Technical University of Crete

Diploma Thesis

Estimating epidemiological parameters on population networks

Author:
Michael Chatzinikolaou

Supervisors:
Dr. Samoladas Vasileios
Dr. Manoussaki Daphne
Comittee member:
Dr. Chalkiadakis Georgios

Ευχαριστίες

Αρχικά θα ήθελα να ευχαριστήσω τα μέλη του διδακτικού προσωπικού Δρ. Σαμολαδά και Δρ. Μανουσάκη, χωρίς τους οποίους δε θα ήταν εφικτή η εκπόνηση της παρούσης εργασίας. Ειδικότερα, την κ. Μανουσάκη για την ακούραστη συνεισφορά της στην επιστημονική επιμέλεια του κειμένου και τον κ. Σαμολαδά για την πολύτιμη καθοδήγησή του σχετικά με τις μεθόδους και την πορεία της εργασίας.

Θα ήθελα επίσης να ευχαριστήσω τους γονείς μου, οι οποίοι με στήριξαν οικονομικά μέχρι να ολοκληρωθεί η εκπόνηση της εργασίας και τους φίλους και συναδέλφους Χρήστο και Σόλωνα, οι οποίοι μου στάθηκαν σαν αδελφοί και οδηγοί έναντι σε κολλήματα που αντιμετώπισα στην πορεία της εκπόνησης της εργασίας μου.

Τέλος, η συνεισφορά στην εκκίνηση της εργασίας των κ. Μουμουτζή Ν. και Σακκά Γ. είναι καθοριστική, και για αυτόν τον λόγο τους είμαι ευγνώμων.

Abstract

The COVID-19 pandemic gave rise to an increase of research related to its epidemiology. Simultaneously, efforts were made to quantify data related to the epidemic.

This thesis is an attempt to model networks of populations as random graphs, with various epidemiological parameters, and try to calculate those parameters given the outcome of the epidemic. The first part of the experiment consists of the implementation of the model, while the second part is the attempt to calculate the parameters on artificial data created by said network. Finally, the sensitivity of the methods is tested against added noise.

Given a network of self-contained populations as nodes, the daily travel between the nodes, and the progress of the epidemic on each node, we can estimate the epidemiological parameters of the node. This shows that interference from travel of infected individuals to nodes affects the estimation of epidemiological parameters of the node only for relatively large values of such individuals.

Table of Contents

List of Figures

1 Introduction

The COVID-19 pandemic gave rise to an increase of research related to epidemiology [1]. From the mathematical and computing point of view, epidemic analysis aims at estimating epidemiological variables such as infections, deaths, immunizations or other attributes of interest on a daily basis or throughout the progress of the epidemic [2]. The aim of this thesis is to model an epidemic on multiple populations on a network of populations, drawing inspiration from the previous work of [3] and [4].

The approach that will be followed is a mathematical approach, also known as the *analytical* approach. This is the compartmental model approach, which employs the use of a system of ordinary differential equations, which, once solved, provide the solutions for the populations that we are interested for. An extensive list of other approaches that can be seen in [5].

The goal of this thesis is to measure at what extend the epidemiological parameters can be estimated on a network, using the SIR model and a simple, yet hopefully useful, random weighted directed graph model.

1.1 Thesis Outline

In the first section we analyze the compartmental models. These are models, initially proposed by Kermak-McKendrick, that have been used thoroughly through the past century and are still being used to model epidemics.

In the second sections, networks and their mathematical basis, graphs, are being presented to the reader.

The third section presents the simulation methods used to create the experiment.

In the fourth and the final sections we see how accurate and tolerant to statistical noise the estimations are, and we discuss the results.

2 The compartmental models

2.1 The Kermack-McKendrick model

One of the oldest approaches to modeling epidemic processes is that of the compartmental models of Kermack and McKendrick [6, 7, 8].

These articles make some assumptions that must be noted:

1. Each individual is equally susceptible to the disease (regardless of age or other factors).

2. The epidemic is short-lived, meaning its duration is short compared to the individual's lifespan.

3. A single infection leads to complete immunity, meaning that once a person is infected, it cannot get infected again.

The paper begins by using a variable $u_{t,\theta}$ to group individuals that have been infected at time t for θ duration of time, with time being discrete. It is then shown that the population can eventually be separated in three variables that denote the fraction of population that belongs to one of three groups on moment t: x_t for those who are not infected as of time t, y_t for those who are currently infected and z_t for those who are removed due to recovery or death.

Under the assumption that the population, as well as and infection and recovery rates remain constant, the paper arrives at the model that later became the SIR model.

It is shown, in the same paper, that under these assumptions the epidemic begins only if the population density is above a certain threshold value and that it terminates, as will be shown in the following section, if a certain relation between population density and other model parameters holds. The size of the epidemic, meaning the total infected individuals and the peak of infections can also be calculated. We can see these calculations and approximations in the later sections, regarding the SIR model.

2.2 The SIR model

In later years, the Kermack McKendrick model became known as the SIR model [9]. The SIR model separates the population into three compartments, S for Susceptible, I for infectious and R for removed individuals (either by death or immunization). The transition of individuals between compartments is modeled as a set of ordinary differential equations (O.D.E.s), describing the dynamic system of the epidemic. Individuals belonging to the S compartment can get infected and therefore become infectious, individuals in the I compartment transmit the disease to S, while R individuals cannot be infected, be it

because of immunity or death.

The SIR model can be described by using the following system of equations:

$$\frac{dS}{dt} = -\beta SI \tag{1}$$
$$\frac{dI}{dt} = \beta SI - \gamma I \tag{2}$$
$$\frac{dR}{dt} = \gamma I \tag{3}$$

Note that each compartment's population is normalized to the total population, which remains constant. This means that S, I and R have values in the range $[0, 1]$. β and γ represent the rate with which individuals transition respectively from the susceptible to the infectious and the infectious to the removed compartments per individual, per day.

We can schematically represent this process (transfer diagram), as such (Fig 1):

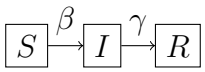

Figure 1: The SIR model.

Solutions to the model above yield results that look like Fig. 2.

A question that was mentioned in the previous section, is whether an epidemic can happen given an initial population I_0 of infected individuals. Only at the start of the epidemic, if we assume that $I_0 > 0$, $S_0 > 0$, $R(t = 0) = 0$, we can deduce that, in order to have an epidemic onset:

$$\frac{dI_0}{dt} > 0 \Leftrightarrow \beta S_0 I_0 - \gamma I_0 > 0 \Leftrightarrow S_0 > \frac{\gamma}{\beta}.$$

In the opposite case, where $\frac{dI_0}{dt} < 0 \Leftrightarrow S_0 < \frac{\gamma}{\beta}$, the epidemic dies out. It is thus derived that the onset of the epidemic depends on the relation between the initial percentage of susceptible individuals, S_0 and the ratio $\frac{\gamma}{\beta}$. This is called a *threshold phenomenon* [10].

There is another useful metric, the ratio between β and γ, better known as $R_0 = \frac{\beta}{\gamma}$, which expresses the rate of new infections for every new individual that enters the *removed* compartment. This is also called the *basic reproduction number*.

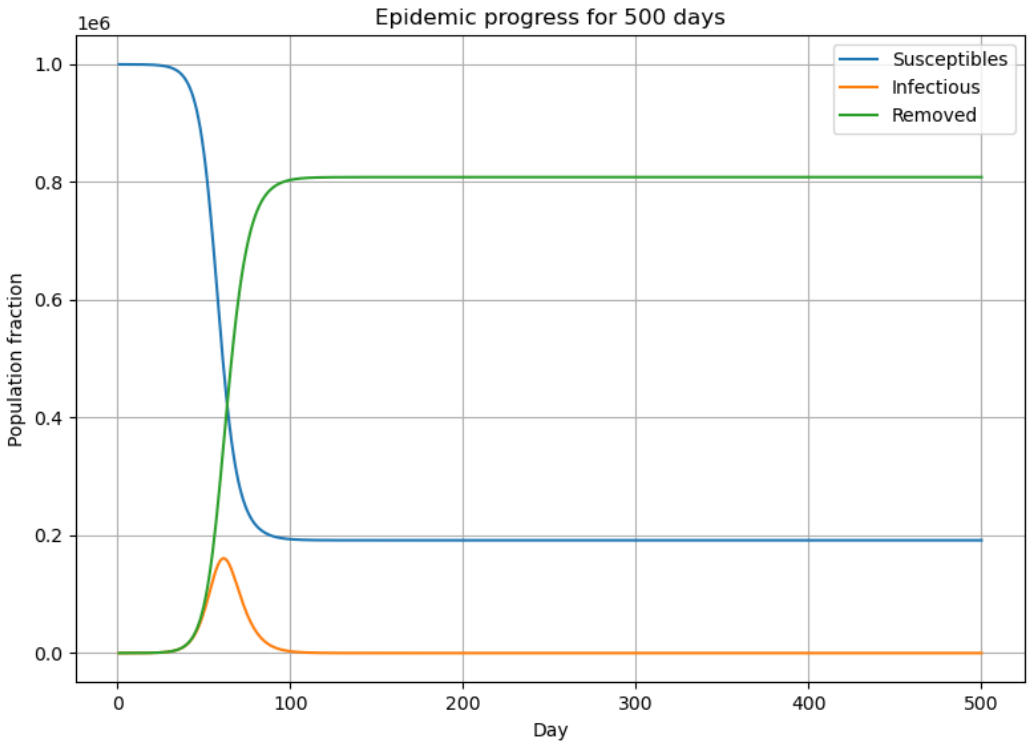

Figure 2: A sample solution to the SIR model. $\beta = 0.4, \gamma = 0.2, R_0 = 2$.

It can be shown from equation (2) that the epidemic reaches this peak for:

$$\frac{dI_m}{dt} = 0 \Leftrightarrow \beta S_m I_m - \gamma I_m = 0 \Leftrightarrow$$
$$I_m(\beta S_m - \gamma) = 0 \Rightarrow I_m = 0 \; or \; S_m = \frac{\gamma}{\beta} = \frac{1}{R_0}, \tag{4}$$

where I_m is the peak I value and S_m the value of S when $I = I_m$.

We obviously keep the second solution, since we must assume that the epidemic(if it happens) is at its peak (thus $I_m \neq 0$). It is also shown in [11] that the peak infected individuals population can be given as:

$$I_m = -S_m + I_0 + S_0 + S_m ln(S_m/S_0) \tag{5}$$

where I_0 and S_0 denote the initially infected and susceptible population, respectively.

Let's suppose we want to know how I_m varies in relation to R_0. We can take the partial derivative of I_m from Eq. (5) in relation to R_0:

$$\begin{aligned} I_m = I_0 + S_0 + S_m ln(S_m/S_0) - S_m &= I_0 + S_0 + \frac{1}{R_0} ln\big(1/(S_0 R_0)\big) - \frac{1}{R_0} \Rightarrow \\ \frac{dI_m}{dR_0} &= -\frac{ln(1/R_0 S_0)}{R_0^2} \end{aligned} \tag{6}$$

We also know that, for an epidemic to happen, $S_0 > \frac{1}{R_0}$. Thus:

$$\begin{aligned} S_0 &> \frac{1}{R_0} \Rightarrow 1/(S_0 R_0) < 1 \Rightarrow ln\big(1/(S_0 R_0)\big) < 0 \Rightarrow -\frac{ln(1/R_0 S_0)}{R_0^2} > 0 \Rightarrow \\ \frac{dI_m}{dR_0} &> 0. \end{aligned} \tag{7}$$

Thus, I_m increases as R_0 increases.

One approximation of the time when the epidemic reaches its maximum number of infected individuals, given that the initial population of infected individuals is small ($i_0 << s_0$)[11], is:

$$t_p = \frac{1}{\gamma} \frac{1}{R_0 - 1} ln\big(R_0 + I_0^{-1}(R_0 - 1)^2\big) \tag{8}$$

which, according to the same paper, is derived when one solves

$$t(S) = \frac{1}{\beta} \int_{S_0}^{S} \frac{1}{S(S - S_m ln(S/S_0) - I_0 - S_0)} dS \tag{9}$$

for $S = Sm$.

More compartmentals can be seen in Appendix A.

2.3 Other methods

Apart from the compartmental models, other methods (mainly agent-based) are utilized for the solution of epidemic systems [12]. The idea of two-level mixing, as can be seen in [3], separates the study of the epidemic in two levels: This of the spread inside an isolated population, and the spread between those isolated populations. The same idea can be seen in the paper [13], which will be analyzed in the next chapter.

So far we've seen how epidemic spreading has been modeled in a single, closed population using the compartmental models. Before combining multiple SIR models in our simulations, we first analyze the basics of graph theory. The main idea behind their employment is the separation of populations in different groups, be it islands, cities, countries etc..

AI advancements including multi-agent approaches to epidemics can also be seen in senior Thesis works, such as the work of Mr. Voloudakis [14].

3 Graph Theory

3.1 Graphs

A *graph* is a set of *vertices* **V** and *edges* $E \subseteq \{\{i, j\} : i, j \in V \text{ and } i \neq j\}$. We denote the graph **G** as $G = (V, E)$. An example of a graph with 6 vertices and 7 edges can be seen in Fig. 3.

