Covid-19 Spread and Evolution

Anas Gauba
Dept. of Computer Science
University of New Mexico
Albuquerque, New Mexico
mgauba@unm.edu

Siri Khalsa
Dept. of Computer Science
University of New Mexico
Albuquerque, New Mexico
skhalsa10@unm.edu

***Abstract*—The existence of viruses has been present as long as life has existed. Understanding the past and future evolution of viruses within its neutral network is a crucial property of a virus. The neutral network allows a virus to be robust, and explore a broader mutational landscape to find a high fitness to survive. In this paper, we studied the evolution and spread of the recently ongoing COVID-19 virus. We studied how variants of viruses attach to humans and mutate. We determine that there is a tiny chance that COVID-19 could evolve into the more deadly SARS variant. The virus will usually mutate and stay within its neutral network, as most mutations are deleterious. Due to the high mutation rate, there is always the possibility a variant will survive and adapt to its circumstances. This brought us to study how a variant of the disease can spread in the population living in a quarantined city. We modeled this scenario by using 2-dimensional Cellular Automata with SIR dynamics using deterministic rules and non-deterministic rules. We also added a second variant of the disease to our toy model and evolved the second variant to coexist with the first variant. We achieved this with the help of a Genetic Algorithm. The results show that modeling epidemic spread with non-deterministic rules gave a realistic look at an epidemic spread. The results also suggest that two disease variants can coexist equivalently in the population. We found out that understanding the evolution of viruses and epidemic spread at the microscopic level can help in understanding at a macroscopic level.**

I. Introduction

The world is currently experiencing an epidemic never before encountered. While it is true that this is not the first time we have experienced a viral pandemic, it has affected society in a way never before written about in history. Across the world, countries are closing schools, businesses, and airports in response to help slow down the spread of this disease. What is at fault for these causal responses in society? The novel coronavirus, COVID-19, and it has declared war.

Viruses have evolved with humans for as long as humans have existed. While we have discovered and learned a lot about them over the years, we still have a lot to learn. We can explore viruses as complex adaptive systems and hopefully gain insights into how they evolve. This paper will explore the COVID-19 virus from this perspective and attempt to understand how this virus evolved and how it has spread.

The ideas behind complex adaptive systems are not simple. There are a lot of different concepts that make up these ideas. This introduction will briefly express some of them in an attempt to ease into the rest of the paper. One of these concepts is **top-down vs. bottom-up causality**. What does this mean though? Often **top-down vs. bottom-up causality** can be subjective to perspective and the definition of the system analyzed. **Bottom-up causality** is the idea that the actions of the entities at the lower level cause behavior at higher levels. [6] For example, A person infected with COVID-19 may not know they are and help transmit the virus. At a higher level, hospitals and governments are starting to recognize something dangerous is spreading. **Top-down causality** usually occurs when an entity emerges from the bottom-up action. The World Health Organization announced on March 11, 2020, [4] that COVID-19 emerged as a pandemic. There is now a **top-down causality**. At a lower level, people, that were once careless with their actions, are now practicing social distancing and self-quarantining.

How exactly did viruses come about? What are they? The goal of this paper is not to teach the details of biology, but some basic ideas could not hurt. The genetic code may be the key to life. It is also a critical aspect of viruses. The cells in our body encapsulate a string of DNA. This DNA contains four different chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). DNA bases pair up with each other, A with T and C with G, to form units called base [15]. These bases get grouped up into sets of three. This set is called a codon, and it is used as instructions to make proteins. This idea is simple and explains the genetic code. Viruses also can contain DNA or RNA, but the concept is the same. Just like humans, they have a genetic code that defines them. One key feature of human cells is that they contain the machinery to replicate themselves. This machinery is what separates them from viruses. Viruses usually need a host to replicate. The origin of genetic code is thought to be a **frozen accident**. The genetic code just happened by **accident**. It may not be the most efficient system, but it has proven functional. It is said to be **frozen** because it happened to change what evolved afterward dramatically. Any change to this system would be extremely disadvantageous [3].

What is the idea of complexity and chaos, and why does it matter when we think about viruses? The **logistic map** is often cited as an example of how complex, chaotic behavior can arise from very simple non-linear systems [25]. It aims to model the population size, given a growth rate. Fascinating behavior can arise based on the growth rate r. When the growth rate r is between 0 and 1, the population dies. When r is between 1 and 3, the population converges to a single value.

When r is between 3 and 3.44949, the population oscillates between two values. As r increases from here, the population oscillates between more and more values until it oscillates between infinite values and is said to be chaotic [25]. Viruses are said to live on the edge of chaos. This edge of chaos may be what gives it the high mutation rate characteristic. This is critical for a virus as it allows it to be extremely adaptable to its environment [14] [5]. There are many different **measures of complexity**. Viruses also have many, but the mutation rate is one way to **measure their complexity**.

The high mutation rate of a virus also allows it to explore its **neutral network**. This neutral network is a landscape of point mutations where genetic variants have the same fitness. It is easy to see how a high mutation rate allows a virus to explore this landscape rather quickly. This ability to explore a neutral landscape leads to robustness and evolvability [14] [22]. This evolvability is critical for a virus to survive. The hosts that they infect are continually fighting back, and this leads to an arms race between host and parasite [14]. Part 1 of this paper explores COVID-19's **neutral network**.

Eventually, the exploration of this **neutral network** will lead to a novel virus, a virus that our bodies are ill-equipped to handle [7], a virus that can lead to **epidemic spread**. In the year 2019, COVID-19 was born. COVID-19 is a disease that, as of this writing, is currently rapidly spreading among large populations of people [14]. As mentioned above, it has been deemed a pandemic, which is an epidemic that has spread to multiple countries. Part 2 of this paper will explore epidemic spread through toy models.

