A Markov Decision Process Approach to Control of Networks and Analysis of North Carolina COVID-19 Spread by a New Compartmental Model

By

Thomas Whitehead

Senior Honors Thesis

Department of Statistics and Operations Research

Univeristy of North Carolina at Chapel Hill

18 October 2020

Approved:

Serhan Ziya, Thesis Advisor

Nilay Argon, Reader

Contents

List of Figures

List of Tables

1. Abstract

Networks play an important role in modelling the interactions of people, and how this can lead to the spread of ideas, influence, products, and disease. For my first semester of research, I focus on modelling social networks as Finite Horizon Markov Decision Processes, so that I can find the optimal set of activations using Backwards Iteration. In particular, I focus on Linear Threshold social networks and Independent Cascade social networks. For the second semester of research, I focus on working with a compartmental model created by Housni et al., to model the spread of Covid-19 in North Carolina [3]. I make three major adjustments to the model: including a parameter that accounts for the possibility of False Negative tests, and eliminating the effects of contact tracing. The former adjustment does not result in any significant changes in outcome, while the latter results in a highly unrealistic predictions.

2. Introduction

In this thesis, I start by giving background information on Discrete Time Markov Chains and Finite Horizon Markov Decision Processes (Finite Horizon MDPs). Then, I describe the typical optimization problem for Finite Horizon MDPs, and how this problem can be solved using Backwards Iteration. I give the Parking Problem as an example of how this process works. Afterwards, I give background information on both Linear Threshold and Independent Cascade social networks. Then, I describe how they can be modelled as Finite Horizon MDPs, and how this set-up can be used to find the optimal set of activations Switching focus, I describe the compartmental model developed by Housni et al. to predict the spread of Covid-19 [3]. Then, I suggest three modifications, in addition to adding new training data. First, I suggest that based on recent events, we should adjust the dates when the social distancing parameters are modified. Second, I suggest that the compartmental model should be adjusted to account for the possiblility of False Negatives. Third, I suggest eliminating the effect of contact tracing from the model. I then give theresults of each of these modifications, and discuss their implications in the Conclusion.

3. Finite Horizon Markov Decision Processes

A. Background

i. Discrete Time Markov Chains

A stochastic process describes a system that evolves randomly over time. Assuming the system is observed at times $n = 0, 1, 2, ...$, then the sequence of random variables $\{X_n | n \geq 0\}$ is known as a discrete time stochastic process [5]. The set of values S that X_n can take is referred to as the state space [5].

A Discrete Time Markov Chain (DTMC) is a discrete time stochastic process on a finite state space S, with the following property [5]:

$$P(X_{n+1} = j | X_n = i, X_{n-1}, ..., X_0) = P(X_{n+1} = j | X_n = i), \ \forall i, j \in S, \ \forall n \geq 0 \tag{1}$$

This means that all that matters in determining the state of the system at time $n + 1$ is the state of the system at time n. A DTMC is time-homogenous if the probability of transitioning between any two states is not dependent on the time n, so [5]:

$$P(X_{n+1} = j | X_n = i) = P(X_1 = j | X_0 = i), \ \forall i, j \in S, \ \forall n \geq 0 \tag{2}$$

ii. Finite Horizon Markov Decision Processes

A Finite Horizon Markov Decision Process with a discrete state space S is a DTMC that is only allowed to occur for a finite number of periods N, known as the time horizon [2]. Also, at each point in time, an action a must be taken, where $a \in A(s)$, the set of admissable actions for state s [2]. As with the state space S , we assume $A(s)$ is discrete. Finite Horizon MDPs have the Markov property in that the state of the system at

time $n+1$ only depends on the state of the system and the action taken at time n [2]. We also assume Finite Horizon MDPs to be time homogenous, where $\forall s, j \in S,\ a \in A(s)$, the probability of transitioning from state s to state j given action a is taken is ${p_{sj}}^a$ [2]. The effective probability is ${q_{sj}}^a = \alpha {q_{sj}}^a$, where $\alpha \in (0, 1]$, and $1 - \alpha$ represents the probability that the process "stops" and yields no returns [2]. The immediate return for taking action a while in state s is $r(s, a)$ [2].

A decision rule δ is a function defined on S that gives the set of appropriate actions for each state, so $\delta(s) \in A(s)$ [2]. Also, $\delta \in \Delta$, the set of all admissible decision rules [2]. A strategy $\pi = (\delta_1, \delta_2, ..., \delta_N)$ gives a sequence of decision rules to use for each period [2]. We define the strategy space $\Pi = \Delta X \Delta X ... \Delta$, as the set of all admissible strategies [2].

iii. Optimization

In a typical optimization problem for Finite Horizon MDPs, you want to find a strategy $\pi \in \Pi$ that maximizes $v_t(\pi, s)$, the expected present values of returns from times $t, t+1, ..., N$, where:

$$v_t(\pi, s) = E\left\{ \sum_{i=t}^{N} \alpha^{i-t} r(s_i, \delta_i(s_i)) \right\} \tag{3}$$

,s_i is the random state of the system at time i, and δ_i is the decision rule prescribed by strategy π for period i [2].

A strategy π^* is optimal for state s in period t, if [2]:

$$v_t(\pi^*, s) \geq v_t(\pi, s),\ \forall \pi \in \Pi \tag{4}$$

A strategy is optimal for period t if it is optimal for state s in period t for all $s \in S$ [2].

If you let,

$$f_t(s) := \max_{\forall \pi \in \Pi} v_t(\pi, s) \tag{5}$$

then we want to find π^* such that $v_t(\pi^*, s) = f_t(s),\ \forall s \in S$ [2].

In the case of a discrete state space, we can write:

$$f_t(s) = \max_{a \in A(s)} r(s, a) + \sum_{j \in S} q_{sj}^a f_{t+1}(j) \tag{6}$$

which are known as the optimality equations [2].

According to the optimality conditions, π^* can be considered the optimal strategy if and only if $\pi_t^*(s)$ attains the maximum in the optimality equations for all t and s [2].

iv. Backwards Iteration

To find the optimal strategy π^*, you start by finding the optimal decision rule δ_N for the one-horizon problem starting at time N. Assuming $f_{N+1}(s) = 0$, the optimality equation for this problem reduces to:

$$f_t(s) = \max_{a \in A(s)} r(s, a) = \max_{a \in A(s)} \sum_{j \in S} q_{sj}^a r_{sj}^a \tag{7}$$

where r_{sj}^a is the return of taking action a in state s and ending up in state j [2]. After finding the optimal solution for the one-period problem starting at period $t = N$, you then find the optimal solution for the

two-period problem starting at period $t = N - 1$. To do this, you find the decision rule δ_{N-1} that satisfies the optimality equation:

$$f_{N-1}(s) = \max_{a \in A(s)} r(s,a) + \sum_{j \in S} q_{sj}^a f_N(s), \tag{8}$$

for all $s \in S$ where $f_N(s)$ can be defined using δ_N. Thus, the optimal strategy for the two-period problem starting at period $t = N - 1$ is (δ_{N-1}, δ_N). In general, for $T \in \{t, t+1, ...N\}$, δ_T can be defined is the decision rule that satisfies the optimality equation:

$$f_T(s) = \max_{a \in A(s)} r(s,a) + \sum_{j \in S} q_{sj}^a f_{T+1}(s) \tag{9}$$

for all $s \in S$, where $f_{T+1}(s)$ can be calculated using δ_{T+1}. By continuing to iterate backwards, you can find the optimal strategy $\pi^* = (\delta_t, .., \delta_N)$ [2].

