

# Modelling the spread of COVID-19

MA4207 - Machine Learning and Network Analysis  
Project

By

Aditya Kumar Singh, Santanu Das, Saharica Balu,  
Siddhikant Mishra

## About COVID-19

- Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).
- An infectious disease outbreak is the occurrence of a disease that is not usually expected and spreads rapidly in a geographical region over a short period of time.

# Objectives...

- Use the SEIR model to predict future trends of the outbreak using past information of the virus spread.
- To analyse the accuracy of the SEIR model in predicting the peak and end of the outbreak, using train and test data.
- To apply the SEIR Model on the network that we have used **Barabasi-Albert** Network to model the real world interactions.
- Discuss shortcomings we faced during our project.

# SEIR Model

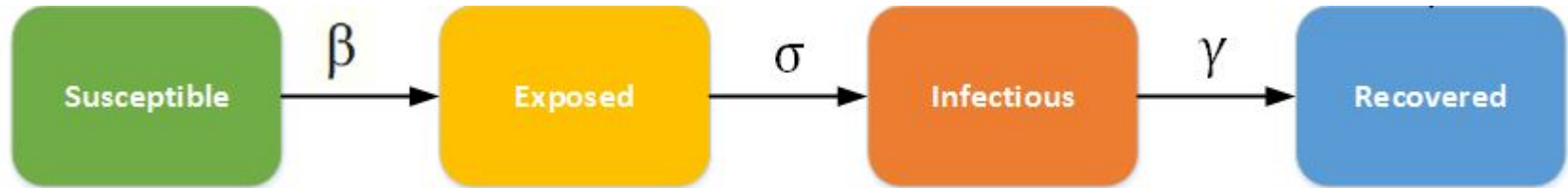
- We divide the total population( $N$ ) in 4 compartments, namely: Susceptible( $S$ ), Exposed( $E$ ), Infected( $I$ ), Removed( $R$ ).
- Key Assumptions:
  - $N$  is constant,  $N = S + E + I + R$
  - SEIR model assumes that the latent and infectious times of the pathogen are exponentially distributed.
  - The individuals in the Exposed state are infected but not yet infectious
  - Once recovered, immune life-long.
  - Well-mixed Population

# Parameters

- Incubation rate  $\sigma$ , is the rate of latent individuals becoming infectious. Given the known average duration of incubation  $Y$ ,  $\sigma = 1/Y$ .
- Recovery rate  $\gamma = 1/D$ , is determined by the average duration of recovery  $D$ , of infection. After this period, they enter the removed phase.
- The infectious rate,  $\beta$ , controls the rate of spread which represents the probability of transmitting disease between a susceptible and an infectious individual.

# Equations

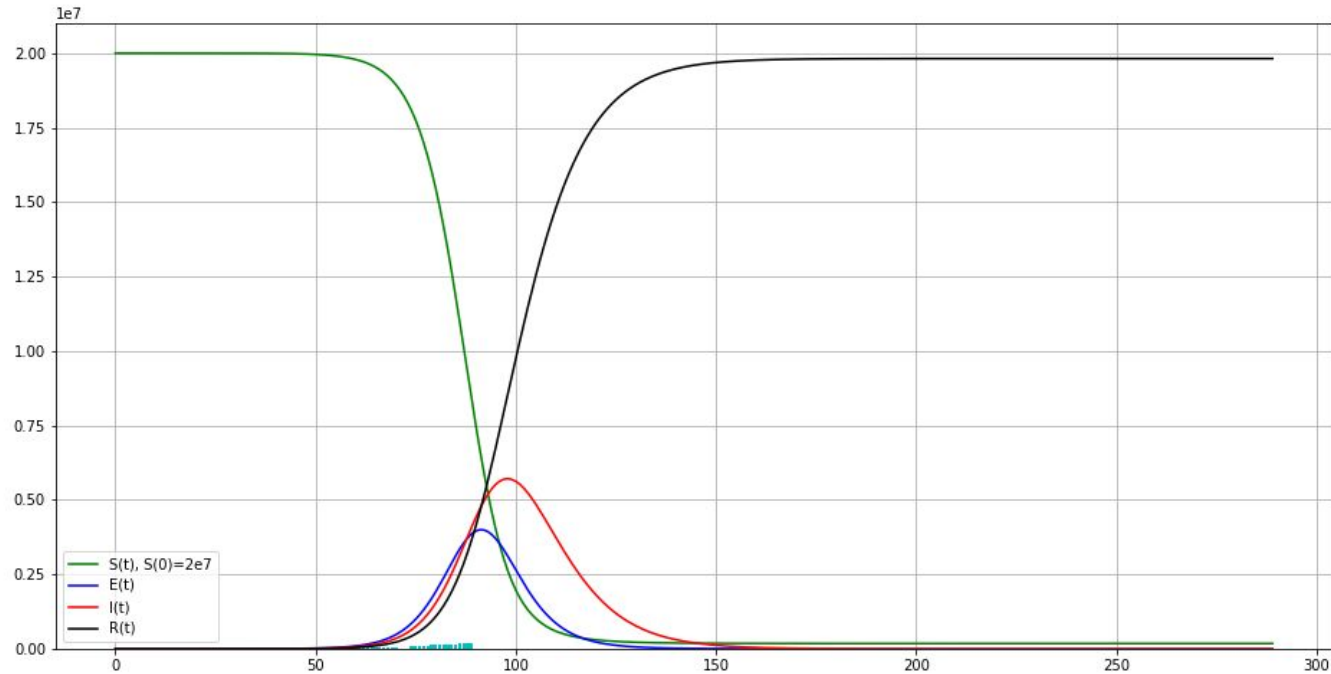
1. Rate of change of S:  $dS/dt = -\beta * S * I / N$
2. Rate of change of E:  $dE/dt = \beta * S * I / N - \sigma * E$
3. Rate of change of I:  $dI/dt = \sigma * E - \gamma * I$
4. Rate of change of R:  $dR/dt = \gamma * I$



# Reproduction Number, $R_0$

- The expected number of secondary cases produced by a single (typical) infection in a completely susceptible population.
- $R_0 = \beta/\gamma$ , which are constants.
- If  $R_0 < 1$ , Epidemic fades out.
- If  $R_0 > 1$ , Epidemic spreads.

# Model Plot



Initial conditions:

$$S(0) = 2e7$$

$$E(0) = 10$$

$$I(0) = 1$$

$$R(0) = 0$$

Parameters:

$$\beta = 1.05$$

$$\sigma = 1/5.2$$

$$\gamma = 1/9.5$$

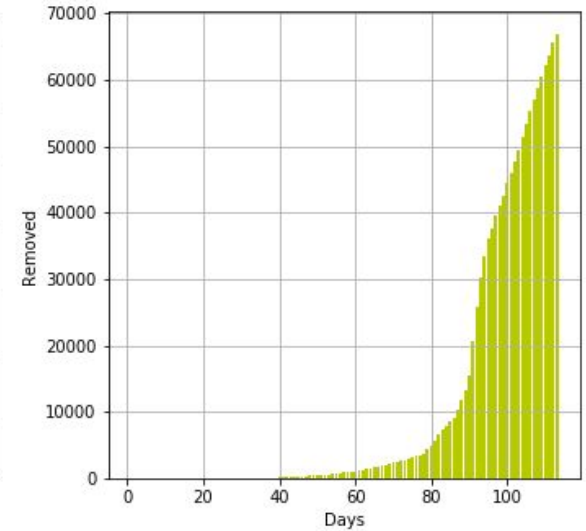
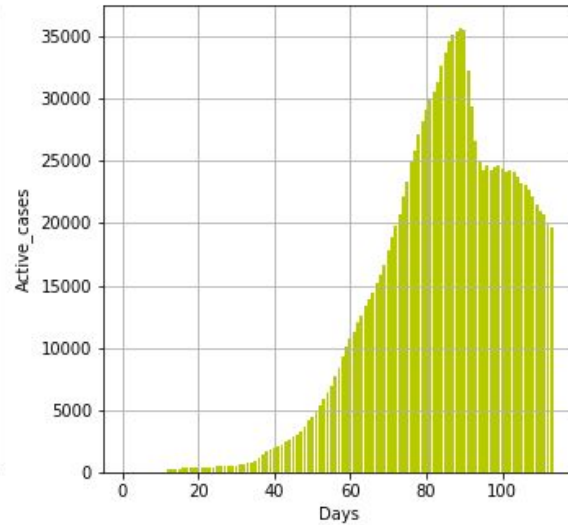
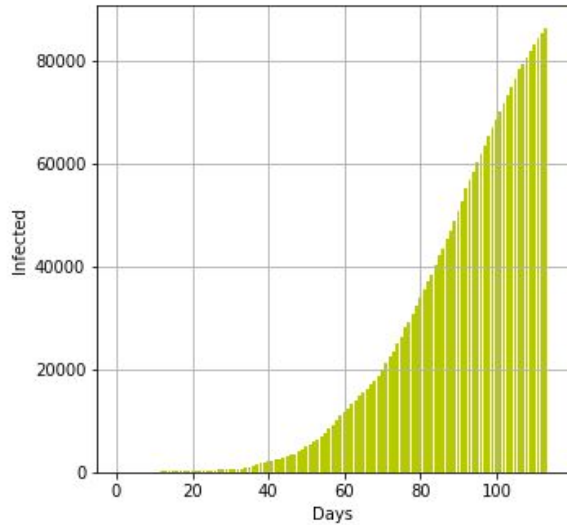


# Parameter trends (increase parameters)

	S	E	I	R
$\beta$	Saturates lower	Left and higher	Left and higher	Saturates higher
$\sigma$	Saturates lower	Left and higher	Left and higher	Saturates higher
$\gamma$	Saturates lower	Left and lower	Left and lower	Saturates higher

- If  $\sigma < \sigma_{\text{threshold}}$ , and  $\gamma > \gamma_{\text{threshold}}$ , then exposed peak is lower as compared to infected and vice-versa.
- And always exposed will peak before infected.

# Data : Qatar



# Parameter Estimation

- Initial reproduction number:  $r_0 = 4$
- $\sigma$  (Mean latent period) =  $1/10$
- $\gamma$  (Mean Infectious Period) =  $1/15$
- $\beta$  (Transmission Rate) =  $(r_0 \gamma)K$
- We know SEIR model, the reproduction number is  $R = \beta / \gamma$
- Since  $\sigma$ ,  $\gamma$  and  $r_0$  are taken to be constant, we fit the data by changing the value of  $K$ .
- Then  $r_{\text{eff}} = r_0 K$ , which is the value reproduction number for a given  $K$
- Keep in mind that any change in beta is because of  $K$ . Changing lock down strictness, changes  $\beta$ .

# Qatar fit data:

Initial conditions:

$S(0)$  = Population of Qatar = 2807805

$E(0) = 0$

$I(0) = 1$

$R(0) = 0$

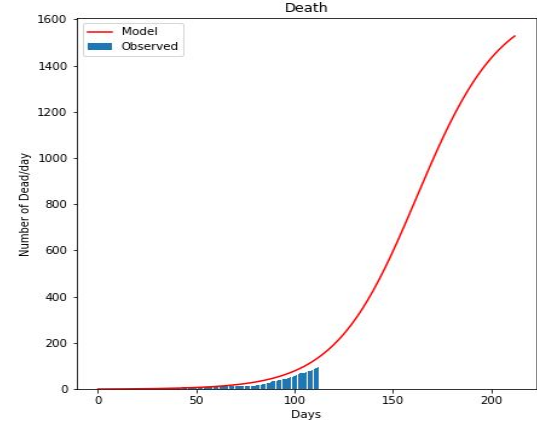
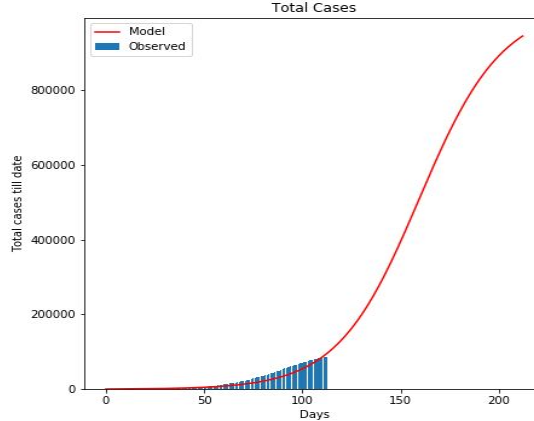
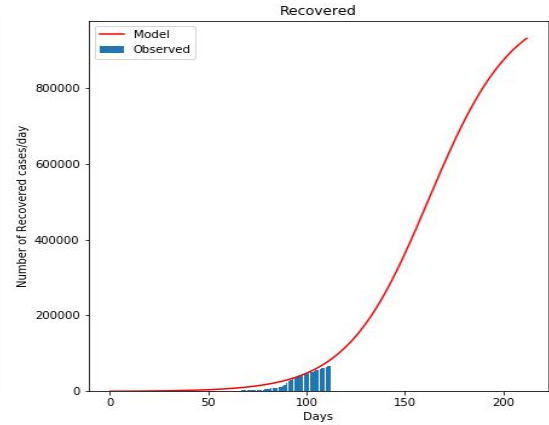
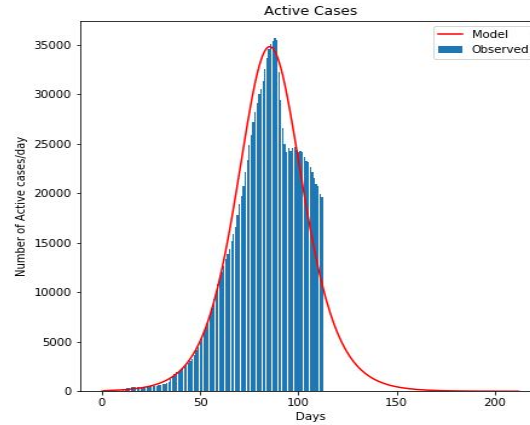
Parameters:

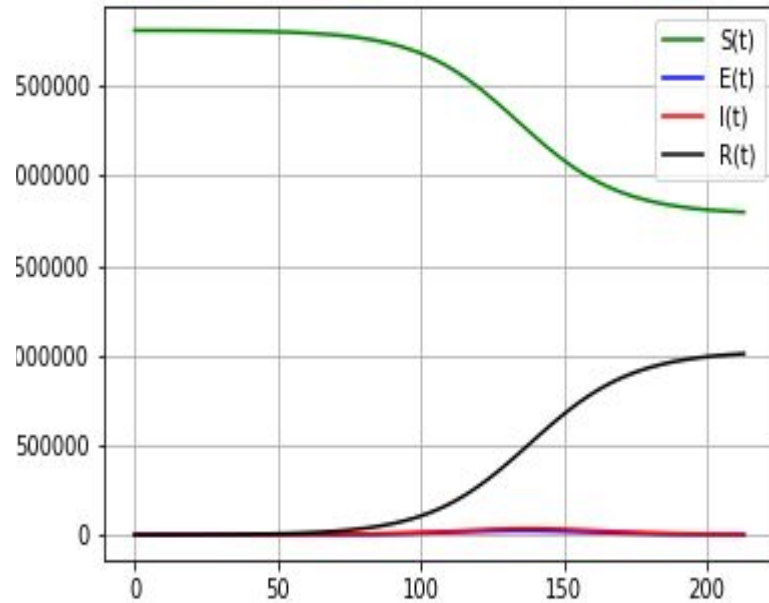
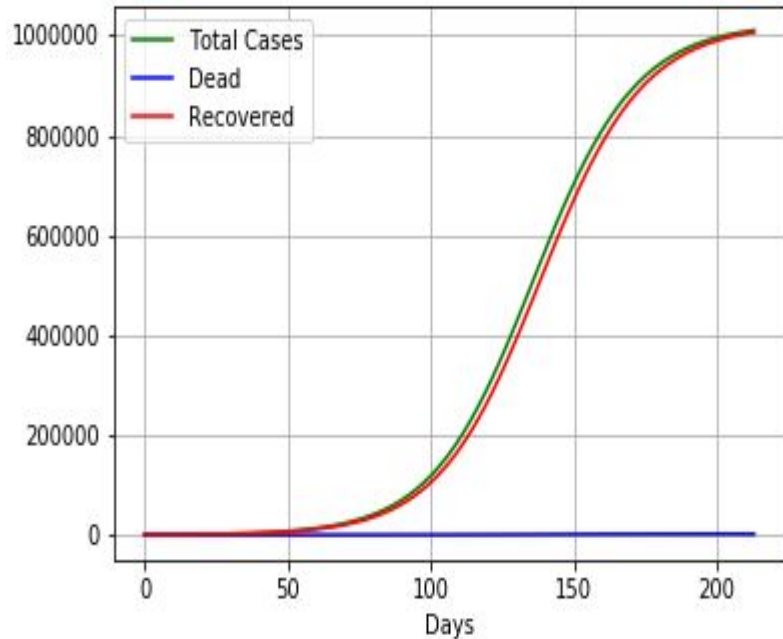
$$\beta = (r_0 \gamma)K = (r_0 \gamma)(0.311) = 0.083$$

$$\sigma = 1/10$$

$$\gamma = 1/15$$

$$r_{\text{eff}} = 1.245$$





Note : The modelled dead curve is got by finding out the fraction of dead cases w.r.t removed cases from real data.

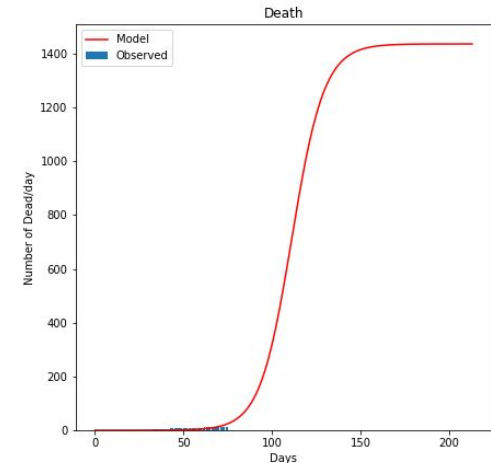
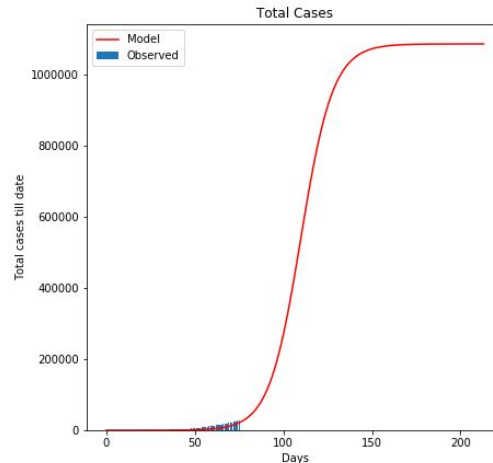
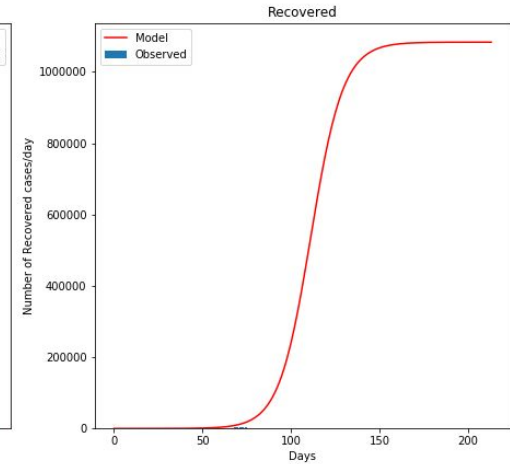
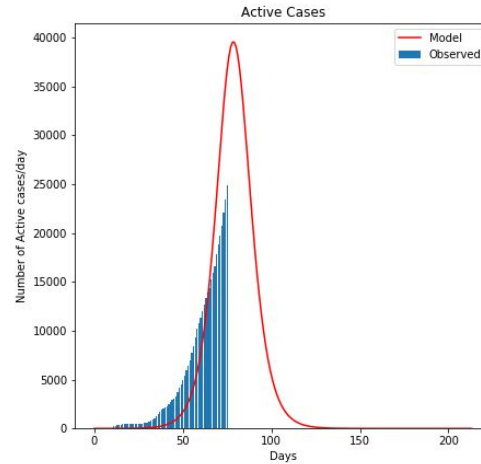
## How does our model predict?

