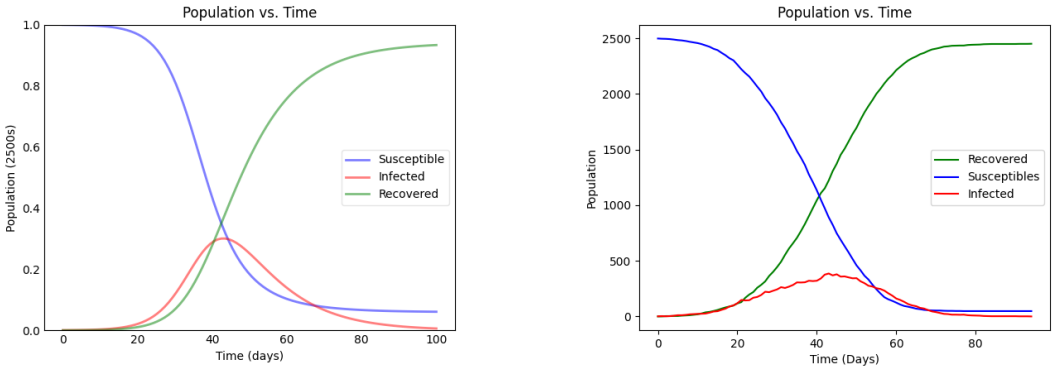
fected only if it comes into direct contact with a infected individual. In other words, infections are only possible within a cell's Moore neighborhood.

- Once individuals have recovered from the disease, they acquire immunity status and cannot be infected again for the remainder of the simulation.

For the sake of comparison, each model will be tested with a population size of 2500 individuals – represented with a 50×50 grid for the cellular automata model – for a time period of 100 days. The code for the models is written completely in Python, using the SciPy, NumPy, and PyGame libraries. The code for the partial-differential model is adapted from the official SciPy website [4], and the cellular automata model from divyeshransariya's repository on GitHub [5].

Results and Discussion

The plot below displays the results of running both types of SIR model with a contact rate of 0.3 and recovery rate of 1/10 days (Figure 3). The general trend for the plot show is that the number of susceptible individuals generally decreases; the number of infected individuals increases, reaches a peak, and then decreases; and that the number of recovered individuals generally increases over time. These plots are very standard for SIR models, but it is significant that the cellular automata model was able to mirror the traditional partial-differential model almost exactly. Indeed, the peak of the infections occurs around the 42 day mark for both models.



(a) Partial Differential SIR Model Plot

(b) Cellular Automata SIR Model Plot

Figure 3: SIR plots side-by-side

That both the cellular automata and mathematical SIR model have very similar predictions proves that cellular automata are a valid method of simulating epidemics compared to the partial-differential model. However, the biggest advantage of using the cellular automata model over the partial-differential model is that it allows us to visually observe the spread of disease with

respect to time in the simulations. Below, The table of four images shows the state of the cellular automata model at t = 0, 25, 50, and 75. It should be noted that the susceptible individuals are represented by the white pixels, the infected individuals by the orange pixels, and the recovered individuals by the dark-gray pixels (Figure 4).

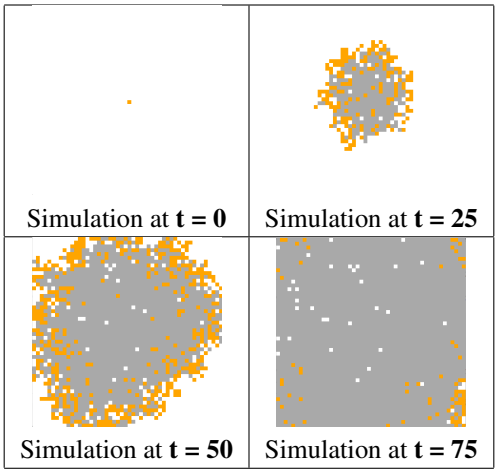


Figure 4: Disease spread at different stages

In the simulation above, the infection is initialized with a single individual at the center of the grid. All the other cells are considered susceptible. As time progresses in the simulation, the infection spreads outwards in a circular fashion. With such images, we can easily imagine how a disease would spread in the real world. Another benefit of using cellular automata is that we can easily modify parameters to achieve complexity in the model. For example, another simulation can be run that randomly selects an individual from the grid instead of initializing the infection from the center. The table of images below shows this simulation, where the infection was placed randomly in the grid (Figure 6). It is important to note that there is a degree of randomness regarding the spread of the disease since the cellular automata rules are based on probabilities.

