Network Dynamics and Learning Homework III

Shannon Mc Mahon
Student id: s289958

***Abstract*—The following document contains the discussion of the solutions of the assigned exercises.**
Solutions of part 3 and 4 discussed with Andrea Rubeis (s290216).
Python implementation is available on *Github*.

I.

This project shows how it is possible to simulate an epidemic on a given graph $G = (V, E)$. The disease propagation model that will be used is a discrete-time simplified version of the SIR compartmental model. At any time $t = 0, 1, ...$ nodes are in state $X_i(t) \in \{S, I, R\}$, where *S* is susceptible, *I* is infected and *R* is recovered. Let β be the probability that the infection is spread from an infected individual to a susceptible one, and ρ be the probability of an infected individual recovering during one time step.
Then we can concisely represent the dynamics between the three compartments *S, I* and *R* as shown in Figure 1 below.

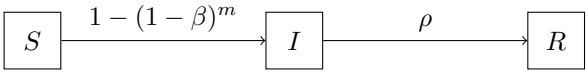

Fig. 1.

The presence of an arc from one category to another means that it is possible for an individual to change their state from the former one to the latter, while the weight of the arc defines the transition probability. For instance, if we consider a node $i \in V$ the probability that $X_i(t+1) = 1$ if $X_i(t) = S$ is given by $1 - (1 - \beta)^m$, where m is the number of infected neighbors of node i. Similarly, the probability that $X_i(t+1) = R$ knowing that $X_i(t) = I$ is by definition ρ.

(a) For our first simulation we use a $k = 4$ regular undirected graph with $|V| = 500$ nodes, and set $\beta = 0.3$, $\rho = 0.7$ We then simulate the evolution of the epidemic on our graph over a 15 week period, with an initial number of infected individuals equal to $I_0 = 10$ (the initial infected nodes are selected at random from the node set).
Figure 2 shows the average behaviour over $N = 100$ iterations of the three curves, with the standard deviation reported as the filled regions, As can be seen, there is an increase of $R(t)$ for the whole duration of the epidemic, as more and more people pass from the infected compartment to the removed one (they either recover or die), while the number of susceptibles $S(t)$ decreases as more and more people contract the disease. As far

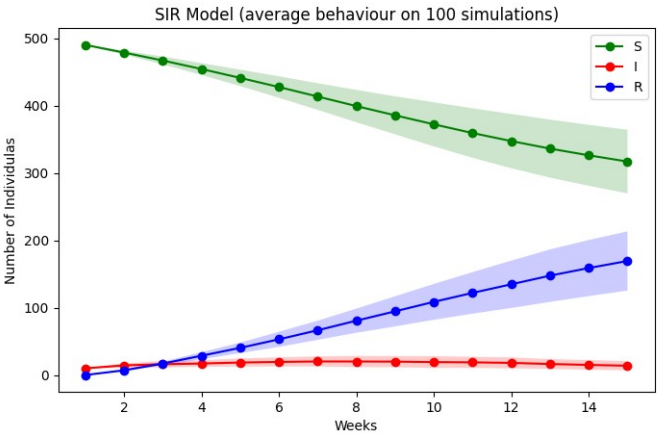

Fig. 2.

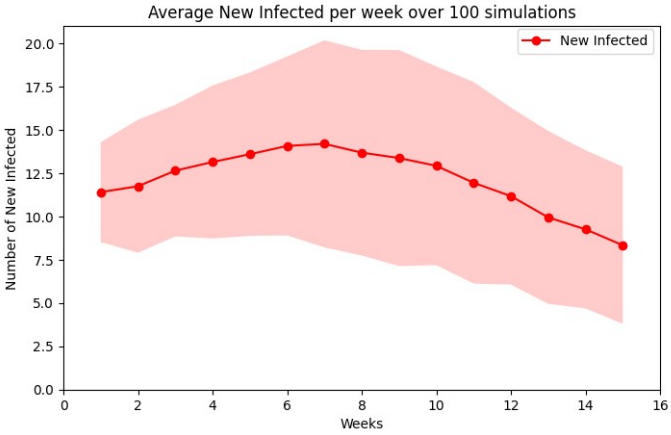

Fig. 3.

the curve $I(t)$ is concerned, the number of infected individuals increases slightly in the first weeks and on average reaches its peak at week 7, only to then decrease gradually (this trend is not reported but if the reader wishes to they can verify this by plotting only the I(t) curve). As we can see 15 weeks are not enough for the disease to die out, however the epidemic is manageable since we do not have an explosion of cases. Note how the simulation tends to a configuration of type $\{S, R\}^{|V|}$ which is absorbing for the SIR model. Figure 3 shows instead the trend related to the newly infected individuals each week (i.e. how many individuals *become* infected

each week).