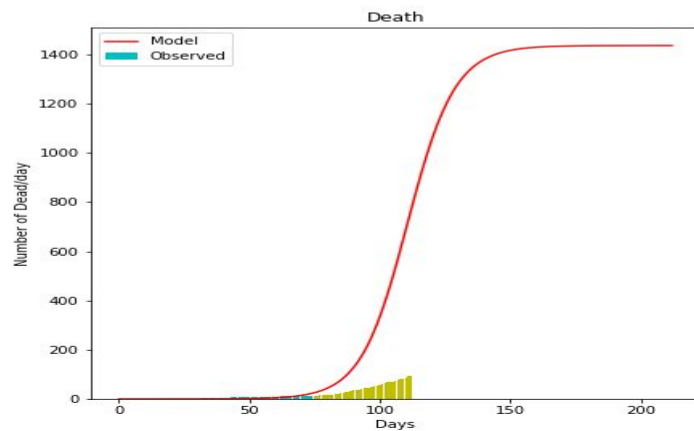
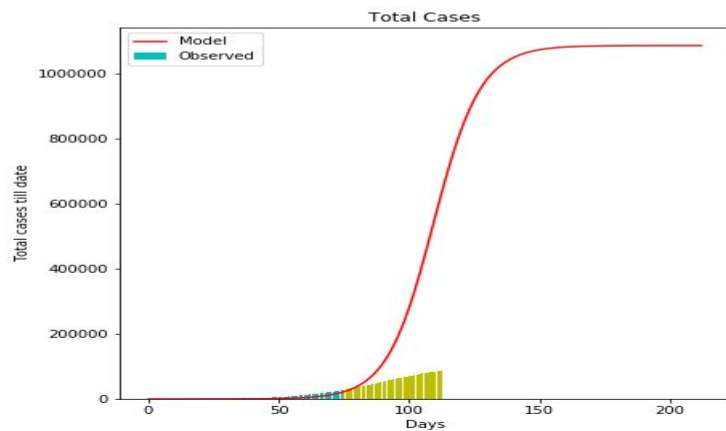
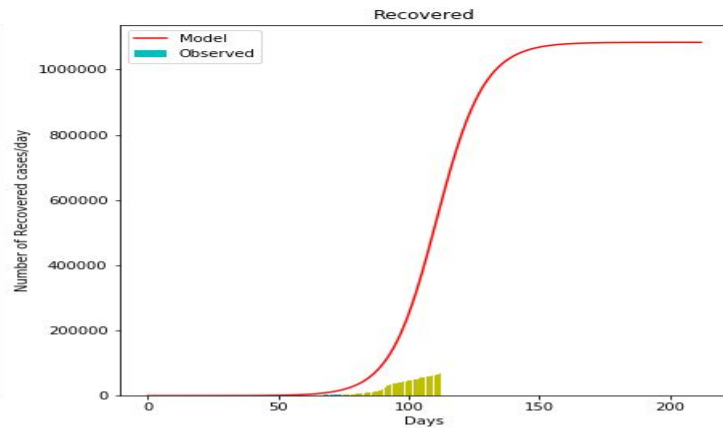
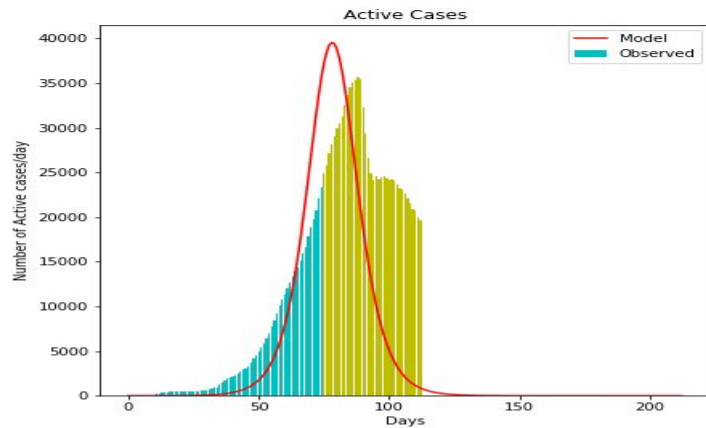
- We took training data : In real world this is the data collected till date. We want to predict what happens next
- Our training data : First 75 days of Qatar data.
- Again we will fit this by changing K value.

Parameters:

$$\beta = (r_0 \gamma)K = (r_0 \gamma)(0.318) = 0.0848$$

$$r_{\text{eff}} = 1.272$$





# Lockdown Analysis

- Effect of lockdown is seen in  $\beta$  . Stricter the lockdown, lesser the transmission rate
- Define 'Lockdown Parameter' ,  $\alpha : 0 \leq \alpha \leq 1$  , where ' $\alpha = 0$ ' is 0% lockdown and ' $\alpha = 1$ ' is 100% lockdown where no connections.
- We need  $\beta(\alpha)$  ?
- Firstly, to calculate different lockdown scenarios, we need
  - Training Data
  - Day on which you want to change  $\alpha$  ,i.e, change lockdown conditions ( $t_{LD}$ )
  - $\alpha$  until the day of lockdown ,  $\alpha_0$  . Note that this cannot be 1. There wouldn't be an epidemic in the first place
- From the data we fit to get  $\beta = \beta_0 = r_0 \gamma \cdot K_0$



## Continued...

- Stricter the lockdown, lesser  $\beta$ .
- Stricter the lockdown, greater  $\alpha$ , so less  $(1-\alpha)$ .
- Observe that  $(1-\alpha)$  changes in the same way as  $\beta$ . Since  $\beta = \text{const} * K$ ,  $K$  changes in the same way as  $(1-\alpha)$
- We know  $K_0$  and  $\alpha_0$ . So for any given  $\alpha$  to be imposed at  $t_{LD}$ , we will have

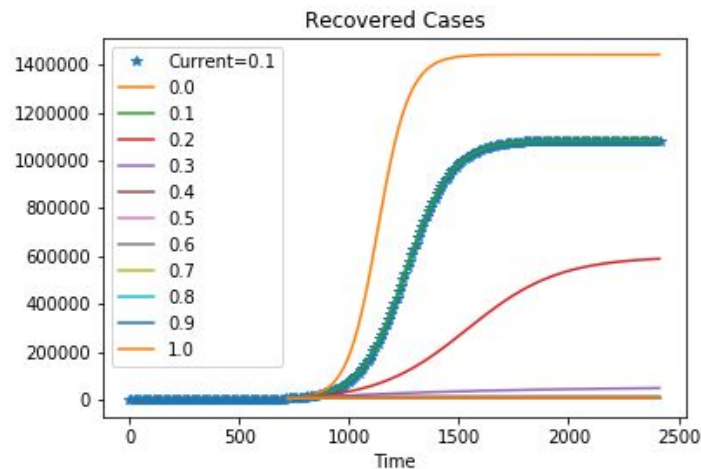
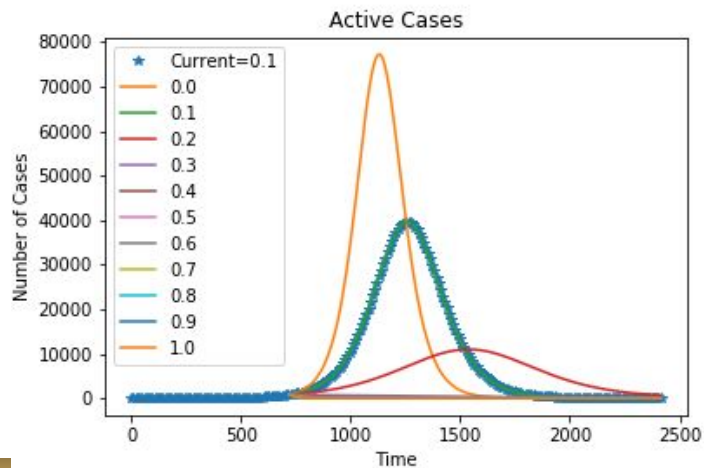
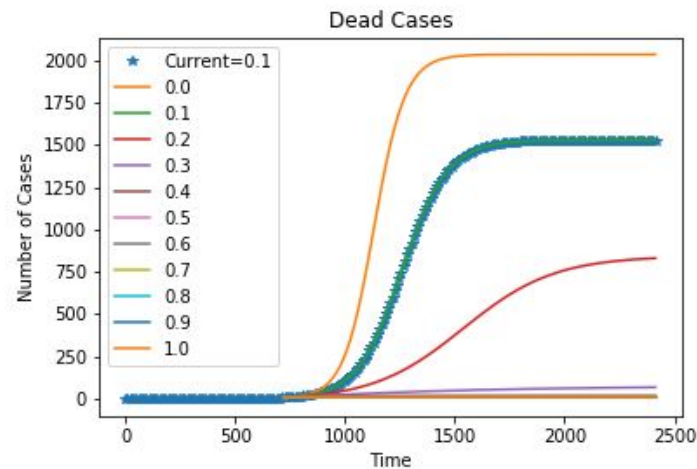
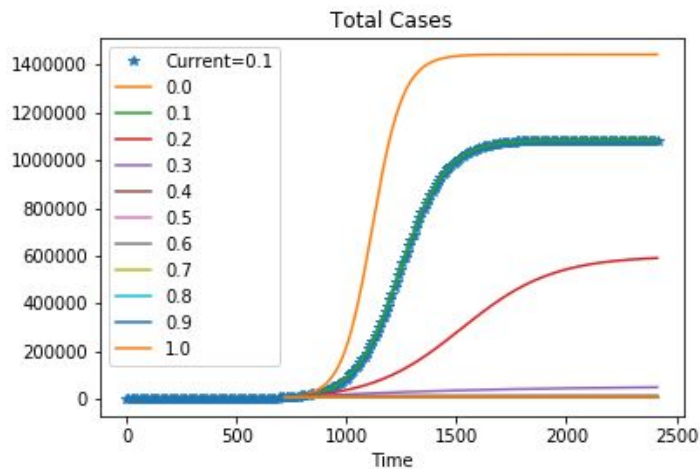
$$K = [K_0 * (1-\alpha)] / [1-\alpha_0] \Rightarrow \beta(\alpha) = r_0 \gamma [K_0 * (1-\alpha)] / [1-\alpha_0]$$