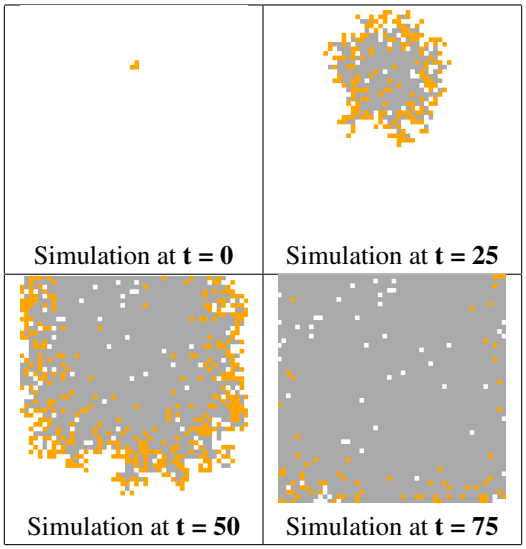


Figure 5: Disease spread with one randomly placed individual

Since the number of infected individuals stays the same in this simulation, it looks quite similar to the first version, which is to be expected. A SIR plot for the simulation looks nearly identical to the previous shown, with the peak of the infection occuring at nearly the exact same time. In the final iteration of the model, I disperse the infection at multiple points randomly selected on the grid to yield a more interesting – and complex – simulation of disease spread. With just a simple modification, the cellular automata model becomes even more complex and even more realistic. The table of images below depicts this iteration (Figure 6).

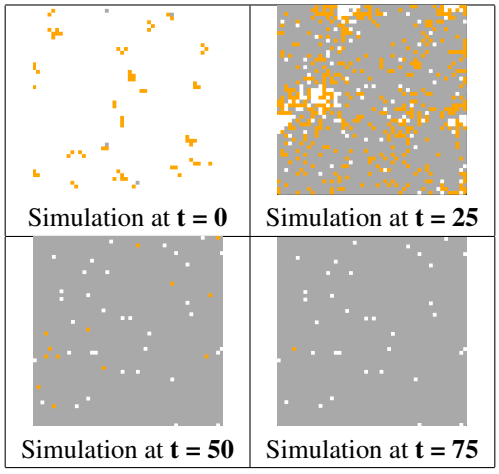


Figure 6: Disease spread with multiple randomly placed individuals

For example, the simulation above could very easily represent the common cold as it spreads in a city or town. Since the infection was initialized at multiple points in the grid, the infection spreads much quicker than the previous models. The SIR plot for the simulation reflects this, with the peak of the infection occurring much earlier in the simulation, around the 17 day mark (Figure 7). The curves for the three different groups are also shifted closer to the left. In the real world, this version of the model could model the disease spread in a city where multiple infected individuals are dispersed among the population.

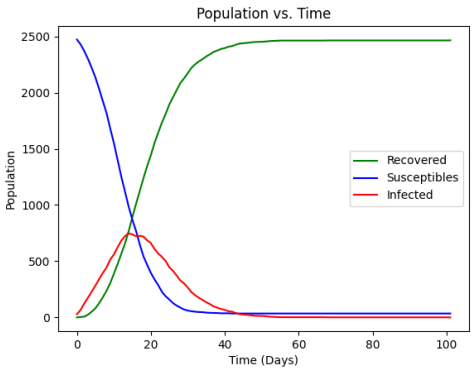


Figure 7: SIR plot for the above simulation

Conclusion

This paper proves that cellular automata are a more viable alternative to traditional partial-differential equation models, primarily because they are better at simulating the spacial and temporal development of epidemics. Furthermore, I demonstrate how cellular automata can scale very well to create more complex models. By increasing the number of cells and/or introducing additional local rules, cellular automata models can become more realistic while still maintaining a high level of interpretability.

Epidemiologists can easily expand on the baseline cellular automata models I have examined in this paper to better predict disease spread in a variety of conditions, which can aid public health officials in formulating more comprehensive plans. Further exploration with cellular automata modeling will lead to more robust, accurate disease spread predictions, which will lead to a lower risk of disease outbreak by bolstering preventative and evacuation measures.

Acknowledgements

The author thanks Mr. Robert Gotwals for assistance with this work. Appreciation is also extended to Jacob Borodovsky from Dartmouth University for introducing me to cellular automata.

References

[1] Mitchell, Melanie. ”Life and Evolution in Computers.” Accessed December 3rd, 2022.

[2] Wolfram, Stephen. ”A New Kind of Science.” `https://www.wolframscience.com/nks/`. Accessed December 3rd, 2022.

[3] Stockton University. ”Mathematical Models of Diseases.” `https://stockton.edu/sciences-math/ezine/fall2021/mathematical-models.html`. Accessed December 3rd, 2022.

[4] SciPy. "The SIR epidemic model." https://scipython.com/book/chapter-8-scipy/additional-examples/the-sir-epidemic-model/. Accessed December 5th, 2022.

[5] divyeshransariya. "sir_model." https://github.com/divyeshransariya/etc/tree/main/sir_model. Accessed December 3rd, 2022.

[6] UCLA University. "Learning Epidemic Models for COVID-19." https://covid19.uclaml.org/model.html. Accessed December 4th, 2022.

[7] S. Hoya White, A. Martín del Rey, G. Rodríguez Sanchez. "Modeling epidemics using cellular automata". https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7127728/. Accessed December 3rd, 2022.