

Homework 3

Andrea Rubeis, s290216

January 15, 2022

Abstract

Here are reported the solutions of the third homework of Network Dynamics and Learning course. Solutions of part 3 and 4 discussed with Shannon Mc Mahon (s289958). Code [here](#).

H1N1 Epidemic in Sweden

During the fall of 2009 there was a large pandemic of the H1N1-virus, commonly known as the swine-flu. During this pandemic it is estimated that about 1.5 million people in Sweden were infected. As an attempt to stop the pandemic and reduce excess mortality the government issued a vaccination program beginning in week 40 of 2009. During the weeks that followed they vaccinated more than 60% of the Swedish population. In this homework, you will simulate the pandemic with the goal of learning the network-structure characteristics and disease-dynamics parameters of the pandemic in Sweden 2009. This task will be divided into 4 parts where the focus of each part is to:

1. get started and learn how to:
 - a simulate a pandemic on a known graph;
 - b generate a random graph;
2. simulate the disease propagation on a random graph without vaccination;
3. simulate disease propagation on a random graph with vaccination;
4. estimate the network-structure characteristics and disease-dynamics parameters for the pandemic in Sweden during the fall of 2009.

All numbers regarding the H1N1 pandemic in Sweden during the fall of 2009 have been taken from the a report by the Swedish Civil Contingencies Agency (Myndigheten for samhällsskydd och beredskap, MSB) and the Swedish Institute for Communicable Disease Control (Smittskyddsinstitutet, SMI).

Preliminary Parts

As a warm-up exercise we will start off by doing two preliminary parts. The first one will involve simulating an epidemic on a given graph, while the second part will be to generate a random graph with preferential attachment.

1.1 Epidemic on a known graph

In this part you will simulate an epidemic on a symmetric k -regular undirected graph with node set $\mathcal{V} = 1, \dots, n$ where every node is directly connected to the $k = 4$ nodes whose index is closest to their own modulo n . See Figure 1 for an example with 8 nodes. The graph that you will simulate the epidemic on will however contain $n = 500$ nodes.

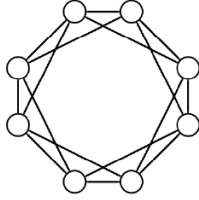


Figure 1: Symmetric k -regular graph

The disease propagation model that you will use to simulate the epidemic is a discrete-time simplified version of the SIR epidemic model. At any time $t = 0, 1, \dots, n$ nodes are in state $X_i(t) \in S, I, R$, where S is susceptible, I is infected and R is recovered. Let $\beta \in [0, 1]$ be the probability that the infection is spread from an infected individual to a susceptible one (given that they are connected by a link) during one time step. Assuming that a susceptible node i has m infected neighbors, this means that the probability that individual i does not get infected by any of the neighbors during one time step is $(1 - \beta)^m$. Thus, the probability that individual i becomes infected by any of its neighbors is $1 - (1 - \beta)^m$. Furthermore, let $\rho \in [0, 1]$ be the probability that an infected individual will recover during one time step. The epidemic is driven by the following transition probabilities

$$\mathbb{P}(X_i(t+1) = I | X_i(t) = S, \sum_{j \in \mathcal{V}} W_{i,j}, \delta_{X_j(t)}^I = m) = (1 - \beta)^m \quad (1)$$

$$\mathbb{P}(X_i(t+1) = R | X_i(t) = I) = \rho \quad (2)$$

where $\sum_{j \in \mathcal{V}} W_{i,j}, \delta_{X_j(t)}^I$ is the number of infected neighbors for node i .

Problem 1.1; You should simulate an epidemic on a symmetric k -regular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = 500$ nodes and $k = 4$. See Figure 1 for an example with

$n = 8$ nodes. Let $\beta = 0.3$ and $\rho = 0.7$. With one week being one unit of time, simulate the epidemic for 15 weeks. You can choose an initial configuration with 10 infected nodes selected at random from the node set \mathcal{V} , or make a different choice of initial configuration (in the latter case, please briefly discuss your motivation). Do this $N = 100$ times and plot the following:

- The average number of newly infected individuals each week. In other words, you should plot how many people *become* infected each week (on the average).
- The average total number of susceptible, infected, and recovered individuals at each week. In other words, you should plot how many individuals *in total that are* susceptible/infected/recovered at each week (on the average).