A *directed* graph is a graph in which an edge is an ordered set, i.e. given the set of vertices V, the edges set is now defined as: $E \subseteq \{(i, j) : i, j \in V \text{ and } i \neq j\}$. Note that now the pair of vertices in the definition is wrapped inside parentheses, instead of brackets.

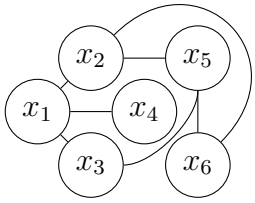

Figure 3: A graph with 6 vertices and 7 edges.

In an undirected graph, the *degree* of a node is the number of edges connected to it. In directed graphs, we distinguish the in-degree as the number of edges pointed *to* the node, versus the out-degree, which is the number of edges pointing *from* the node of reference to other nodes.

The mean degree of a graph is the mean degree of its nodes, or

$$[k] = \sum_{i \in [1,N]} \frac{deg_i}{N}$$

.

The edges of a graph can also have attributes related to them. One such attribute is the *weight*, which is a number describing something useful. In this thesis, the weight of the edges is the number of people that travel between nodes.

A *random graph* $G(n, p)$ is a graph with n nodes in which each possible edge between two nodes exists with probability p. This graph can be created using a computer, with the help of randomized processes.

According to [15], there are three different types of random graphs when it comes to connectivity, which depends on the value of $[k]$: Critical ($[k] \approx 1$), where the graph is either connected or not connected depending on the experiment run, sub-critical ($[k] < 1$) where

the graph is not connected and super-critical ($[k] > 1$). This depends on the *degree* of the random network, which can be calculated to be $[k] = p(N-1)$.

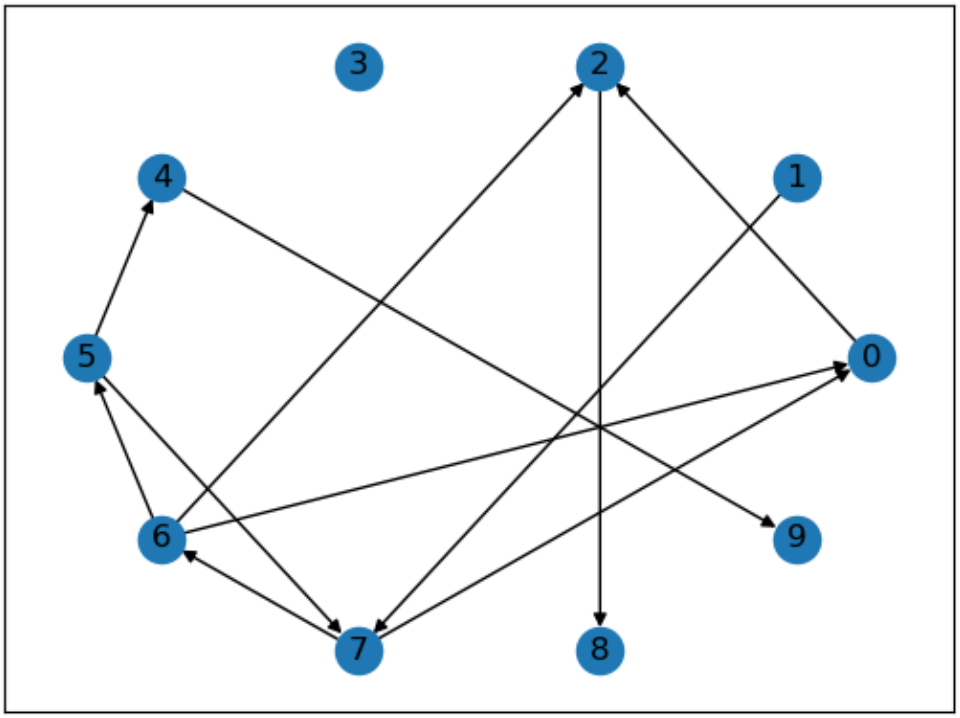

Figure 4: A random directed graph with $n = 10$ and $p = 0.1$

3.2 Epidemics on networks

Networks are being used extensively in epidemiology [5].

One aspect of modeling epidemics on a network is the connection between human communities. Those networks' description is no simple task and requires a cross-disciplinary approach [16].

4 Multiple node simulation results

Following are experiments with other graph sizes and degrees:

The mean degree of the nodes varies and can be found on each figure's title. For each node: $\beta = 0.5, \gamma = 0.05, I_0 = 10\text{idv}, R_0 = 0$. The population of each node i is $N_i = \frac{10^7}{\#\text{nodes}}$

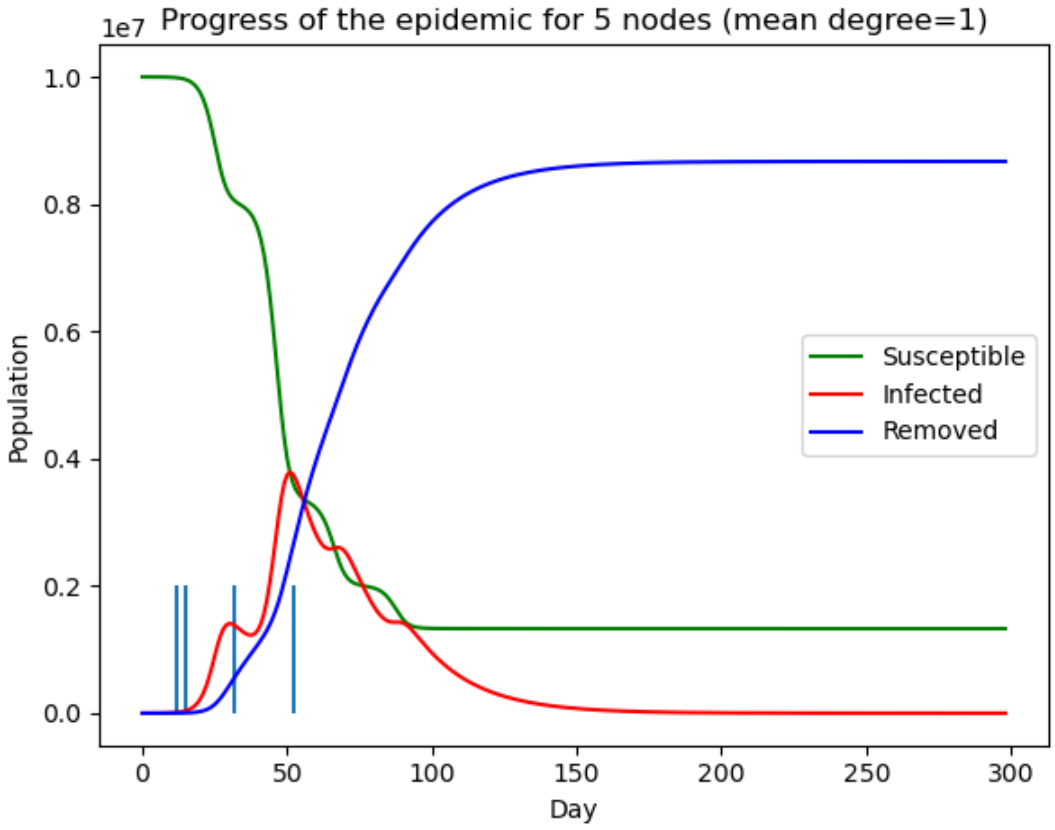

Figure 5: Epidemic progress on 5 nodes with mean degree 1. Critical region.

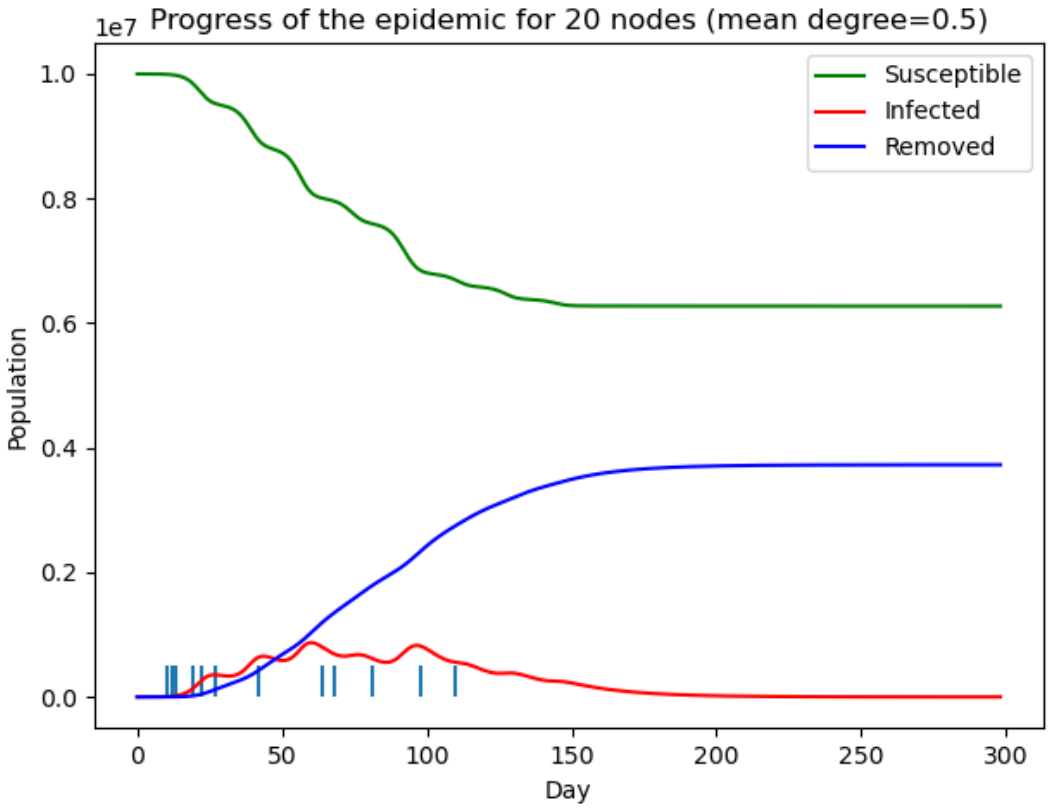

Figure 6: Epidemic progress on 20 nodes with mean degree 0.5. Sub-critical region.

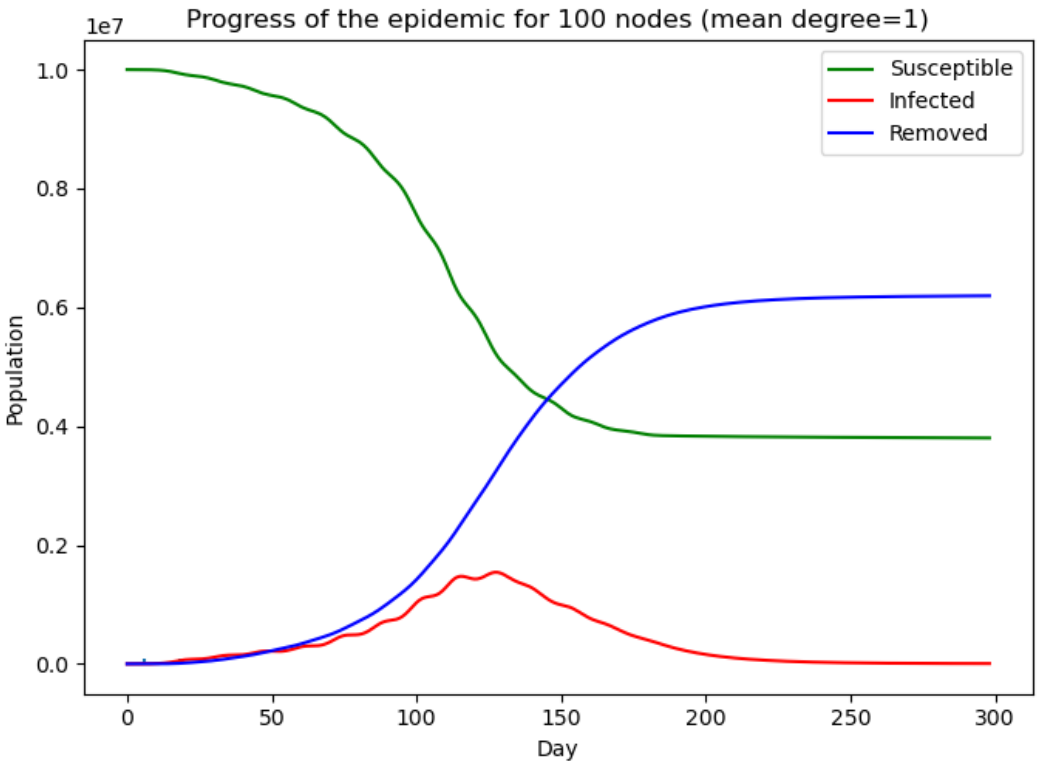

Figure 7: Epidemic progress on 100 nodes with mean degree 1.

