Report homework 3
Network Dynamics and Learning

Matteo Merlo s287576

January 15, 2022

0 Introduction

This is the report of the third homework of the course Network Dynamics and Learning 2021/2022. Topic: Epidemics and Random graphs.

1 Epidemic on a known graph

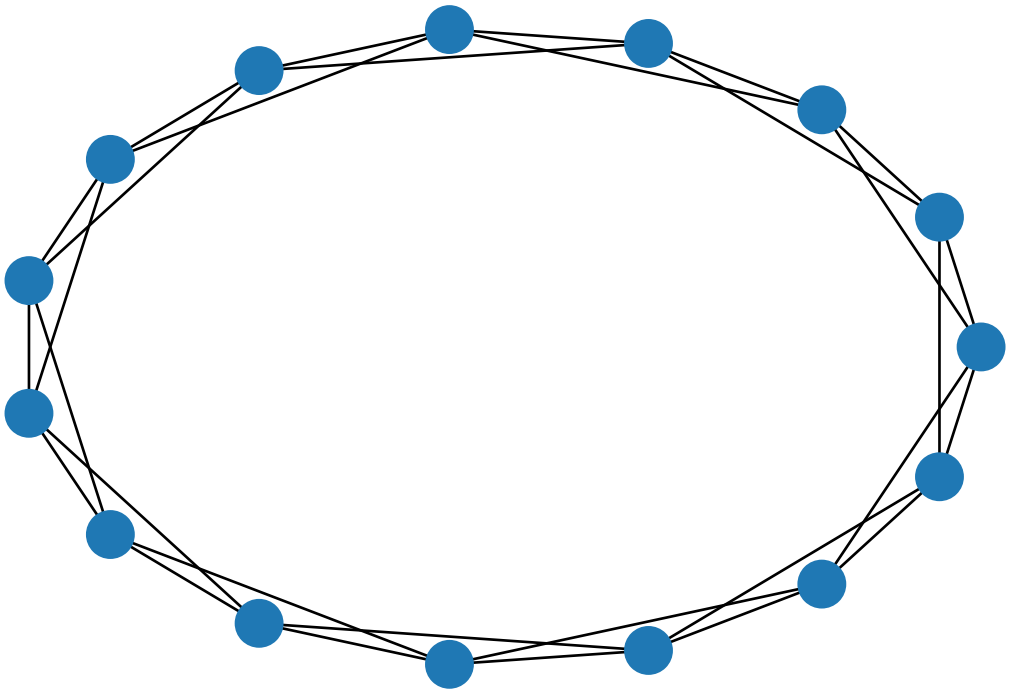

Figure 1: Regular Graph

In the first problem an epidemic is simulated on a symmetric k-regular undirected graph $\mathcal{G}$ with $|\mathcal{V}| = 500$ nodes and $k = 4$ for 15 weeks (steps). In this kind of graph, every node is connected to the k nodes whose index is closest to their own modulo n. The graph is unweighted and undirected. In Figure 1 is showed a regular graph N = 15 and k = 4.

Then, we define a discrete-time simplified version of the SIR epidemic model. Our network system simulates pairwise interactions and spontaneous mutations.

Each individual is called **agent** and each of them is associated with a state related to the H1N1-virus infection. Consider the state space $\mathcal{A} = \{0, 1, 2\}$ is corresponding respectively to:

- **S** (susceptible)
- **I** (infected)
- **R** (recovered)

We can describe our system through a discrete-time Markov chain $X(t)$ on the configuration space $\mathcal{X} = \mathcal{A}^{\mathcal{V}}$ defined by an initial distribution x and a transition probability matrix P.

We define the following objects.

- The probability $\beta \in [0, 1]$ that the infection is spread from an infected individual **I** to a susceptible one **S**, given that they are connected by a link, during one time step.
- The probability $\rho \in [0, 1]$ that an infected individual **I** will recover, i.e., mutate its state into **R**, during one time step.

The epidemic is driven by the following transition probabilities.
Consider a node $i \in \mathcal{V}$.