Hint: Since we use a fairly large amount of nodes for this simulation it is a good idea to use sparse matrices for this and the following problems.

Solution: According to the instructions above, we simulate the Sewdsh H1N1 epidemic using the *SIR* model on a symmetric k -regular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = 500$ nodes and $k = 4$ with β and ρ equal to 0.3 and 0.7 respectively. We start with 10 infected agents selected at random from \mathcal{V} and simulate $N = 100$ times. The images below illustrate the average number of newly infected individuals and the average total number of susceptible, infected and recovered each week.

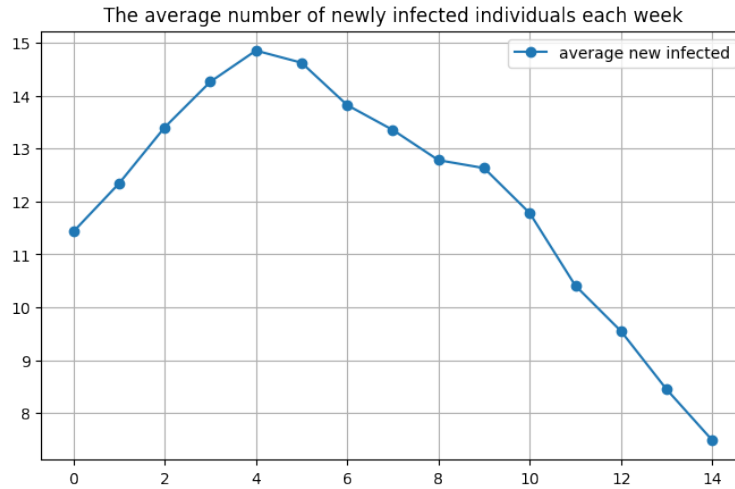


Figure 2: Average newly infected, $k = 4$

As we can see from Figure 2, the average newly infected has an increasing trend in the first weeks reaching a peak close 15, but then from week 4 to the last week simulated we have a decreasing trend.

The average total number of susceptible, infected, and recovered individuals at each week

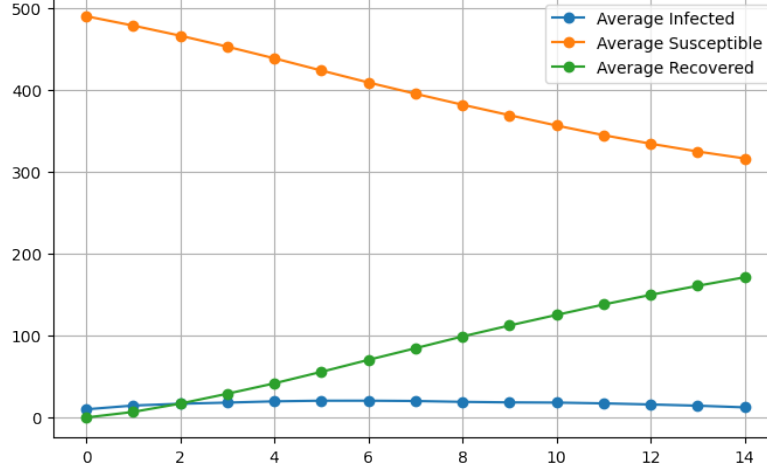


Figure 3: Average total susceptibles, infected and recovered, $k = 4$

In Figure 3 instead the average total number of infected (blue), susceptibles (yellow) and recovered (green) are reported. The average of infected people remains quiet stable and close to 10 for the whole simulation, which means that the epidemics is under control given by the pretty high value of ρ (0.7). The susceptible agents instead have a decreasing trend as long as people get infected. On the opposite, the recovered ones have an increasing trend given by the fact that infected people can recover from the disease. We can notice that the overall system tends to the absorbing configuration $\{S, R\}^{|\mathcal{V}|}$, indeed from the theory we know that in *SIR* modelling if we start with few infected agents at time 0 for t that goes to ∞ we will have the absorbing configuration $\{S, R\}^{|\mathcal{V}|}$. However, 15 weeks are not enough to reach it.