B. Example: Parking Problem

As an illustration of how Backwards Iteration can be used to solve optimization problems for Finite Horizon MDPs, I consider the Parking Problem. In the Parking Problem, you are driving to work and start to pass parking spots, each of which are open with probability p [2]. At each spot, if that spot is open, you have the choice of either parking in that spot, or continuing on to find a better spot. The cost of parking x spots away from work is x [2]. If you reach work without having found a spot, you have to park in the parking deck, which has cost $c > 1$ [2]. The goal is to find a strategy that minimizes the expected cost of parking.

In order to make this problem a Finite Horizon MDP, you define the state $s = (x, i)$, where $x \geq 0$ is the number of the of the space you are approaching, and $i = 0$ if the space is available and $i = 1$ if the space is not available [2]. The possible actions a are 0 = park or 1= move on [2]. If the space is open, you can take both of these actions, so $A_{(x,0)} = \{0, 1\}$; if the space is not open, then you must move on, so $A_{(x,1)} = \{1\}$ [2].

If $f(s) = f(x, i)$ gives the minimum expected cost of parking starting at space x with availability i, then $f(0, i) = c$ [2]. For $x = 1$ [2]:

$$f(1,i) = \begin{cases} min(1,c) = 1 & \text{if } i = 0 \\ c & \text{if } i = 1 \end{cases}$$

If $i = 0$, then spot 1 is available, so you have the choice of parking there, incurring cost 1, or moving on to the parking lot, incurring cost c. Since $c > 1$, then $f(1, 0) = 1$. If $i = 1$, then spot 1 is occupied, so you must park in the parking lot, incurring cost $f(1, 1) = c$. You can define:

$$F(x) = pf(x,0) + qf(x,1) \tag{10}$$

where $q := 1 - p$ [2]. The function F gives the minimum expected cost of parking when you are approaching space x, but don't know whether or not it is occupied [2]. You can define f using F, where [2]:

$$f(x,i) = \begin{cases} min(x, F(x-1)) & \text{if } i = 0 \\ F(x-1) & \text{if } i = 1 \end{cases}$$

If $i = 0$, then space x is open, so you have the choice of either parking there, incurring cost x, or moving on to the next spot, incurring a minimum expected cost of $F(x - 1)$. If $i = 1$, then space x is not open, so you must move on to the next space, incurring a minimum expected cost $F(x - 1)$. Using the above equation, you can redefine $F(x)$ in terms of $F(x - 1)$ as [2]:

$$F(x) = p \min(x, F(x-1)) + qF(x-1) \tag{11}$$

In order to determine the optimal strategy for choosing a parking spot, you define [2]:

$$g(x) = F(x-1) - x \tag{12}$$

Using the above equation, you can re-write f as [2]:

$$f(x,i) = \begin{cases} x + \min(0, g(x)) & \text{if } i = 0 \\ x + g(x) & \text{if } i = 1 \end{cases}$$

If $g(x) \geq 0$, then $F(x-1) \geq x$, so the minimum expected cost of moving on to spot $x-1$ is greater than the cost at parking at spot x. Thus, if space x open, then you should park in it, incurring cost x. Conversely, if $g(x) < 0$, then the minimum expected cost of moving on to spot $x-1$ is less than the cost at parking at x. Thus, even if space x open, you should move on to the next spot, incurring a minimum expected cost of $F(x-1)$.

Using $g(1) \geq 0$, $g(x) < 0 \ \forall x \geq c$, and that g is a strictly decreasing function of x, then there exists $S \geq 1$ such that $g(S) \geq 0$ and $g(S+1) < 0$ [2]. Thus, $\forall x > S$, $g(x) < 0$, so the optimal decision would be to move on from spot x even it is open. By similar reasoning, $\forall x \leq S$, $g(S) \geq 0$, so the optimal decision would be to park in these spots if possible. Thus, the optimal strategy in the parking problem is to ignore all spots $x > S$ where $g(x) < 0$, and park in the first available spot x where $g(x) \geq 0$ [2]. Using $F(0) = c$, and the recursive formula for $F(x)$ in Equation 11, you can calculate $\max_{\{g(S) \geq 0\}} S$, the spot where a person should first start looking to park. The value of S for various values of p and C can be seen in Table 1 below.

$p \vert c$	5	10	100
0.25	3	5	12
0.5	2	3	6
0.75	2	2	4

Table 1: Solution S to Parking Problem for $p = 0.25, 0.5, 0.75$ and $c = 5, 10, 100$

As can be seen in Table 1, S is directly proportional to c, because as c increases, the cost of not finding a parking spot increases, so you should take a more risk averse strategy and start looking for spots earlier. Also, S is inversely proportional to p, because as p increases, the risk of not finding an open spot decreases, so you can start looking for spots closer to work.

C. Modelling Social Networks as Finite Horizon MDPs

i. Background: Social Networks

Social networks are defined as graphs that can be used to model the interactions of groups of individuals [1]. They have a wide variety of applications including modelling the spread of infectious disease or determining how a product should be marketed so as to maximize the number of people that purchase it. In my research, I focus on two fundamental types of social networks: Linear Threshhold and the Independent Cascade [1]. In both, you define a directed graph G, where each node is either active or not active [1]. Additionally, I only consider the progressive case, where nodes can be switched from being inactive to active, but not vice versa [1]. In both of these networks, a node v will start out inactive, but as more neighbors of v become active over time, v will eventually become active.

a. Linear Threshold

In a Linear Threshold social network, a node v is influenced by its neighbors w according to the weight $b_{v,w}$, where [1]:

$$\sum_{w \ neighbor \ of \ v} b_{v,w} \leq 1 \tag{13}$$

Each node v also has threshhold $\theta_v \sim \ Uniform[0,1]$ [1]. In the model, you start out with a set of initially active nodes A_0, and then iterate through the finite horizon via the following process [1]:

1. All nodes that are active at time $t-1$ remain active at time t

2. All nodes that are inactive at time $t-1$ only become active at t if:

$$\sum_{w \ active \ neighbor \ of \ v} b_{v,w} \geq \theta_v \tag{14}$$

b. Independent Cascade

As with the Linear Threshhold social network, in the Independent Cascade social network, you start with an initial set of activated nodes A_0 [1]. However, in this case, when a node v becomes activated at time t, it has exactly one chance to activate each of its neighbors w with probability $p_{v,w}$ [1]. If v succeeds, then w becomes activated at time $t+1$, but regardless of outcome, v cannot make any further attempts to activate w [1].

ii. Problem Set-Up

For both Linear Threshhold and Independent Cascade social networks, a typical optimization problem is to find the most "influential" nodes, i.e. the set of nodes A_0 such that the number of activated nodes at the end of the horizon is maximized [1]. Because of the probabilistic nature of node infection in both types of networks, you can solve this problem by considering social networks to be Finite Horizon MDPs, with some qualifications and modifications. First, I decide to make it possible for the user to activate a node at any period within the time horizon, so that I can define an action for any period within the time horizon. Also, I set a limit on the number of nodes the user can activate themselves; otherwise, the problem becomes trivial.

a. Problem Set-Up- Linear Threshhold

To start, I assume there are N nodes, I available activations, and a horizon H. Then, for $v, w \in \{1, 2, ..., N\}$, I define $b_{v,w}$ that satisfies Equation (13), and the random variables $\theta_v \sim Uniform(0, 1)$. The random nature of θ_v is what allows Linear Threshhold social networks to be modelled as Finite Horizon MDPs.