II. Methods & Results

A. Part 1 - How can neutral networks help us to understand the past and potential future evolution of COVID-19?

There are many types of viruses. There are DNA viruses, RNA viruses, and Retro-transcribing viruses. Some are small, and some are big. [14] When they infect their host, they can do it in a multitude of ways. They have been evolving along with us for as long as life itself. Retro-transcribing viruses even have the capability of changing our DNA. In a sense, viruses helped create us. The SARS-CoV-2 virus is an RNA virus that causes COVID-19. DNA viruses, which have built-in error-correcting, are more stable than RNA viruses. Due to the lack of error-correction, RNA viruses mutate often. [28] This part will explore neutral networks to help us understand the past and potential future evolution of SARS-CoV-2.

To begin to understand neutral networks, we need to understand mutations of a strand of RNA. RNA contains a string of nucleotides. A nucleotide comprises one of the four different chemical bases: adenine (A), guanine (G), cytosine (C), and thymine (T). Three nucleotides are called a codon, and it is a code to 1 of 20 amino acids. How many possible codons are there? There are $4 * 4 * 4 = 64$. There are way more codons than amino acids. This imbalance means that there are redundancy and some codons map to the same amino acid. We can determine, on average, how many codons map to the same amino acids. We know that $\frac{64}{20} = 3.2$. We can simplify this to 3 and say that, on average, 3 codons map to the same amino acid. We can say that these 3 codons are neutral with each other. A mutation occurs when a nucleotide changes bases when replicated. It is considered a **neutral mutation** when, after an RNA mutates, all codons result in the same amino acids as the original. A **neutral network** consists of all RNA strands that would result in the same amino acids as the original. The RNA strands in this network connect via neutral mutations, and because the current codon accounts for 1 of the 3 codons that are neutral, it is only possible to mutate to 2 of the remaining.

We will now explore the scope of all possible RNA strands that can be created by mutations, and the fraction of these RNA strands that make up the neutral network for the original RNA strand. We will start by understanding how many total RNA strands exist. We will assume that SARS-CoV-2 has 30,000 nucleotides. First, let us determine how many RNA strands exist after k mutations away. The following equation can calculate this:

$$3^k \binom{n}{k} = \frac{3^k n!}{k!(n-k)!} \tag{1}$$

$k =$ number of mutations and $n =$ number of nucleotides. The total possible RNA strands are:

$$\sum_{k=0}^{n} \left(3^k \binom{n}{k} = \frac{3^k n!}{k!(n-k)!} \right) = 4^n \tag{2}$$

We can plug in $k = 1$ and $n = 30000$ in equation 1 to calculate the number of RNA strands one mutation away,

$$3^1 \binom{30000}{1} = \frac{3^1 * 30000!}{1!(30000-1)!} = 90000 \tag{3}$$

and the total possible RNA strands are 4^{30000}.

Now, we will determine how many strands are neutral out of the total. We will start by analyzing a smaller example. We will assume n = 3. We already know from above that this is one codon, and the original RNA would have about 3 neutral RNA Strands. This means that $\frac{\#\, Neutral}{Total} = \frac{3}{64}$ is the fraction of neutral RNA strands. Furthermore, we can mutate from 1 of them to only 2 others. Now let us assume n = 6. This length implies that the RNA strand is two codons. Each codon has, on average, 2 neutral mutations. The fraction of neutral RNA strands is $\frac{2^2}{64^2}$, which can be generalized as $(\frac{2}{64})^{n/3}$, assuming n is divisible by 3.

The fraction of the neutral network gets smaller as the RNA strand gets longer. The fraction of the neutral network for the SARS-CoV-2 is $(\frac{2}{64})^{10000}$. This neutral network is minuscule but is traversable by crawling the network from the original RNA strand. It gives the virus high robustness, which allows it to explore many non-neutral mutations than would be possible without the neutral network.

The scope of possible mutations is vast, and the neutral network is incredibly small in comparison. However, it is easy to traverse the neutral network by making single mutations as such mutations connect the network.

We will now focus on the spike protein. The spike protein is the molecule of a coronavirus that allows it to bind to a receptor on a host cell and then fusing the viral and host membrane [21]. It makes up a small subset of the 30000 RNA strand. We will explore spike proteins that are 15 nucleotides long and code for 5 amino acids. The total possible genomes that can be created with 15 nucleotides is $4^{15} = 1,073,741,824$. Figure 1 shows how the number of possible unique genomes increases as we mutate the original genome more and more times.

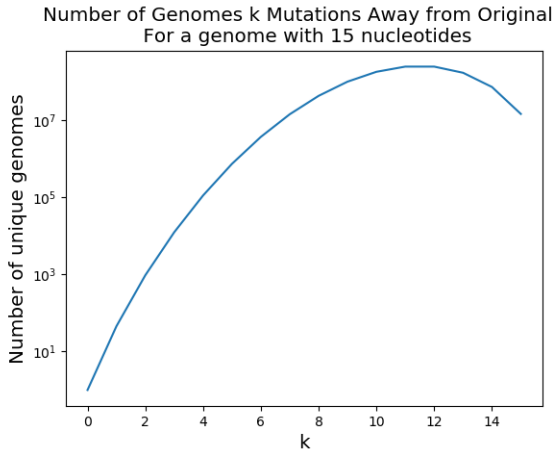

Fig. 1. The figure showcase the number genomes that exist k mutations away for a genome 15 nucleotides in length. This information is useful to understand the spike protein talked about in this section. The number of unique genomes is calculated using equation 1 in the text above.

Fig. 2. This chart expressed the genetic code and what amino acids a codon maps too. This chart was taken from the Project 2 assignment. A citation was not found there.