1.2 Generate a random graph

In this part you will generate a random graph according to the *preferential attachment model*. The goal is to have a randomly generated graph with average degree close to k . The idea is the following: at time $t = 1$ we start with an initial graph \mathcal{G}_1 , that is *complete* with $k + 1$ nodes. Then at every time $t \geq 2$, create a new graph $\mathcal{G}_t = (\mathcal{V}_t, \mathcal{E}_t)$ by adding a new node to \mathcal{G}_{t-1} and connect it to some of the existing nodes \mathcal{V}_{t-1} of \mathcal{G}_{t-1} chosen according to some stochastic rule. The rule by which the new node add links to the nodes of \mathcal{G}_{t-1} is *preferential attachment*. This means that at every time-step $t \geq 2$, every new node added at time t will have a degree $w_t(t) = c = k/2$. Hence, it should add c undirected links to the existing graph \mathcal{G}_{t-1} . It decides which of the nodes in \mathcal{V}_{t-1} it should connect to based on some probability that is proportional to the current degree

of the node it is connecting to. In other words, if we denote the new node n_t , the probability that there will be a link between node n_t and node $i \in \mathcal{V}_{t-1}$ is:

$$\mathbb{P}(W_{n_t,i}(t) = W_{i,n_t}(t) = 1 | \mathcal{G}_{t-1} = (\mathcal{V}_{t-1}, \mathcal{E}_{t-1})) = \frac{w_i(t-1)}{\sum_{j \in \mathcal{V}_{t-1}} w_j(t-1)}, i \in \mathcal{V}_{t-1} \quad (3)$$

where $W(t)$ is the adjacency matrix for the next time-step t and $w_i(t-1)$ is the degree of node i prior to adding the new node. Some care should be taken here so that you do not add multiple links to the same node.

You should also note that if k is odd, it is a bit trickier to generate the random graph such that the average degree will be k . If k happens to be an odd number, then $c = k/2$ will not be an integer. However, it is still possible to achieve an average degree of k when you add a large number of nodes. This can be done by alternating between adding $\lfloor k/2 \rfloor$ and $\lceil k/2 \rceil$ links when adding a new node to the graph.

Problem 1.2 Your goal is to, by using preferential attachment, generate a random graph of a large size (at least 900 nodes) with average degree $k \in \mathbb{Z}^+$. Let the initial graph $\mathcal{G}_1 = (\mathcal{V}_1; \mathcal{E}_1)$ be a complete graph with $|\mathcal{V}_1| = k_0 = k + 1$ nodes.

Note that the goal here is to implement a fairly general algorithm where it is very easy to change the average degree. It should be possible to change the average degree by only changing the value of k in your algorithm. This algorithm will then be used in Section 4.

Solution: According to what asked above, we start by generating a random complete graph $\mathcal{G}_1 = (\mathcal{V}_1; \mathcal{E}_1)$ with $|\mathcal{V}_1| = k_0 = k + 1$ nodes and $k = 6$. At every time step we add a node according to the stochastic rule 3 until reaching 1000 final desired nodes. Notice that, before adding a node, $c = \frac{k}{2}$ are sampled from the graph without replacement to avoid the repetitions of edges to same agents.

2 Simulate a Pandemic without vaccination

In this part you will be using the graph generated in Section 1.2 and then simulate an epidemic on it. The disease propagation model is again the discrete-time version of the SIR epidemic model used in Section 1.1

Problem 2: Using the methods developed in Section 1, generate a preferential attachment random graph $\mathcal{G} = (\mathcal{V}; \mathcal{E})$, with $|\mathcal{V}| = 500$ nodes. The average degree should be $k = 6$. Let $\beta = 0.3$ and $\rho = 0.7$. With one week being one unit of time, simulate the epidemic for 15 weeks. You can choose an initial configuration with 10 infected nodes selected at random from the node set \mathcal{V} , or make a different choice of initial configuration (in the latter case, please briefly discuss your motivation). Do this $N = 100$ times and plot the following:

- The average number of newly infected individuals each week. In other words, you should plot how many individuals *become* infected each week (on the average).
- The average total number of susceptible, infected, and recovered individuals at each week. In other words, you should plot how many individuals *in total that are* susceptible/infected/recovered at each week (on the average).

Hint: Remember to continue using sparse matrices for this and following problems.

