Modelling the Spread of COVID-19 in Alberta using an SIR Model and Simulating the Disease using a Spatial Network

ABRAHAM HUSSEIN[1]

[1] *Mathematics and Statistics Department, University of Calgary, MS 476, 2500 University Drive NW, Calgary AB, T2N 1N4, Canada*

Compiled June 5, 2021

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) first emerged in Wuhan, China November 2019 and has since spread in an unprecedented rate to exist throughout most countries in the world. The disease is forecasted to have a lasting impact to the global economy and to public health measures. The epidemiology of the virus is being actively studied [1]. This paper develops a SIR Model of the coronavirus disease (COVID-19) outbreak in Alberta based on data collected by the Albertan government. Using the SIR model's estimated infection and removal rate, simulations of the disease's spread will be done on a spatial network. The simulations suggest the majority of Albertans will contract the disease if self isolation and distancing measures are ignored and infected individuals are able to maintain their social contacts.

1. INTRODUCTION

The world underestimated the infectious capabilities of COVID-19. The virus spread beyond ground zero in Wuhan to infect the world even as many countries closed their borders and restricted businesses to slow the spread of the virus. Alberta was proactive in their response to COVID-19 and has kept their number of confirmed cases fairly low compared to the rest of the country [2]. SIR models have been used to study the spread of the SARS outbreak [3], a pre-

vious coronavirus strain that first emerged out in Foshan, China and persisted from November 2002 to May 2004. An SIR model is based off 3 distinct categories [4]: Susceptible individuals $S(t)$, infected individuals $I(t)$, and removed individuals $R(t)$. From those categories the SIR model can be defined as the following system of differential equations:

$$\frac{dS}{dt} = -kSI,$$
$$\frac{dI}{dt} = kSI - \gamma I,$$
$$\frac{dR}{dt} = \gamma I,$$

where $k > 0$ is the infection rate and $\gamma > 0$ is the removal rate. Using the data provided from the Albertan government we can estimate the value for k and γ. Using Alberta specific infection and removal rate we can simulate the spread of the virus in Alberta through a spatial network.

A. Data

The data-set for this report was obtained from Alberta's provincial government [A], ranging

from the first reported provincial infection on March 6th, 2020 to April 13th, 2020 ($t = 40$ days). The data-set includes the daily total of confirmed cases, probable cases, recoveries, and deaths. During this time frame, 1198 confirm cases of the virus were reported. In addition, 534 cases were considered probable during the period, 48 deaths were recorded, and 875 individuals recovered.
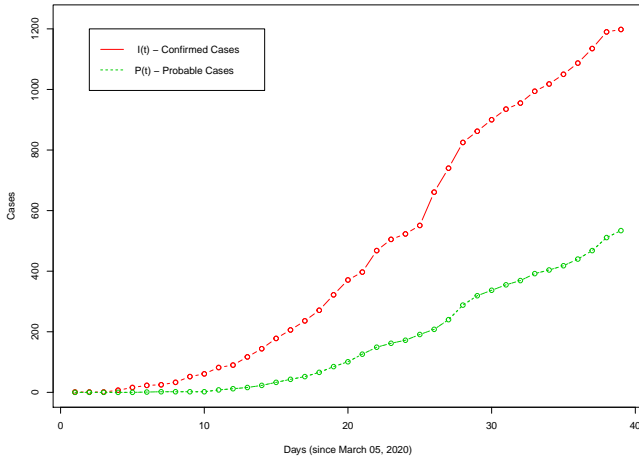

Fig. 1. Daily number of COVID-19 cases; March 3, 2020 to April 13, 2020.