1. If node i is susceptible **S** at time t, then:
   - $P(X_i(t+1) = \mathrm{S} \mid X_i(t) = \mathrm{S},\ \sum_{j \in \mathcal{V}} W_{ij} \delta^I_{X_j(t)} = m) = (1-\beta)^m$, where:
     - $\sum_{j \in \mathcal{V}} W_{ij} \delta^I_{X_j(t)}$ is the number of infected neighbors for node i
     - $(1-\beta)^m$ is the probability that node i does **not** get infected by any of the neighbors during one time step
   - $P(X_i(t+1) = \mathrm{I} \mid X_i(t) = \mathrm{S},\ \sum_{j \in \mathcal{V}} W_{ij} \delta^I_{X_j(t)} = m) = 1 - (1-\beta)^m$
     - $1 - (1-\beta)^m$ instead is the probability that node i get infected by any of the neighbors during one time step
   - $P(X_i(t+1) = \mathrm{R} \mid X_i(t) = \mathrm{S}) = 0$
2. If node i is infected **I** at time t, then:
   - $P(X_i(t+1) = \mathrm{S} \mid X_i(t) = \mathrm{I}) = 0$
   - $P(X_i(t+1) = \mathrm{I} \mid X_i(t) = \mathrm{I}) = 1 - \rho$
   - $P(X_i(t+1) = \mathrm{R} \mid X_i(t) = \mathrm{I}) = \rho$
3. If node i is recovered **R** at time t, then:
   - $P(X_i(t+1) = \mathrm{S} \mid X_i(t) = \mathrm{R}) = 0$
   - $P(X_i(t+1) = \mathrm{I} \mid X_i(t) = \mathrm{R}) = 0$
   - $P(X_i(t+1) = \mathrm{R} \mid X_i(t) = \mathrm{R}) = 1$

The proposed implementation of the discrete-time SIR model takes any graph $\mathcal{G}$ and runs an epidemic simulation with parameters β, ρ and number of initially infected individuals for a given number of steps.

The time interval considered between each discrete step is one week.

The discrete time starts at $t = 0$ and the simulation runs until the end at week $t = 15$, for a total of 16 discrete time steps.

The model will be running a the discrete-time simplified version of the SIR model 100 times for 15 weeks, with parameters $\beta = 0.3$ and $\rho = 0.7$. For each of these simulations, a random different initial configuration with 10 different infected nodes is chosen on $\mathcal{G}$ according to a uniform distribution. The other nodes are all susceptible.