Solution: Again we execute for 15 weeks our simulation $N = 100$ times with 10 initial infected agents randomly chosen with $\beta = 0.3$ and $\rho = 0.7$, but now the graph is generated according to the preferential attachment expressed in point 1.2 with $n = 500$ nodes and average degree $k = 6$. We report below the average number of newly infected and average total number of susceptible, infected and recovered at each week for this case.

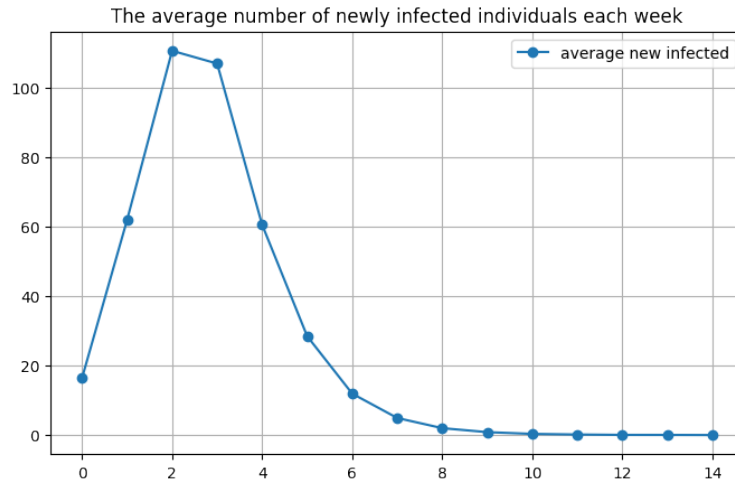


Figure 4: Average newly infected, $k = 6$

With respect to Figure 2, in Figure 4 the average newly infected assume much more higher values, in particular reaches a peak of approximately 110 in the second week then rapidly decreases.

The average total number of susceptible, infected, and recovered individuals at each week

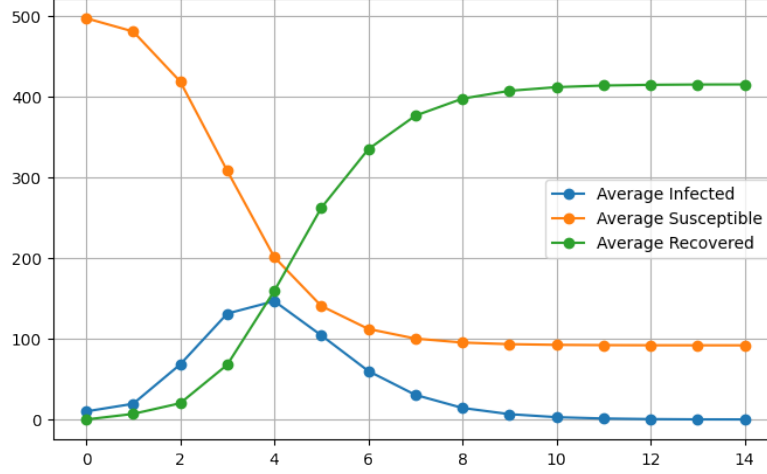


Figure 5: Average total susceptibles, infected and recovered, $k = 6$

Also in Figure 5 comparing the blue line with the one in Figure 3, we can see that this time the curve assumes values much more higher. The growth of curves related to infected agents is due to the fact that, in the previous case, the average degree was equal to 4, while now it is 6, meaning an expansion of the number of interactions an individual can have, so a major diffusion of the disease.

Regarding the susceptibles and recovered agents, they seem to converge around 100 and 410 respectively.

3. Simulate a pandemic with vaccination

In this part you will essentially do the same thing as before, but you will also try to take some action to slow down the epidemic. This is normally done using vaccination. Therefore, during each week, some parts of the population will receive vaccination. *Once a person is vaccinated it cannot be infected.* Furthermore, *the vaccination is assumed to take effect immediately once given*, i.e. if person a is vaccinated in week 10, then a is no longer susceptible during that week, and can therefore not infect any other individual.

You should once again simulate the disease propagation for 15 weeks, but you should now also distribute vaccination to the population. This should be done such that the total fraction of population that has received vaccination by each week is according to:

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

$Vacc(t)$ should be interpreted as 55% of the population has received vaccination *by* week 7, and 5% received vaccination *during* week 7.