2. METHODS

A. The SIR Model

The SIR model used for COVID-19 in Alberta is a discrete model based on a model previously used for the SARS outbreak in China [3]. The SIR model for Alberta has 3 groups based on the published data: The first group is the susceptible population $S(t)$. The second groups encompasses the infected population which will consist of the confirmed cases $C(t)$ and probable cases $P(t)$ reported by Albertan government. The third group is the removed population formed from the recorded recovered individuals $H(t)$ and the reported deaths to COVID-19 $D(t)$. The infection rate for the confirmed cases will be denoted β_t. The removal rate, or the probability of recovering during the day, will be denoted γ_t. This approach for constructing the model limits our analysis. Rather than assess the rates for $I(t)$, $P(t)$, $H(t)$, and $D(t)$ separately, we will combined them to alleviate the computational burden in simulations. The system of differential equations for the SIR model is the following:

$$\frac{dS}{dt} = -\beta_t(C(t) + P(t))S, \tag{1}$$

$$\frac{dI}{dt} = \beta_t(C(t) + P(t))S - \gamma(C(t) + P(t)), \tag{2}$$

$$\frac{dR}{dt} = \gamma_t(C(t) + P(t)), \tag{3}$$

where $I(t) = C(t) + P(t)$ is the infected population at time t, $R(t) = H(t) + D(t)$, is the recovered population at time t, $\beta_t > 0$ is the infection rate, and γ_t is the recovery rate.

A.1. Parameter Estimates

Estimation of β_t and γ_t was done by calculating the changing rate for $S(t)$, $I(t)$, and $R(t)$ of changes of the the system of differential equations, as opposed to an optimization approach. Similar methods have been used for the SARS outbreak [3]. If we consider the 1-step jump from day $t-1$ to t, then we know the changes for our system during the 1-step jump is:

$$\frac{dS}{dt} = S(t) - S(t-1), \tag{4}$$

$$\frac{dI}{dt} = C(t) + P(t) - C(t-1) - P(t-1), \tag{5}$$

$$\frac{dR}{dt} = H(t) + D(t) - H(t-1) + D(t-1). \tag{6}$$

Therefore, a 1-step change in our system can be observed and we can substitute our changing rates back into (1) and (3) to solve for β_t and γ_t, respectively. For a 1-step change we have,

$$\beta_t = \frac{-S(t) + S(t-1)}{(C(t-1) + P(t-1))S(t-1)},$$
$$\gamma_t = \frac{H(t) + D(t) - H(t-1) + D(t-1)}{C(t-1) + P(t-1)}.$$

The infection rate β_t and recovery rate γ_t only reflect the change from one day to next and do not reflect the overall rates for the disease in Alberta. Through this method we will generate two samples for the parameters of interest, $X : \{\beta_1, ..., \beta_{39}\}$ and $Y : \{\gamma_1, ..., \gamma_{39}\}$. The final estimate for β_t and γ_t will be the sample mean $\overline{\beta_t}$ and $\overline{\gamma_t}$.

B. The Spacial Network

The spacial network developed consists of nodes arranged in an symmetric grid. Each node was connected to the nodes that are directly adjacent to it. As result, the amount of connections a node could have ranged from 3 to 8. Each node had a total of 3 states: susceptible, infected, and removed. All nodes are initially made susceptible except the initial disease carrier. A susceptible node could only be infected if an adjacent node has the disease. The probability a susceptible node being infected is the same as the estimated infection rate from the SIR model, multiplied by contact factor. Similarly, the probability of an infected node being removed is the estimated removal rate; recurrent infections are not considered and once a node is removed they cannot be infectious again.

Simulations have been used to model an epidemic based on the disease's movements and contact data [5][6]. There is a lack of literature surrounding social contacts and its impact on infectious diseases in Canada. As such, the simulation for Alberta will use the social contacts and mixing patterns from the Netherlands [7]. The spread of COVID-19 in Canada and the Netherlands are reportedly similar.