$$r_{\text{eff}} = r_0 [K_0 * (1-\alpha)] / [1-\alpha_0] = r_0 * K$$

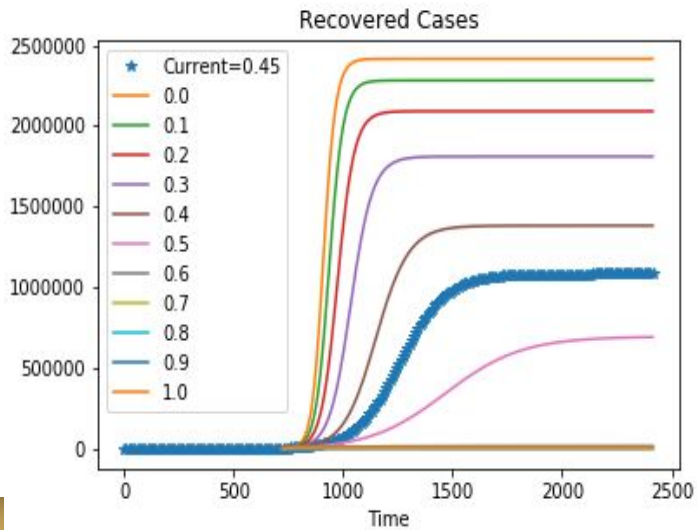
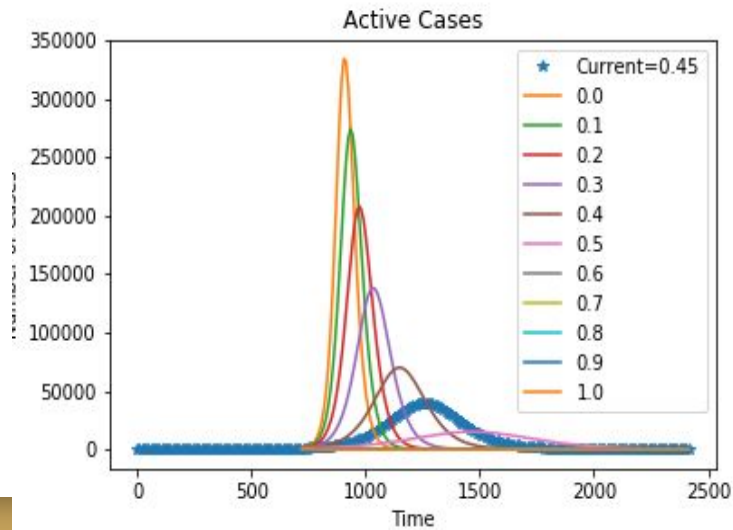
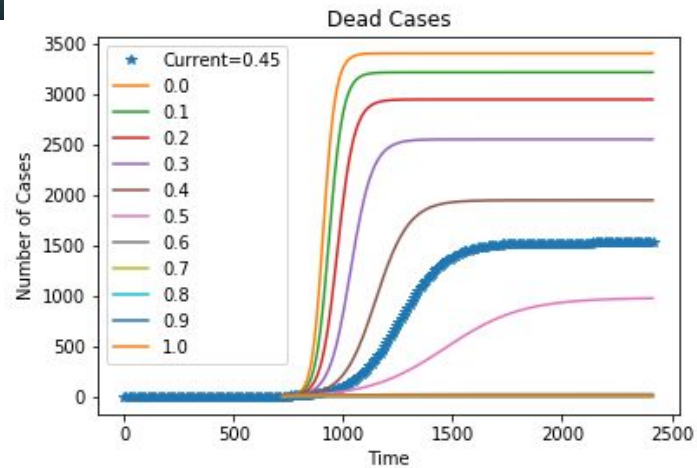
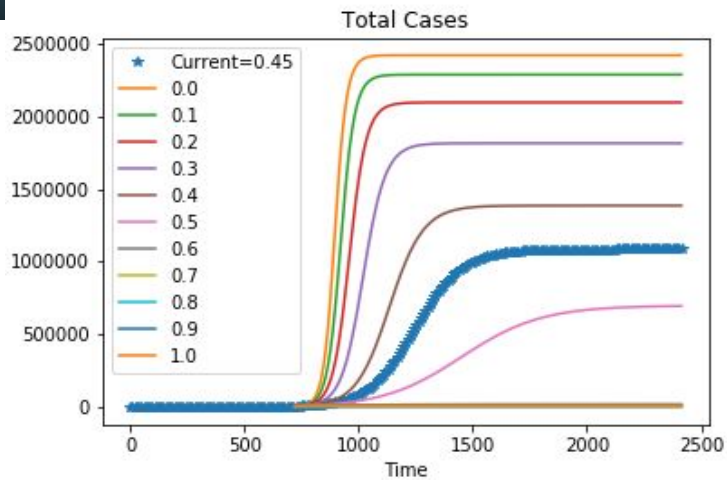
This reproduction number is achieved when lockdown with  $\alpha$ -strictness is imposed. We will plot for different  $\alpha$  from  $t_{LD}$  onward.

We took the same training data (so  $K_0$  is known by fitting it) and we gave it some  $\alpha_0 = 0.1, 0.45, 0.85$  and plotted for various  $\alpha$ .