The following table contains the spike proteins for various coronaviruses. The spike proteins are similar except COVID-2019 vs. SARS-2002:

SARS-2002	R	Y	L	N	Y	T
Civet-2002	R	Y	L	K	Y	S
Bat-2013	R	S	F	N	Y	N
Covid-2019	N	L	F	Q	Q	N
position	1	2	3	4	5	6

Looking at this table and the genetic code chart in figure 2, we can determine the minimum amount of mutations needed for the COVID-2019 spike protein to become the more deadly SARS-2002 one. To do this, we will ignore position 1 and add up the total amount of mutations to get from the amino acid in one spike to the other. The following is what we would get:

$$\begin{aligned} L &\implies Y = 2\, mutations \\ F &\implies L = 1\, mutations \\ Q &\implies N = 2\, mutations \\ Q &\implies Y = 2\, mutations \\ N &\implies T = 1\, mutations \\ TOTAL &= 8\, mutations \end{aligned}$$

What is the chance that one of the above would happen, and that an amino acid could mutate into another amino acid? Let us assume that any of the 30000 nucleotides have an equally likely chance of mutating. Let us also assume that a base that is currently set as 1 of the 4 possible bases has an equally likely chance of mutating into one of the 3 remaining bases. These assumptions are not exactly true but help to simplify the problem. The probability that $F \implies L$ would be $\frac{1}{30000} * \frac{1}{3} = \frac{1}{90000}$. The probability for just this one is small, but the probability that one out of the five amino acids to change goes up slightly. If we assume that every amino acid takes two mutations, then the probability would be about $\frac{10}{30000}$ to pick any of the available nucleotides and then a $\frac{1}{30000}$ to mutate the second nucleotide for that appropriate amino acid. The final probability would be $\frac{1}{30000} * \frac{10}{30000} * \frac{1}{9}$. If we assume, based on Bedford's blog, that, on average, 1 of the 30000 nucleotides mutates upon transmission every 7 days, then it would take a minimum of 8 weeks for the COVID-19 spike to mutate into the SARS-2002 spike. This time would be a best-case scenario where every mutation leads to the new spike. The probability is tiny and would be the following:

$$\prod_{m=1}^{8} \frac{m}{90000} \tag{4}$$

We will now focus on the 15 nucleotides of the spike protein. This focus will allow us to use a computer to explore the neutral network. We will also make some new assumptions here to help us explore a path from the current COVID-19 spike protein to the SARS-2002. What is the likelihood that this will happen? If it did happen, how long would it take? These are some questions we will try to answer in the computer simulation. For the simulation to be successful, we need to make some assumptions about the neutral network. At the beginning of this section, the neutral network was defined as any mutation that does not lead to change in the amino

acids. Since we are analyzing the path from two different sets of amino acids, we need to make some new assumptions on an acceptable neutral network that is traversable via mutations.

To build this network, we made sure there was at least one possible path from the amino acids in each spike protein. Using the table above, we have assumed that each position could have a set of amino acids. Any permutation of these is considered part of the neutral network.

$$
\begin{aligned}
&Position\ 2\ : L\ ,\ Y\ ,\ F\ ,\ Q\ ,\ H\\
&Position\ 3\ : F\ ,\ L\\
&Position\ 4\ : Q\ ,\ N\ ,\ K\ ,\ H\\
&Position\ 5\ : Q\ ,\ Y\ ,\ H\ ,\ N\\
&Position\ 6\ : N\ ,\ S\ ,\ T
\end{aligned}
$$

These assumptions define our neutral network. The simulation that we run will take 100 COVID-19 spike proteins and randomly mutate each one. After each mutation, we will check to see if it is still in the neutral network defined above. If it is not, then the variant will die and get replaced with a new COVID-19 spike. Eventually, there will be a variant that mutates to SARS. The following is the data from the simulation that runs until 5 SARS variants have been discovered:

- Total SARS variants: 5
- Total dead variants: 73676819
- The chance a variant will become SARS: 0.000000068
- Max mutations needed to get to SARS: 12
- Min mutations needed to get to SARS: 10
- Average mutations of the dead are: 1.657
- The average mutation needed per SARS: 10.6

The simulation shows that it is improbable that a COVID-19 spike protein will mutate into the SARS version. If it does, on average, it will take about 10.6 weeks to mutate into the SARS variant.

Now, we will perform the same simulation with the spike protein **LLYYD**, BAT-RAT13 [27]. To analyze this one, we again need to make some assumptions about the neutral network. Much like was done above, we define the approved amino acids for each position. The neutral network contains every possible permutation for the positions.

$$
\begin{aligned}
&Position\ 2\ : L\\
&Position\ 3\ : F\ ,\ L\\
&Position\ 4\ : Q\ ,\ H\ ,\ Y\\
&Position\ 5\ : Q\ ,\ H\ ,\ L\ ,\ Y\\
&Position\ 6\ : N\ ,\ D
\end{aligned}
$$

The output of this simulation ran with the following results:

- Total SARS variants: 5
- Total dead variants: 2852679
- The chance a variant will become BAT-RAT13: 0.000001753
- Max mutations needed to get to BAT-RAT13: 7
- Min mutations needed to get to BAT-RAT13: 6
- Average mutations of the dead are: 1.503
- The average mutation needed per BAT-RAT13: 6.4

We can see that the COVID-19 spike has a way better chance to become BAT-RAT13. The chances are still not likely. We can see that, on average, this would only take about 6.4 mutations and about 6.4 weeks to get there. In this section, we have explored how small the neutral network is compared to the total possible network. Because the current genome is part of the neutral network, it can travel this network, even though it is small. It also means that it is much more likely for a mutation to lead to deleterious results. Mutations usually are not useful, but without mutations, the evolution of viruses will not happen. Viruses rely on mutations in their arms race to survive against their hosts. This mutation characteristic sometimes leads to a pandemic of epic proportions. A situation where this small little virus tells its host, "check. It is your move." The next part of this paper will explore epidemic spread using a toy model. What can we learn from these simple models? Will it help us determine what move we will make next?

B. Part 2 - What can we understand about the spread of Covid-19?

After discussing the history and future evolution of COVID-19 with SARS and other similar viruses that came before and looking at the neutral space of COVID-19, we now focus on how COVID-19 can spread among species, humans in particular. To understand the spread of COVID-19, we built a simple 3 state, 2-dimensional Cellular Automata (CA) with the SIR dynamics to model this epidemic spread. For our 2d CA, we chose our board size to be 50x50. Modeling the epidemic spread with the help of CA can help model a scenario of an already quarantined city or a population on a cruise ship where this epidemic is spreading. Due to this assumption, modeling the epidemic spread with CA makes sense here as the cells on 2d CA do not move, and the edges and corners have few neighbors. In our 2d CA toy model, each center cell represents a human being with eight surrounding cells representing other human beings in the population as neighbors. We also include the center cell when discussing neighbors for a total of 9 neighbors. This radius keeps the neighborhood size of a cell to one since a cell can only interact with its closest neighbors in our model. This neighborhood function is called a Moore neighborhood [27].