B.1. Spacial Network Algorithm

The algorithm for the spacial network simulation was as follows:

Algorithm 1. Spacial Network Simulation

```
 1: procedure SPATIAL.SIM(γ, β,C,Days)
 2:     Generate nxn matrix with nodes = 0
 3:     Infect node I_0 ← 1
 4:     for < t in 1:Days> do
 5:         Infected = vector of infections
 6:         for < j in 1:Infected> do
 7:             Generate C∼ N(13.85, 10.54²)
 8:             Ensure 0 ≤C≤ 8
 9:             N = Adjacent for Infected[j]
10:             if C/8 · β_t > runif(N,0,1) then
11:                 Infect Adjacent[N]
12:         Store S(t)
13:         Store I(t)
14:         if γ >runif(0,1) then
15:             R = I(t)
16:         Store R(t)
17:     return data.frame(S(t),I(t),R(t))
```

3. RESULTS

Table 1. Results of sample ($n = 39$) for β and γ

Parameter	Mean (Median)	Standard Deviation
β_t	0.3216 (0.1278)	0.9599
γ_t	0.0256 (0.0281)	0.0199

The estimated value for $\overline{\beta_t}$ was 0.322 and the estimated value for $\overline{\gamma}$ was 0.0256. The summary statistics are summarized on Table 1. The density curves are seen on Figure 2 and have been built in consideration of Alberta's last census count of 4,371,316 individuals [8]. The model does not consider the geographically wide spread of Albertans and from existing data we know the majority of COVID-19 cases originate from dense urban centres such as Calgary and Edmonton.
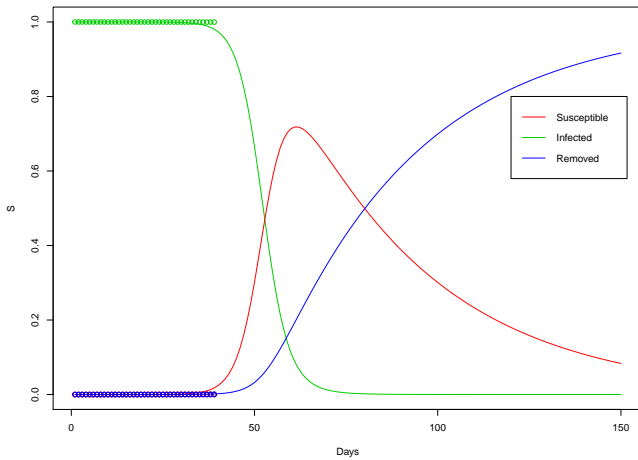

Fig. 2. The SIR Model; points on the curve are the observed proportions.

Simulations were done for a 200x200 grid, resulting in 40,000 nodes. The movement of the disease was observed for 200 days. The number of daily contacts were randomly generated from a normal distribution $\mathcal{N}(13.85, 10.54^2)$ [7]. Every simulation resulted in all nodes being infected and the time to total infection was dependent on the network's size. Figure 3 displays the spread of infection where grey nodes are susceptible, red nodes are infected, and blue nodes are removed.
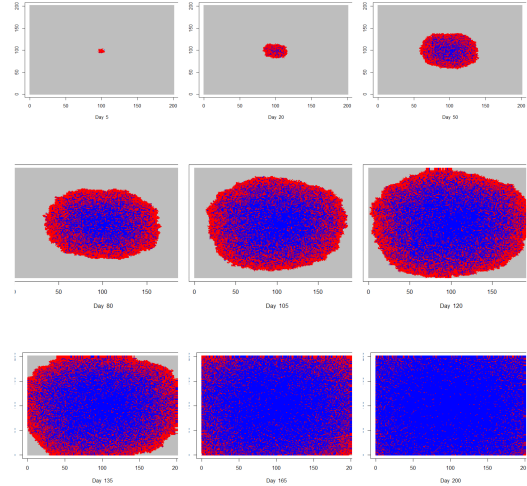