I define the state space $S \subseteq \mathbb{R}^{N+1}$, where:

$$s(i) = \begin{cases} 1 & \text{if node } i \text{ is active} \\ 0 & \text{if node } i \text{ is inactive} \end{cases}$$

$\forall i \in \{1, 2, ..., N\}$, and $s(N+1)$ is the number of available activations.

For each node , I define the action $a_i = 0$ to be to not activate node i, and $a_i = 1$, to be to activate node i. If node i is inactive, then $a_i \in \{0, 1\}$; if node i is active, then $a_i \in \{0\}$. Thus, for each period t, I define the action a as an N-vector such that $a(i) = a_i$, and:

$$\sum_{i=1}^{n} a_i \leq s(N+1) \tag{15}$$

namely, that the total number of nodes activated cannot exceed the number of activations available.

To define the transition probabilities, I consider the possible transitions for each node v:

1. If $s(v) = 1$ at $t-1$, then $s(v) = 1$ at t with probability 1.
2. If $s(v) = 0$ and $a(v) = 1$ at $t-1$, then $s(v) = 1$ at t with probability 1.
3. If $s(v) = 0$ and $a(v) = 0$ at $t-1$, then $s(v) = 1$ at t with probability:

$$P(\sum_{w \ neighbor \ of \ v, \ s(w)=1} b_{v,w} \geq \theta_v) = \sum_{w \ neighbor \ of \ v, \ s(w)=1} b_{v,w} \tag{16}$$

4. (Not node related) If $s(N+1) = l$ at $t-1$, then $s(N+1) = l - \sum_{i=1}^{n} a_i$ at time t with probability 1.

Using the above probabilities, you can calculate $p^a_{sj},\ \forall a \in A,\ \forall s, j \in S$. I define $r(s,a)$ as the expected number of newly activated nodes after taking action a while in state s, and:

$$r(s,a) = \sum_{i=1}^{N} a_i + \sum_{v,s(v)=0,a(v)=0} \sum_{w\ neighbor\ of\ v,s(w)=1} b_{v,w} \tag{17}$$

the number of nodes that are manually activated plus the expected number of inactive nodes that are activated by their neighbors. Using $\alpha = 1$, I can now define:

$$v_t(\pi, s) = E\left\{ \sum_{i=1}^{H} r(s_i, \delta_i(s_i)) \right\} \tag{18}$$

as the expected number of activations at the end of the horizon, using strategy π and starting state s. I want to find a strategy π^* that maximizes v_t for all initial states s. I can do this using Backwards Iteration as described above, working back from the one-horizon problem to the H-horizon problem.

b. Problem Set-Up Independent Cascade

For Independent Cascade social networks, I define N, I, and H the same as for Linear Threshhold social networks. I also define an NXN matrix P, where P_{ij} is the probability the node i infects node j. We define the state space $S \subseteq \mathbb{R}^{N+1}$:

$$s(i) = \begin{cases} 0 & \text{if node } i \text{ is inactive} \\ 1 & \text{if node } i \text{ is active and can activate other nodes} \\ 2 & \text{if node } i \text{ is active and cannot activate other nodes} \end{cases}$$

where $\forall i \in \{1, 2, ..N\}$ and $s(N+1)$ is the number of infections remaining. The action space is the same as that for the Linear Threshhold social networks, with the same restrictions. To define the transition probabilities, I consider the possible transitions for each node v:

1. If $s(v) = 2$ at time $t-1$, then $s(v) = 2$ at time t with probability 1.
2. If $s(v) = 1$ at time $t-1$, then $s(v) = 2$ at time t with probability 1.
3. If $s(v) = 0$ and $a(v) = 1$, then $s(v) = 2$ at time t with probability 1.
4. If $s(v) = 0$ and $a(v) = 0$, then $s(v) = 1$ with probability:

$$1 - \prod_{w\ neighbor\ of\ v,\ s(w)=1} 1 - P_{wv} \tag{19}$$

   and $s(v) = 0$ with probability

$$\prod_{w\ neighbor\ of\ v,\ s(w)=1} 1 - P_{wv} \tag{20}$$

5. (Not node related) If $s(N+1) = l$ at $t-1$, then $s(N+1) = l - \sum_{i=1}^{N} a_i$ at time t with probability 1.

Using these five statements, you can calculate $p^a_{sj},\ \forall a \in A,\ \forall s, j \in S$. If you define $r(s,a)$ the same as above, then you have

$$r(s,a) = \sum_{i=1}^{N} a_i + \sum_{v,\ s(v)=0,\ a(v)=0} 1 - \prod_{w\ neighbor\ of\ v,\ s(w)=1} 1 - p_{wv} \tag{21}$$

I can define v_t using Equation 18, with my new definition for $r(s, a)$ in Equation 21, and again solve the maximization problem using Backwards Iteration.

4. Covid-19 Compartmental Model

A. Background

In this model, the entire population of North Carolina ($N = 10,490,000$), is split up into 13 compartments. Nine of these compartments are created by classifying individuals through two dimensions. For the first dimension, individuals are classified as [3]:

1. Unknown Not Infected (N)
2. Unknown Infected (I)
3. Unknown Recovered (R)

For the second dimension, individuals are classified as [3]:

1. Symptomatic Isolated (Si)- individual shows symptoms of some illness, may be COVID-19, and self-isolates
2. Asymptomatic Isolated (Ai)- individual does not show symptoms of any illness, but self-isolates due to potential contact with infected individuals
3. Asymptomatic Nonisolated (An)- individual does not show symptoms of any illness, does not self-isolate

Thus, an example of a compartment is I_{An}, which includes individuals who are infected with Covid-19, but have not yet begun to show any symptoms, and are not self-isolating.

The other four compartments are [3]:

1. Death (D)
2. Hospitalization (H)
3. Known Infected (KI)- tests positive, but not yet hospitalized
4. Known Recovered (KR)- tests positive, but recovers.

Five of these compartments have sub-compartments [3]:

1. I_{Si}
   - $I_{Si}(recovered)$-will recover naturally, with no hospitalization
   - $I_{Si}(hospitalized)$- will require hospitalization
   - $I_{Si}(death)$- will die without access to Covid-19 diagnostic test
2. I_{Ai}
   - $I_{Ai}(recovered)$- will never develop Covid-19 symptoms
   - $I_{Ai}(show\ symptoms)$- will show symptoms, currently pre-symptomatic
3. I_{An}- same as I_{Ai}
4. H
   - $H(die)$- will eventually die
   - $H(recovered)$- will eventually recover
5. KI
   - $KI(hospitalized)$- will require hospitalization

- $KI(recovered)$ - will recover without visiting hospital

An explanation of the flow between compartments can be seen below in the Appendix (Section 6).

B. Modifications

The parameter β is defined as the contact rate between infected and non-infected individuals [3].

In the compartmental model, Housni et al. defines two values for β:

1. $\beta_h = \frac{R_0}{N * sympToRecoveryTime}$
2. $\beta_l = \frac{2}{3}\beta_h$

where R_0 is the number of new infections produced by each infected individual, $N = 10,490,000$ is the total population of North Carolina, and $sympToRecoveryTime$ is the amount of time it takes a symptomatic individual to recover from Covid-19 [3]. The parameter β_h gives the contact rate between infected and non-infected individuals, both non-isolating, while β_l gives the contact rate between infected and non-infected individuals, where at least one of them is self-isolating [3].