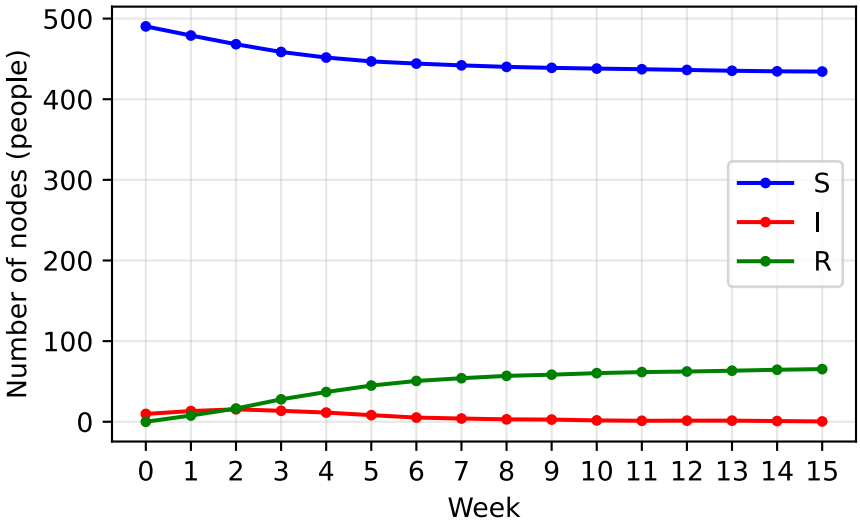

Figure 2: SIR model result of first simulation

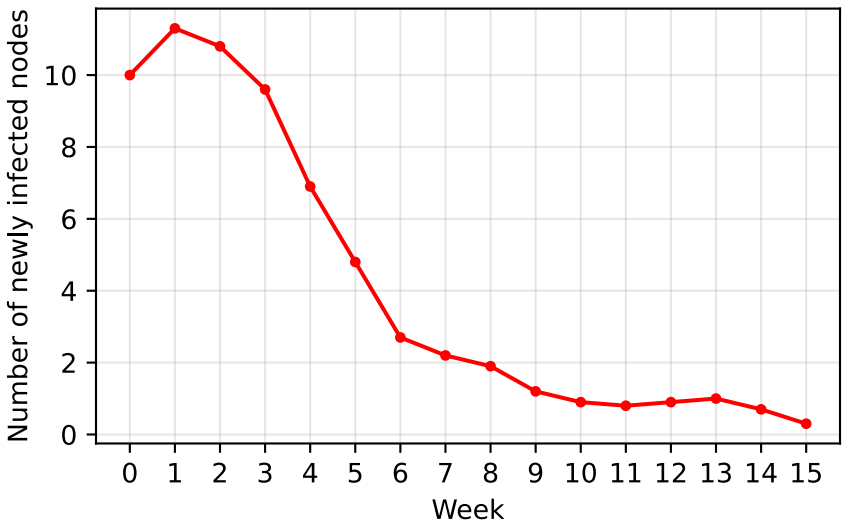

Figure 3: New infected each week in the first simulation

The previous plot shows how nodes that are initialized randomly on the graph propagate the infection. The epidemic slows down early and does not reach new nodes fast enough to infect most individuals in the population. That because of the graph structure of the simulation almost reaches the absorbing configuration where susceptible represent the 90% of the population whereas the other 10% is represented by recovered (or dead there is no such distinction in this case) people.

2 Simulate a pandemic without vaccination

The second problem consists in simulating an epidemic based on the discrete-time SIR model using a preferential attachment random graph.
In Figure 4 is showed a model of random graph with N = 15 and k = 5 with preferential attachment.

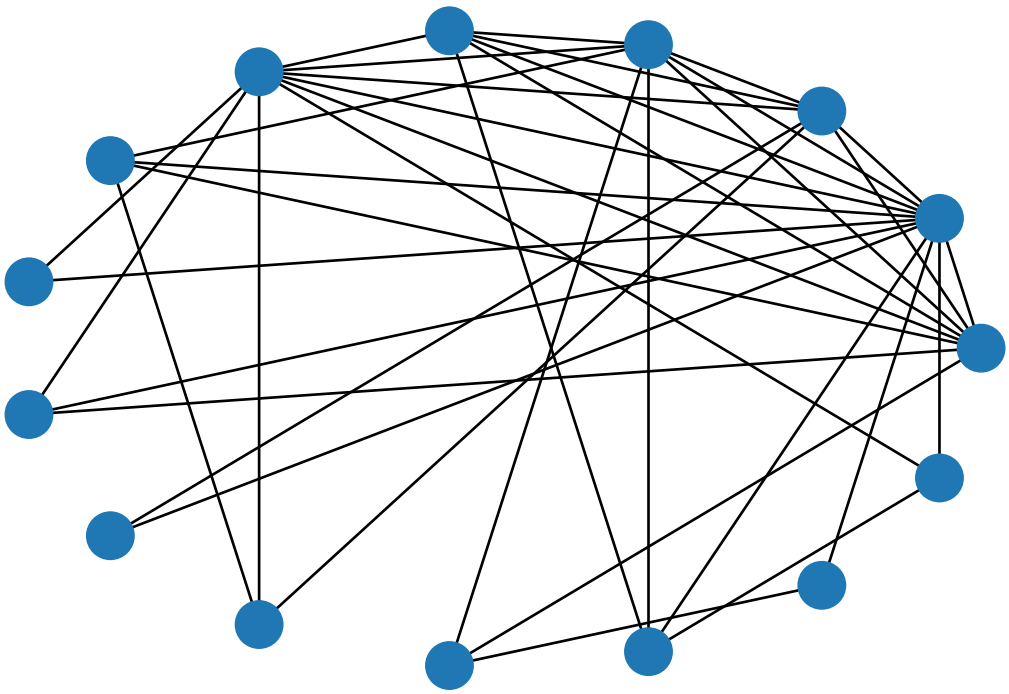

Figure 4: Random graph N = 15, k = 5 with referential attachment

In this case 100 simulations are performed as before and plot the average total number of susceptible, infected and recovered individuals each week and the average number of newly infected individuals each week.

The preferential attachment graph idea is the following:

- at time $t = 0$ with an initial undirected graph $\mathcal{G}_0$;
- at each time $t \in \{1, 2, ...\}$, we create a new graph $\mathcal{G}_t$ obtained by adding a new node to $\mathcal{G}_{t-1}$, and connecting such a node to c nodes randomly selected, according to a preferential attachment rule.

This rule says that at each time step t, the new node n_t will have a degree $w_{n_t}(t) = \frac{k}{2}$. The nodes to which n_t should connect are chosen according to the following probability rule.

The probability that there will be a link between node n_t and node $i \in \mathcal{V}_{t-1}$ is:

$$P(W_{n_t,i}(t) = W_{i,n_t}(t) = 1 \mid \mathcal{G}_{t-1} = (\mathcal{V}_{t1}, \mathcal{E}_{t1})) = \frac{w_i(t-1)}{\sum_{j \in \mathcal{V}_{t-1}} w_j(t-1)}, \qquad i \in \mathcal{V}_{t-1}$$

Each of the 100 simulations consists in the following steps.