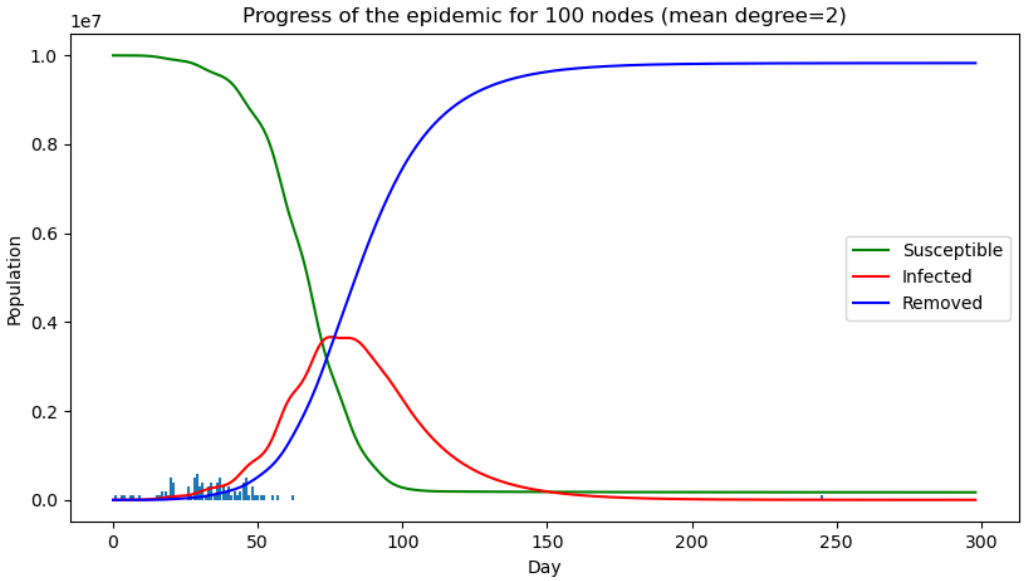

Figure 8: Epidemic progress on 100 nodes with mean degree 2. Super-critical region.

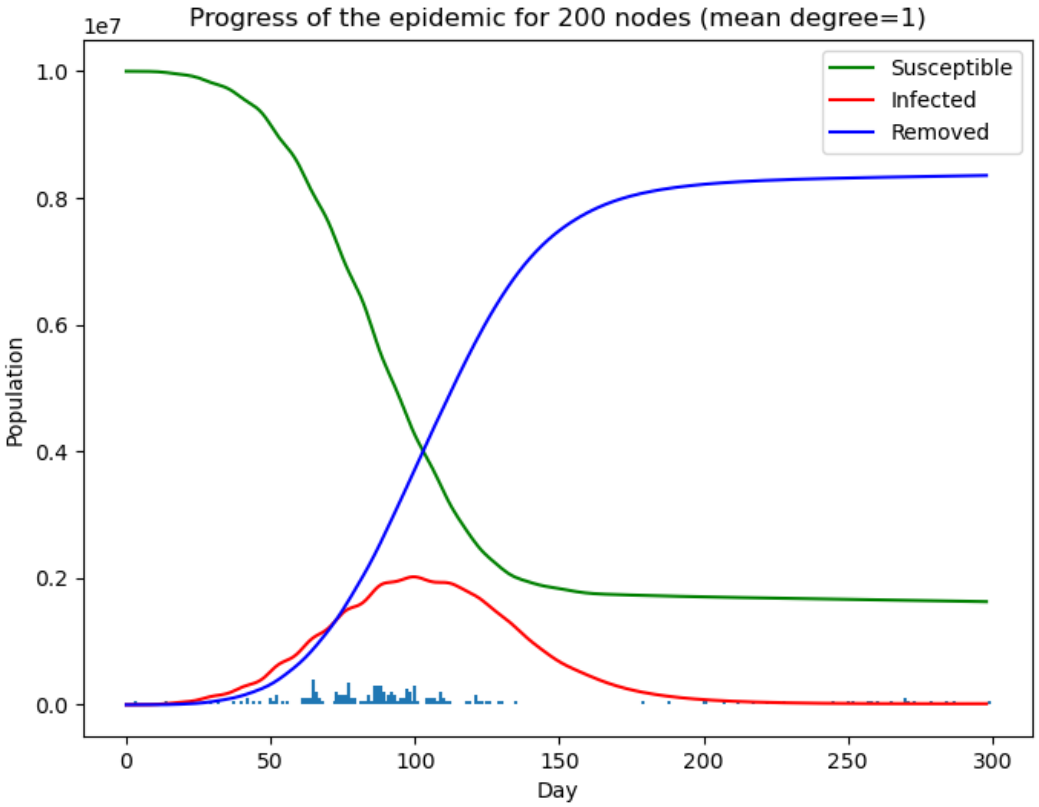

Figure 9: Epidemic progress on 200 nodes with mean degree 1.

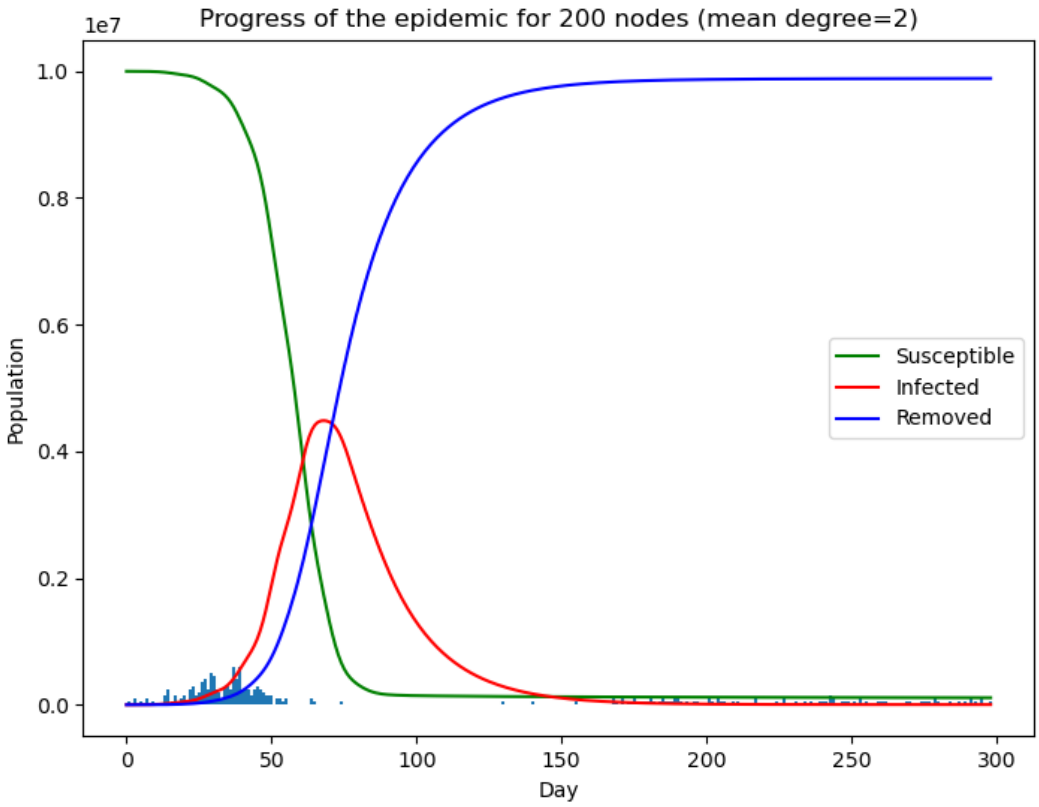

Figure 10: Epidemic progress on 200 nodes with mean degree 2.

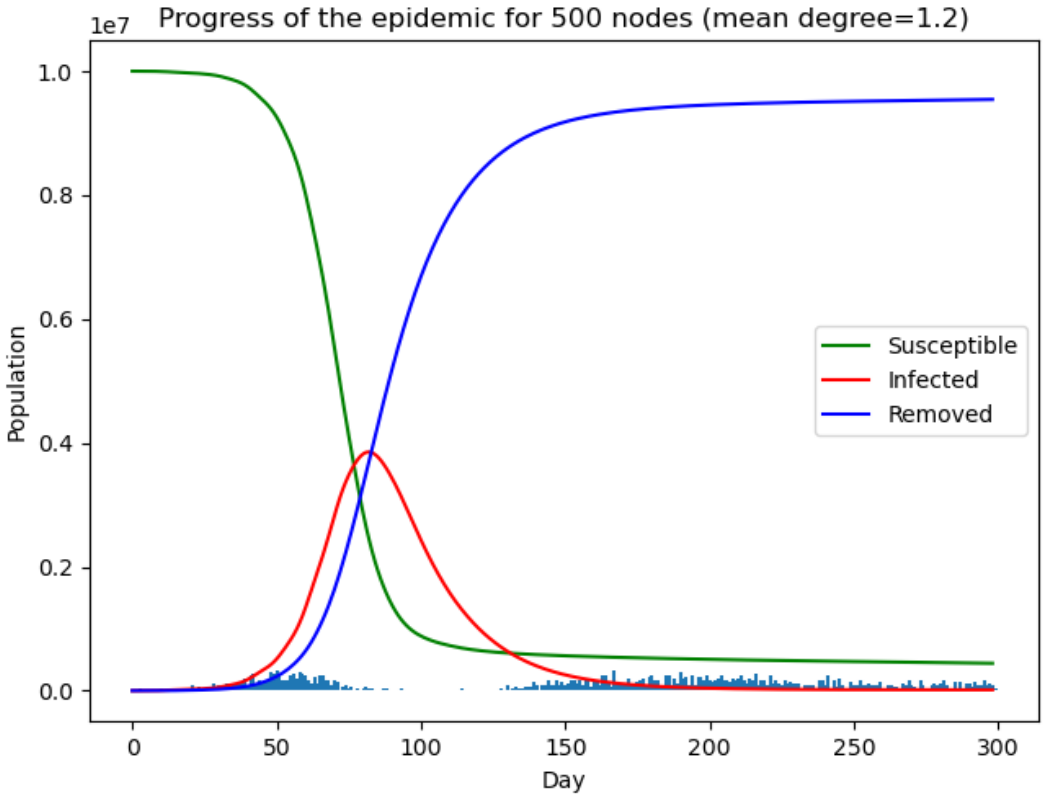

Figure 11: Epidemic progress on 500 nodes with mean degree 1.2.

It is clear that high degrees of nodes lead to a steeper peak of the epidemic, while sparser networks lead to waves, and a flatter curve. One can also observe the threshold phenomenon, since for lower graph connectivity we have almost no epidemic, while for higher values we have a full epidemic.

4.1 An SEIR example

In [4], the authors explore a network using the SEIR model, combined with population migration between infected cities.

During the outbreak of the late 2019-early 2020 coronavirus epidemic in mainland China, the Chinese government took measures such as quarantine and travel restrictions to minimize the spread of the epidemic. On this paper, the researchers evaluate the effect these measures had in restricting the spread and total size of the epidemic, using two different methods: solving the epidemic numerically using an extended SEIR model and training a Neural Network on the 2003 SARS epidemic and then using it to predict this epidemic's

outcome. For the purpose of modeling the epidemic, the researchers use an extended SEIR model which includes population migration between provinces. In the following equations, the additional $S_{in/out}(t)$ and $E_{in/out}(t)$ variables model the inflow and outflow of migrant workers from and to other provinces. The system is described as such:

$$
\begin{aligned}
S[t+1] &= S[t] + S_{in}[t] - S_{out} - \frac{\beta_1 \times r[t] \times I[t] \times S[t]}{N[t]} - \frac{\beta_2 \times r[t] \times E[t] \times S[t]}{N[t]} \\
E[t+1] &= E[t] + E_{in}[t] - E_{out} + \frac{\beta_1 \times r[t] \times I[t] \times S[t]}{N[t]} + \frac{\beta_2 \times r[t] \times E[t] \times S[t]}{N[t]} - \sigma E[t] \\
I[t+1] &= I[t] + \sigma E[t] - \gamma I[t] \\
R[t+1] &= \gamma I[t] + R[t] \\
S_{in} &= In[t] \times (1 - P_{out}[t]) \\
S_{out} &= Out[t] \times (1 - P_{out}[t]) \\
E_{in}[t] &= In[t] \times P_{out}[t] \\
E_{in}[t] &= Out[t] \times P_{out}[t]
\end{aligned}
$$

Here, β denotes the frequency of transition from susceptible to exposed via contact with exposed people (who can transmit the virus but do not have symptoms), γ the rate of illness resolution (either death or recovery) and σ the incubation period. $r(t)$ denotes the contacts per person of each individual per day. With $P_{out}[t]$ the authors denote the probability of the out-flowing people being exposed. Thus, $In[t]$ stands for the total incoming people, while $Out[t]$ for the total out-flowing people.

This model takes into consideration the movement between different provinces. The system is not "closed", i.e. the population on each province does not remain constant, but the In and Out variables denote movement to and fro the province they reference.

Drawing inspiration from the aforementioned paper [4], we use computer assisted calculations to generalize this model, using SIR for simplicity's sake, and examine how:

1. the connectivity of a network affects the progress of an epidemic, instead of the parameters.

2. one can estimate the parameters on each population, without knowing the traffic between the populations.

In the next chapter, we explain how this model is setup using Python.