To simulate the actual vaccination you should, at the beginning of each week, find the correct number of individuals to vaccinate according to $Vacc(t)$. You should then find individuals to vaccinate. These individuals should be selected uniformly at random from the population that *has not yet received vaccination*. This means that *an infected individual might receive vaccination as well*. The reason behind this is that some people were not able to tell whether they had the H1N1-virus or just the common cold. If an infected individual becomes vaccinated it is assumed that she will not be able to infect another individual. In other words, we assume that regardless of the state of an individual prior to the vaccination, *once vaccinated the individual will not be able to become infected nor infect any other individuals*.

Problem 3: Using the method developed in the previous section, generate a random graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with $|\mathcal{V}| = 500$ nodes. The average degree should be $k = 6$. Let $\beta = 0.3$ and $\rho = 0.7$. With one week being one unit of time, simulate the epidemic *with vaccination* for 15 weeks, using the vaccination scheme $Vacc(t)$ above. You can choose an initial configuration with 10 infected nodes selected at random from the node set \mathcal{V} or make a different choice of initial configuration (in the latter case, please briefly discuss your motivation).

Do this $N = 100$ times and plot the following:

- The average number of newly infected and newly vaccinated individuals each week.
- The average total number of susceptible, infected, recovered and vaccinated individuals at each week.

Solution: We repeat our simulation with 10 initial infected individuals for $N = 100$ times for 15 weeks. Again the graph is generated according to what was done in point 1.2 with 500 nodes and $k = 6$ with $\beta = 0.3$ and $\rho = 0.7$. This case differs from the previous one by the introduction of the vaccine, which drastically reduces the infection among the population. So now we have susceptibles, infected, recovered and vaccinated agents. Notice that once an agent receives the vaccine it cannot change its state anymore. We randomly select the people to vaccinate among infected and susceptible people and we report here the average of newly infected people and the average total number of infected, susceptibles, recovered and vaccinated.

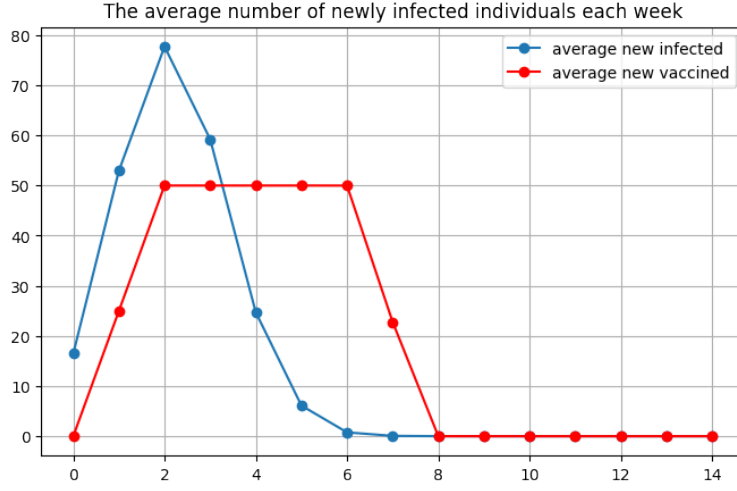


Figure 6: Comparison of average newly infected and vaccinated, $k = 6$

From Figure 6 we can clearly see how much the vaccine contrasts the diffusion of the epidemics. Indeed comparing Figure 6 with Figure 4, the peak around 110 in Figure 4 reduces to a peak close to 80 in Figure 6. The average newly vaccinated goes up until it reaches a value close to 50, remains stable until week 7 and then tends to 0 for the remaining weeks.

