Network Dynamics: Homework III

Andrea Silvi
Politecnico di Torino
andrea.silvi@studenti.polito.it

1. Preliminary parts

We start off by simulating 15 weeks of an epidemic on a given symmetric k-regular graph and on both a random graph generated according to the *preferential attachment* model. The epidemic is simulated by leveraging a simplified SIR model in discrete time, where 1 week is equal to 1 time unit. It evolves following two simple rules:

- a susceptible individual s, with m infected neighbouring nodes, gets infected in week j with a probability of $1-(1-\beta)^m$, where $\beta \in [0,1]$ is the probability that the infection is spread from an infected individual to a susceptible neighbour during one time step;
- an infected individual i recovers from illness in week j with a probability of ρ, with $\rho \in [0,1]$.

1.1. Problem 1.1

In this first part, in order to have a graph on which we can simulate the epidemic, we start off by generating a symmetric k-regular graph $\mathcal{G}$ with $|\mathcal{V}| = 500$, $k = 4$, meaning that a node is connected only to the 4 nodes closest to it. Afterwards, we select 10 of its nodes to be the first infected. Note that the other 490 nodes are considered susceptible. Then, by following the two rules reported in the previous section, for 15 weeks we simulate each week the new populations of susceptibles (the previous susceptibles that do not get infected), infected (the previous infected that do not recover plus the previous susceptibles that get infected by their infected neighbours), and the recovered (the previous recovered plus the previous infected that recover). Note that we use here and afterwards (unless specified differently) $\beta = 0.3$ and $\rho = 0.7$.

We simulate this process for 100 simulations and we average the results in terms of infected, susceptibles and recovered each week in order to try to obtain a not so noisy estimation. We obtain the results reported in Figure (1) in terms of newly infected per week and of total infected, susceptibles and recovered per week. As we can see, the number of newly infected tends to die down fairly quickly and so the disease cannot spread much around the population. If we consider the probability of transmission β and the probability of recovery ρ we know that the coefficient $R_0 = \frac{\beta}{\rho}$ represents if the epidemics in the long run will survive or die off. Since we have $R_0 < 1$, we expect the disease to die off and the results confirm this.

1.2. Problem 1.2

We now want to be able to generate a random graph with n nodes and average node degree equal to k by following the preferential attachment model: we start from a complete graph with $k+1$ nodes, and then we add 1 node at the time until we reach the desired population n, by connecting it to $k/2$ other nodes, randomly selected proportionally to their degree distribution. Whenever k is odd, we just alternate between randomly selecting $k/2+1$ and $k/2-1$ nodes to link to a new one.

We then use these two algorithms afterwards in order to generate random graphs and simulate different rounds of epidemics on them.

2. Simulate an epidemic without vaccination

We repeat the simulations done in Section (1.1), but instead of using a given graph, each time we generate a new random graph using the algorithm defined in Section (1.2) with $k = 6$. We again use as parameters $|\mathcal{V}| = 500$, $\beta = 0.3$ and $\rho = 0.7$, and we start each time with 10 randomly selected nodes as infected. The results are reported in Figure (2). As we can see, given the same epidemic parameters as in Section (1.2), the different graph model gives us different results: the epidemic is able to spread more and the peak is 3 weeks after and 1 order of magnitude higher. More than 50% of the populations gets infected but eventually in the 15 weeks the epidemic dies off and every infected recovers. Of course, we also need to consider that the average node degree of the graph in Section (1.1) was 4, while now we are using $k = 6$ in these simulations, thus nodes now are a bit more connected and the epidemic finds it easier to expand.