For this model, which starts with $t = 1$ being March 13, the β parameters are modified as follows [3]:

1. March 27- β_h and β_l are reduced by a factor of $\frac{1}{2}$ in correspondence with stay at home measures.
2. May 8- Phase 1 relaxes social distancing to 25 percent of stay at home levels
3. May 22- Phase 2 relaxes social distancing to 50 percent of stay at home levels

To update the model to the current date August 22, I keep social distancing at 25% of stay at home levels on May 22, and add a fourth date, September 11, when social distancing is relaxed to 50% of stay at home levels, in accordance with moving to Phase 3. Additionally, while the model is currently trained on data from only up until June 15, with the new model, I train on data from up until August 10.

Also, in the original compartmental model, it is assumed that the Covid-19 Diagnostic Tests have 100% accuracy, which ignores the possibility of false negatives. Studies indicate that the False Negative Rate (FNR) for Covid-19 diagnostic tests ranges from about 2% to 30% [4].This means that there is a segment of the population unaccounted for in the original model who are infected, but test negative, and thus maintain normal interactions. If this is the case, then the original model should be underestimating the number of infections, and thus the number of deaths. To correct for this, I include a parameter in the model for the Sensitivity ($Sens = 1 - FNR$) of the Covid-19 diagnostic tests, and adjust the update equations so that the proportion of postive tests for the infected compartments are:

1. I_{Si}-$\pi_{Si}Sens$
2. I_{An}-$\pi_{An}Sens$
3. I_{Ai}-$\pi_{Ai}Sens$

where π_{Si}, π_{An}, and π_{Ai} are the proportion of the Symptomatic Isolated, Asymptomatic Non-Isolated, and Asymptomatic Isolated populations respectively that are tested [3].

To examine the effects of contact tracing, I remove the three compartments I_{Ai}, R_{Ai}, N_{Ai}, so now individuals can only be Asymptomatic Non-Isolated (An) or Symptomatic Isolated (Si). Thus, indivuduals will now only isolate themselves if they develop Covid 19-like symptoms.

C. Results

i. Finding R_0 and InitialInf

The values for R_0 and $InitialInf$ (initial amount of individuals infected), are determined by minimizing:

$$\sum_{i=1}^{T} 5 * (newDeath[i] - historicAvgDeath[i])^2 + (H[i] - historicHosp[i])^2 \tag{22}$$

where T is the final day of training data, $NewDeath[i]$ is the predicted deaths on day i, $historicAvgDeath[i]$ is the actual number of deaths on day i, $H[i]$ is the number of predicted hospitalizations on day i, and $historicHosp[i]$ is the actual number of hospitalizations on day i [3]. Thus, we find the values of R_0 and $InitialInf$ that minimize the weighted sum of the square error loss in predicting the number of deaths and hospitalizations, with most of the weight placed on accurately predicting the number of deaths.

ii. New Training Data and Beta-Factor Update Dates

When adjusting for the new data up to August 10, and for keeping social distancing at 25% of stay at home levels until September 11, you have $R_0 = 3.15$ and $InitialInf = 547$. The resulting predictions can be seen in Figures 1 and 2, and the predicted cumulative totals for each compartment can be seen in Table 2. In Figure 1, the daily number of individuals in each compartment only increases gradually until September 11, when social distancing is relaxed to 50% of stay at home levels. At this point, you start seeing a particularly sharp exponential increase in both the Infected Asymptomatic and Known Infected compartments. There is also a noticeable, though less pronounced increased in the daily number of hospitalized and Infected Symptomatic individuals.

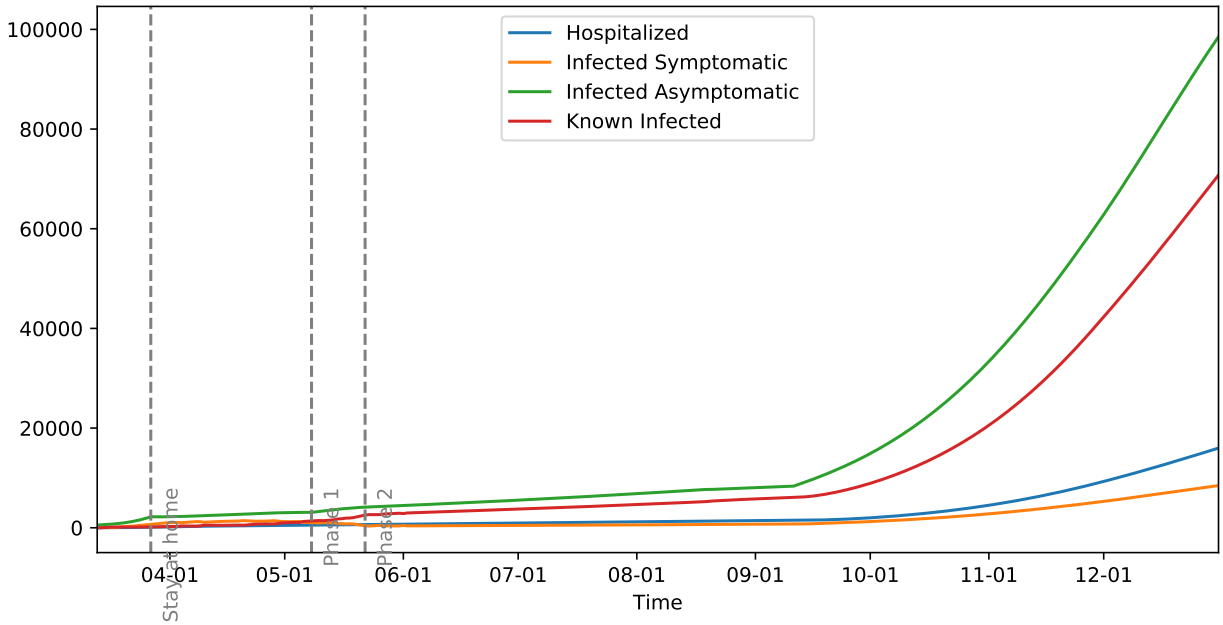

Figure 1: Daily Projections $R_0 = 3.15$ and $InitialInf = 547$

In Figure 2, the graphs of the predicted cumulative deaths versus actual cumulative deaths match well until May 22nd. Then, the predicted cumulative deaths continues to increase exponentially, while the actual cumulative deaths continues to increase in a linear manner.

iii. Sensitivity

When refitting the model using a sensitivity of 70%, I obtain an $R_0 = 3.07$ and $InitialInf = 547$. The results can be seen in Figures 3 and 4 below, and the predicted cumulative totals for each compartment can be seen in Table 3.