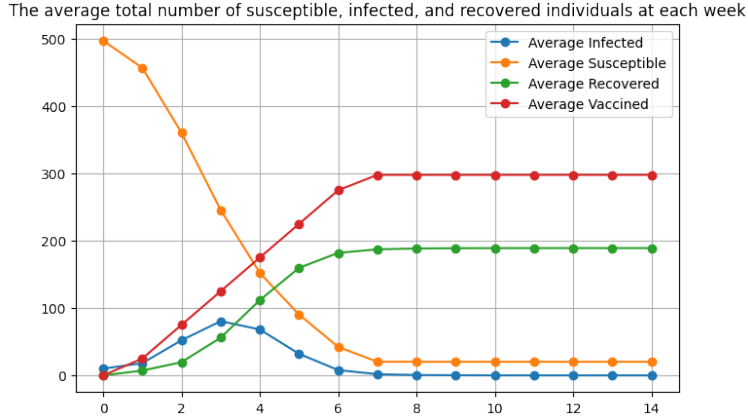


Figure 7: Average total susceptibles, infected and recovered, $k = 6$

Comparing Figure 7 with Figure 5 the blue line of infected agents is lower as expected. Both the yellow susceptible curves converge around week 7, but at different values, the former around 100 while the latter around 20. The green recovered lines have the same behaviour but the former converges approximately

to 400 around week 11, while the latter converges around about week 7 to a value close to 200. The new red curve of vaccinated instead seems to converge to 300 at week 7.

4. The H1N1 pandemic in Sweden 2009

In this part you will use all the previous parts in order to estimate the social structure of the Swedish population and the disease-spread parameter during the H1N1 pandemic. As mentioned before, during the fall of 2009 about 1.5 million people out of a total population of 9 million were infected with H1N1, and about 60% of the population received vaccination.

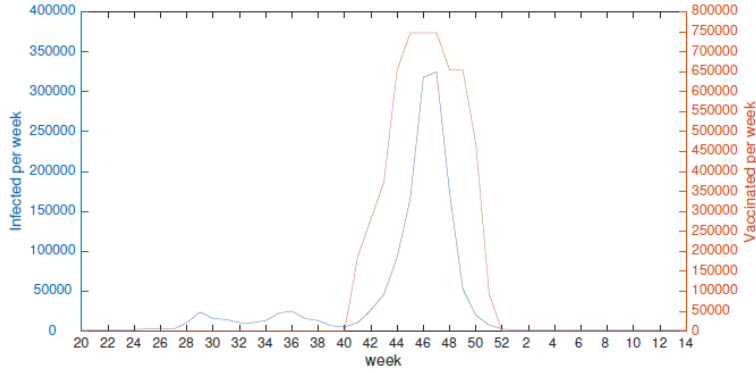


Figure 8: H1N1 Pandemic in Sweden during the fall of 2009.

Figure 8 illustrates the number of newly infected (and vaccinated) individuals each week. We will simulate the pandemic between week 42, 2009 and week 5, 2010. During these weeks, the fraction of population that had received vaccination was:

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

In order to not spend too much time running simulations, we will scale down the population of Sweden by a factor of 10^4 . This means that the population during the simulation will be $n = |\mathcal{V}| = 934$. For the scaled version, the number of newly infected individuals each week in the period between week 42, 2009 and week 5, 2010 was:

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

The following algorithm will do a gradient-based search over the parameter space of k , β and ρ in order to find the set of parameters that best matches the real pandemic.

Algorithm: Start with an initial guess of the parameters, k_0 , β_0 , and ρ_0 (here one could use $k_0 = 10$, $\beta_0 = 0.3$, $\rho_0 = 0.6$ as an initial guess) along with some Δk , $\Delta \beta$, and $\Delta \rho$ (here one might use $\Delta k = 1$, $\Delta \beta = 0.1$, and $\Delta \rho = 0.1$).

1. For each set of parameters (k, β, ρ) in the parameter-space $k \in \{k_0 - \Delta k, k_0, k_0 + \Delta k\}$, $\beta \in \{\beta_0 - \Delta\beta, \beta_0, \beta_0 + \Delta\beta\}$ and $\rho \in \{\rho_0 - \Delta\rho, \rho_0, \rho_0 + \Delta\rho\}$:
 - a Generate a random graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ using the preferential attachment model developed in Section 1.2. The average degree should be k , and there should be $|\mathcal{V}| = 934$ nodes in the graph.
 - b Starting from week 42, simulate the pandemic for 15 weeks on \mathcal{G} . You should use the method developed in Section 3 with the vaccination scheme described above. Do this $N = 10$ times, and compute the *average* number of newly infected individuals each week, $I(t)$.
 - c Compute the root-mean-square error (RMSE) between the simulation and the real pandemic:

$$RMSE = \sqrt{\frac{1}{15} \sum_{t=1}^{15} (I(t) - I_0(t))^2}$$

where $I(t)$ is the average number of newly infected individuals each week in the simulation and $I_0(t)$ is the true value of newly infected individuals each week.