- Generate a random graph according to the *preferential attachment model*, with average degree close to k.
- Simulate an epidemic according to the simplified version of the SIR model.

The parameters of the simulations k, number of nodes, β, ρ, number of steps and number of initially infected individuals are fixed for all 100 simulations.

In our simulations, a graph is generated with an average degree of $k = 6$ and 500 nodes. $\beta = 0.3$ and $\rho = 0.7$ are set as required by delivery, and carry out each simulation for 15 weeks, with an initial configuration of 10 infected nodes.

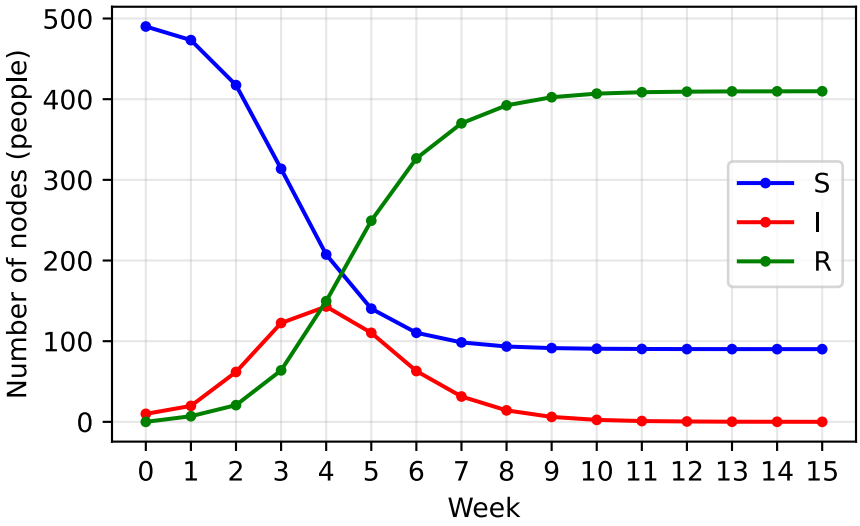

Figure 5: SIR model result without vaccination

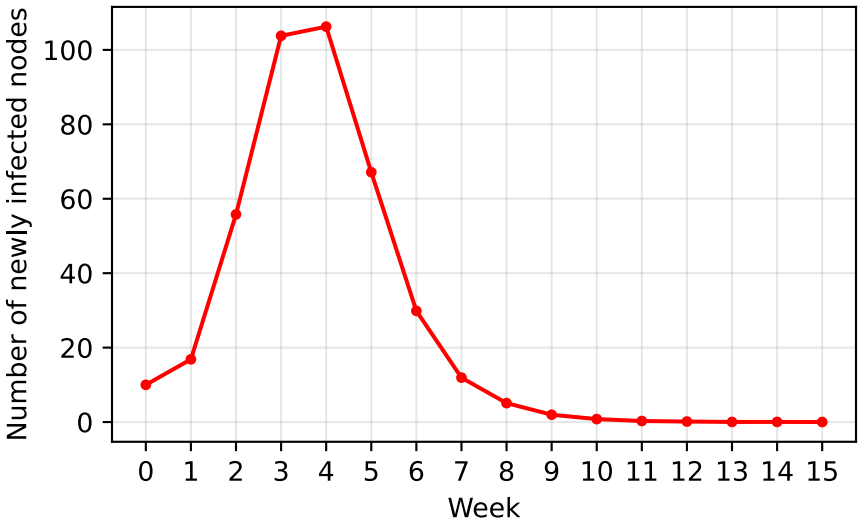

Figure 6: New infected each week without vaccination

The infection rate grows faster and both the infection cases and the newly infections present a more defined bell shape, while the susceptible decrease drastically approximately to the 30% in 6 weeks, anyway the recovery rate is pretty high and in the end the new infections rate is 0, which could lead towards a stable configuration.

3 Simulate a pandemic with vaccination

In this part a SIR model takes into account **vaccinated individuals** as a separate state **V**. Vaccinated individuals are not able to become infected nor infect any other individual.

The new state space becomes $\mathcal{A} = \{0, 1, 2, 3\}$ corresponding to:

- **S** (susceptible)
- **I** (infected)
- **R** (recovered)
- **V** (vaccinated) states.