5 The simulation

For the needs of this thesis, a Python program was implemented that simulates the spread of the epidemic on a network of populations. The network is simulated using a weighted, directed graph. Each node of the graph simulates a population, with its respective parameters. The class diagram for this program is shown below:

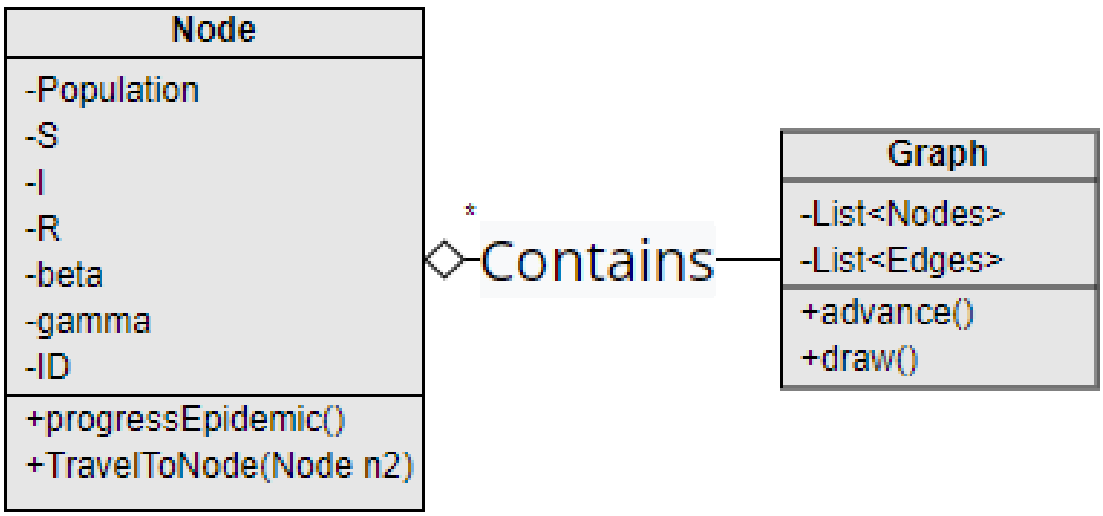

Figure 12: The class diagram for the simulation's main classes.

A single population node is modeled using the system of equations mentioned in equations (1)-(3). The node is a Python class object with attributes Node(**id**,S,I,R,β,γ,Population). S, I and R are decimal numbers in the range $[0, 1]$ representing the quotient of the population belonging to the respective compartment. By solving the equations (1)-(3), we update the node's population up to a specified date.

Running a single node with various R_0s yields the following results. We can see that as the R_0 rises, the maximum simultaneously infected individuals rise, and the duration of the epidemic is shorter, as per Eq. 5 and 8.

5.1 Simulating a single node

A single node is simulated using the Euler's method [17] on Eq. 1- 3:

$$S[t] = S[t-h] + h\frac{dS(t)}{dt} \tag{10}$$

$$I[t] = I[t-h] + h\frac{dI(t)}{dt} \tag{11}$$

$$R[t] = R[t-h] + h\frac{dR(t)}{dt} \tag{12}$$

The method's local truncation error (LTE) can be calculated as:

$$LTE = \frac{1}{2}h^2 y''(t_0) + O(h^3) \tag{13}$$

In order to apply(13) to (10)-(12) we must first calculate the second-order derivatives in relation to time for S, I, and R:

1. For S:

$$\frac{d^2S}{dt^2} = \frac{d(-\beta SI)}{dt} = -\beta(S'I + I'S) = \beta^2 SI^2 - \beta^2 S^2 I + \beta\gamma SI = \beta SI(\beta I - \beta S + \gamma) \tag{14}$$

2. For I:

$$\begin{aligned} \frac{d^2I}{dt^2} &= \frac{d(\beta SI)}{dt} - \frac{d(\gamma I)}{dt} = -\beta SI(\beta I - \beta S + \gamma) - \gamma(\beta SI - \gamma I) = \\ &= -\beta SI(I + S + 2\gamma) + \gamma^2 I \end{aligned} \tag{15}$$

3. For R:

$$\frac{d^2R}{dt^2} = \frac{d(\gamma I)}{dt} = \gamma(\beta SI - \gamma I)$$

Thus, the LTE for each compartment's approximation becomes:

1.

$$LTE_S = \frac{1}{2}h^2 S''(t_0) + O(h^3) = \frac{1}{2}h^2 \beta SI(\beta I - \beta S + \gamma) + O(h^3) \tag{16}$$

2.

$$LTE_I = \frac{1}{2}h^2 I''(t_0) + O(h^3) = \frac{1}{2}h^2\Big(-\beta SI(I + S + 2\gamma) + \gamma^2 I\Big) + O(h^3) \tag{17}$$

3.

$$LTE_R = \frac{1}{2}h^2 R''(t_0) + O(h^3) = \frac{1}{2}h^2\gamma(\beta SI - \gamma I) + O(h^3) \tag{18}$$

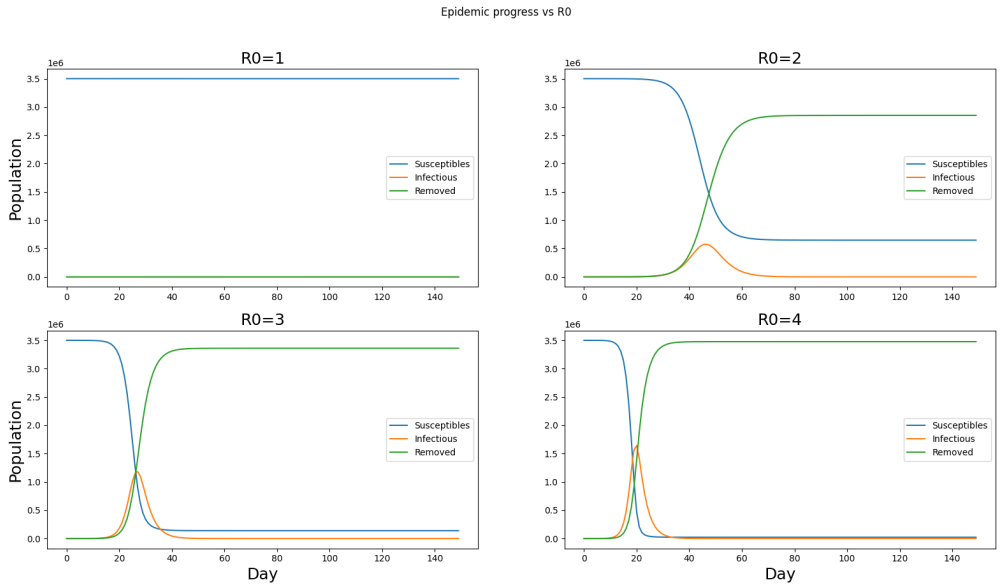

Figure 13: Comparative progress of the epidemic for different R_0 values. $S_0 = 1$. $\gamma = 0.3$ and $\beta = R_0\gamma$.

5.2 Multiple nodes simulation

The graph contains multiple such nodes, as well as the edges (connections/links) between them. Those edges connect two different nodes with a probability p, resulting in a random graph as seen in Ch. 3.

The graph's edges simulate the travel between cities. Thus, an edge e_{ij} connects nodes i and j with some weight, w_{ij}. This means that a total of w individuals will travel from node i to node j [18]. (In the simulation, travel between the nodes is calculated using the *multivariate hyper-geometric distribution*[19]. This happens because there are three groups of individuals: S, I, R and we are drawing $w_{i,j}$ of them without replacement).

Let $b_{Ci,j}$ the number of individuals belonging to compartment C that travel between nodes i, j, from node i to node j. Additionally, let b_{Cj} be the sum of people belonging in compartment C that arrive minus the people that leave node j, i.e. $b_{Cj}(t) = \sum_{i \in N} \left(b_{Ci,j}(t) - b_{Cj,i}(t) \right)$ (N is the set of nodes in the network).

Each node i has a set of variables, S_i, I_i, R_i that correspond to the Susceptible, Infected

and Removed compartments respectively. On the first step of the iterative algorithm, the node solves the equations of (19)-(21) by the preset β and γ values and the sum of the individuals that arrive to minus the individuals that travel from the node.

Thus, on each node i, the system of the epidemic is described by:

$$\frac{dS_i}{dt} = -\frac{\beta_i}{N_i} S_i I_i + b_{Si} \tag{19}$$

$$\frac{dI_i}{dt} = \frac{\beta_i}{N_i} S_i I_i - \gamma_i I_i + b_{Ii} \tag{20}$$

$$\frac{dR_i}{dt} = \gamma_i I_i + b_{Ri} \tag{21}$$

In order to continue, we must first non-dimensionalize the previous equations. This is done by setting new variables as such:

$$S_t^* = \frac{S_t}{N_i} \tag{22}$$

$$I_t^* = \frac{I_t}{N_i} \tag{23}$$

$$R_t^* = \frac{R_t}{N_i} \tag{24}$$

where N is the total population of the network, which remains constant (assuming no births or deaths).

Then, equations (19) to (21) become:

$$\frac{dS_i^*}{dt} = -\beta_i S_i^* I_i^* + \frac{b_{Si}}{N_i} \tag{25}$$

$$\frac{dI_i^*}{dt} = \beta_i S_i^* I_i^* - \gamma_i I_i^* + \frac{b_{Ii}}{N_i} \tag{26}$$

$$\frac{dR_i^*}{dt} = \gamma_i I_i^* + \frac{b_{Ri}}{N_i} \tag{27}$$

Adding for each node i, and assuming populations are equal on each node and the graph is bi-directional Eq. (25) (27) becomes:

$$\sum \frac{dS_i^*}{dt} \approx -\sum \beta_i S_i^* I_i^* \quad (28)$$

$$\sum \frac{dI_i^*}{dt} \approx \sum \beta_i S_i^* I_i^* - \sum \gamma_i I_i^* \quad (29)$$

$$\sum \frac{dR_i^*}{dt} \approx \sum \gamma_i I_i^* \quad (30)$$

or:

$$\frac{d}{dt}\sum S_i^* \approx -\sum \beta_i S_i^* I_i^* \quad (31)$$

$$\frac{d}{dt}\sum I_i^* \approx \sum \beta_i S_i^* I_i^* - \sum \gamma_i I_i^* \quad (32)$$

$$\frac{d}{dt}\sum R_i^* \approx \sum \gamma_i I_i^* \quad (33)$$

5.3 Euler's method on the whole network

By applying Euler's method on equations 25 to 27 we get:

$$\begin{aligned} S_i^*[t+h] - S_i^*[t] &= -h\beta_i S_i^*[t] I_i^*[t] + h\frac{b_{Si}}{N_i} \\ I_i^*[t+h] - I_i^*[t] &= h\beta_i S_i^*[t] I_i^*[t] - h\gamma_i I_i^*[t] + h\frac{b_{Ii}}{N_i} \\ R_i^*[t+h] - R_i^*[t] &= h\gamma_i I_i^*[t] + h\frac{b_{Ri}}{N_i} \end{aligned}$$

or:

$$\begin{aligned} \Delta S_i^*[t] &= -h\beta_i S_i^*[t] I_i^*[t] + h\frac{b_{Si}}{N_i} \\ \Delta I_i^*[t] &= h\beta_i S_i^*[t] I_i^*[t] - h\gamma_i I_i^*[t] + h\frac{b_{Ii}}{N_i} \\ \Delta R_i^*[t] &= h\gamma_i I_i^*[t] + h\frac{b_{Ri}}{N_i} \end{aligned}$$

Setting, for example $h = 1$, which was used in the simulations:

$$
\begin{aligned}
\Delta S_i^*[t] = S_i^*[t+1] - S_i^*[t] &= -\beta_i S_i^*[t] I_i^*[t] + \frac{b_{Si}}{N_i} \\
\Delta I_i^*[t] = I_i^*[t+1] - I_i^*[t] &= \beta_i S_i^*[t] I_i^*[t] - \gamma_i I_i^*[t] + \frac{b_{Ii}}{N_i} \\
\Delta R_i^*[t] = R_i^*[t+1] - R_i^*[t] &= \gamma_i I_i^*[t] + \frac{b_{Ri}}{N_i}
\end{aligned}
$$

Adding across all nodes, the resulting equations approximate the total change on the whole graph:

$$
\begin{aligned}
\Delta S^*[t] = \sum \Delta S_i^*[t] &= -\sum \beta_i S_i^*[t] I_i^*[t] && (34) \\
\Delta I^*[t] = \sum \Delta I_i^*[t] &= \sum \beta_i S_i^*[t] I_i^*[t] - \sum \gamma_i I_i^*[t] && (35) \\
\Delta R^*[t] = \sum \Delta R_i^*[t] &= \sum \gamma_i I_i^*[t] && (36)
\end{aligned}
$$

where $\Delta^*[\cdot][t]$ denotes the change across the whole network on *discrete* time t.