(b) We now define a different graph structure which will be used in the following points. Specifically, we generate a random graph with average degree close to a chosen k by use of the preferential attachment model. The latter involves starting with a graph which is complete with k+1 nodes. Then iteratively a node is added to the graph, and connected to $c = \frac{k}{2}$ nodes of the original graph (if k is odd then alternate between adding $\lceil \frac{k}{2} \rceil$ and $\lceil \frac{k}{2} \rceil$ nodes). The probability that there will be a link between node n_t and node $i \in V_{t-1}$ is used to define where to add the links, and is given by

$$\mathbb{P}(W_{n_t,i}(t) = W_{i,n_t}(t) = 1 | G_{t-1} = (V_{t-1}, E_{t-1})) = \\ = \frac{w_i(t-1)}{\sum_{j \in V_{t-1}} w_j(t-1)} \quad (1)$$

where $W(t)$ is the adjacency matrix for the next time step t and $w_i(t-1)$ is the degree of the node i prior to adding the new node. To illustrate this method suppose we wish to obtain a graph with 8 nodes and average degree $k = 4$. We can start from a complete graph of degree $k + 1 = 5$ and gradually add nodes and their edges according to the rule stated above. The final graph we obtain is shown in Figure 4 and has nodes 0 and 3 with degree 6, nodes 1 and 4 with degree 5, nodes 5, 6 and 7 with degree 2 and node 2 with degree 4. This yields an average of 4, as desired. We will use graphs generated by means of the preferential attachment method in the following exercises.

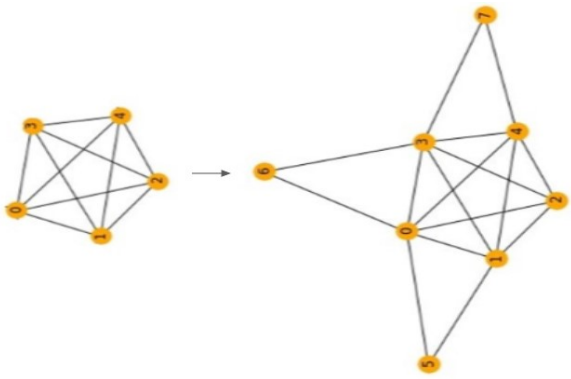

Fig. 4.

II.

In this section we once more simulate an epidemic $N = 100$ times for a period of 15 weeks, starting with $I_0 = 10$ infected. This time however we use a random graph G with 500 nodes and an average degree $k = 6$ obtained by means of the preferential attachment method previously illustrated (β and ρ maintain their values of 0.3 and 0.7 respectively). Figure 5 shows the average trends of *S(t), I(t), R(t)* If we compare these curves with the ones obtained in the previous section we can see that there is a larger diffusion of the disease with respect to the previous scenario, as can be deduced from the $I(t)$ curve. This is due to the 50% increase in the average degree of nodes (from 4 to 6) which results in more connections in the graph, (i.e. the disease spreads more easily). In figure 6 instead we see that the number of newly infected individuals reaches its peak around week 4, only to then decrease rapidly to zero, converging once again to a configuration of type $\{S, R\}^{|V|}$

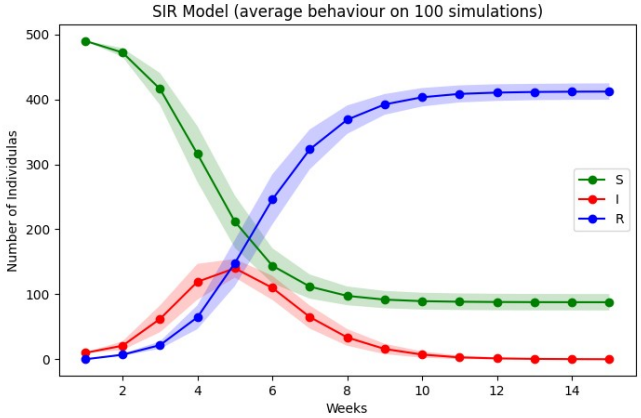

Fig. 5.