In order to explore epidemic spread, determining which rules to use are very important, and understanding SIR dynamics also becomes essential. In the study of epidemic spread, SIR dynamics is crucial because it defines the state of the cell in every possible situation. A cell can either be Susceptible

(S), Infected (I), or Recovered (R) [24]. The neighboring cells determine the next state. This constraint gives us the ability to build a rule table that accounts for how a cell in state S can become I, and how a cell in the state I can become R. When a cell recovers, it is considered immune to the disease.

Keeping SIR dynamics and neighborhood function in mind, let us understand how many possible rules there are in a rule table. Since our toy model encodes a scenario of a quarantined city, there are three possible cases here to determine rules: normal rule entries, edge rule entries, and corner rule entries. In order to understand each of these rule entries for our toy model, let us first understand some basic principles: A cell follows SIR dynamics, so it has 3 states. The neighborhood function tells us that a cell has 9 neighbors, including itself. Keeping these principles in mind, understanding normal rule entries is trivial, as its just permutation of SIR in 9 places since there are 9 neighbors, So, normal rule entries = 3^9. The edge and corner rule entries are special rule entries since the cells on the edge or corner have fewer neighbors. If we talk about the cells on the edges or corners, they do not have all 9 neighbors in their neighborhood. To be precise, the cells on the edge have 6 neighbors in bounds, and the cells on the corner have 4 neighbors in bounds, and there are 4 edges and corners of 2d CA. Thus, edge rule entries = $4 * 3^6$ and corner rule entries = $4 * 3^4$. In total, there are 3^9 + $4 * 3^6$ + $4 * 3^4$ rule entries in the rule table.

Now that we know how many rules there are, it is time to determine how we can simulate the epidemic spread by transitioning cell states using SIR dynamics. In order to study the epidemic spread, we came up with several rules to simulate the epidemic spread. These include both deterministic and non-deterministic rules. We first build a deterministic rule mapping to simulate the epidemic spread. The deterministic rule mapping for a given cell on the board that we chose is as follows:

- If a cell is in state S, it goes to state I only if there exists at least one neighbor who is in state I. Otherwise, it will stay in state S.
- If a cell is in state I, it goes to state R only if all of its neighbors are in state I. Otherwise, it will stay in state I,
- If a cell is in state R, it always remains in state R.

Using these deterministic rules for simulating the epidemic spread, we initialized all cells in our CA to state S except for one cell in the middle to be in state I. As the disease spreads, some cells become infected while some cells recover. Figure 3 shows how the disease spreads using deterministic rules that we build above:

The figure 3 shows that the disease is spreading in a spiral shape. As the disease spreads from the center of the population, the innermost cells are getting recovered as all of their neighbors become infected, and outer cells are getting infected as soon as one or more of its neighbors become infected (see Appendix A for seeing additional figures showing how the disease spreads using deterministic rules). The thing noticeable here is that the disease spreads rapidly, which is how an epidemic spreads. After just 25 iterations, we can see

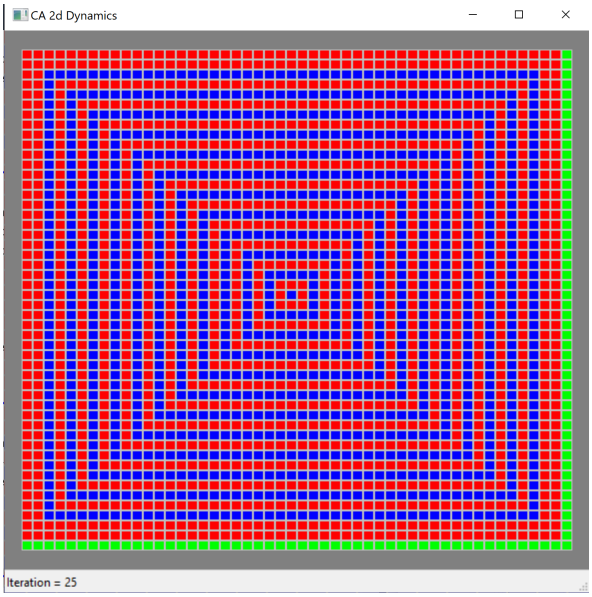

Fig. 3. The figure shows the epidemic spread after 25 iterations of CA using deterministic rules. After 25 iterations, most cells are either infected (red) or recovered (blue) and some cells are still susceptible (green).

that all cells except some on the edge become infected, and some recovered and become immune to the disease. However, the results we see from the deterministic rules are not realistic. They are not realistic because every time we simulate this epidemic spread, we get the same results showing only one way of how an epidemic spreads. In the real world, it is hard to predict how an epidemic spreads because the spreading of an infectious disease is not entirely deterministic. For a real epidemic spread, there are more factors involved, such as analyzing how many people will get infected? How many will recover, and how many will die from the epidemic spread [26]? What is the likelihood other humans surround a human, and how many of them are infected? To answer this, we need to define non-deterministic rules based on a probability of how likely a cell can get infected due to infected neighbors.

In order to model an epidemic spread that behaves non-deterministically, we come up with non-deterministic rules. These rules attach a probability to whether a cell becomes infected. Therefore, we revised our rule mapping from being deterministic to non-deterministic. The non-deterministic rule mapping for a given cell on the board is as follows:

- If a cell is in state S, the probability of being infected is based on the infected neighbors.
- If a cell is in state I, the probability of being recovered is fixed to 10%.
- If a cell is in state R, there is 0% probability that it will change. Therefore, it remains R.