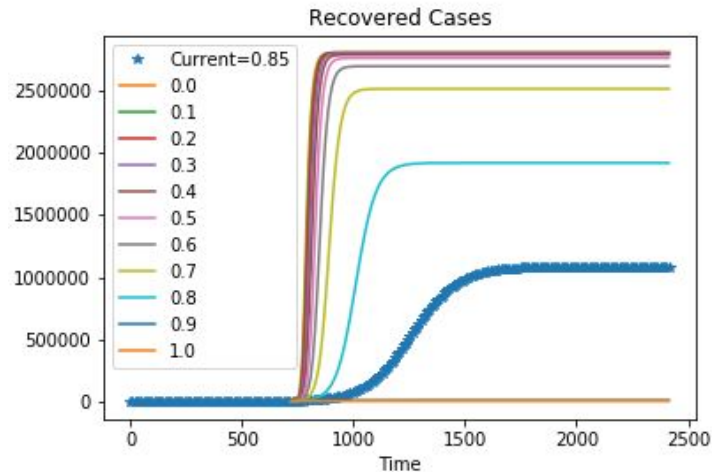
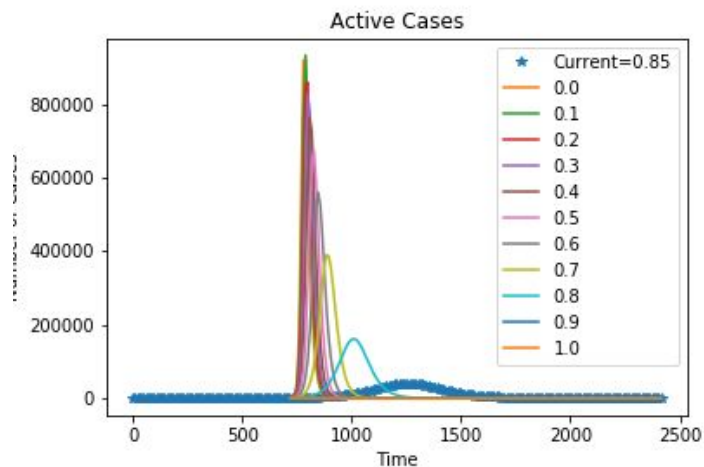
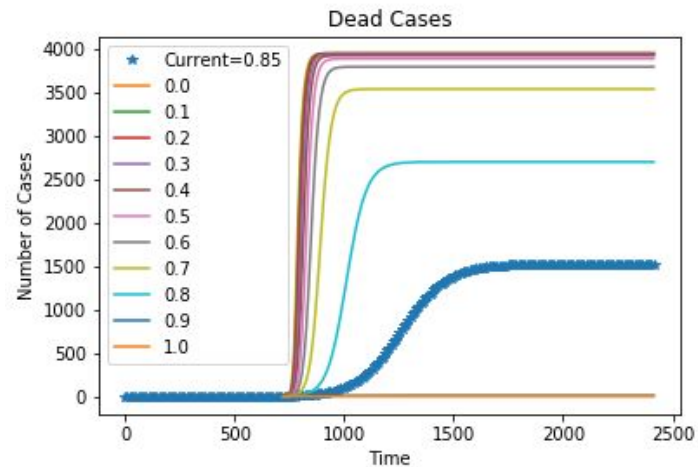
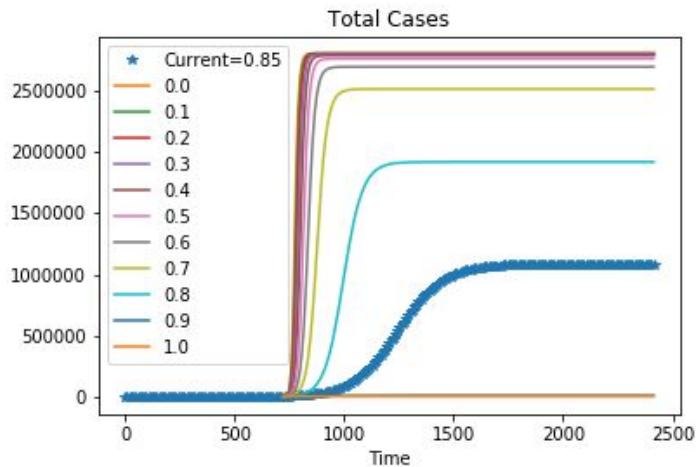
$$\alpha_0 = 0.1$$



$$\alpha_0 = 0.45$$



$$\alpha_0 = 0.85$$



# Shortcomings..

- Since this mathematical model has parameters which are based on the **average** only, it **can not really capture human interaction**. We are working with a collection of average values of parameter. So, it is bound to have error.
- Measuring  $\alpha$  is problem. We can't know what exactly  $\alpha$  means, policy wise.
- Implementation of the lockdown.  $\alpha$  has to be continuously monitored.
- Since the pandemic is going on, we do not have definite data to test and use for future pandemics.
- $N$  isn't constant for a given country.
- The parameters affecting  $\beta$  aren't exhaustive.

# SEIR Model on Barabasi-Albert Network

- We take a Barabasi-Albert Network of “ $n$ ” nodes with “ $m$ ” as no. edges that come with new vertex added to the existing network each time and it connects preferentially to those nodes having higher degree(follows preferential attachment algorithm).

## Idea:

- We will name our nodes(indexing) and divide them into 4 compartments(list) namely, Exposed(E), Infected(I), Removed(R), Susceptible(S) and track them eventually with time.

# Contd...

## Initialization:

1. Indexing of nodes of network.
2. Randomly choose one node and put it in **I**.
3. At time = 0, take no. of nodes in **E** = 0.
4. Also no. of individuals in **R** = 0.
5. **S** = Indexing set - (**E** + **I** + **R**).
6. Time = 0.

# Algorithm...

while (no. of Infected  $\neq$  0): {

1. if (no. of exposed  $\neq$  0)
  - Randomly choose nodes(person) from **E** with probability being different for each node(**sigma-influence** model).
  - Put selected ones in **I**.
2. Randomly choose individuals from **I** with some probability threshold(**gamma**) and put it in **R**.
3. Now put all those nodes, connected to nodes in **I**, in **E**.
4. Set **S** = **Indexing Set** - **E** - **I** - **R**.
5. Time += 1.