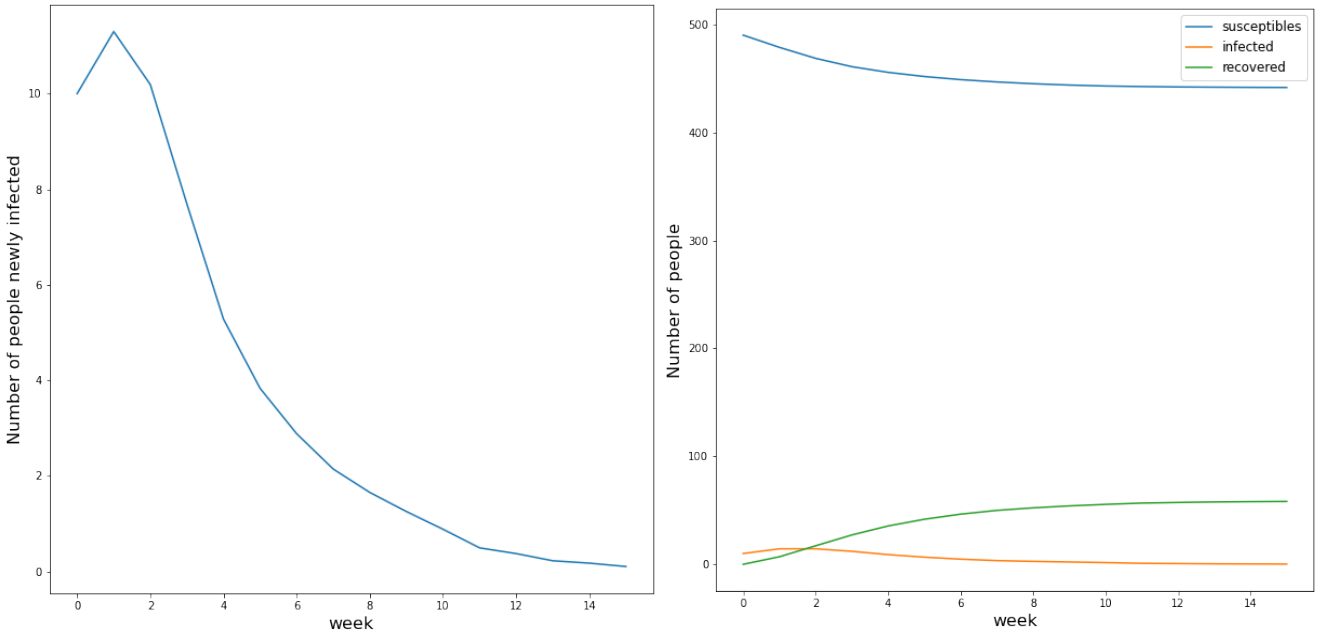

Figure 1: The average evolution of newly infected each week and of total susceptibles, infected and recovered on a symmetric 4-regular graph with 500 nodes, starting with 10 infected random nodes.