For simulating the epidemic spread using non-deterministic rules, we kept the same initial configuration for the CA board that we used in the deterministic version. The only thing we changed was to use the non-deterministic rules to showcase how a disease spreads. Figure 4 shows the epidemic

spread using non-deterministic rules. The simulation shows that the disease spreads slowly initially. Then the number of cells infected increases sharply, creating, what seems to be, a realistic disease spread (see Appendix B for seeing additional figures showing the epidemic spread using non-deterministic rules). The disease spreads gradually, and after a while, the whole population gets affected by the disease, as can be seen in figure 4. We can see that using non-deterministic rules creates a more organic looking scenario of an epidemic spread because a cell follows probabilistic factors.

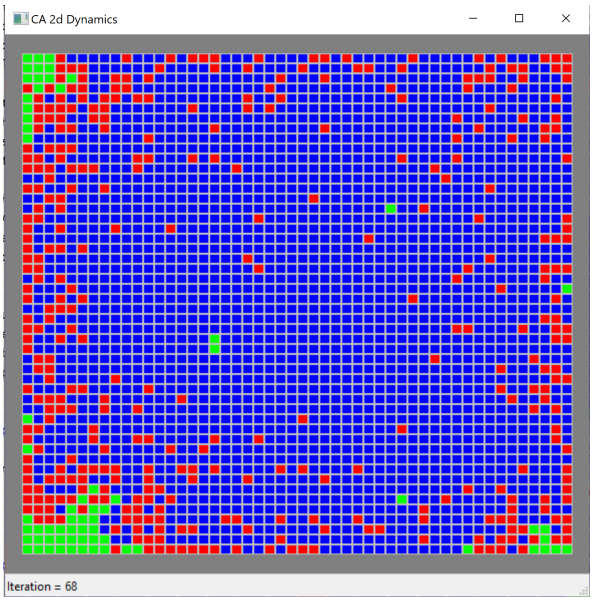

Fig. 4. The figure shows the epidemic spread after 68 iterations. Almost all cells have become infected and a lot of the cells have become immune to the disease and have recovered.

After simulating epidemic spread using both deterministic and non-deterministic rules, we continue our exploration of how an epidemic spreads. So far, our model has only been simulating the epidemic spread using a single disease variant of COVID-19. We have seen how an epidemic spreads if there is a single variant. However, in order to make our exploration interesting, we introduce a second variant of the disease to our toy model. This second disease variant uses the same SIR dynamics as the first variant used and is non-deterministic. For simplicity's sake, instead of denoting infected and recovered cells as I and R respectively, we denote the cells in the second disease variant as I' and R'. For the second variant, we introduce two additional probabilities for a given cell in state S and state I'. To make it consistent with the first variant, once the cell is in state R', it stays R' as the cell has recovered from this second disease variant. To include both variants in our CA, we also initialized the cells in our CA randomly, where all cells are in state S except for two random cells, which are I and I' respectively.

Can both disease variants coexist equally in a population? To answer this, we initialize our two additional probabilities for the second variant to be values chosen at uniform random between [0,1]. To make both disease variants coexist equally, we ran a Genetic Algorithm (GA) against our toy model to evolve both probabilities of transitioning from state S to I' and state I' to R'. The goal of the GA here is to make sure that the probabilities of the second disease variant evolve to coexist with the first disease variant. We achieve the coexisting of two diseases in a population by adding two additional probabilities in each of the 100 populations of CA in our GA. We then let our GA perform multiple runs, where a run consists of iterating over each CA's board until there are no infected cells, I or I' left for both variants. We let GA perform runs until the second disease variant evolves and coexist equally with the first variant, meaning the recovery rate for both disease variants is approximately equal. Both R and R' are relatively close to each other. The closer the recovering rate is for both diseases, the lower and better fitness value we get for the CA. This gives us the ability to simulate both diseases in our 2d CA toy model.

Once the GA gives us the evolved probabilities for the second variant, we simulate the epidemic spread using both variants. Figure 5 shows how both diseases coexist equally and how both diseases spread in the population. The simulation shows both variants of the disease start spreading in the population at approximately an equal rate and coexist together [see Appendix C for seeing additional figures]. Hence, the results in figure 5 satisfies the question we brought up earlier and shows us that two disease variants can coexist equally in a population.

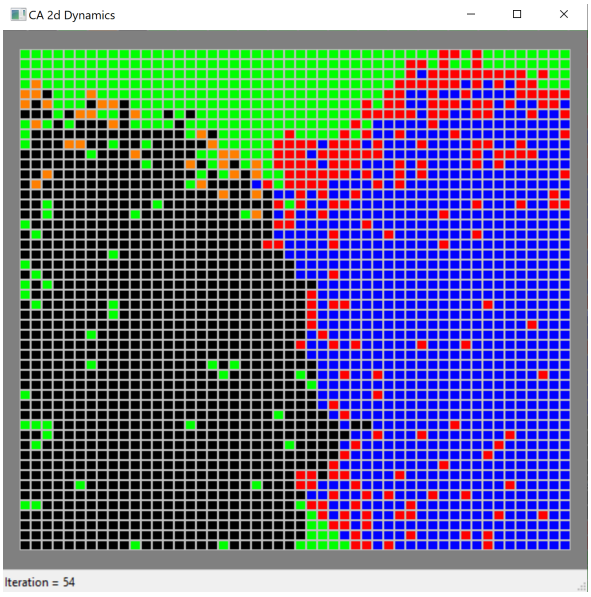

Fig. 5. The figure shows spreading of two variants of the disease among the population. After 54 iterations, both disease variants coexist equally. The cells in I' and R' state can be seen in orange and black respectively and cells in I and R state can be seen in red and blue respectively. The remaining cells are in state S (green).

With the help of using 2d CA toy models with SIR dynamics, we explored epidemic spread by using different rules. The first model used deterministic rules and expressed a non-

organic outcome. This model did not seem to effectively model how an epidemic might spread. We improved this model by utilizing probabilities in our second model, which used non-deterministic rules. This change ultimately leads to a more natural spread. We further explored the epidemic spread by introducing a second variant that competes with the first one. To determine if they could coexist equivalently, we used a GA to determine the rules for the second variant. It turns out both variants could coexist and infect the population equivalently. This exploration does not accurately model a real population. It creates more questions. Does self-isolation help? How does hygiene change the infection rate? Ultimately these models begin the process of understanding a complex phenomenon and should be explored further.