2. Update k_0, β_0 and ρ_0 to the set of parameters yielding the lowest RMSE. If the result is was the same set of parameters, the algorithm should stop.

Problem 4: Using the algorithm above, estimate the average degree k and the disease-spread parameters β and ρ for the pandemic.

Once you have found the best estimate, report what parameters you got. You should also show the following plots:

- The average number of newly infected each week.
- The average total number of susceptible, infected, recovered and vaccinated individuals at each week.

Hint: The algorithm will be somewhat slow to execute as it needs to run the method developed in Section 3 multiple times before updating k_0, β_0, ρ_0 . It might be a good idea to play around with the values of Δk , $\Delta\beta$, and $\Delta\rho$. For instance, you can start with large values and then decrease them when you cannot find a better set of parameters (for example, you can reduce them by half).

Solution: As suggested above, to find out which are the best parameters that model the Swedish H1N1-pandemic in 2009, we generate the random graph \mathcal{G} through our attachment in point 1.2 with 934 nodes for each set of the parameters (k, β, ρ) in the parameter space $k \in \{k_0 - \Delta k, k_0, k_0 + \Delta k\}$, $\beta \in \{\beta_0 - \Delta\beta, \beta_0, \beta_0 + \Delta\beta\}$ and $\rho \in \{\rho_0 - \Delta\rho, \rho_0, \rho_0 + \Delta\rho\}$. We start with an initial guess $k_0 = 10, \beta_0 = 0.3, \rho_0 = 0.6$ along with $\Delta k = 1, \Delta\beta = 0.1$, and $\Delta\rho = 0.1$.

We simulate for 15 weeks as before, but this time we simulate for $N = 10$ times and start with 1 single initial infected each time. At each step of the gradient search we compute the RMSE - i.e. the difference between the simulated newly infected and the real ones. The parameters's optimization is computed by taking the parameters corresponding to the best RMSE obtained for each step of the gradient search and collect them in the list *solutions*. After simulating 10 times the gradient search, we take the solutions collected so far and finally compute the average to get the best parameters $best_k$, $best_beta$ and $best_rho$ in order to have more reliable results.

We obtain the following results: $best_k = 9.0$, $best_beta = 0.22$ and $best_rho = 0.62$.

As before we report below the average number of newly infected individuals and the average total number of susceptible, infected, recovered and vaccinated individuals at each week according to the model. Finally we compare our estimation with the real curve of infected.

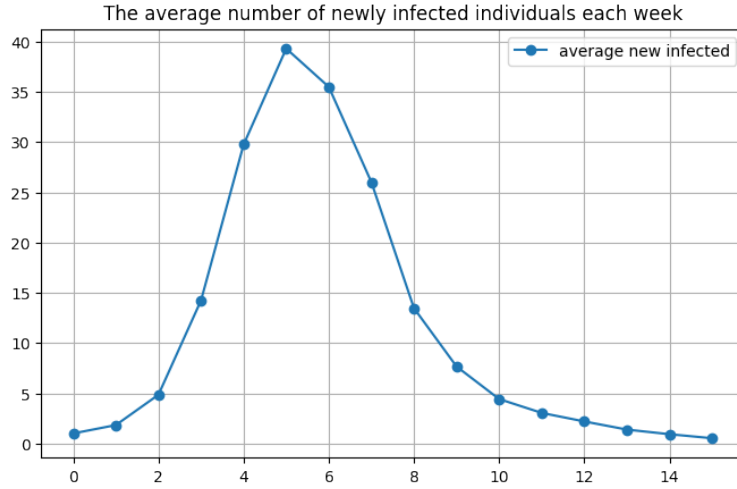


Figure 9: Single instance of the average newly infected obtained with best parameters (k, β, ρ)