The number of nodes to be vaccinated is determined by the array $\text{Vacc}(t)$. This array represents the total fraction of population that has received vaccination by the end of that week. It starts from $t = 1$ and the first element of the array refers to the fraction of population that has received vaccination *by* week 1. The vaccination array given is:

$$\text{Vacc}(t) = [5, 15, 25, 35, 45, 55, 60, 60, 60, 60, 60, 60, 60, 60, 60]$$

To clarify the idea, according to this array, 35% of the population has received vaccination *by* week 4. In the time period that goes from the instant in which the simulation begins to the last instant of the 5th week, and 10% of the population has received vaccination during week 4. Since we assume that the vaccination takes effect immediately once given, the nodes that receive vaccination during week 4 will not be able to become infected nor infect any other node in the future.

An epidemic model based on the discrete-time SIR with vaccination (SIRV) is simulated. 100 simulations of this type are performed and a plot will show the average total number of susceptible, infected, recovered and vaccinated individuals each week and the average number of newly infected and newly vaccinated individuals each week.

In each of the 100 simulations, a graph is generated with average degree $k = 6$ and 500 nodes and simulated for 15 weeks with parameters $\beta = 0.3$, $\rho = 0.7$ and an initial configuration of 10 randomly chosen infected nodes. These settings are exactly the same as the previous simulation without vaccination, so results are more easily comparable.

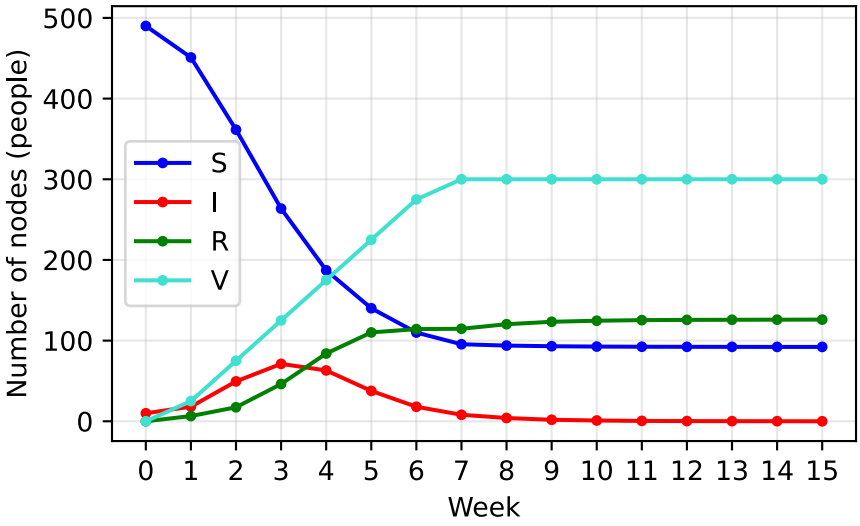

Figure 7: SIR model result with vaccination

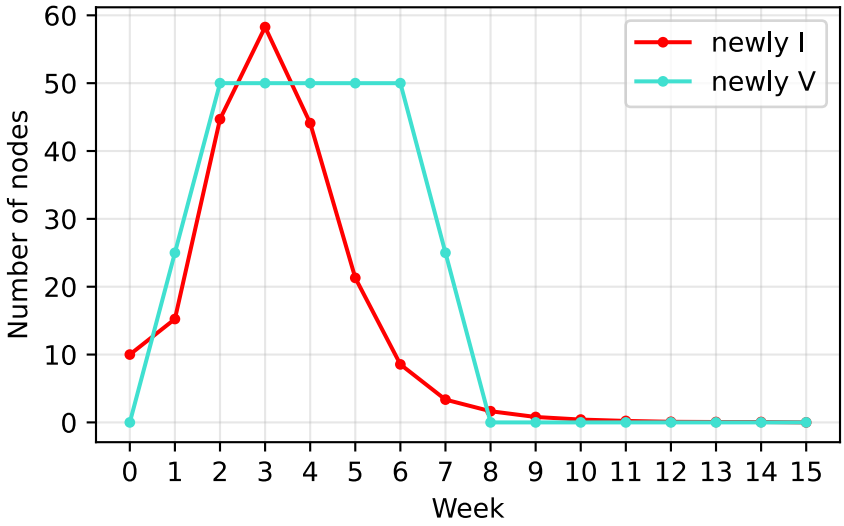

Figure 8: New infected and new vaccinated each week

We can observe that the vaccination campaign made a huge difference, the infection rate decreases rapidly, and the final state of recovered + infected population is around 20% (against 70/80% of the previous experiment).