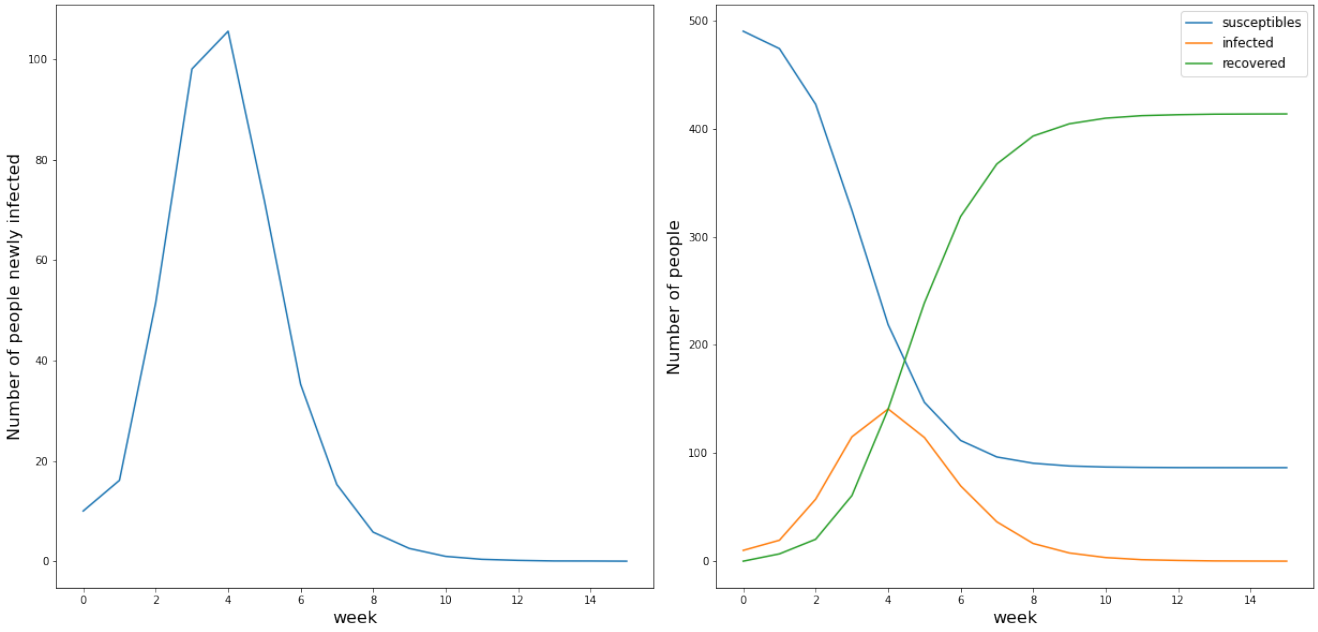

Figure 2: The average evolution of newly infected each week and of total susceptibles, infected and recovered on a graph randomly generated following the preferential attachment model with 500 nodes and average degree $k = 6$, starting with 10 infected random nodes.

3. Simulate an epidemic with vaccination

We now expand the model by introducing the concept of vaccinations: at the start of each week, a subset of the population gets randomly selected to get vaccinated. Whenever a person gets vaccinated, she is automatically considered recovered. Note that we consider as eligible to vaccination all people that have not yet been vaccinated. This means that also infected and already recovered can be vaccinated: whenever an infected gets randomly selected to be vaccinated, it represents those cases where a person does not know that she has the disease and gets vaccinated, while whenever an already recovered gets selected to be vaccinated, it represents those cases where a person have already contracted the illness and recovered without ever knowing

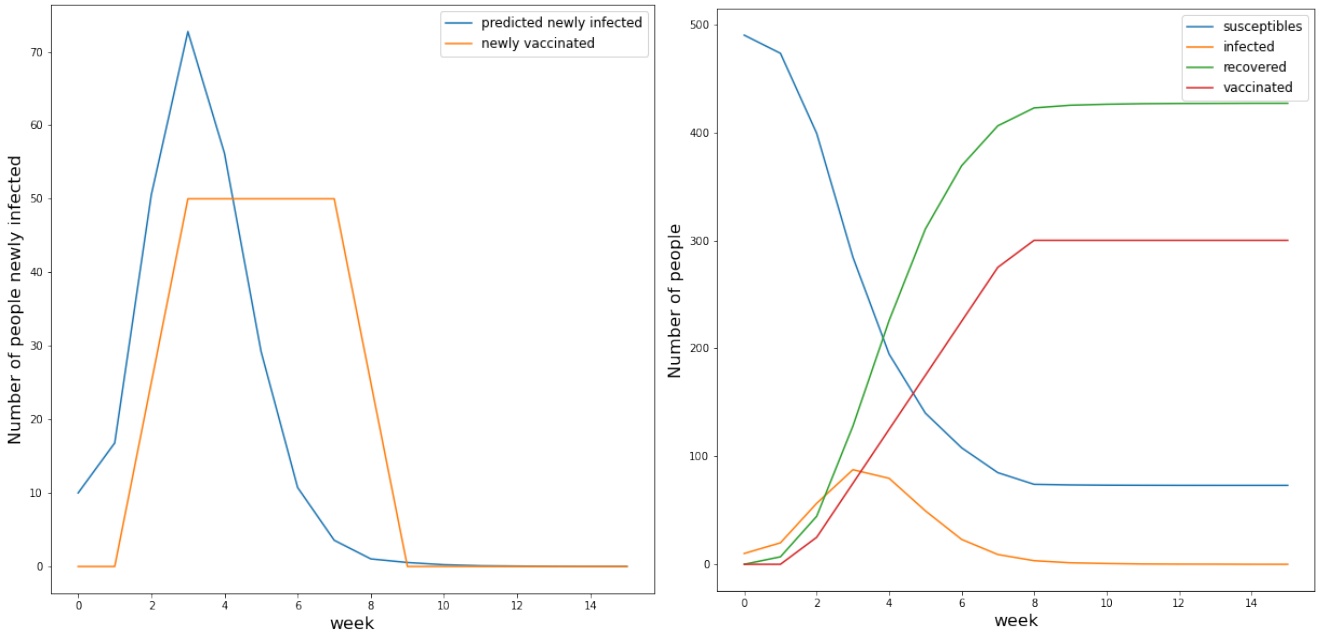

Figure 3: The average evolution of newly infected each week against the number of newly vaccinated, and the average evolution of total susceptibles, infected, recovered and vaccinated on a graph randomly generated following the preferential attachment model with 500 nodes and average degree $k = 6$, starting with 10 infected random nodes.

they had it (probably because they were symptomless for the whole illness duration).