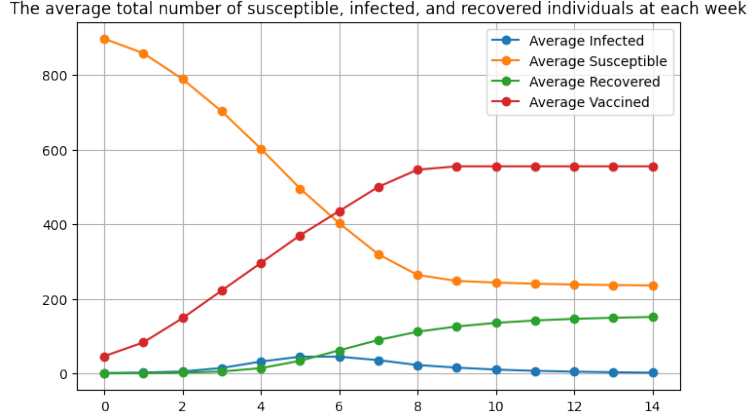


Figure 10: Average total susceptibles, infected recovered and vaccinated obtained with best parameters (k, β, ρ)

All the curves above are obtained after generating the graph according to our attachment in point 1.2 with 934 nodes and $k = best_k$. We pass $best_beta$ and $best_rho$ to our *averages_newly_infected_and_sir_week* function used to solve every point discussed so far and we obtain a loss equal to 8.34.

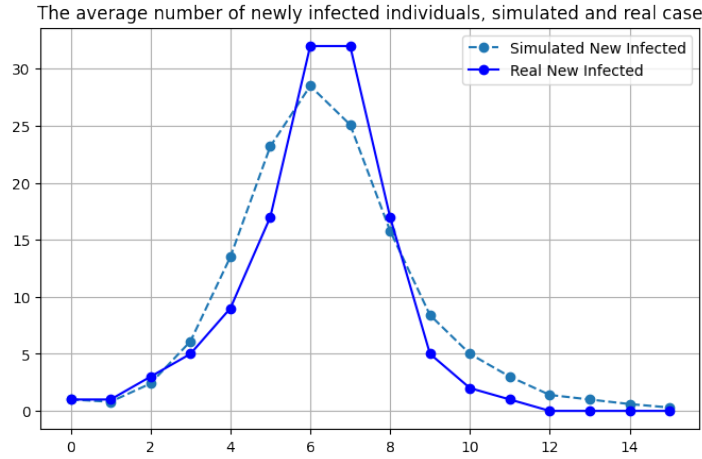


Figure 11: Comparison between the average of the 10 average newly infected simulated obtained with best parameters (k, β, ρ) and the real infected

As the caption says, Figure 11 shows the comparison between the average of the 10 average newly infected curves obtained and the real curve. The RMSE obtained from our simulation in this case is equal to 4.01 which is lower than the one related to the single instance in Figure 9 as expected.

5 Challenge (optional)

Try to find a better random graph (i.e. one that does not use preferential attachment) to represent the network for the pandemic. Try to also find a better algorithm to estimate the parameters.

Solution: In this case we change our attachment with the Erdos-Renyi model to see if we reach better results. The algorithm is the same of point 4. The simulation is represented by the violet curve in the picture below. We obtain the following results: $best_k = 9.0$, $best_beta = 0.14$ and $best_rho = 0.73$.

Optimization: The average number of newly infected individuals, simulated and real case

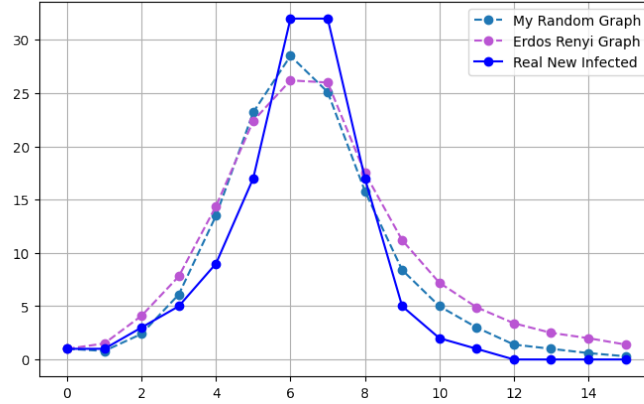


Figure 12: Comparison between the average of the 10 average newly infected simulated with our usual random graph and Erdos-Renyi graph obtained with best parameters (k, β, ρ) and the real infected

As we can see from the figure above, the Erdos-Renyi model does not seem to improve the results found at previous point. The RMSE obtained equal to 4.31 which is quiet close to 4.01 found at point 4 and indeed the light-blue and violet curves are more or less comparable.