}



# Briefing of terms...

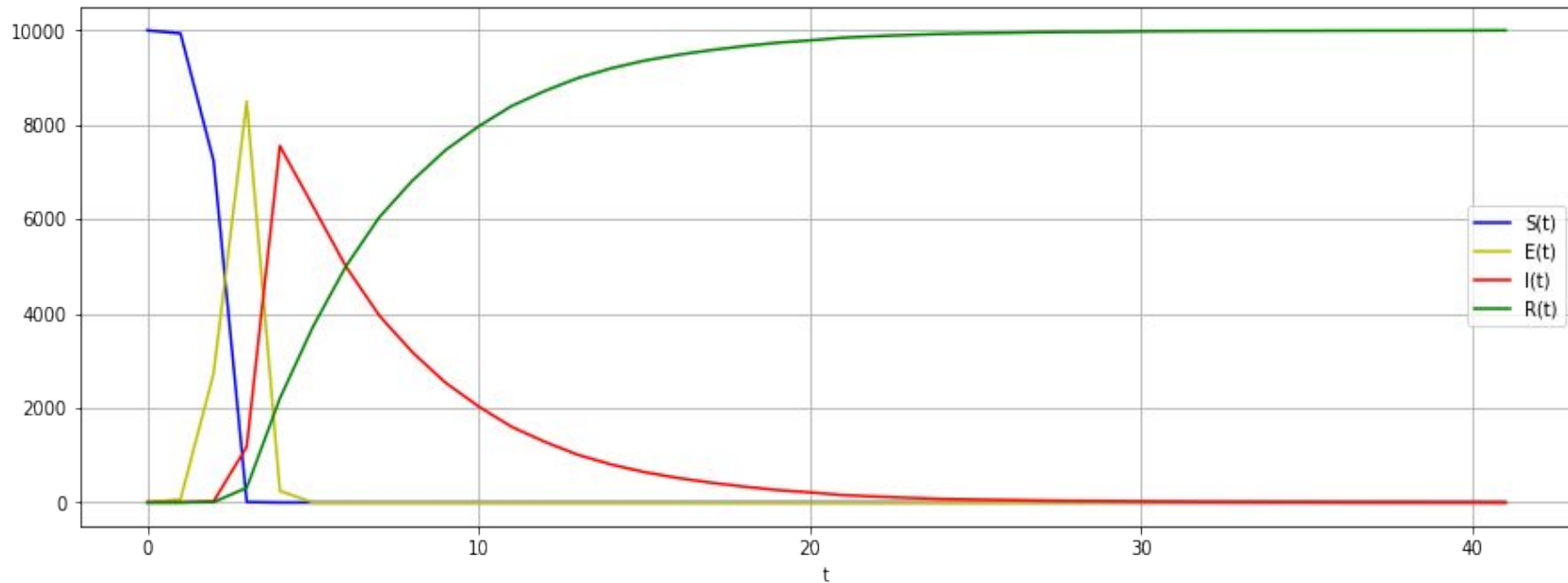
1. **Sigma( $\sigma$ )-Influence Model** : It shows the fact that the infection probability increases faster than a linear infection. In  $\sigma$ -influence model, the probability that a exposed node will be infected by one of its infected neighbors is  $\sigma$ , and the infections from all its infected neighbors are independent. So if a exposed node has  $n$  infected nodes at time 't', the probability that it will be infected at the next time by its  $n$  infected neighbors is

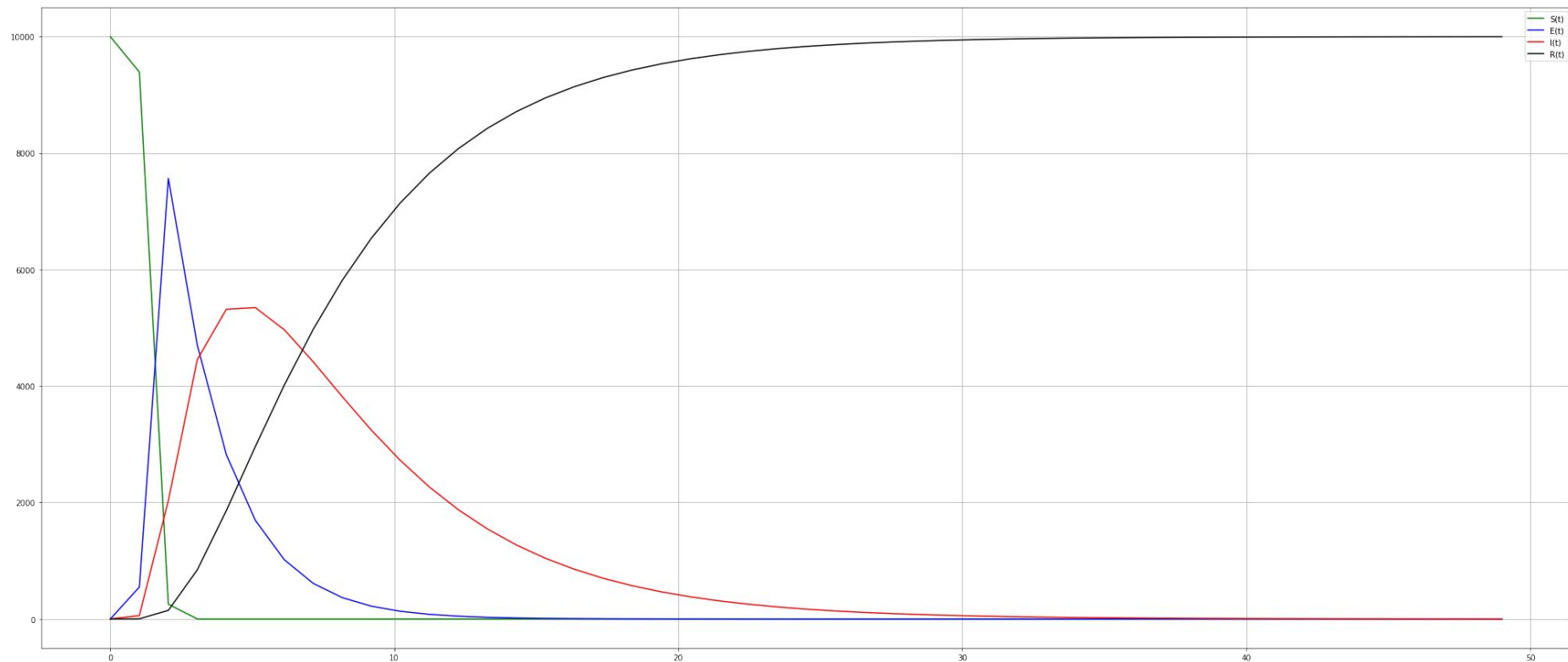
$$\text{Probability} = 1-(1-\sigma)^n$$

2. **Gamma( $\gamma$ )** : This gives us the “chance” of **Infected** entering into **Removed** group. It is a threshold kept between 0 and 1 and for each infected individual we generate a random number between 0 and 1. If that is less than the **threshold**, then infected goes in removed group.