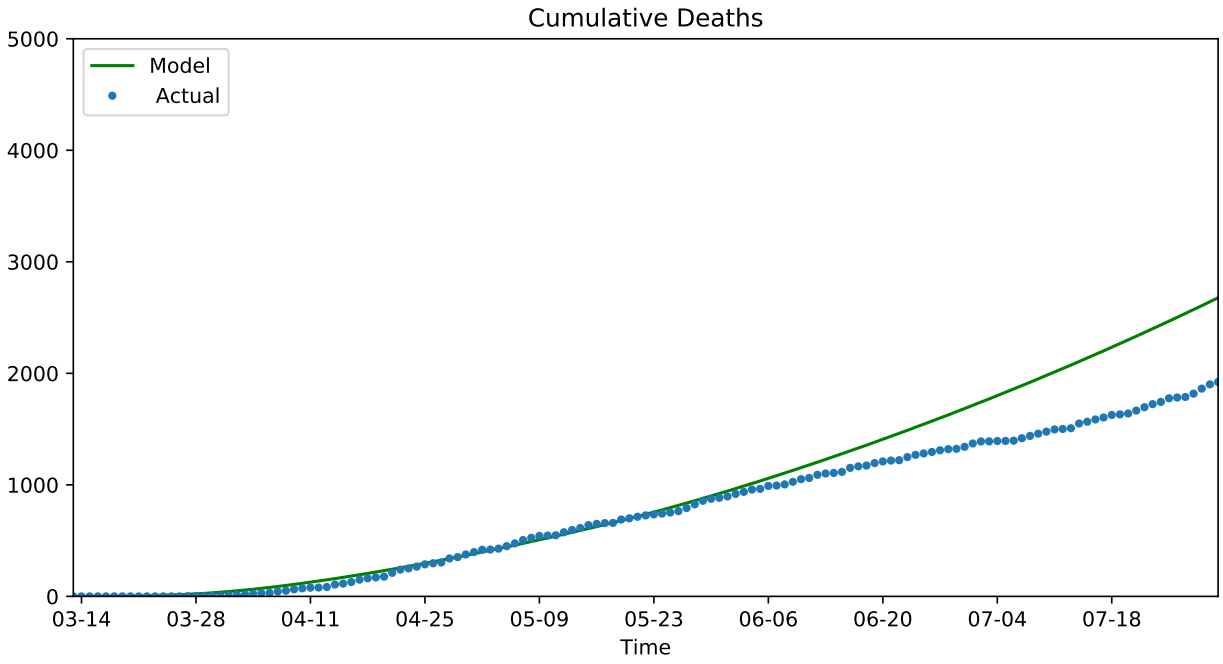

Figure 2: Predicted Cumulative Deaths vs. Actual Cumulative Deaths $R_0 = 3.15$ and $InitialInf = 547$

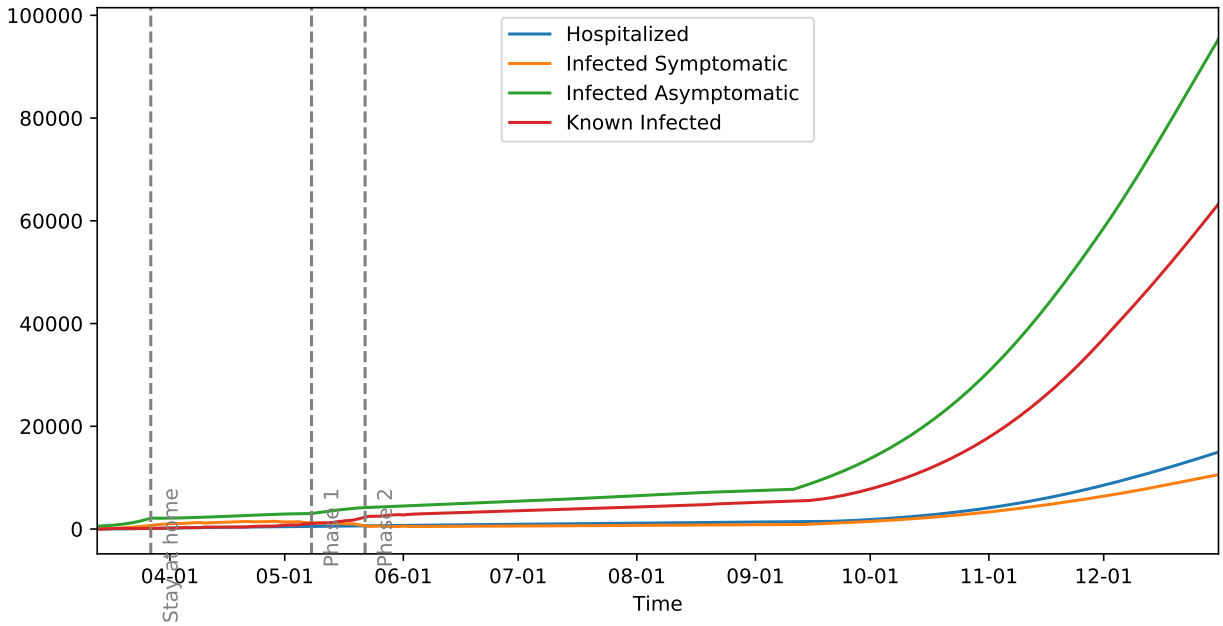

Figure 3: Daily Projections $R_0 = 3.07$ and $InitialInf = 547$, Sens=0.7

Compartment	Cumulative
Hospitalized	871296
Infected Symptomatic	541117
Infected Asymptomatic	5810419
Known Infected	3817062
Deaths	26815

Table 2: Cumulative predicted number of individuals in each compartment for entire year, $R_0 = 3.15$ and $InitialInf = 547$

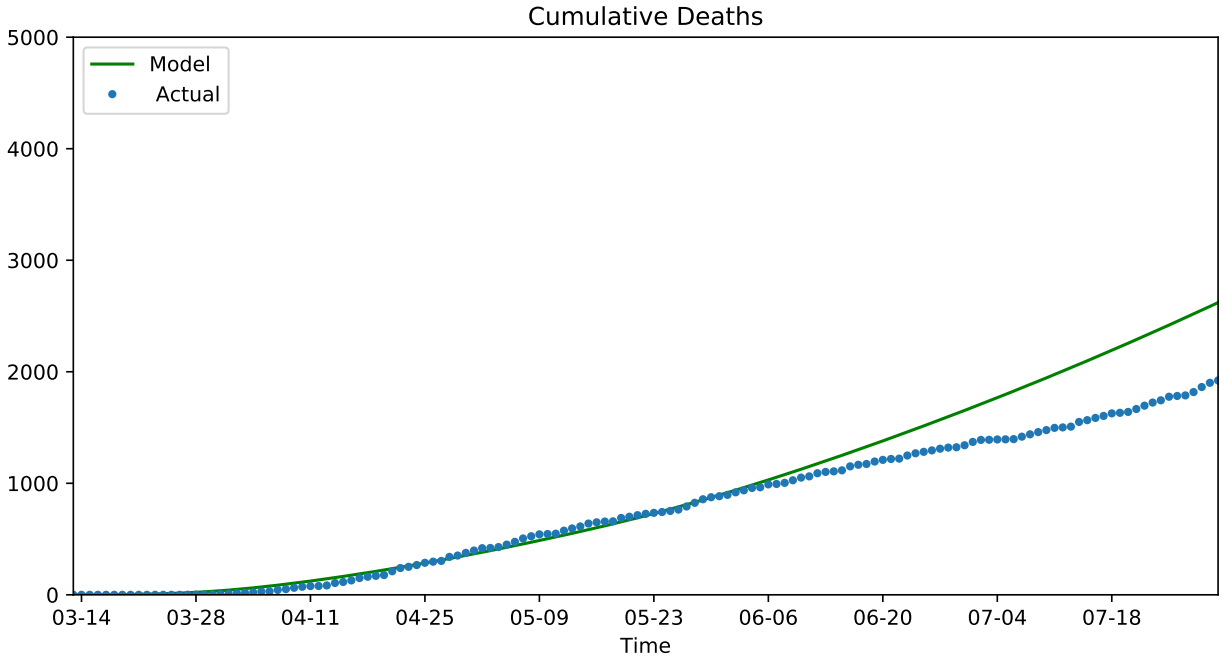

Figure 4: Predicted vs Actual Deaths $R_0 = 3.07$, $InitialInf = 547$, Sens=0.7

Compartment	Cumulative	Diff. from Sens.=1
Hospitalized	809286	-62010
Infected Symptomatic	650652	+109535
Infected Asymptomatic	5465766	-344653
Known Infected	3374302	-442760
Deaths	24955	-1860

Table 3: Cumulative predicted number of individuals in each compartment for entire year, $R_0 = 3.07$, $InitialInf = 547$, $Sens. = 0.7$

The only two compartments that show a significant difference when the sensitivity is decreased to 0.7 are the Infected Symptomatic compartments and the Known Infected compartments. The Infected Symptomatic compartments shows a 20% increase in the number of individuals, which makes sense intuitively, as individuals from this compartment are less likely to leave the compartment for testing positive. The Known Infected compartments shows a 12% decrease, which makes sense, as infected individuals are less likely to test positive. However, the total number of people infected asymptomatic, hospitalized, or dead all slightly decrease, which does not make sense, as one would expect all of these values to increase if we were to allow more infected people to interact normally with the population

iv. No Contact Tracing

After eliminating the Asymptomatic Isolated compartments and all contact tracing parameters, I refit the model with $Sens = 1$, obtaining $R_0 = 5$ and $InitialInf = 2347$.