4 The H1N1 pandemic in Sweden 2009

Our goal is to estimate parameters k, β and ρ for our model which best match the **H1N1 pandemic in Sweden in 2009**, which we simulate for 15 weeks. In the following vector is reported the fraction of population that had received vaccination during these weeks.

$$\text{Vacc}(t) = [5, 9, 16, 24, 32, 40, 47, 54, 59, 60, 60, 60, 60, 60, 60]$$

The population of Sweden [1] for computational purposes is scaled down to 934 nodes. The number of newly infected individuals each week in the 15-week period from week 42, 2009 to week 5, 2010 is reported in the following vector.

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

To estimate the parameters that best match the real pandemic, we define an algorithm which performs a grid-search over the parameter space of k, β and ρ. The distance between our simulation and the real pandemic is measured through the root-mean-square error (RMSE) between the number of infected individuals each week $I(t)$ of the simulation and the number of infected individuals each week $I_0(t)$ of the real pandemic. This could be interpreted as the *cost function* of our algorithm.

$$\text{RMSE} = \sqrt{\frac{1}{n_{\text{weeks}}} \sum_{t=1}^{n_{\text{weeks}}} \left(I(t) - I_0(t)\right)^2} \qquad \text{with } n_{\text{weeks}} = 15 \text{ steps of the simulation}$$

Algorithm
The passages of algorithm are the following.

1. Fix an initial guess of the parameters $k_0 = 10$, $\beta_0 = 0.3$, $\rho_0 = 0.6$ and define $\Delta k = 1$, $\Delta\beta = 0.1$, $\Delta\rho = 0.1$ (then other initial combination are tried).

2. Construct the search space as a list of $3 * 3 * 3 = 27$ parameter sets, given by $\{k_0 - \Delta k, k_0, k_0 + \Delta k\} \times \{\beta_0 - \Delta\beta, \beta_0, \beta_0 + \Delta\beta\} \times \{\rho_0 - \Delta\rho, \rho_0, \rho_0 + \Delta\rho\}$.

3. For each parameter set (k, β, ρ):
   (a) Generate a random graph using the preferential attachment model with average degree k and 934 nodes.
   (b) Simulate the pandemic 10 times, using the SIRV model with the given parameters.
   (c) Compute the average number of newly infected individuals each week, $NI(t)$.
   (d) Compute the RMSE between the simulation and the real pandemic.
4. Update k_0, β_0, ρ_0 to the set of parameters yielding the lowest RMSE.
5. If the best result achieve has the same set of parameters from the previous k_0, β_0, ρ_0 it stop and exit. Otherwise, return to 2.

To increase the variability of results an integer α variable is added to multiply the center of intervals k_0, β_0 and ρ_0. The basic idea is escape from a local minima and try to find a new parameters that could improve the general result [4].

For the simulation the starting parameters are the following instead:

- $k_o = 8$, $\Delta k = 1$,
- $\beta_o = 0.3$, $\Delta\beta = 0.025$,
- $\rho_o = 0.6$, $\Delta\rho = 0.025$,

Then our goal is to understand how well the model with optimal parameters actually fits the real epidemic. Therefore, we perform 100 simulations using the best set of parameters, and plot the average number of newly infected individuals each week compared to the true value of newly infected individuals each week $I_0(t)$, as well as the average total number of susceptible, infected, recovered and vaccinated individuals each week.

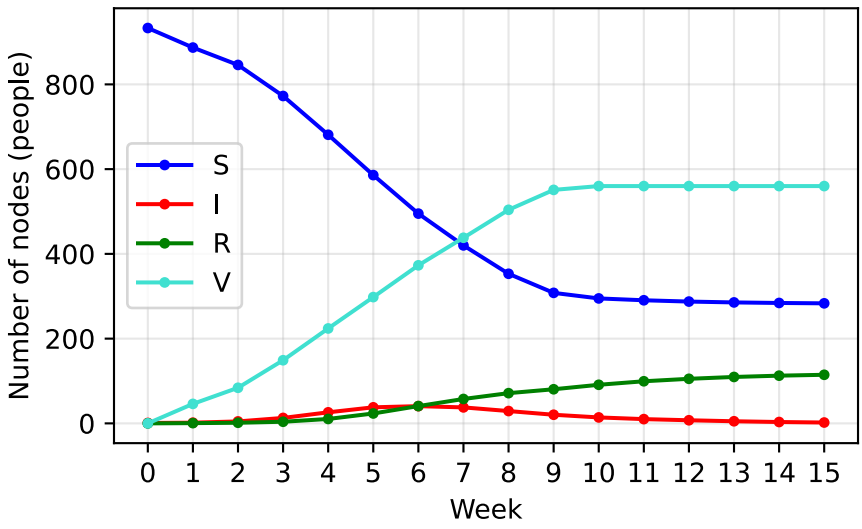

Figure 9: SIR Sweden model result with preferential attachment

From the result we can observe, the RMSE cost generally tries to get lower as more steps are taken. However, it remains still quite high despite the algorithm stops, indicating that the set of parameters that we found does not fit very well the real data.