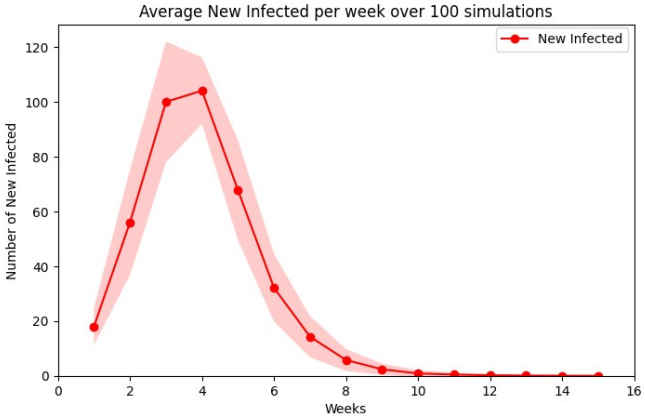

Fig. 6.

III.

We now introduce an additional compartment to the model, that of vaccinated individuals. We are given the following vector containing the fraction of population that *so far* has received vaccination for each week:

$$V(t) = [0, 5, 15, 25, 35, 45, 55, 60, 60, 60, 60, 60, 60, 60, 60]$$

From this data, knowing that the total population is of $N = 500$ individuals, we can calculate the number of individuals that should be vaccinated each week in the algorithm. For example in the third week 15% of the population has been vaccinated, with $V(3) - V(2) = 10\%$ receiving it *during* week 3, i.e. 50 people. Note that since this quantity

is fixed for each week and known a priori, the standard deviation for the vaccinated compartment will be null. We also underline how infected people are valid candidates for receiving vaccination since they may not be aware they have contracted the disease and get vaccinated all the same (e.g. asymptomatic individuals).
Figures 7 and 8 show the outcome of the simulations of the epidemics with vaccination.

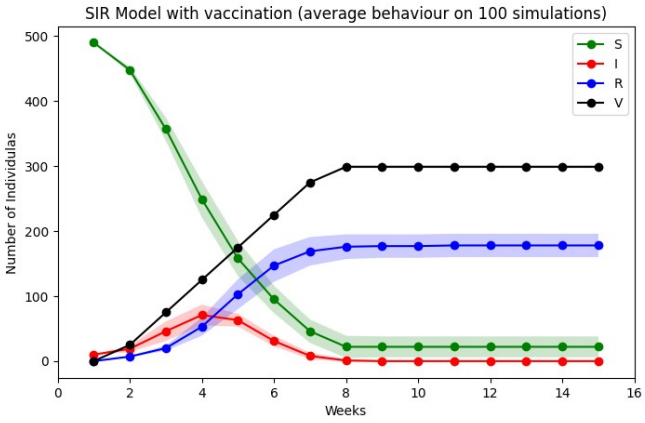

Fig. 7.

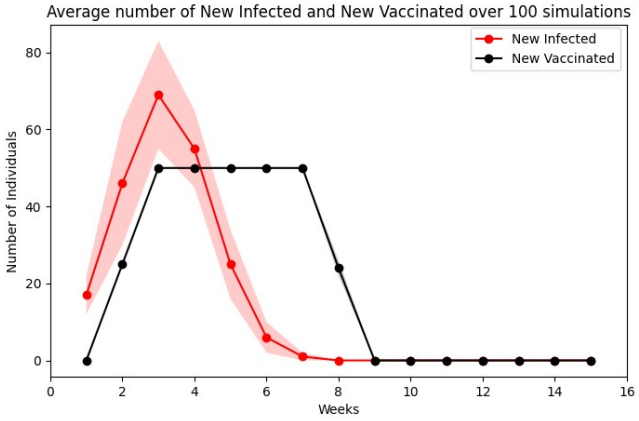

Fig. 8.

If we compare Figure 8 with Figure 6 we see that the introduction of vaccination reduces on average the peak of new infected individuals by 30% (from 100 to 70 approximately), which results in an epidemic that is much more manageable. The effect of vaccination is thus comparable to social distancing, since it reduces the probability of getting infected. This is visible also looking at the $S(t)$ curve, which decreases more rapidly and reaches smaller values: at week eight without vaccination we have approximately 100 susceptible individuals versus 25 with vaccination. Furthermore we see that with vaccination also the number of recovered individuals is significantly lower.

IV.