Fig. 3. The simulated spread of COVID-19 from one person captured at (top-left→bottom-right): Day 5, 20, 50, 80, 105, 120, 135, 165, 200.
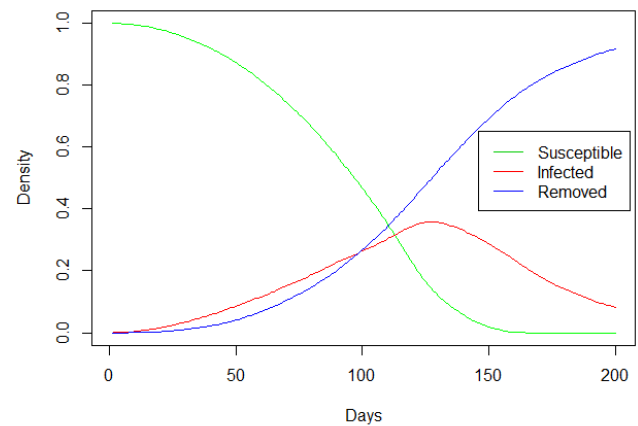

Fig. 4. SIR Model of the simulation.

4. DISCUSSION AND PROJECT CONSIDERATIONS

From repeated simulations, the greatest controllable variable in the spread of COVID-19 is reducing social contacts. Ideally an infected individual would reduce their daily contact to 0. Implementing self-isolation, as advised by the provincial and federal government, would considerably 'flatten the curve' of infected nodes. Self-isolation is the simplest measure Albertans can to reduce the provincial impact of COVID-19. Figure 5 reflects the changes to a simulation if self isolation is followed by all Albertans.

Fig. 5. A simulation when self isolation is implemented.

The estimation for the parameters β_t and γ_t could be done through optimization instead of using an adapted Euler's method. However, utilizing Euler's method introduces the option to treat the parameters as random variables. As such, a Bayesian approach could be used to estimate the parameters. MCMC methods could have been utilized to estimate the parameters of interest, as the data was readily available and the parameters should be considered random. Implementing MCMC methods proved challenging and instead a simpler analysis was done to

direct more time into developing the spatial network algorithm.

Distance between adjacent nodes was not considered and the distance between all adjacent nodes was considered identical. This assumption does not reflect real social connections, for example loved ones would be considered 'closer' than coworkers and peers. Adding noise to the calculation for infection between adjacent nodes could reflect real differences between social contacts and modelled human contacts more dynamically. However, it is unclear if adding noise would increase or decrease the spread of COVID-19.

Running longer simulations took a substantial amount of time. The largest simulations done were for 40,000 nodes and took approximately 48 hrs to complete. The computational time increased considerably as more nodes became infected and there was no method available to trim calculation time. For future considerations, a simple check skip the infection step if all adjacent nodes are infected would shave time on simulations. It would not have been possible to simulate enough nodes for all Albertans.

A. Dataset Citation

A. No author, "COVID-19 in Alberta," [retrieved 14 April 2020], https://covid19stats.alberta.ca/

REFERENCES

1. H. A. Rothan and S. N. Byrareddy, "The epidemiology and pathogenesis of coronavirus disease (COVID-19) outbreak," (2020).
2. Province of Alberta, "COVID-19 Update," Tech. rep. (2020).
3. Z. Zhang, "The outbreak pattern of SARS cases in China as revealed by a mathematical model," Ecol. Model. **204**, 420–426 (2007).
4. V. Capasso, "A Generalization of the Kermack-McKendrick Deterministic Epidemic Model*," Tech. rep.
5. V. K. Nguyen, R. Mikolajczyk, and E. A. Hernandez-Vargas, "High-

resolution epidemic simulation using within-host infection and contact data," BMC Public Heal. **18** (2018).

6.   Z. Cao, P. Zhao, J. Liu, and W. Zhong, "A spatial point pattern analysis of the 2003 SARS Epidemic in Beijing," in *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on the Use of GIS in Emergency Management, EM-GIS 2017,* (Association for Computing Machinery, Inc, 2017).

7.   J. L. Mossong, N. Hens, M. Jit, P. Beutels, K. Auranen, R. Mikolajczyk, M. Massari, S. Salmaso, G. Scalia Tomba, J. Wallinga, J. Heijne, M. Sadkowska-Todys, M. Rosinska, W. J. Edmunds, and S. Riley, "Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases," Tech. rep. (2008).

8.   Statistics Canada, "Quarterly Population Report," Tech. rep. (2019).