In Figure 5, you see the daily number of infected asymptomatic, infected symptomatic, and hospitalized people peaking at the beginning of the second phase. The daily number of known infected peaks about a month later. By September 1st, all of these compartments have flatlined at 0.

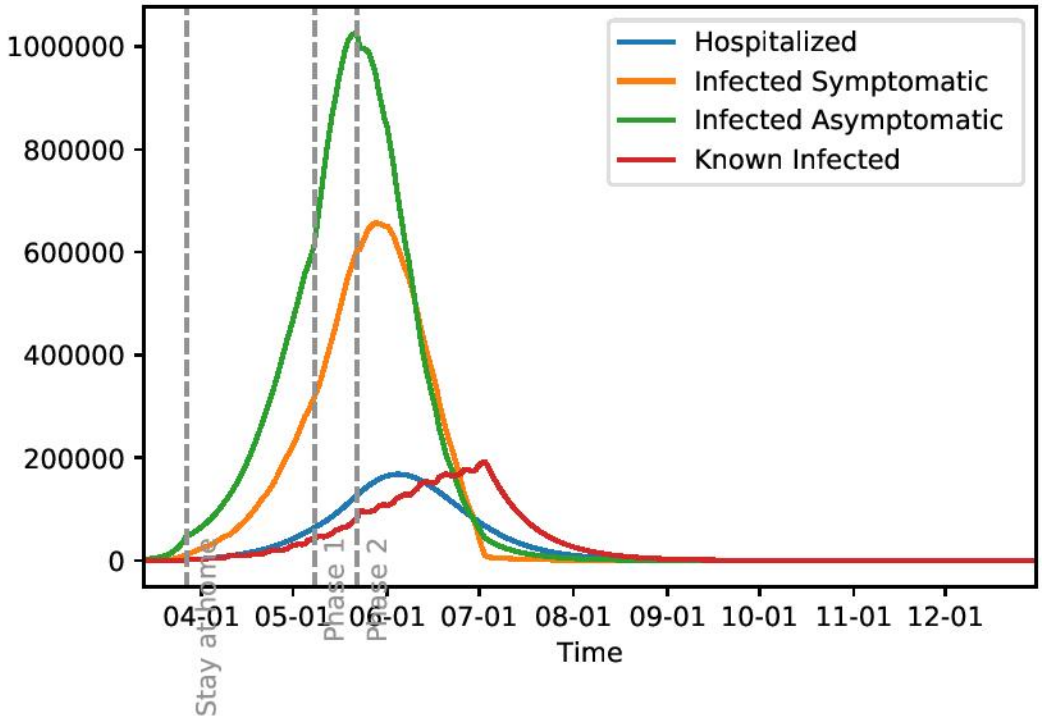

Figure 5: Projections $Sens = 1$, $R_0 = 5$, and $InitialInf = 2347$

In Figure 6, you see the number of deaths follows a sharp exponential increase, exiting the chart at about April 1.

Compartment	Cumulative
Hospitalized	8920658
Infected Symptomatic	28678315
Infected Asymptomatic	44155091
Known Infected	10190576
Deaths	276112

Table 4: Cumulative predicted number of individuals in each compartment for entire year, $Sens. = 1$, $R_0 = 5$, and $InitialInf = 2347$

From Table 4, you can see the model predicts that the cumulative number of Infected Symptomatic and Infected Asymptomatic Individuals will both exceed the total population of North Carolina, 85% of the population will be hospitalized, 97% will be known infected, and 2.6% will die from Covid-19.

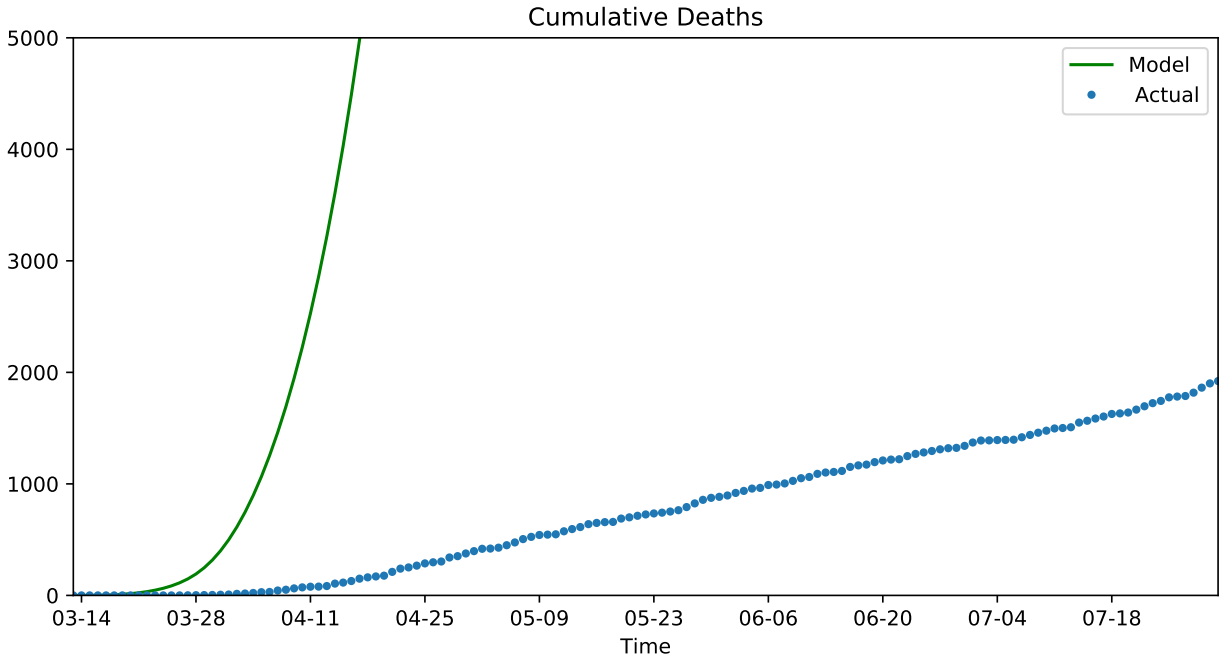

Figure 6: Predicted Deaths vs Actual Deaths $Sens = 1$, $R_0 = 5$, and $InitialInf = 2347$

5. Conclusion

When adding the additional training data, one of the main things that I found was that in the middle of May, the number of deaths per day stopped increasing. This may be in part due to increased social distancing measures, beyond that which was recommended by the government. Thus, perhaps in order to achieve better predictive power for the model, I should further decrease the social distancing parameter β during this period of time, in order to decrease the amount of predicted deaths. However, doing so may cause the model to underpedict the number of infections. Thus, I should also consider other ways to decrease the amount of predicted deaths without decreasing the predicted amount infected, that is also in accordance with some recent observed trend in the spread of the disease. For example, I could decrease the parameters $deathOutofSympFrac$ and $deathoutofHospFrac$ (See Appendix), if I find that the medical community has become better equipped to handle the Covid-19 pandemic.