We replicate the same experiments as in Section (2) but we also introduce vaccinations, according to

$$vacc = [0, 5, 15, 25, 35, 45, 55, 60, 60, 60, 60, 60, 60, 60, 60]$$

where $vacc(i)$ is the total fraction of population which has been vaccinated up to week i.

We report the results obtained by averaging again the outcomes of 100 simulations in Figure (3). As we can see, the peak of newly infected is lower than without vaccinations, and the curve of infected does not reach the 20% of total population anymore. This confirms that vaccinations are still somewhat effective in attenuating the spread of an epidemic, even if only a fraction of the population gets them.

4. The H1N1 epidemic in Sweden, 2009

We now consider the real life example of the Influenza epidemic in Sweden during 2009 in order to estimate the social structure of the Swedish population and the epidemic parameters, by leveraging the algorithms reported in the previous sections and gradient-based search.

In order to reduce the computational power needed we scale down the Swedish population and the effects of the H1N1 epidemic on it by a factor of 10^4. We then obtain a population size of $|\mathcal{V}| = 934$.

We consider a 15 week period between week 42 of 2009 and week 5 of 2010, over which in real life both the H1N1 influenza and vaccines administrations were at their peak. During this period, the fraction of population vaccinated evolved according to

$$v = [5, 9, 16, 24, 32, 40, 47, 54, 59, 60, 60, 60, 60, 60, 60, 60]$$

while the scaled number of newly infected per week was

$$I_{true} = [1, 1, 3, 5, 9, 17, 32, 32, 17, 5, 2, 1, 0, 0, 0, 0] \quad (1)$$

In order to find out what the best triple of parameters (k, β, ρ) is for this real life example, we start from a given point ($k_0 = 10$, $\beta_0 = 0.3$, $\rho_0 = 0.6$) and we simulate 50 times for each triple of parameters (k_i, β_i, ρ_i) obtained from the product space $\{k, k+\Delta k, k-\Delta k\} \otimes \{\beta, \beta+\Delta\beta, \beta-\Delta\beta\} \otimes \{\rho, \rho+\Delta\rho, \rho-\Delta\rho\}$, where $\Delta k = 3$, $\Delta\beta = \Delta\rho = 0.2$. For each simulation, we firstly generate a random preferential attachment graph with average degree equal to k_i, we randomly select the initial infected in accordance with $I_{true}(0)$, then we simulate 15 weeks of an epidemic using as parameters β_i and ρ_i, administering vaccines to non vaccinated according to vector v. We average the evolution of newly people infected from the 50 runs and we compare it with the actual vector of newly infected reported in (1), leveraging the root mean square error:

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

We save the triple of parameters (k^*, β^*, ρ^*) that gives us the lowest RMSE and generate again the product space with