III. Discussion & Conclusions

This paper has explored just a small piece of a very complex world involving viruses and the current epidemic that is spreading across the world. The paper has focused on attempting to understand **neutral networks** and **epidemic spreads**. We explored it at an elementary level. Many questions are generated just from this basic exploration. The neutral network is an interesting property of viruses. It is a tiny part of the entire possible network, but it exists close together. This allows it to be traversed. How can a **neutral networks** help us map out the movement of a virus over time and geographical locations? Can it help us understand the origin of viruses? More importantly, can it tell us what all current viruses will become in the future? If we understand how viruses evolve, we may be able to stop them proactively instead of reacting to them.

Understanding epidemic spread is critical if we must react to viruses. They can help us react more intelligently to reduce the negative impact of an epidemic. We explored epidemic spread at a microscopic level and assumed a situation of a population existing in an already quarantined city. Although that is not the case everywhere, and the population tends to move, whether it be through travel or being in a mass gathering at certain events. However, understanding the epidemic spread at a microscopic level can help us understand epidemic spread at a macroscopic level. With the help of this tiny model we built for studying epidemic spread, we can see, even if a population is in a quarantined city and living closely together, the chances of epidemic spread are still higher.

References

[1] K. G. Andersen, A. Rambaut, W. I. Lipkin, E. C. Holmes, and R. F. Garry, "The proximal origin of SARS-CoV-2," Nature Medicine, 2020.

[2] Asia's SCARIEST Meat Market! Dog, Cat, Rat, Bat and more at Tomohon Market in North Sulawesi. Best Ever Food Review Show, 2018.

[3] F. H. Ckrick, "The origin of the Genetic Code," The Origin of the Genetic Code. [Online]. Available: https://www.bibliotecapleyades.net/ciencia/ciencia_adn09.htm. [Accessed: 11-Apr-2020].

[4] "Coronavirus Disease (COVID-19) - events as they happen," World Health Organization. [Online]. Available: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/events-as-they-happen. [Accessed: 10-Apr-2020].

[5] Edge of Chaos. Systems Innovation, 2016.

[6] G. F. R. Ellis, "Top-down causation and emergence: some comments on mechanisms," Interface Focus, vol. 2, no. 1, pp. 126–140, 2011.

[7] Explained, Season 2, Episode The Next Pandemic. Netflix, 2018.

[8] E. C. Holmes, The evolution and emergence of RNA viruses. Oxford: Oxford University Press, 2011.

[9] E. Koonin, "Frozen Accident Pushing 50: Stereochemistry, Expansion, and Chance in the Evolution of the Genetic Code," Life, vol. 7, no. 2, p. 22, 2017.

[10] D. Leister, "Thawing out frozen metabolic accidents," BMC Biology, vol. 17, no. 1, 2019.

[11] D. M. Morens and J. K. Taubenberger, "Influenza Cataclysm, 1918," New England Journal of Medicine, vol. 379, no. 24, pp. 2285–2287, 2018.

[12] D. M. Morens, P. Daszak, and J. K. Taubenberger, "Escaping Pandora's Box — Another Novel Coronavirus," New England Journal of Medicine, vol. 382, no. 14, pp. 1293–1295, Feb. 2020.

[13] D. Söll and U. L. Rajbhandary, "The genetic code - Thawing the 'frozen accident,'" Journal of Biosciences, vol. 31, no. 4, pp. 459–463, 2006.

[14] V. Solé Ricard and S. F. Elena, Viruses as complex adaptive systems. Princeton, NJ: Princeton University Press, 2019.

[15] "What is DNA? - Genetics Home Reference - NIH," U.S. National Library of Medicine, 30-Mar-2020. [Online]. Available: https://ghr.nlm.nih.gov/primer/basics/dna. [Accessed: 10-Apr-2020].

[16] 0:01 / 8:36 The next outbreak? We're not ready — Bill Gates. TED, 2015.

[17] B. Alicea, "Neutral networks: a critical review," Synthetic Daisies, 01-Jan-1970. [Online]. Available: http://syntheticdaisies.blogspot.com/2011/11/neutral-networks-critical-review.html. [Accessed: 10-Apr-2020].

[18] A. Frank, "How Does The World Work: Top-Down or Bottom-Up?," NPR, 29-Sep-2013. [Online]. Available: https://www.npr.org/sections/13.7/2013/09/29/225359504/how-does-the-world-work-top-down-or-bottom-up. [Accessed: 10-Apr-2020].

[19] How pandemics spread. TED-Ed, 2012.

[20] How wildlife trade is linked to coronavirus. Vox, 2020.

[21] F. Li, "Structure, Function, and Evolution of Coronavirus Spike Proteins," Annual Review of Virology, vol. 3, no. 1, pp. 237–261, 2016.

[22] "Neutral network (evolution)," Wikipedia, 06-Apr-2020. [Online]. Available: https://en.wikipedia.org/wiki/Neutral_network_(evolution). [Accessed: 10-Apr-2020].

[23] K. M. Peck and A. S. Lauring, "Complexities of Viral Mutation Rates," Journal of Virology, vol. 92, no. 14, Feb. 2018.

[24] "The SIR Model for Spread of Disease - The Differential Equation Model," The SIR Model for Spread of Disease - The Differential Equation Model — Mathematical Association of America. [Online]. Available: https://www.maa.org/press/periodicals/loci/joma/the-sir-model-for-spread-of-disease-the-differential-equation-model. [Accessed: 11-Apr-2020].

[25] "Logistic map," Wikipedia, 21-Mar-2020. [Online]. Available: https://en.wikipedia.org/wiki/Logistic_map. [Accessed: 11-Apr-2020].