I observe a number of seemingly counterintuitive results when I decrease the sensitivity of the Covid-19 tests. First, I observe a decrease in the number of hospitalizations, which may be due to reduced inflow from the KI compartment. Second, I observe opposite effects for the Infected Symptomatic and Infected Asymptomatic compartments, despite the fact that both groups get tested at the same rate. This result may be related to the differing recovery times for the two groups, which is 14 days for the Symp. group and 10 days for the Asymp. group. Thus, for individuals in the Asymp. group it is possible that by decreasing the outflow to the KI compartment, the outflow to the Recovery compartments also increases. While this may also occur for the Symp. compartment, since the recovery time is longer, in this case, the outflow to the Recovery compartment may not be enough to counterract the additional inflow of infected individuals and the continued presence of infected individuals who test negative.

The removal of the Asymptomatic Isolated compartments and any form of contact tracing results in an unrealistic model. Of particular note is how the model predicts that in both the Symptomatic and Asymptomatic compartments, there are more individuals infected than the entire population of North Carolina. This indicates that this model is allowing individuals to be infected twice, which should be impossible. While I did expect that eliminating contact tracing to increase the potency of the disease, I did not necessarily expect it to do so to this degree.

6. Appendix

A. Flow between Compartments

The inflow and outflow of each compartment is as follows:

1. I_{Si}

   Inflow:

   - $N_{Si} \rightarrow I_{Si}$- not tested, becomes infected
   - $I_{An} \rightarrow I_{Si}$- becomes symptomatic
   - $I_{Ai} \rightarrow I_{Si}$- becomes symptomatic

   Outflow:

   - $I_{Si} \rightarrow R_{An}$- recovers
   - $I_{Si} \rightarrow H$- becomes hospitalized
   - $I_{Si} \rightarrow D$- dies
   - $I_{Si} \rightarrow KI$- tests positive

2. I_{Ai}

   Inflow:

   - $N_{Ai} \rightarrow I_{Ai}$- not tested, becomes infected
   - $I_{An} \rightarrow I_{Ai}$- identified as close contact

   Outflow:

   - $I_{Ai} \rightarrow R_{An}$- recovers
   - $I_{Ai} \rightarrow I_{Si}$- starts showing symptoms
   - $I_{Ai} \rightarrow KI$- tests positive

3. I_{An}

   Inflow:

   - $N_{An} \rightarrow I_{An}$- not tested, becomes infected

   Outflow:

   - $I_{An} \rightarrow R_{An}$- recovers
   - $I_{An} \rightarrow I_{Si}$- becomes symptomatic
   - $I_{An} \rightarrow I_{Ai}$- identified as close contact
   - $I_{An} \rightarrow KI$- tests positive

4. N_{Si}

   Inflow:

   - $N_{An} \rightarrow N_{Si}$- not tested, develops symptoms of other condition or disease
   - $N_{Ai} \rightarrow N_{Si}$- develops symptoms of other condition or disease

   Outflow:

   - $N_{Si} \rightarrow N_{An}$- tests negative, or leaves self-quarantine
   - $N_{Si} \rightarrow I_{Si}$- not tested, becomes infected

- $N_{Si} \rightarrow N_{Ai}$- leaves self-quarantine

5. N_{Ai}

   Inflow:

   - $N_{An} \rightarrow N_{Ai}$- identified as close contact

   Outflow:

   - $N_{Ai} \rightarrow N_{An}$- tests negative, or leaves self-quarantine
   - $N_{Ai} \rightarrow I_{Ai}$- not tested, becomes infected
   - $N_{Ai} \rightarrow N_{Si}$- develops non-Covid 19 symptoms

6. N_{An}

   Inflow:

   - $N_{Si} \rightarrow N_{An}$- recovers from non-Covid 19 symptoms, or tests negative
   - $N_{Ai} \rightarrow N_{An}$- leaves self-quarantine, or tests negative

   Outflow:

   - $N_{An} \rightarrow N_{Ai}$- identified as close contact
   - $N_{An} \rightarrow N_{Si}$- develops non-Covid 19 symptoms
   - $N_{An} \rightarrow I_{An}$- becomes infected

7. R_{Si}

   Inflow:

   - $R_{An} \rightarrow R_{Si}$- not tested, develops non-Covid 19 symptoms
   - $R_{Ai} \rightarrow R_{Si}$- not tested, develops non-Covid 19 symptoms

   Outflow:

   - $R_{Si} \rightarrow R_{An}$- recovers from non-Covid 19 illness

8. R_{Si}

   Inflow:

   - $R_{An} \rightarrow R_{Ai}$- identified as close contact

   Outflow:

   - $R_{Ai} \rightarrow R_{An}$-leaves self-quarantine

   $R_{Ai} \rightarrow R_{Si}$- not tested, develops non-Covid 19 symptoms

9. R_{An}

   Inflow:

   - $I_{Si} \rightarrow R_{An}$-recovers naturally
   - $I_{Ai} \rightarrow R_{An}$-recovers naturally
   - $I_{An} \rightarrow R_{An}$-recovers naturally
   - $R_{Ai} \rightarrow R_{An}$- tests negative or leaves self-quarantine
   - $R_{Si} \rightarrow R_{An}$- tests negative or recovers from non-Covid 19 symptoms

   Outflow:

- $R_{An} \rightarrow R_{Ai}$- identified as close contact
- $R_{An} \rightarrow R_{Si}$- develops non-Covid 19 illness

10. KI

    Inflow:

    - $I_{Si} \rightarrow KI$-tests positive
    - $I_{Ai} \rightarrow KI$-tests positive
    - $I_{An} \rightarrow KI$-tests positive

    Outflow:

    - $KI \rightarrow H$- hospitalized
    - $KI \rightarrow R$-recovers

11. H

    Inflow:

    - $I_{Si} \rightarrow H$- hospitalized
    - $KI \rightarrow H$- hospitalized

    Outflow:

    - $H \rightarrow D$- dies
    - $H \rightarrow KR$- recovers

12. KR

    Inflow:

    - $KI \rightarrow KR$- recovers
    - $H \rightarrow KR$- recovers

13. D

    Inflow:

    - $H \rightarrow D$- dies
    - $I_{Si} \rightarrow D$-dies

B. Update Equations and Parameters-No Contact Tracing

1. I_{Si}

$$I_{Si}^{t+1}(recovered) = I_{Si}^{t}(recovered)(1-\pi_{Si}Sens)(1-\frac{1}{sympToRecoveryTime})$$

$$+recoverOutOfSympFrac[\beta_l(I_{An}(1-\pi_{An}*Sens)+I_{Si}(1-\pi_{Si}Sens)]N_{Si}^{t}(1-\pi_{Si})+\frac{I_{An}(show\ symptom)(1-\pi_{An}Sens)}{infToSympTime}]$$