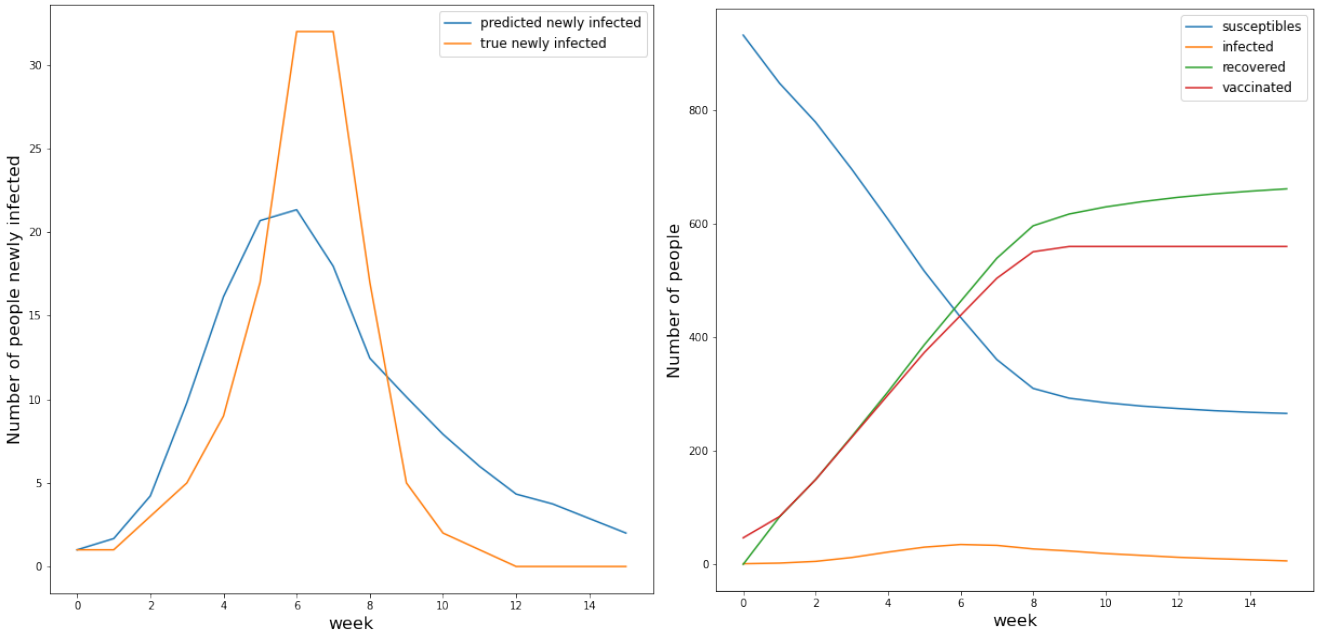

Figure 4: The average evolution of newly infected each week against the true number of newly infected between week 42, 2009 and week 5, 2010 of the H1N1 epidemic in Sweden, and the average evolution of total susceptibles, infected, recovered and vaccinated on a graph randomly generated following the preferential attachment model with 934 nodes and parameters ($k^* = 16$, $\beta^* = 0.1$, $\rho^* = 0.5$).

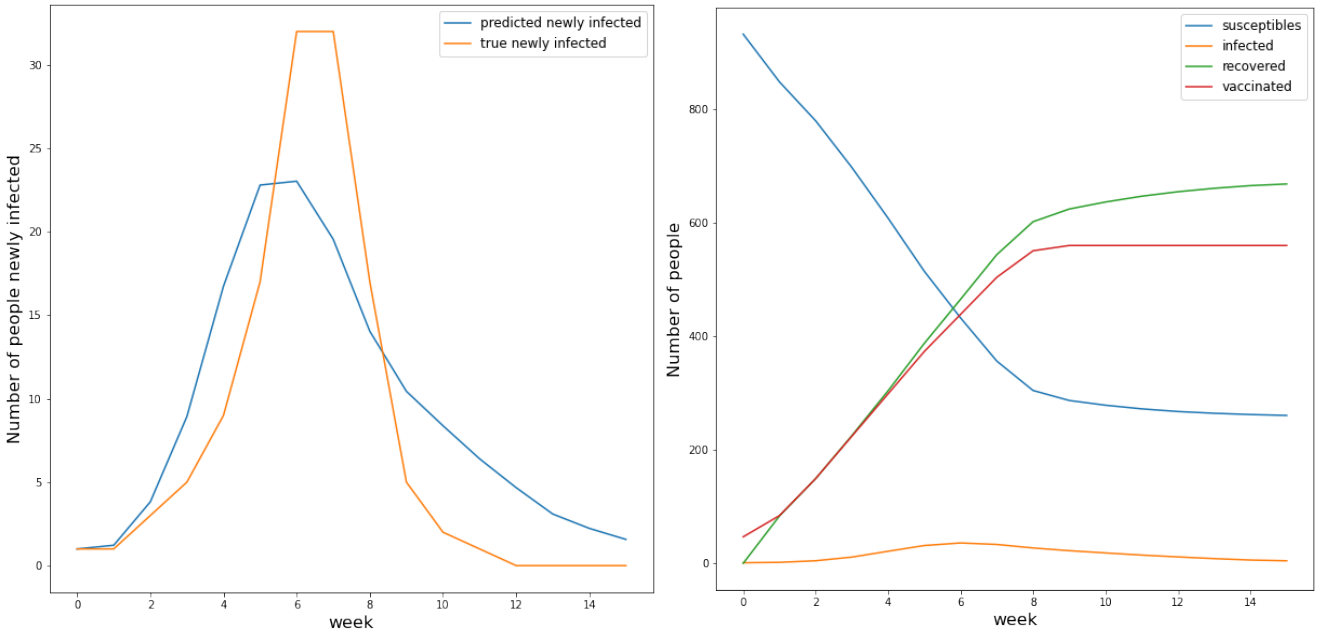

Figure 5: The average evolution of newly infected each week against the true number of newly infected between week 42, 2009 and week 5, 2010 of the H1N1 epidemic in Sweden, and the average evolution of total susceptibles, infected, recovered and vaccinated on a graph randomly generated following the preferential attachment model with 934 nodes and parameters ($k^* = 17$, $\beta^* = 0.10347$, $\rho^* = 0.55$).