In this part we analyse the Swedish H1N1 epidemic of 2009. Specifically the goal is to try to estimate the disease-spread parameters k, β and ρ between week 42, 2009 and week 5, 2010.
In order not to spend too much time running simulations, we scale down the population of Sweden by a factor of 10^4. By doing so, we have a population of $N = 934$. For this scaled version, the number of newly infected individuals each week is provided and given by

$$I_0(t) = [1, 1, 3, 5, 9, 17, 32, 32, 17, 5, 2, 1, 0, 0, 0, 0]$$

while the fraction of population that had received vaccination is

$$V(t) = [5, 9, 16, 24, 32, 40, 47, 54, 59, 60, 60, 60, 60, 60, 60, 60]$$

To obtain the values that best represent the real pandemic we simulate a gradient-based search over the parameter space of k, β, ρ. Firstly we define an initial guess of parameters k_0, β_0, ρ_0, along with some $\Delta k, \Delta\beta, \Delta\rho$. For each set of parameters (k, β, ρ) in the parameter-space $k \in \{k_0 - \Delta k, \beta_0, k_0 + \Delta k\}$, $\beta \in \{\beta_0 - \Delta\beta, \beta_0, \beta_0 + \Delta\beta\}$ and $\rho \in \{\rho_0 - \Delta\rho, \rho_0, \rho_0 + \Delta\rho\}$ we generate a random graph with average degree k (using the preferential attachment model) and number of nodes equal to the total population $N = 934$. We then simulate the pandemic for 15 weeks on G using the previous method, and do so $N = 10$ times. Once this step is complete the root-mean-square error (RMSE) between the real number of newly infected $I_0(t)$ and that obtained by means of the simulation $I_{new}(t)$ is computed.

Once the RMSE has been calculated for all possible combinations we pick the one that yields the lowest RMSE and update k_0, β_0, ρ_0 with such optimal parameters. Furthermore, to improve the optimisation we also half $\Delta k, \Delta\beta, \Delta\rho$ so as to narrow down the neighborhood being analysed (since we should be getting closer to the optimal configuration). A delicate point in this process is choosing the initial values of the parameters. It is a good idea to explore the parameter-space a bit before starting the process, in order to start off already with a good set of values (i.e. whose related loss is not too high). In this case, k_0, β_0, ρ_0 were set to $13, 0.3, 0.6$ respectively, while $\Delta k = 1, \Delta\beta = 0.1, \Delta\rho = 0.1$.
The whole process was repeated a total of 10 times in order to obtain a more robust estimate of the best parameters, calculated as the average and coinciding with $k = 12$, $\beta = 0.15$ and $\rho = 0.68$ and with associated loss 3.88 ± 1.09. Figure 10 compares the predicted number of new infected $I_{new}(t)$ over the 10 simulations with $I_0(t)$ (the real number of infected individuals). To evaluate how these parameters perform we also ran 20 simulations of an epidemic with the optimal k, β and ρ, each repeated on a given graph 100 times, and found an average validation loss of 7.72. We suspect this deterioration to be due to two factors: i) the number of simulations used to determine the optimal parameters was set to 10 for execution time's sake, which is limiting ii) the starting graphs are always different due to the randomness of the process. Figure 9

shows the *S(t), I(t), R(t)* curves obtained with the best set of parameters.

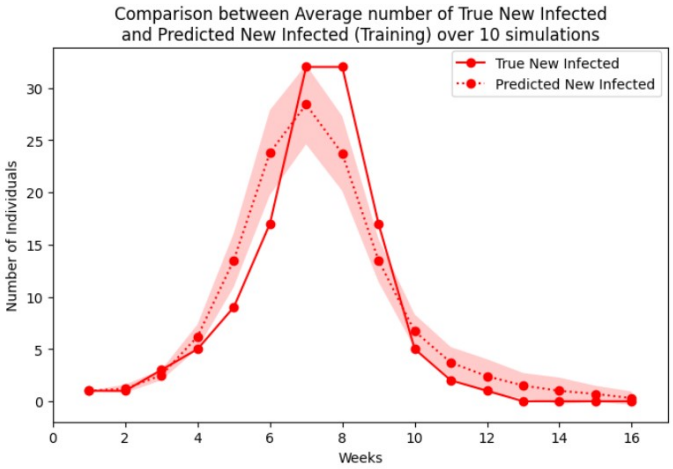

Fig. 9.