Moreover, a huge number of parameters have been tried. This indicates that the variability of the experiment is quite high, and even by initializing the parameters to a fixed value and running several simulations per parameter set, we still have several sources of randomness. Finally, if we observe that it is evident how the RMSE loss oscillates and in general, the algorithm has some problem in

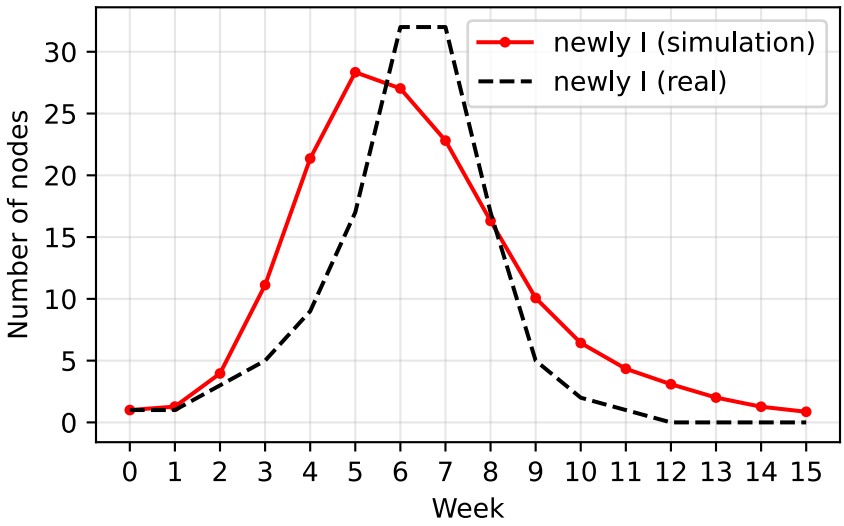

Figure 10: Comparison with new infected simulated (red) and new vaccinated real in Sweden each week

converging. This may be due to the variability of simulations, and also due to the algorithm finding a local minimum from which it is not able to escape.

Nonetheless, the best set of parameters found is $k = 9$, $\beta = 0.175$, $\rho = 0.6$ associated with the lowest RMSE score, 2.64.

In figure 10 clearly shows how the optimal parameter-set found does not fit very well the real epidemic curve. The most evident issue is that the number of newly infected individuals rises faster then real one, reaching a peak at week 5. It subsequently drops too early without ever reaching the actual number of newly infected of the real pandemic. Moreover, at the end from week 10, the number of newly infected nodes remains too high as the epidemic fades, in contrast with the more severe drop of the real epidemic.

5 Challenge

A potential problem with our algorithm may be the partial inadequacy of the preferential attachment graph. The social structure and the interactions between individuals of the population of Sweden may not be modelled very well with this kind of graph. In an attempt to solve this problem, another type of random graph is explored and a new run to search new best parameter will be done, with the goal of further lowering the RMSE score and find a better fit for the real epidemic.

The graph model choosed is the **Newman-Watts-Strogatz small-world model** [2]. This model generates a random graph as follows:

1. Create a ring over the n nodes of the graph, each node connected to k neighbors, $\frac{k}{2}$ on each side. We denote this underlying graph with $\mathcal{G}_u$.

2. For each edge (i, j) in the underlying graph $\mathcal{G}_u$, it adds an edge (i, k) with a randomly chosen existing node w with probability p.

3. Generate a random graph $\mathcal{G}$.

In this case, the parameter grid is composed by four different parameters, namely k (number of nearest neighbors of the underlying graph), p (probability of creating a new edge), β and ρ. This means that the search space will be a list of $3*3*3*3 = 81$ parameter sets, given by $\{k_0 - \Delta k, k_0, k_0 + \Delta k\} \times \{p_0 - \Delta p, p_0, p_0 + \Delta p\} \times \{\beta_0 - \Delta\beta, \beta_0, \beta_0 + \Delta\beta\} \times \{\rho_0 - \Delta\rho, \rho_0, \rho_0 + \Delta\rho\}$.