$k = k^*$, $\beta = \beta^*$ and $\rho = \rho^*$. We do this for 20 epochs, mainly because of computation time, and at the end we consider as most fitting parameters the ones that generate the results with the lowest RMSE.

Note that whenever the algorithm cannot find a new best set of parameters in an epoch, we divide by 2 both $\Delta\beta$ and $\Delta\rho$, until they are equal to 0.025, and we calculate as the new Δk the closest integer larger than $\frac{\Delta k}{2}$.

We obtain as best parameters the triple ($k^* = 16$, $\beta^* = 0.1$, $\rho^* = 0.5$), with a minimum RMSE $\simeq 5.294$. In or-

der to not report biased results, since we are always picking the parameters that give us the minimum RMSE from a fixed number of simulations, which is not a deterministic value but rather a quite noisy estimator, we then simulate again 100 times with the same parameters and we obtain an RMSE $\simeq 6.067$ and the graphs reported in Figure (4). As we can see, the shape of the newly infected curve is quite similar to the actual one, its peak happens at the same time as the true peak of the epidemic but the model under estimate the peak of newly infected by a factor of 30% which could be quite a problem if we used this model for predictions. Also the true infection dies down fairly more quickly than the results obtained from our simulations.

In order to estimate how noisy the previous parameters search is, we try to rerun the same algorithm as before starting from a new initial point ($k_0 = 14$, $\beta_0 = 0.1$, $\rho_0 = 0.5$). Since our results are averages of simulated stochastic processes on randomly generated graphs, the RMSE is not deterministic, so we want to understand if, by starting from a different point, the algorithm wanders off and finds a different optimum or reaches one close to the previous. The first case could possibly mean that either or both the previous and the new optimum are local optimums. It could also mean that the search is falsely lead by low results in terms of RMSE that are actually lucky results obtained from noisy simulations. We also remove the lower bound limitation for $\Delta\beta$ and $\Delta\rho$ to see if taking smaller steps than 0.025 can help us reach better results.

We obtain as best set of parameters ($k^* = 17$, $\beta^* = 0.10347$, $\rho^* = 0.55$), which are actually fairly close to the ones we obtained previously. This makes us suggest that we are actually in a subspace of the product space that gives us consistently good values close to a theoretical global optimum.

The results of 100 simulations leveraging this new set of parameters are reported in Figure (5). As we can see from the graphs, they are fairly close to the ones generated from the previously found best parameters, but the new parameters give us a closer estimate to reality in terms of the peak of newly infected. We get also an RMS $\simeq 5.72824$. Since we are using a simplified model of both the Swedish population (the random preferential attachment graph) and of the epidemic itself (the SIR model in discrete time), we consider ourselves satisfied with these results.

5. Small World random graph in epidemic

We introduce a new type of random graph in order to try to better estimate what the structure of the Swedish population is. While leveraging a graph generated following the preferential attachment model gives us the advantage of being able to control the average degree of its nodes (which represents on average how many people in one week one person has a long interaction with), the resulting network structure has a small number of triangles while in real life networks, if a node a shares a link with two nodes b and c, then with high probability also b and c interact regularly and thus should share a link in the network.

We introduce then a new random graph model called *Small World*, which depends now on two parameters k and p:

- k is the number of closest nodes each node is connected to;
- p is the probability that a long range connection between two random nodes i and j in the graph exists.

The generation of a Small World graph starts from a symmetric k-regular graph as in (1.1) with k=k given as parameter, and then for each couple of nodes in a graph a link gets added with probability p.

We hypothesise that a Small World random graph can represent better how interactions in a society work: the initial symmetric k-regular graph represents the close interactions that happen frequently between neighbours in the graph, which represent for example the people that live or work together everyday, while the randomly added long range connections stand for those casual encounters between people that do not usually find themselves in the same real life situations.