Running the simulation with 3 nodes, yields the following result:

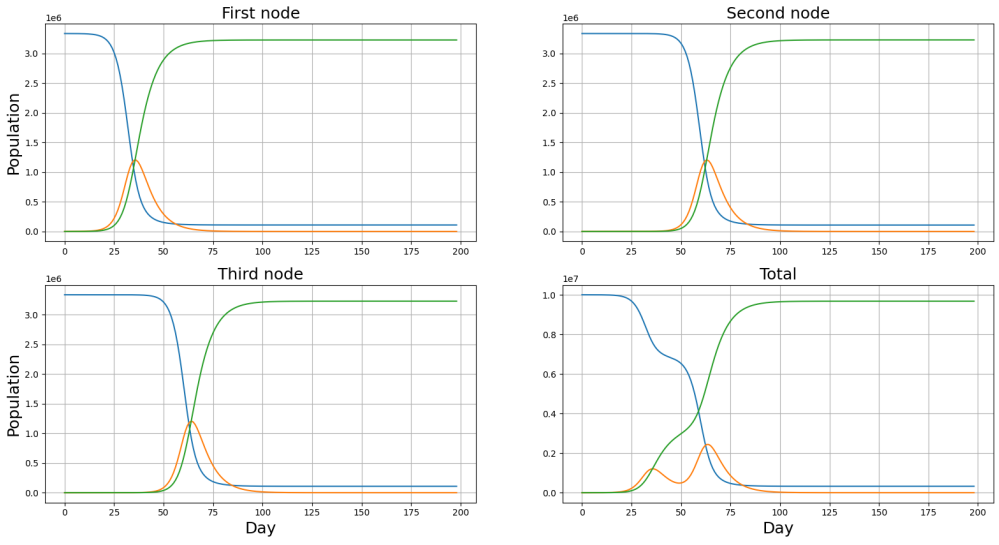

Figure 14: Cumulative progress of the epidemic on 3 population nodes. The mean degree of the nodes is 1. For each node: $\beta = 0.5, \gamma = 0.1, I_0 = 10\text{idv}, R_0 = 5$. The population of each node i is $N_i = 10^7/3$

5.4 Simulation Results

An interesting result is that while the compartmental model simulations create two main outcomes (no epidemic, the curves are mainly flat) or an epidemic (the outcomes are the classical SIR style curves), this model creates a curve with waves of different amplitudes. Sub-critical graphs create localized epidemics, while super-critical (or highly connected) graphs create epidemics that resemble the SIR model (let's remind the reader that the SIR model is a model derived from the assumption that physical human communication is a chemical reaction). The middle case (critical models) create the wave-like curves.

As it is shown in section 6.1, when a graph is sub-critical no epidemic happens, while where a graph is critical an epidemic may or may not happen (this is a random event). When the graph is super-critical the epidemic always happens. It is also shown that the connectivity of the graph affects the outlook of the epidemic regarding the peaks. The more connected a graph is, the less peaks there are (more concentrated transmission). These results can be shown in 4.

We can observe two local maxima in the infected population. Node 0 gets infected in the beginning of the simulation. Nodes 1 and 2 get infected on the days that can be seen from the lines on the graph. The first peak in time is the result of the peak of the epidemic on node 0, and the other peak is caused by the almost simultaneous peak of the epidemic on nodes 1 and 2. If a node gets infected on day t_i, the peak will happen in day $t_i + t_p$, for t_p the time seen in Eq. 8, with the equation having the parameters of node i. E.g., if two different nodes i and j receive an infected individual on day t, and their peak infected populations are I_{m1} and I_{m2} respectively, the amount of infected individuals owed to these two nodes on time $t + t_p$ will be $I_{m1} + I_{m2}$. Since all nodes (for simplicity's sake) have the same parameters, this results in the peaks of nodes 1 and 2 coinciding, resulting in a peak infected population of $S_{m1} + S_{m2}$.

The parameters associated with this process are:

- The number of individuals traveling between the nodes.
- The existence (or not) of an edge between two nodes.
- The epidemiological parameters of each node.

6 Estimation Methods

Following are some methods that can be used to derive the epidemiological parameters simply by having measurements of the compartmental populations. These are algebraic methods, are derived from the Euler's method equations, solving for the parameters in question [20].

Since the simulation is discrete, assuming a time step of one day. There is, of course, error in these methods, since the Euler's method is an approximation. The following estimation approximates every node's population $N_i \approx N$, and the graph is bidirectional.

6.1 Multiple nodes estimation

In order to estimate the β and γ values for each node, we can utilize the equations seen in section 5.3. Specifically, one can rewrite Eq. 34 as such:

$$\Delta S^*[t_j] = -\overrightarrow{\beta} \cdot \overrightarrow{S^*I^*}[t_j] \tag{37}$$

where $\overrightarrow{\beta}$ is the $1 \times (\#\text{ Nodes})$ matrix containing β_i values for each node i, and $\overrightarrow{S^*I^*}[t_j]$ the vector containing the products $S_i^* I_i^*$ of the susceptible and infected individuals of all nodes i at time t_j. N denotes the total population of the graph, across all nodes, which remains constant.

If we create m measurements on many t_js, where j is the incremental number of the measurement on all nodes simultaneously (i.e. the j-th 'day'), and change the notion of the vectors to a bold font, we get:

$$\begin{aligned}
\Delta S^*{}_{t_1} &= -\boldsymbol{\beta} \cdot (\boldsymbol{S^*I^*})_{t_1} \\
\Delta S^*{}_{t_2} &= -\boldsymbol{\beta} \cdot (\boldsymbol{S^*I^*})_{t_2} \\
&\vdots \\
\Delta S^*{}_{t_m} &= -\boldsymbol{\beta} \cdot (\boldsymbol{S^*I^*})_{t_m}
\end{aligned}$$

Or, in matrix form:

$$\boldsymbol{M\beta} = \boldsymbol{\Delta S} \Leftrightarrow$$

$$\begin{bmatrix} -S_1^*[t_1]I_1^*[t_1] & \cdots & -S_n^*[t_1]I_n^*[t_1] \\ -S_1^*[t_2]I_1^*[t_2] & \cdots & -S_n^*[t_2]I_n^*[t_2] \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ -S_1^*[t_m]I_1^*[t_m] & \cdots & -S_n^*[t_m]I_n^*[t_m] \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} = \begin{bmatrix} \Delta S^*[t_1] \\ \Delta S^*[t_2] \\ \cdot \\ \cdot \\ \cdot \\ \Delta S^*[t_m] \end{bmatrix}$$

Each measurement we make, adds a row on the matrix M.

By replacing Eq. 22 and 23 above, we get:

$$\begin{bmatrix} -S_1[t_1]I1[t_1] & \cdots & -S_n[t_1]I_n[t_1] \\ -S_1[t_2]I_1[t_2] & \cdots & -S_n[t_2]I_n[t_2] \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ -S_1[t_m]I_1[t_m] & \cdots & -S_n[t_m]I_n[t_m] \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{bmatrix} = \begin{bmatrix} \Delta S[t_1]N \\ \Delta S[t_2]N \\ \cdot \\ \cdot \\ \cdot \\ \Delta S[t_m]N \end{bmatrix}$$

In order to calculate the betas one can either use the formula $\boldsymbol{\beta} = \boldsymbol{M}^+\boldsymbol{\Delta S}$, if the system is solvable, or minimize the expression $||M\beta - \Delta S||_2$ subject to $\boldsymbol{\beta} >= 0$. Here, M^+ denotes the Moore-Penrose pseudo-inverse matrix of M [21].

For this reason, the **scipy** python library's *nnls* (which stands for non-negative least squares solver) function was used [22]. Here we must also note that the Moore-Penrose pseudoinverse provides a least squares solution to a system of linear equations. The issue with the solution it can have negative values, which is impossible given that we need to estimate a non-negative β.

The first method results in negative βs, while the second forces some to zero.

6.2 Calculating γ's

In order to estimate the γ's of the network, we use equation 36. In a same manner as the previous subsection, we can model the equation as:

$$\Delta R^*[t_j] = \overrightarrow{\gamma} \cdot \overrightarrow{I^*}[t_j] \tag{38}$$

Taking many measurements on different t_js:

$$\begin{aligned}
\Delta {R^*}_{t_1} &= \boldsymbol{\gamma} \cdot \boldsymbol{I}^*_{t_1} \\
\Delta {R^*}_{t_2} &= \boldsymbol{\gamma} \cdot \boldsymbol{I}^*_{t_2} \\
&\cdot \\
&\cdot \\
&\cdot \\
\Delta {R^*}_{t_m} &= \boldsymbol{\gamma} \cdot \boldsymbol{I}^*_{t_m}
\end{aligned}$$

or:

$$\begin{bmatrix} I^*_1[t_1] & \cdots & I^*_n[t_1] \\ I^*_1[t_2] & \cdots & I^*_n[t_2] \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ I^*_1[t_m] & \cdots & I^*_n[t_m] \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_n \end{bmatrix} = \begin{bmatrix} \Delta R^*[t_1] \\ \Delta R^*[t_2] \\ \cdot \\ \cdot \\ \cdot \\ \Delta R^*[t_m] \end{bmatrix}$$

Since $I^* = \frac{I}{N}$ and $R^* = \frac{R}{N}$ the above matrix becomes:

$$\begin{bmatrix} I_1[t_1] & \cdots & I_n[t_1] \\ I_1[t_2] & \cdots & I_n[t_2] \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ I_1[t_m] & \cdots & I_n[t_m] \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_n \end{bmatrix} = \begin{bmatrix} \Delta R[t_1] \\ \Delta R[t_2] \\ \cdot \\ \cdot \\ \cdot \\ \Delta R[t_m] \end{bmatrix}$$

Finding the γs again becomes a minimization problem, as seen in the previous subsection.

7 Estimation Results

Following are some results for different total nodes. The simulated β and γ are randomly generated following the uniform distribution. Note that the populations are equal on every node. β values range in $[0.1, 0.2]$ while γ in $[0.01, 0.05]$, yielding an R_0 range of $[2, 10]$. The s*cipy.nnls()*method is used on each iteration. The graph is on the critical region, meaning that each node is at average connected to 1 other node. The graph's horizontal axis denotes the number of nodes that got eventually infected, regardless of the initial network node number. Not all nodes get infected on every run of the experiment. The vertical lines denote the maximum error.

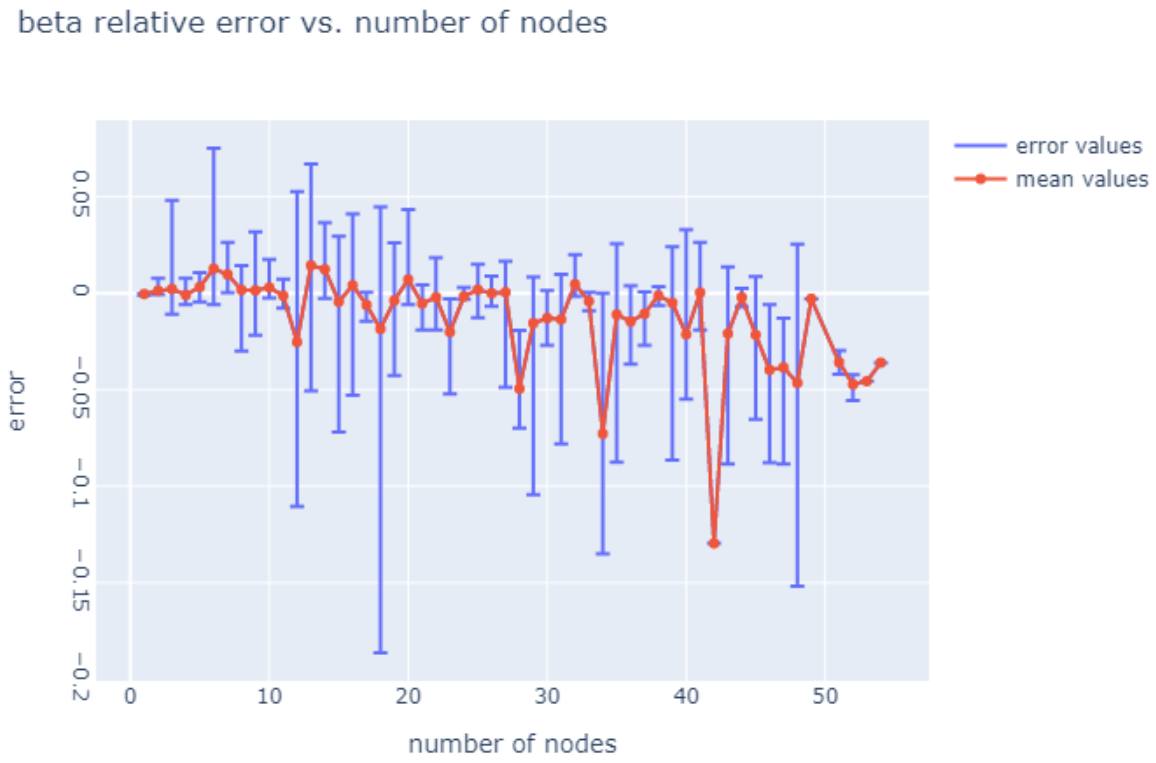

Figure 15: NNLS β error for 1 to 43 nodes. Duration $<$ 2000 days. mean degree $= 1$