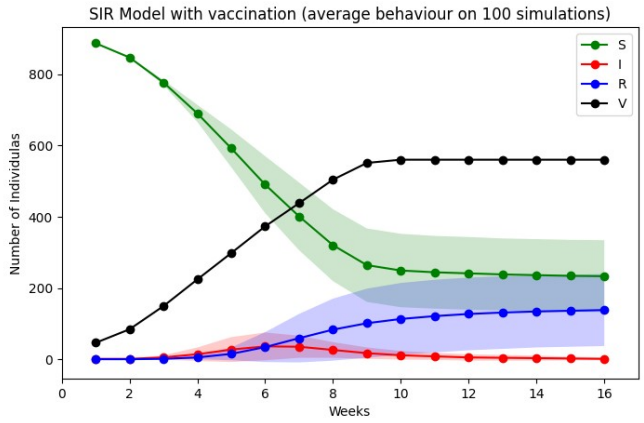

Fig. 10.

V.

In this section we present a few ways to possibly improve the performances obtained so far. Firstly we try and generate a better random graph, and to do so use the Holme and Kim algorithm for growing graphs with powerlaw degree distribution and approximate average clustering. Defining a graph with this method requires setting 3 parameters: the number of nodes $|V| = 934$, the number of random edges to add for each new node (which is none other than k) and finally the probability p of adding a triangle after adding a random edge. In other words, the difference with respect to the graph obtained with preferential attachment is that we allow for the construction of a triangle with a given probability, which in practical terms has the effect of increasing the interactions. It follows that in addition to k, β and ρ we must also optimize p. By using the gradient descent method defined in section IV we find $k = 4.0$, $\beta = 0.29$, $\rho = 0.67$ and $p = 0.43$ which allow us to slightly improve performances by granting us a loss of 3.68 ± 0.51. Figures 11 and 12 show the relative plots.

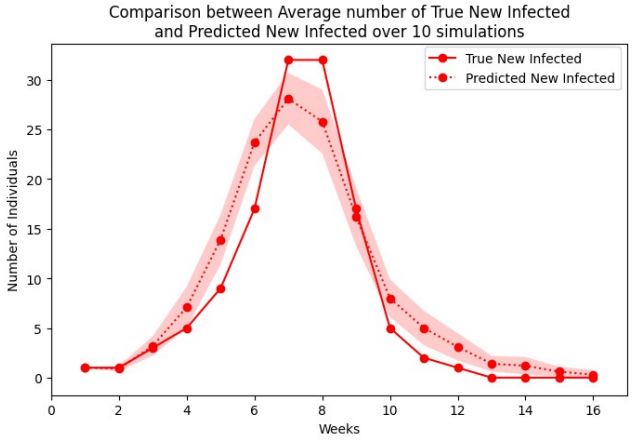

Fig. 11.

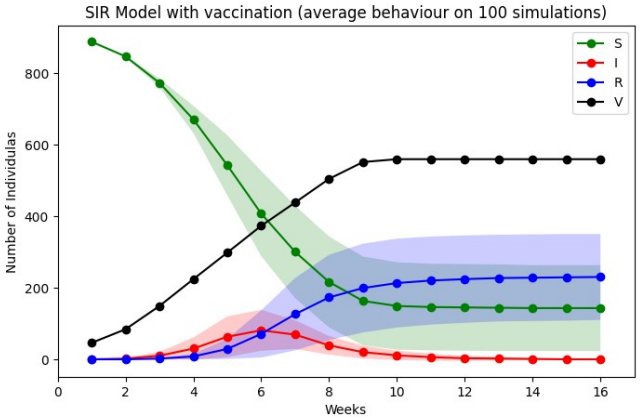

Fig. 12.

Moreover, although not applicable in this context since Sweden is known to avoid draconian measures, a possible modification could consist in redefining the parameter β as follows:

$$\beta(t) = \begin{cases} \beta_0 & t \leq t_{lock} \\ \beta_0 e^{\frac{-t+t_{lock}}{\tau_\beta}} & t > t_{lock} \end{cases} \tag{2}$$

This piecewise function takes into account the fact that once lockdown is imposed at time $t = t_{lock}$ the transmission rate of the disease will decrease since there are fewer interactions, making β a time dependant variable which decreases exponentially (note that in this case we would also have to determine the optimal parameter τ_β)
To conclude we also note that if data for the other compartments were available we could redefine our loss for the optimisation function in such a way as to take into account all categories (for example we could set it equal to the sum of RMSEs calculated on each compartment).