# Plots for Networked-SEIR Model

$n = 10000$ ,  $m = 25$ ,  $\sigma = 0.5$ ,  $\gamma = 0.2$

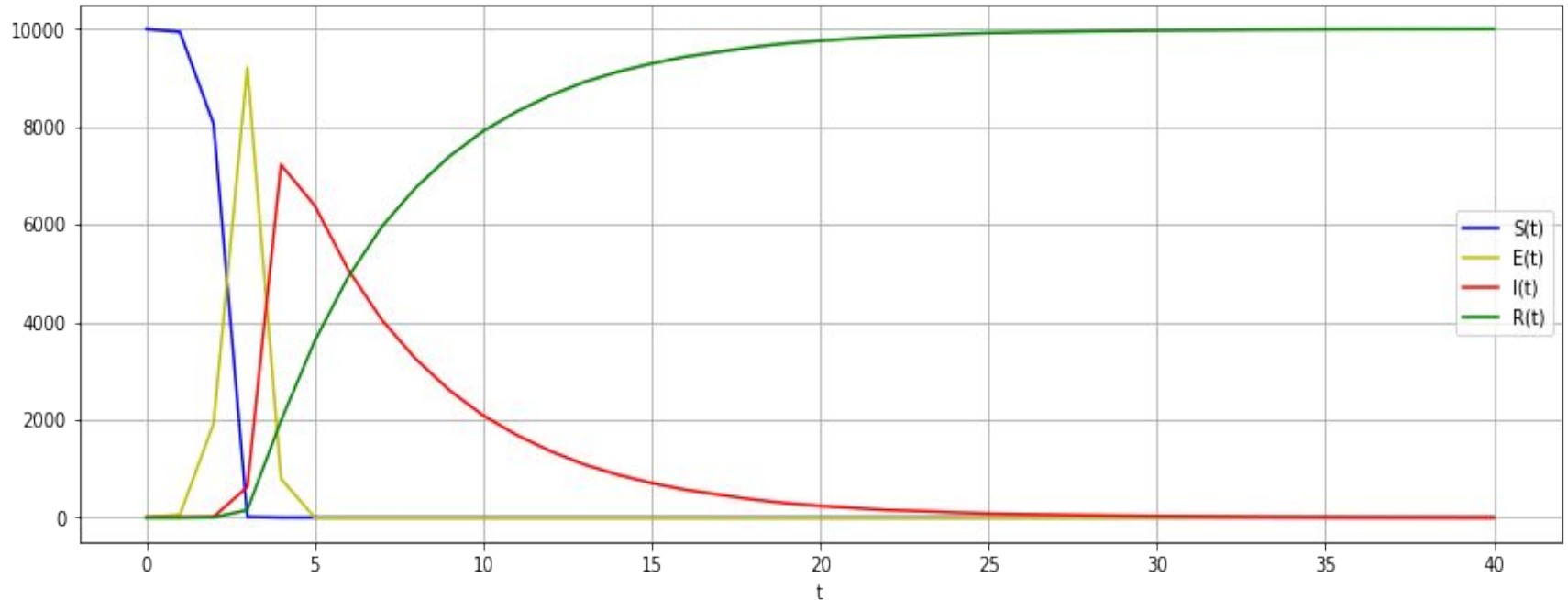


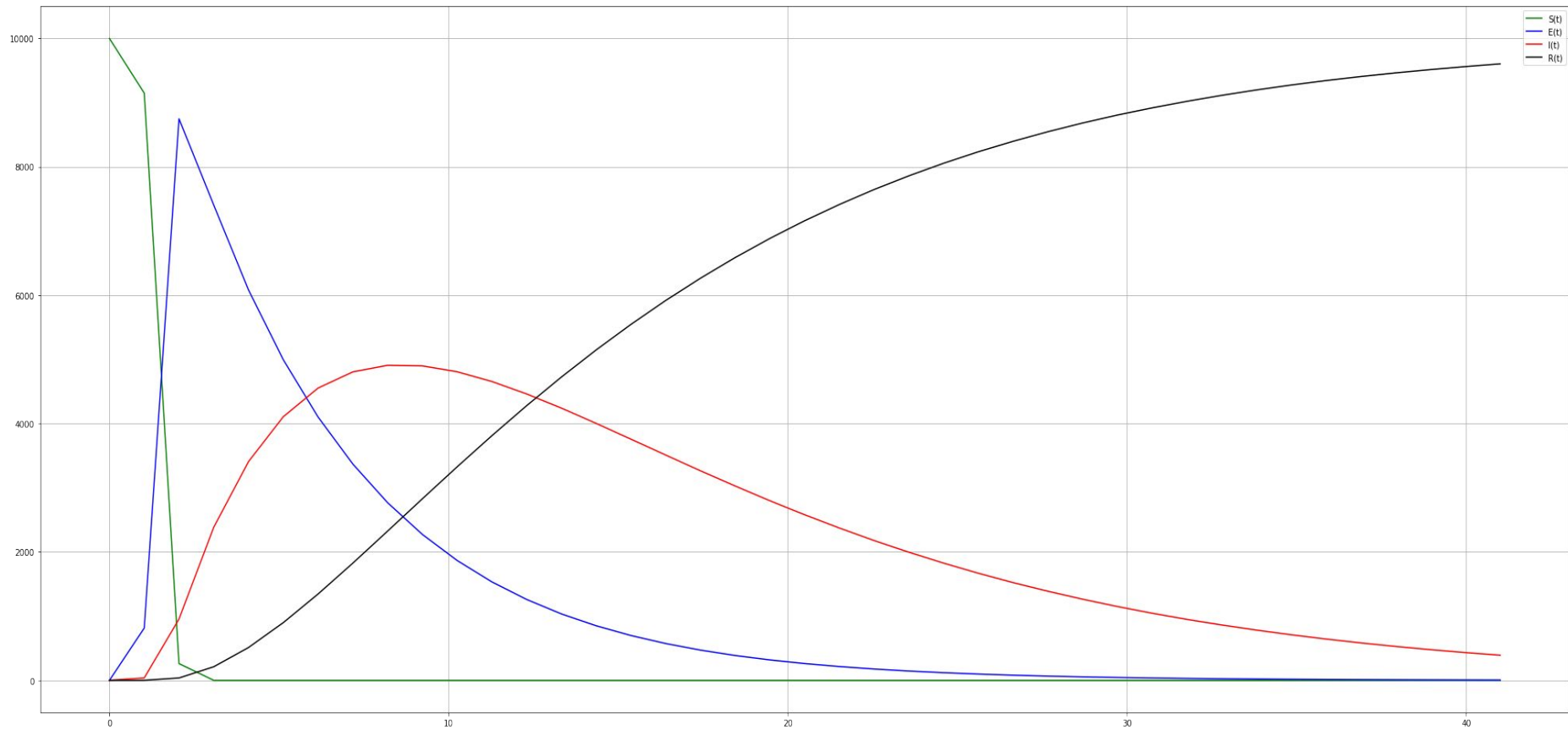


**Corresponding plot from solution for ODE with same parameters.**

# Another plot for different parameters...

$n = 10000$ ,  $m = 50$ ,  $\sigma = 0.35$ ,  $\gamma = 0.2$





**Corresponding plot from solution for ODE with same parameters.**

# Shortcomings...

1. Large country's data pose many more complexities while fitting because it involves the following reasons:
  - ❑ Large countries consists of large no. of **Communities**, where some of them are isolated and we take them into account while constructing **Susceptible** group. So our assumption of well-mixed fails.
  - ❑ Each region/province is not equally affected which again inculcates non-uniformity in data.
  - ❑ Pandemic is not over yet, and from the data we have, since it follows a non-linear pattern, it's hard to say when it will over. So fitting our model on **Real-Time** data put a big challenge before us. If we have complete data, then we are sure to capture non-linearity in our model by adjusting parameters accordingly.
2. Now coming to Network Part:
  - ❑ The Barabasi-Albert network we have used represents a single well-mixed community, which don't capture the whole city/country human interactions due to different communities. Hence our effort to fit **NY** data, which involves high density cases, fails.
  - ❑ And to take multiple **Barabasi-Albert** model in order to represent human-society, we're not fortunate enough to have sufficient computational power. Hardly we can do computations for **10,000** nodes, which again takes 20 min on average.

# Conclusion...

- ❑ In pandemic like this, providing timely information to the public is paramount. A better modelling of COVID-19 will assist govt. and authorities to disseminate verified articles, provide updates to the situation and advocate good personal hygiene to the people.
- ❑ And also this will spread awareness to the common people by providing scientific-based data analysis, prediction and verified news.
- ❑ Coming up with good models is the need of the hour. One has to have a sound knowledge of Economics, Public-Policy along with the necessary tools from Machine Learning and Network analysis to improve the existing ones.

# References...

- ❖ **Data Source(Kaggle):**

[Novel Corona Virus 2019 Data set](#)

- ❖ [Corona Tracker: World-wide COVID-19 Outbreak Data Analysis and Prediction](#)
- ❖ Reproduction number analysis follows from article below.

[Reproduction numbers of infectious disease models](#)

- ❖ Sigma-Influence Model we used on Barabasi-Albert network is inspired from following article.

<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5051690/#:~:text=Social%20contact%20networks%20are%20the,interactions%20which%20may%20cause%20infections.>





*THANK YOU*