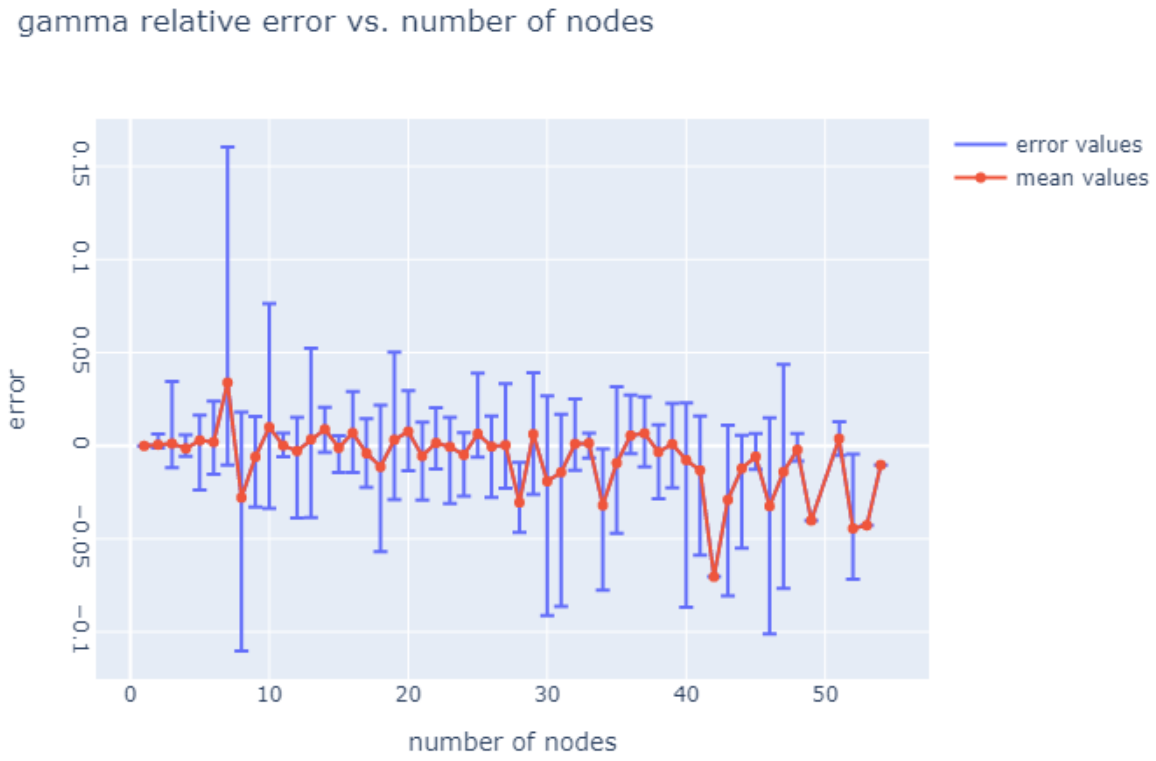

Figure 16: NNLS γ error for 1 to 43 nodes. Duration $<$ 2000 days. mean degree $= 1$

We can see that the algorithm is more precise when calculating β than γ.

7.1 Adding noise

In order to model uncertainty in our measurements of the compartments, we added random noise on a single node. The noise was generated drawing samples from the exponential distribution [23]:

$$Pr\{sample = x\} = \lambda e^{-\lambda x} \tag{39}$$

where e is the physical constant and λ is a parameter.

The mean value of x given a set λ is: $E[x|\lambda] = \frac{1}{\lambda}$. This means that $E[error] = \frac{1}{\lambda} I(t)$.

The samples that were created were then added to the I compartment for each day of the simulation. The rationale behind this choice was the fact that the detected individuals

in the I compartment are always fewer than or equal to the population of the truly infected individuals.

We denote the signal of the infected individuals as

$$\overline{I}_{t-1} = I_{t-1} + n_I(t-1) \tag{40}$$

and the noisy S signal with $\overline{S}$. Since we know that $N = S + I + R$, $\overline{S}$ can then be derived from:

$$\overline{S}(t) = N(t) - \overline{I}(t) - \overline{R}(t) = N - \Big(I(t) + n_I(t)\Big) - R(t) = S(t) - n_I(t) \tag{41}$$

The error on every point is calculated as:

$$error = \frac{\beta_{est} - \beta_{true}}{\beta_{true}} \tag{42}$$

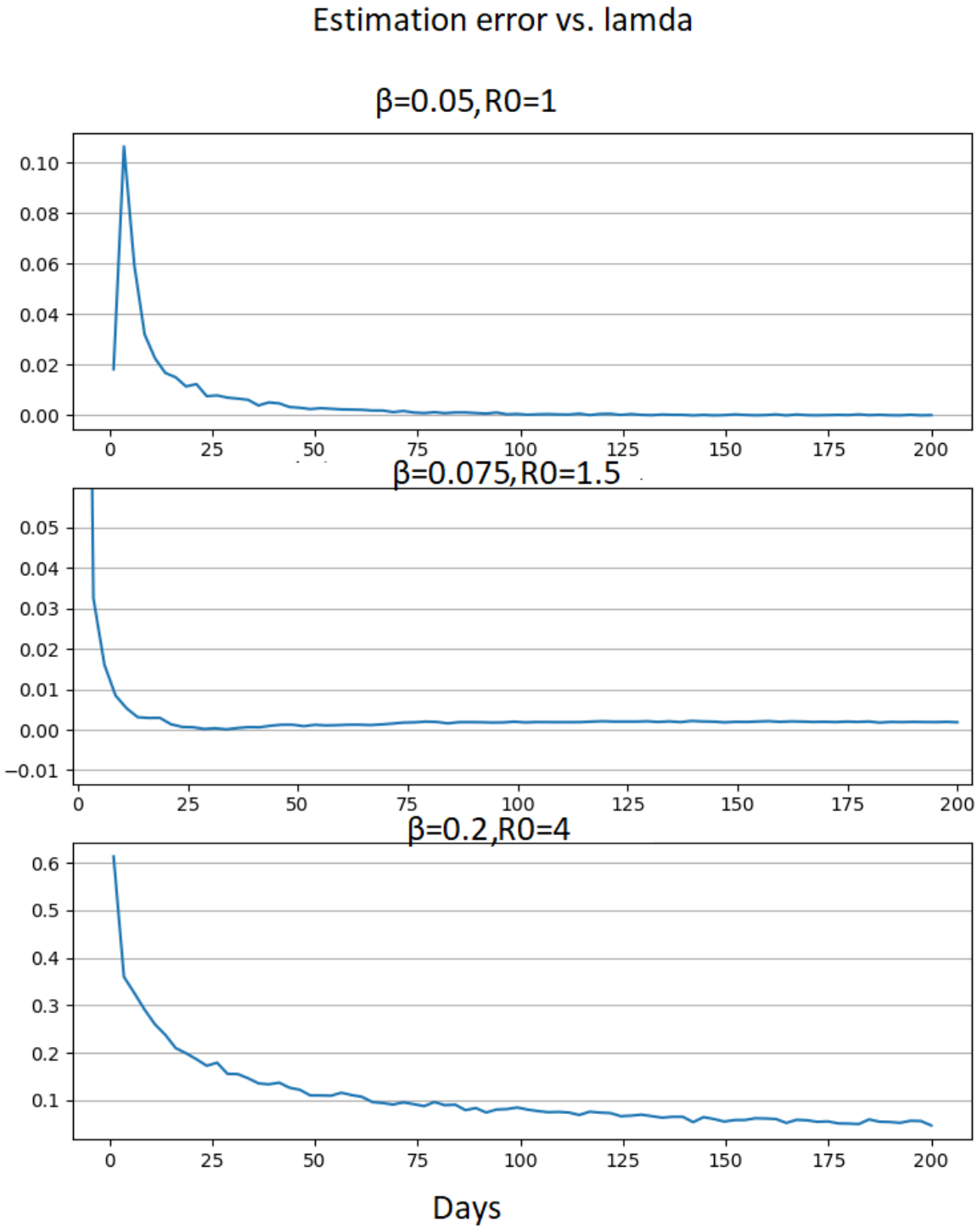

Figure 17: Estimation error vs. λ value.

We see that even for a single node and relatively small values of λ there is some considerable error.

Following are the errors for different values of λ:

1. $\lambda = 100$

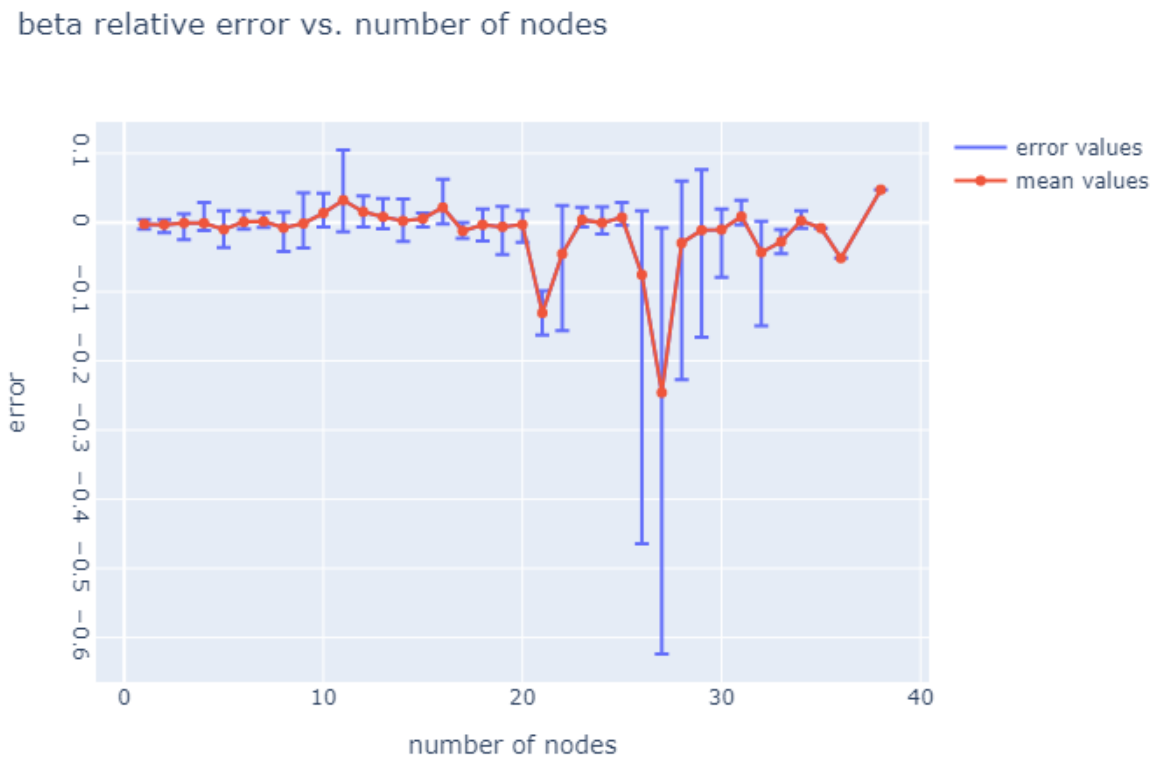

Figure 18: β estimation error for $\lambda = 100$.

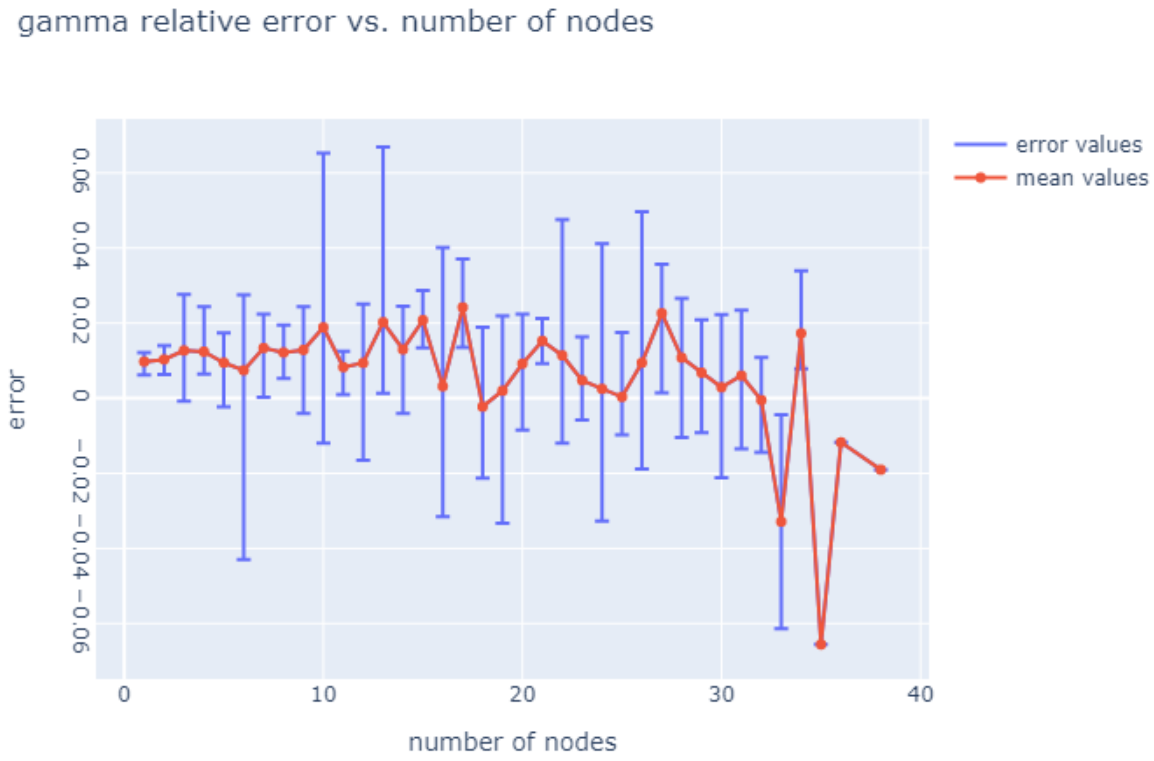

Figure 19: γ estimation error for $\lambda = 100$.