[26] "Modelers Struggle to Predict the Future of the COVID-19 Pandemic," The Scientist Magazine®. [Online]. Available: https://www.the-scientist.com/news-opinion/modelers-struggle-to-predict-the-future-of-the-covid-19-pandemic-67261. [Accessed: 11-Apr-2020].

[27] "Moore neighborhood," Wikipedia, 16-Sep-2019. [Online]. Available: https://en.wikipedia.org/wiki/Moore_neighborhood. [Accessed: 11-Apr-2020].

[28] "Viruses and Evolution," History of Vaccines. [Online]. Available: https://www.historyofvaccines.org/content/articles/viruses-and-evolution. [Accessed: 11-Apr-2020].

Contributions

Siri Khalsa wrote the code for Part 1. He also wrote the methods and results for Part 1, introduction, and parts of the references of this paper. Anas Gauba wrote the code for Part 2 and wrote the associated methods and results paragraphs for Part 2. He also wrote the abstract and added appendix section. Both authors collaborated to write discussion and conclusions section of this paper.

APPENDIX A
ADDITIONAL FIGURES SHOWING THE SIMULATION OF CA USING DETERMINISTIC RULES

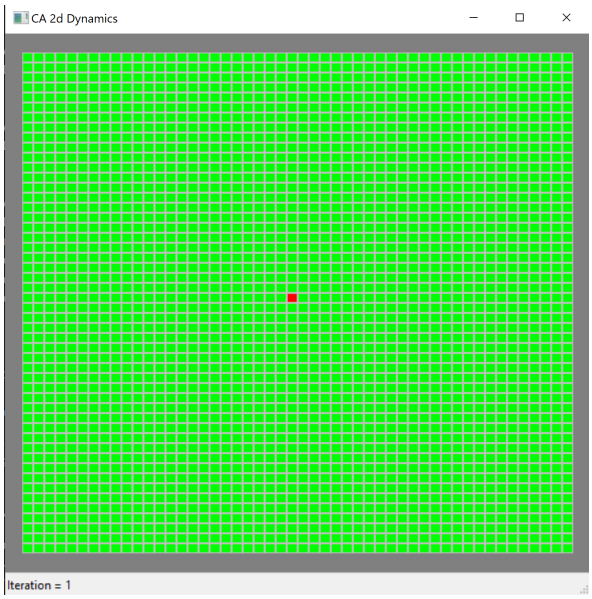

Fig. 6. The initial configuration of CA using deterministic rules. The cells in green are Susceptible and the cell in red is Infected.

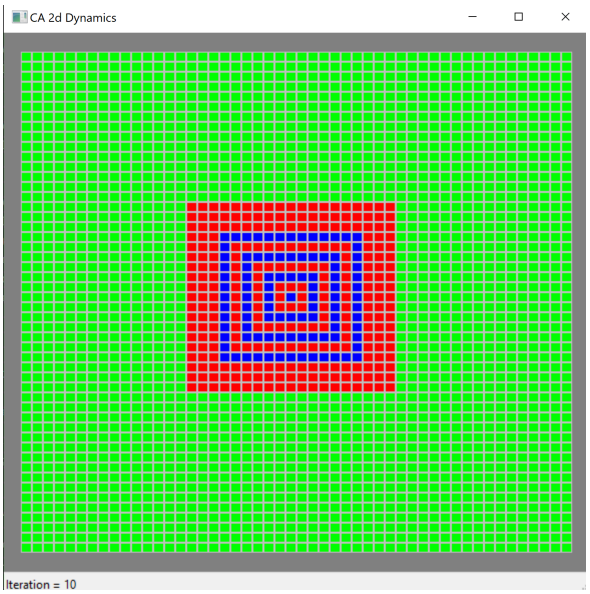

Fig. 7. The figure shows the epidemic spread after 10 iterations of CA using deterministic rules. The recovered cells are shown in blue.

APPENDIX B
ADDITIONAL FIGURES SHOWING THE SIMULATION OF CA USING NON-DETERMINISTIC RULES

The figures 8 and 9 show a sharp increase in the epidemic spread after an initial slow start. As the disease spreads, the probability of a cell in state S becoming infected increases as more neighbors get infected.

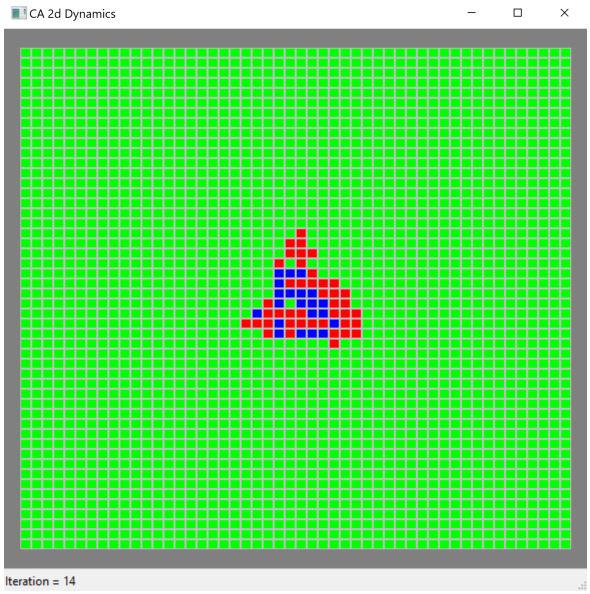

Fig. 8. The figure shows the disease spreading gradually after 14 iterations. Using non-deterministic rules, the figure shows some cells recovering as the disease spreads.