- $I_{Si}^{t+1}(hospitalized)$- same as $I_{Si}^{t+1}(recovered)$, except $recoverOutOfSympFrac$ becomes $hospOutOfSympFrac$ and $sympToRecoveryTime$ becomes $sympToHospTime$
- $I_{Si}^{t+1}(death)$- same as $I_{Si}^{t+1}(recovered)$, except $recoverOutOfSympFrac$ becomes $deathOutOfSympFrac$ and $sympToRecoveryTime$ becomes $sympToDeathTime$

2. I_{An}

$$I_{An}^{t+1}(recovered) = I_{An}^{t}(recovered)(1-\pi_{An}Sens)(1-\frac{1}{asympToRecoveryTime})$$

$$+(1-symptomFrac)[(\beta_h I_{An}^{+}(1-\pi_{An}Sens)+\beta_l I_{Si}(1-\pi_{Si}Sens))(N_{An}^{+}(1-\pi_{An}))]$$

- $I_{An}^{t+1}(showsymptoms)$- same except replace $asympToRecoveryTime$ with $infToSympTime$ and $1-symptomFrac$ with $symptomFrac$

3. N_{Si}

$$N_{Si}^{t+1} = N_{Si}^{t}(1-\pi_{Si})(1-\frac{1}{selfQuarTime}-\beta_l(I_{An}^{t}(1-\pi_{An}Sens)+I_{Si}^{t}(1-\pi_{Si}Sens)))$$

$$+nonCOVIDSymptRate(N_{Si}^{t}\pi_{Si}+N_{An}^{t}\pi_{An})$$

4. N_{An}

$$N_{An}^{t+1} = (N_{An}^{t}\pi_{An}+N_{Si}^{t}\pi_{Si})(1-nonCOVIDSymptRate-(\beta_h I_{An}^{t}(1-\pi_{An}Sens)+\beta_l I_{Si}^{t}(1-\pi_{Si}Sens)))$$

$$+\frac{N_{Si}^{t}(1-\pi_{Si})}{selfQuarTime}$$

5. R_{Si}

$$R_{Si}^{t+1} = R_{Si}^{t}(1-\pi_{Si})(1-\frac{1}{selfQuarTime})+nonCOVIDSymptRate(R_{Si}^{t}\pi_{Si}+R_{An}^{t}\pi_{An})$$

6. R_{An}

$$R_{An}^{t+1} = (R_{An}\pi_{An}+R_{Si}\pi_{Si})(1-nonCOVIDSymptRate)+\frac{R_{Si}^{t}(1-\pi_{Si})}{selfQuarTime}+\frac{I_{Si}^{t}(1-\pi_{Si}Sens)}{sympToRecoveryTime}$$

$$+\frac{I_{An}^{t}(1-\pi_{An}Sens)}{asympToRecoveryTime}$$

7. KI

$$KI^{t+1}(recovered) = (KI^{t}(recovered)+I_{Si}^{t}(recovered)\pi_{Si}Sens+I_{An}^{t}(recovered)\pi_{An}Sens$$

$$+recoverOutOfSymptFrac(I_{An}^{t}(sympt)\pi_{An}Sens))(1-\frac{1}{sympToRecoveryTime})$$

$$KI^{t+1}(hospitalized) = (KI^{t}(hospitalized)+I_{Si}^{t}(hospitalized)\pi_{Si}Sens$$

$$+hospOutOfSymptFrac[I_{An}^{t}(sympt.)\pi_{An}Sens])(1-\frac{1}{infToHospTime})$$

8. H

$$H^{t+1}(die) = H^{t}(die)(1-\frac{1}{hospToDeathTime})+\frac{deathFrac}{infToHospTime}(I_{Si}^{t}(hosp)+KI^{t}(hosp.))$$

$$+hospOutOfSymptFrac[I_{An}^{t}(sympt.)\pi_{An}Sens])$$

- $H^{t+1}(recovered)$- same format as above, replace $hospToDeathTime$ with $hospToRecoveryTime$, and $deathFrac$ with $recFrac$

9. KR

$$KR^{t+1} = KR^t + \frac{H^t(rec.)}{hospToRecoveryTime} + \frac{KI^t(rec.)}{sympToRecoveryTime} + \frac{Sens(I^t_{Si}(rec.)\pi_{Si} + I^t_{An}(rec.)\pi_{An})}{sympToRecoveryTime} + \frac{recoverOutOfSymptFrac[I^t_{An}(sympt.)\pi_{An}Sens]}{sympToRecoveryTime}$$

10. D

$$D^{t+1} = D^t + \frac{H^t(die)}{hospToDeathTime} + \frac{I^t_{Si}(death)}{sympToDeathTime}$$

Parameters:

Testing:

1. π_{Si}: fraction of symptomatic isolated population that is tested
2. π_{An}: fraction of asymptomatic non-isolated population that is tested
3. $Sens$: sensitivity of COVID-19 tests

Rates:

1. $sympToRecoveryTime = 14$: average number of days for symptomatic person to recover from disease
2. $infToSympTime = 5$: average number of days until infected person shows symptoms
3. $sympToHospTime = 5$: average number of days from onset of symptoms to hospitalization
4. $sympToDeathTime = 14$: average time from symptoms onset to death
5. $asympToRecoveryTime = 10$: average time from infection to recovery
6. $selfQuarTime = 10$: average time that individual self-isolates
7. $nonCovidSymptRate = \frac{1}{1200}$: rate than individual develops symptoms to non-COVID-19 disease
8. $infToHospTime = 5$: average time until an infected person requires hospitalization
9. $hospToDeathTime = 14$: average time between hospitalization and death
10. $hospToRecoveryTime = 14$: average time between hospitalization and recovery

Fractions:

1. $recoverOutOfSympFrac = 0.78$: fraction of inflow to I_{Si} that will recover at home
2. $hospOutOfSympFrac = 0.2$: fraction of inflow to I_{Si} compartment that will require hospitalization
3. $deathOutOfSympFrac = 0.02$: fraction of inflow to I_{Si} compartment that will die without being tested
4. $symptomFrac = 0.5$: fraction of inflow to I_{An} compartment that will develop symptoms
5. $recoverOutOfSympFrac = 0.80$: fraction of individuals from $I_{An}(sympt.)$ that will recover naturally without any hospitalization
6. $hospOutOfSymptFrac = 0.20$: fraction of individuals from $I_{An}(sympt.)$ that will require hospitalization
7. $deathFrac = \frac{1}{3}$: fraction of inflow to H that will die
8. $recFrac = \frac{2}{3}$: fraction of inflow to H that will recover

References

1. David Kempe, Jon Kleinberg, and Eva Tardos, "Maximizing the Spread of Influence through a Social Network," *Theory of Computing*, vol. 11, pp. 105-147, Apr. 2015, doi: 10.4086/toc.2015.v011a004.

2. Evan L. Porteus, *Foundations of Stochastic Inventory Theory.* Stanford, California, USA: Stanford University Press, 2002.

3. Omar El Housni, Mika Sumida, Paat Rusmevichientong, Huseyin Topaloglu, and Serhan Ziya,"Future Evolution of Covid-19 Pandemic in North Carolina: Can we Flatten the Curve?" http://ziya.web.unc.edu/files/2020/07/CovidModelandAnalysis-NC-Version-July-2.pdf.

4. Steven Woloshin, Neeraj Patel, and Aaron S. Kesselheim, "False Negative Tests for SARs-CoV-2 Infection- Challenges and Implications," *The New England Journal of Medicine*, vol. 383, ed. 38, Aug. 2020, doi: 10.1056/NEJMp2015897.

5. V.G. Kulkarni, *Modeling, Analysis, Design, and Control of Stochastic Systems.* New York, NY, USA: Springer, 1999.