2. $\lambda = 50$

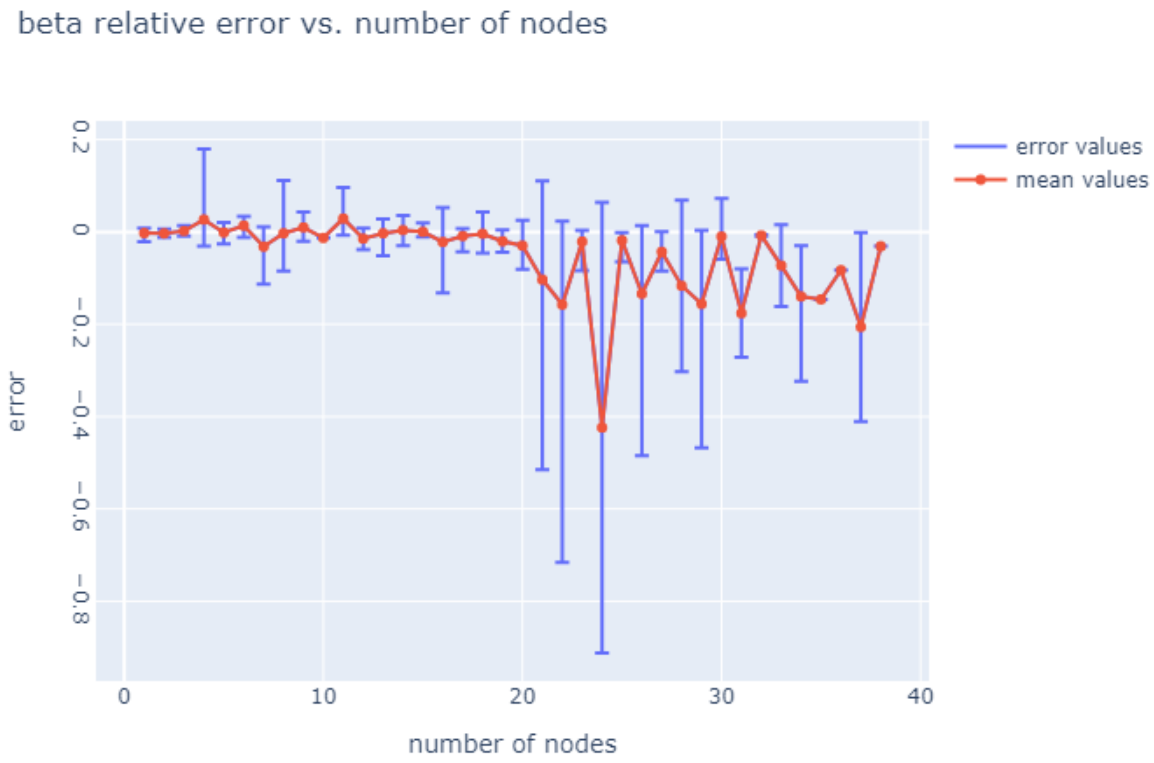

Figure 20: β estimation error for $\lambda = 50$.

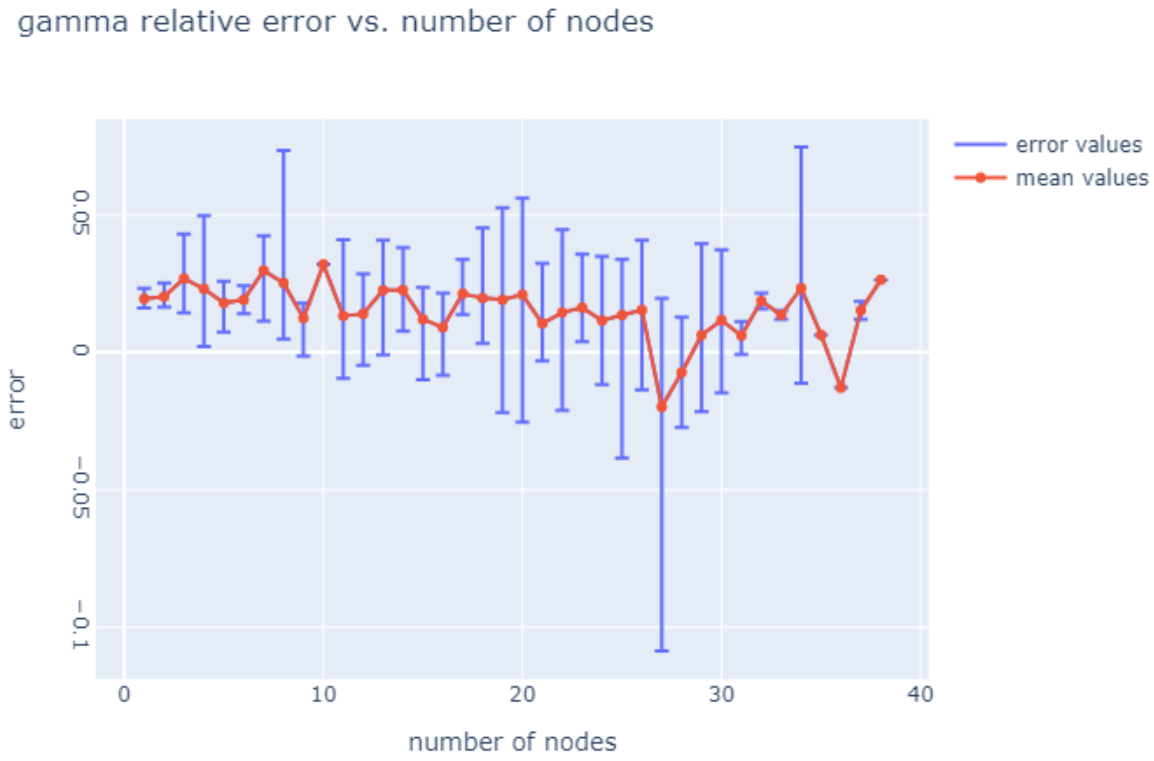

Figure 21: γ estimation error for $\lambda = 50$.

3. $\lambda = 10$

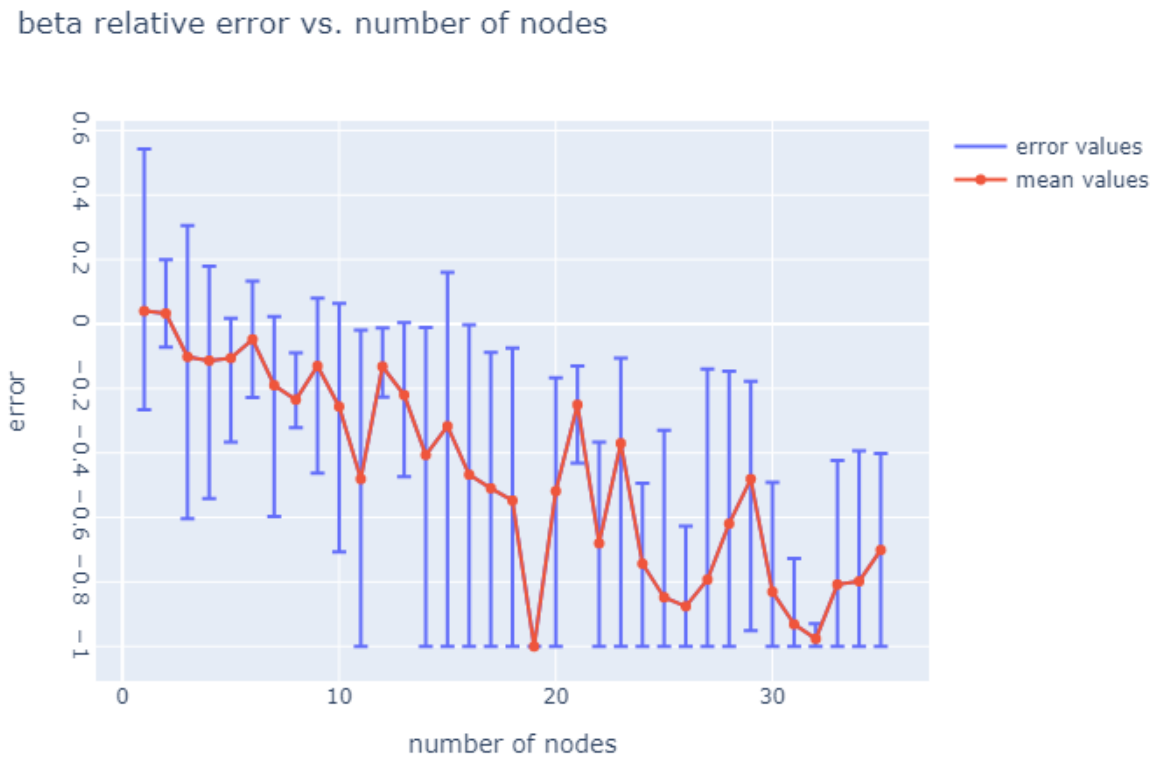

Figure 22: β estimation error for $\lambda = 10$.

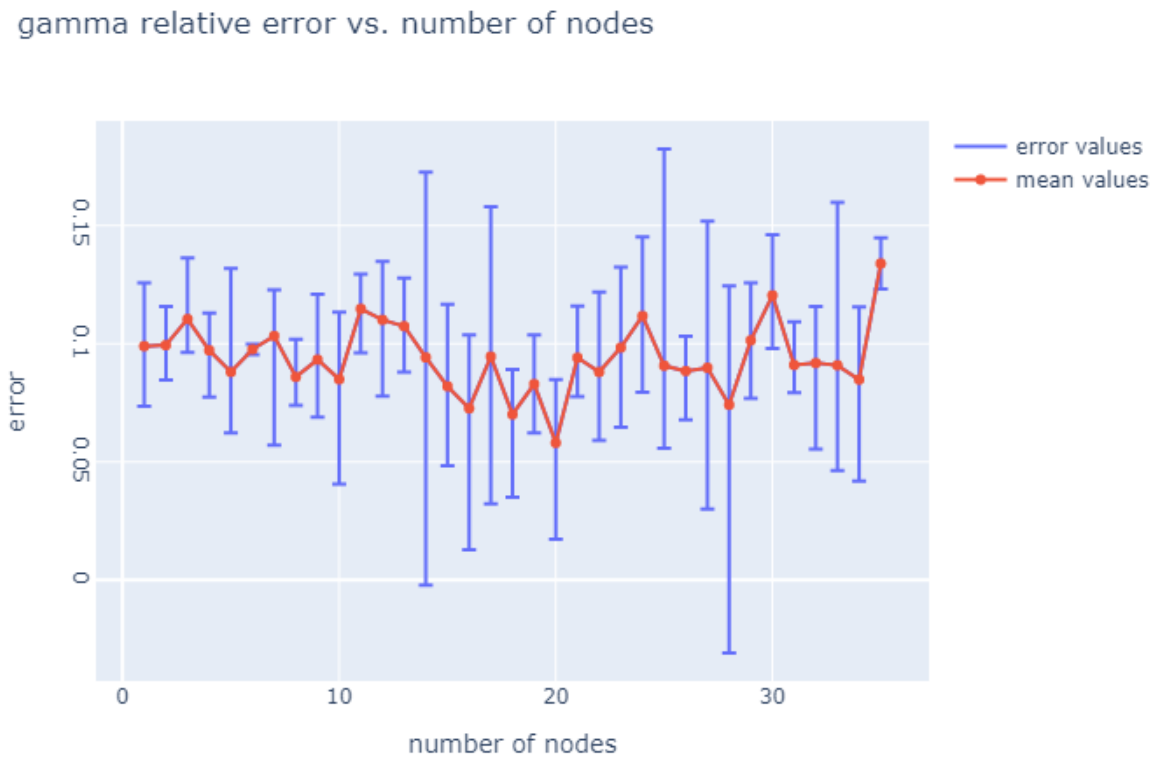

Figure 23: γ estimation error for $\lambda = 10$.