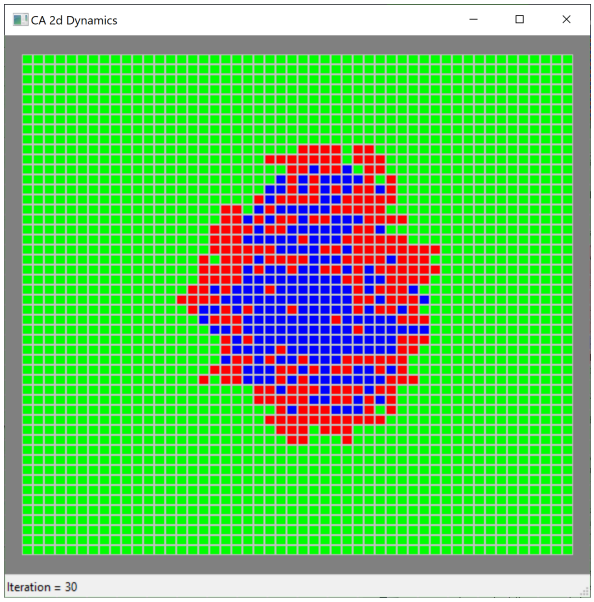

Fig. 9. The figure shows the epidemic spread after 30 iterations. The amount of cells infected is 3x more than it was after 14 iterations.

APPENDIX C
ADDITIONAL FIGURES SHOWING THE SIMULATION OF CA USING TWO VARIANTS

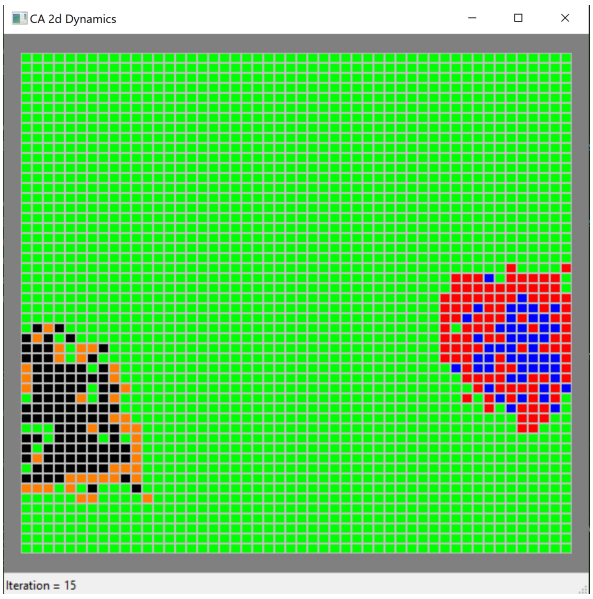

Fig. 10. The figure shows spreading of two variants of the disease after 15 iterations.

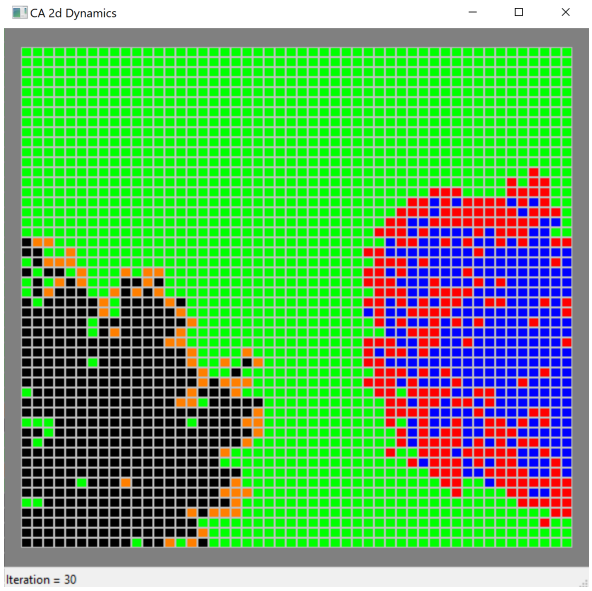

Fig. 11. The figure shows spreading of two variants of the disease after 30 iterations. Both variants are trying to infect the population equivalently.

PROJECT 3 PROPOSAL

Project title: Asymptomatic Screening to Reduce Epidemic Spread

Research question and hypothesis: Research Question: Can a more effective screening process that includes asymptomatic carriers decrease epidemic spread? Can this eliminate the need for a lockdown? Hypothesis: We think there could be a more effective screening process that can allow the economy to continue as usual, or at least reduce the need to close everything besides essential businesses. We think that, if there were no shortages on tests, then there could exist a scenario where everyone gets tested weekly. They will quarantine if sick or continue life as usual. All businesses can stay open. We will explore this hypothesis.

Motivation: Project 2 was mainly an introduction to epidemic spread modeling. The SIR models are fascinating, but they leave a lot to be desired. We want to effectively model the movement of people and how this affects the simulation. We feel that asymptomatic people could be a factor as to why this epidemic spread so forcefully.

There is an article called Aerosol and *Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1*. The goal of this paper is to see if COVID-19 stays around longer on surfaces compared to the original SARS. The paper found that they are similar. It determined that there must be a different factor that determines why COVID-19 is more contagious. One of their thoughts is due to asymptomatic patients.

There is another article called *How can an asymptomatic person still spread the coronavirus?*. This article explains how the immune system works and how it is possible that even though a person does not show symptoms, they can still be infectious. They are walking around entering places as if they are healthy. People react as if they are healthy. As soon as there is a sneeze or a cough to clear the throat, the virus spreads.

The last paper that we thought was interesting is *Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19*. This article found that the current screening procedure misses more than half of the infected screened. The article estimates that when a person could show symptoms based on the exposure time and how a traveling screening program will miss %50 to %75 of the infected individuals. They determined that a symptom and risk screening programs.

Proposed research: A big portion of this paper will be about the current screening process. What does this involve? How is it done? One of the papers above is an excellent start to answering this. We could also analyze historical screening processes. How were past viral outbreaks addressed? We will need to understand the current state to introduce the paper and why we need to try another approach. This research could take about a week or two.

Next, we need to start collecting as much data as possible to build an agent-based model on the epidemic spread of COVID-19. The data that we would be collecting is information like incubation period, percentage of asymptomatic carriers, how infections are asymptomatic carriers compared to symptomatic

ones, how long does the virus survive on a surface. These are all needed to help build a model. This research could take anywhere from 1 week to 2 weeks.

A majority of the time will be needed to code up an agent based model. There may be research involved in learning agent based modeling and how to best model a scenario. Most likely, the time will be spent on learning yet another new programming language. This will take about 2 weeks giving time for debugging.