For the simulation the starting parameters are the following:

- $k_o = 9$, $\Delta k = 1$,
- $\beta_o = 0.25$, $\Delta\beta = 0.025$,
- $\rho_o = 0.95$, $\Delta\rho = 0.025$

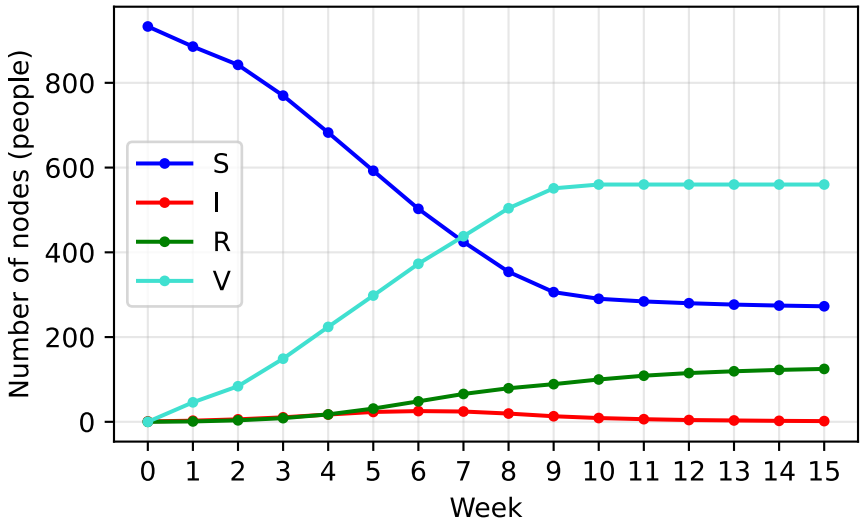

Figure 11: SIR Sweden model result with Newman-Watts-Strogatz small-world model

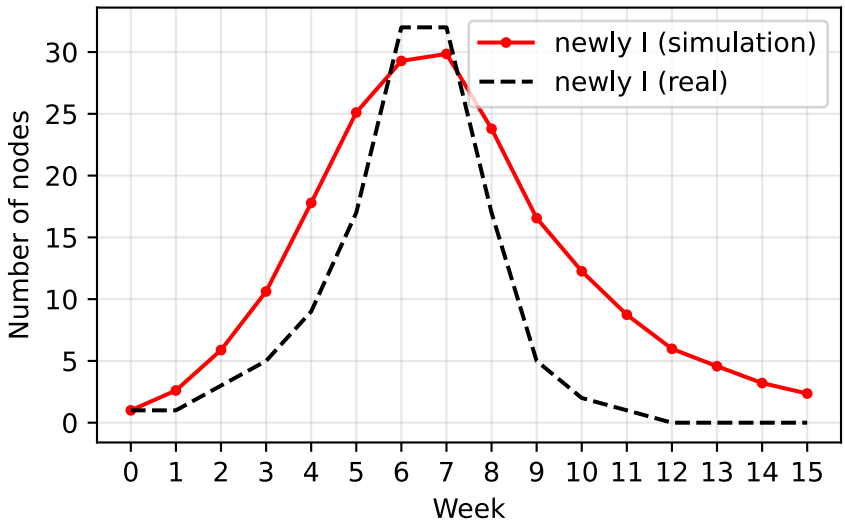

Figure 12: Comparison with new infected simulated (red) and new vaccinated real in Sweden each week

From the multiple run, the best set of parameters found is $k = 9$, $\beta = 0.25$, $\rho = 1$ and *probability* = 0.5 associated with the lowest RMSE score, 3.73. In Figure 12 it shows that the Newman-Watts-Strogatz model does not fit the real epidemic trend better than the preferential attachment model. The curve does not change so much. While it partially solves the problem of the epidemic peaking too early, the number of newly infected remains excessively high before and in particular after the peak number of newly infected. Most importantly, the peak number of newly infected is never reached, but remains well under the peak of the real epidemic.

So we may assume that:

- Probably the RMSE function may be inadequate as loss function for this type of models.
- Isolated people with small contact each other are not considered. Sweden is a huge country with a low density.

In conclusion, we acknowledge that both models have problems in fitting well the real epidemic, and a more extensive research is needed to find a model or a better combination of parameters.

6 Collaborations

The results, especially for exercise 4 and 5, are compared with:

- Matteo Giardino
- Andrea Tampellini

References

[1] Sweden. https://en.wikipedia.org/wiki/SwedenDemographics.
[2] Watts–Strogatz model. https://en.wikipedia.org/wiki/Watts-Strogatz_model.
[3] Fabio Fagnani Giacomo Como. Lecture Notes of Network Dynamics and Learning.
[4] R. Tadei, F. Della Croce, and A. Grosso. Fondamenti di Ottimizzazione. Esculapio, 2005