4. $\lambda = 5$

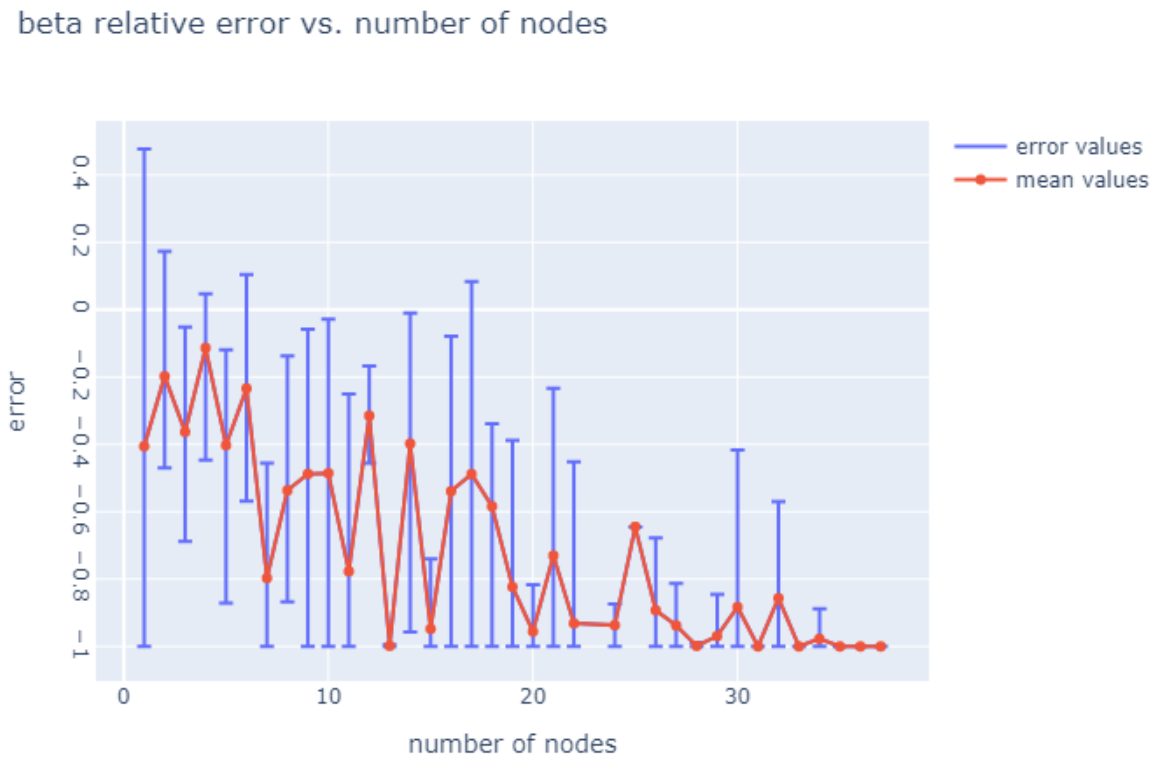

Figure 24: β estimation error for $\lambda = 10$.

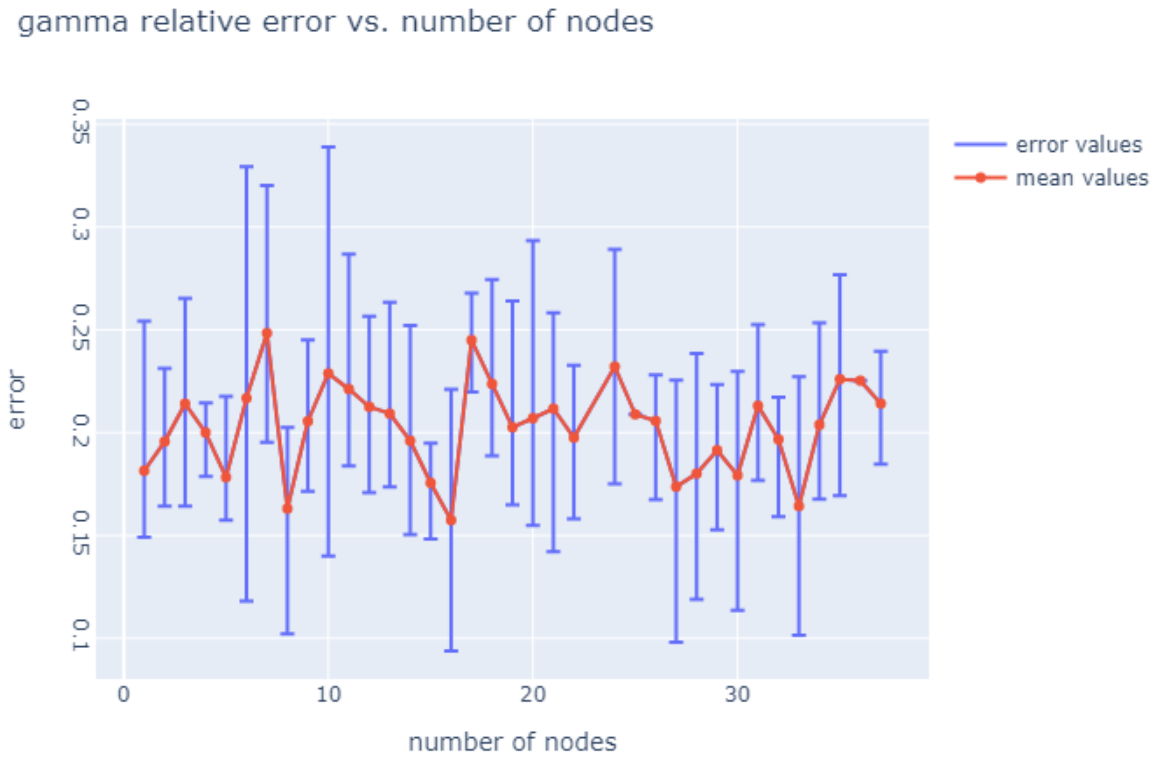

Figure 25: γ estimation error for $\lambda = 5$.

For lambdas less than 10, the error gets significant even for small numbers of nodes.

8 Future Work

The following topics are of interest, in order to measure the model's effectiveness and compare it to real world phenomena:

1. Using a Machine Learning algorithm to estimate parameters on these networks. That is, extending and comparing the algorithm used in [13] to the model presented here.

2. How does the topology of the network can affect outcomes? In this thesis, the network used was mainly a random graph of mean degree 1. One could examine how different graph topologies affect the accuracy of the model.

3. How Can we control the epidemic by changing the traffic between the nodes by scheduling the biases of the edges?

4. In the algorithm used in this thesis, the approach was the Euler method. Runge-Kutta method, which is generally more accurate but more computationally complex, could be used to possibly improve the accuracy of the model and algorithm.

References

[1] Jane Radecki and Roger Schonfeld. The Impacts of COVID-19 on the Research Enterprise. Technical report, Ithaka S+R, October 2020.

[2] M. Choisy, J.-F. Guégan, and P. Rohani. *Mathematical Modeling of Infectious Diseases Dynamics*, chapter 22, pages 379–404. John Wiley & Sons, Ltd, 2007.

[3] Nikolaos Demiris. *Bayesian Inference for Stochastic Epidemic Models using Markov chain Monte Carlo Methods*. PhD thesis, 2004.

[4] Zifeng Yang, Zhiqi Zeng, Ke Wang, Sook-San Wong, Wenhua Liang, Mark Zanin, Peng Liu, Xudong Cao, Zhongqiang Gao, Zhitong Mai, Jingyi Liang, Xiaoqing Liu, Shiyue Li, Yimin Li, Feng Ye, Weijie Guan, Yifan Yang, Fei Li, Shengmei Luo, Yuqi Xie, Bin Liu, Zhoulang Wang, Shaobo Zhang, Yaonan Wang, Nanshan Zhong, and Jianxing He. Modified SEIR and AI prediction of the epidemics trend of covid-19 in china under public health interventions. *Journal of Thoracic Disease*, 12(3), 2020.

[5] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of Modern Physics*, 87(3):925–979, Aug 2015.

[6] W O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics–i. 1927. *Bull. Math. Biol.*, 53(1-2):33–55, 1991.

[7] W O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics–II. the problem of endemicity.1932. *Bull. Math. Biol.*, 53(1-2):57–87, 1991.

[8] W O Kermack and A G McKendrick. Contributions to the mathematical theory of epidemics–III. further studies of the problem of endemicity. 1933. *Bull. Math. Biol.*, 53(1-2):89–118, 1991.

[9] Fred Brauer. The kermack–mckendrick epidemic model revisited. *Mathematical Biosciences*, 198(2):119–131, 2005.

[10] James D. Murray. *Mathematical Biology*. Springer-Verlag New York, 2002.

[11] Mustafa Turkyilmazoglu. Explicit formulae for the peak time of an epidemic from the SIR model. *Physica D*, 422(132902):132902, August 2021.

[12] Cliff C. Kerr, Robyn M. Stuart, Dina Mistry, Romesh G. Abeysuriya, Katherine Rosenfeld, Gregory R. Hart, Rafael C. Núñez, Jamie A. Cohen, Prashanth Selvaraj, Brittany Hagedorn, Lauren George, Michał Jastrzębski, Amanda S. Izzo, Greer Fowler, Anna Palmer, Dominic Delport, Nick Scott, Sherrie L. Kelly, Caroline S. Bennette, Bradley G. Wagner, Stewart T. Chang, Assaf P. Oron, Edward A. Wenger,

Jasmina Panovska-Griffiths, Michael Famulare, and Daniel J. Klein. Covasim: An agent-based model of covid-19 dynamics and interventions. *PLOS Computational Biology*, 17(7):1–32, 07 2021.

[13] Yong Yang, Peter M. Atkinson, and Dick Ettema. Analysis of cdc social control measures using an agent-based simulation of an influenza epidemic in a city. *BMC Infectious Diseases*, 11(1):199, Jul 2011.

[14] Voloudakis Konstantinos. Implementation of a software server and of an autonomous agent for the cooperative strategic game "pandemic". 2021.

[15] Albert-László Barabási. *Network Science*. Cambridge, 2016.

[16] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Rev. Mod. Phys.*, 80:1275–1335, Oct 2008.

[17] *Numerical Differential Equation Methods*, chapter 2, pages 55–142. John Wiley and Sons, Ltd, 2016.

[18] Arnaud Casteigts, Paola Flocchini, Walter Quattrociocchi, and Nicola Santoro. Time-varying graphs and dynamic networks, 2010.

[19] Vose Software. Multivariate hypergeometric distribution.

[20] Numerical methods for differential equations. http://faculty.olin.edu/bstorey/Notes/DiffEq.pdf.

[21] Moore–Penrose inverse, February 2022. Page Version ID: 1073211427.

[22] Scipy.optimize.nnls.

[23] Numpy.random.exponential.

[24] Nakul Chitnis. Introduction to seir models. In *Workshop on mathematical models of climate variability, environmental change and infectious diseases, Trieste, Italy*, 2017.

[25] Cheng Cheng, DongDong Zhang, Dejian Dang, Juan Geng, Peiyu Zhu, Mingzhu Yuan, Ruonan Liang, Haiyan Yang, Yuefei Jin, Jing Xie, Shuaiyin Chen, and Guangcai Duan. The incubation period of covid-19: a global meta-analysis of 53 studies and a chinese observation study of 11 545 patients. *Infectious Diseases of Poverty*, 10(1):119, Sep 2021.

A Additional compartmental models

Apart from the SIR model, which was the basic model, other models have since been developed [24]. Following is a brief explanation of some of the most widely spread.

A.1 The SI model

The SI model assumes permanent infection (e.g. as seen in the HIV virus). Here, the compartments are reduced to just two (Susceptible and Infected). The equations that describe them are seen in the following equations and Fig. 26.

$$\begin{aligned} \frac{dS}{dt} &= -r\beta S\frac{I}{N} \\ \frac{dI}{dt} &= r\beta S\frac{I}{N} \end{aligned}$$

Here, β denotes the probability of disease transmission per contact and r denotes the contact rate, or the number of contacts per unit time. N is the total population, or $S + I$.

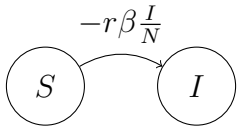

Figure 26: The SI transfer diagram.

A.2 The SIS model

In the SIR model, we assumed that the Removed compartment implies some form of permanent immunity. The SIS model assumes no immunity, while the SIRS model assumes temporary immunity. Thus, individuals move only from the S to the I compartment, and after some time return to the S compartment. This is a model resembling the common cold or, as some could argue, a simplified description of the current pandemic (COVID-19). The equations and the transfer diagram can be seen in equations 43 and 44, as well as in Fig. 27.

$$\frac{dS}{dt} = -r\beta S\frac{I}{N} \tag{43}$$

$$\frac{dI}{dt} = r\beta S\frac{I}{N} \tag{44}$$

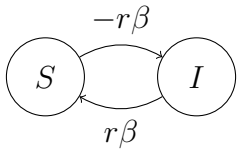

Figure 27: The SIS transfer diagram.

A.3 The SEIR model

A new idea is introduced in the **SEIR** model, which assumes a new compartment, the *exposed* (E). This compartment includes the individuals that were exposed to the disease, but are in the *incubation period*. The incubation period is the period during which the individual is infected but not infectious [25].

$$\begin{aligned}
\frac{dS}{dt} &= \Lambda - r\beta S\frac{I}{N} - \mu S \\
\frac{dE}{dt} &= r\beta S\frac{I}{N} - \epsilon E \\
\frac{dI}{dt} &= \epsilon E - \gamma I - \mu I \\
\frac{dR}{dt} &= \gamma I - \mu R
\end{aligned}$$

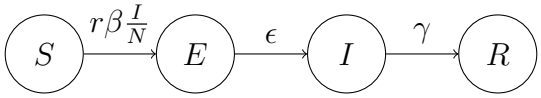

Figure 28: The SEIR transfer diagram.

Here, μ denotes the death rate for each compartment.