We rerun the experiments of Section (4) but we use the Small World model to generate each simulation the random graph instead. We also need to consider a new parameter p in the parameters search, meaning that at each epoch we evaluate in the worst case 81 different sets of parameters. To save some computational time, we let the parameter search run for only 10 epochs. We start from the initial point ($k_0 = 6$, $\beta_0 = 0.1$, $\rho_0 = 0.5$, $p = 0.02$) and to obtain the product space of all possible parameters we use $\Delta k = 2$, $\Delta\beta = \Delta\rho = 0.2$ and $\Delta p = 0.01$. Whenever we cannot find a better estimate in terms of RMSE, we divide all of these values by 2. We stop dividing $\Delta\beta$ and $\Delta\rho$ whenever they get equal to 0.025, Δp whenever it gets equal to 0.0025. We keep $\Delta k = 2$ throughout the search, since we want a symmetric k-regular graph at the start of the graph generation and an odd k would not hold symmetry.

As best parameters we obtain the quadruple ($k^* = 6$, $\beta^* = 0.2$, $\rho^* = 1$, $p^* = 0.005$). It is very interesting to note that the change in the random graph model results in very different values as best ρ and β. Especially $\rho = 1$ means that now whenever a person contracts the virus one week from one of its neighbours, he is then sure to recover from it the next week. This big discrepancy suggests us that probably both the preferential attachment and the small world models are not optimal for representing the interactions of the Swedish society, thus the epidemic parameters β and ρ adapts accordingly to make up for the models shortcomings. We run 100 simulations with the new set of best

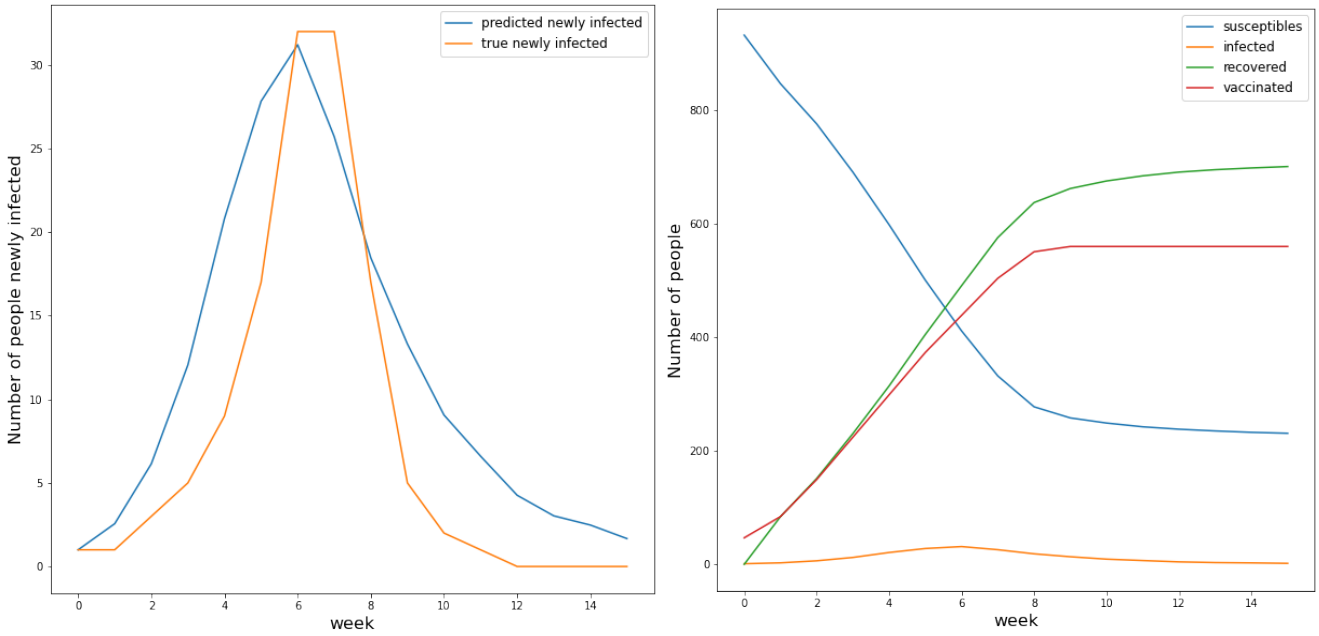

Figure 6: The average evolution of newly infected each week against the true number of newly infected between week 42, 2009 and week 5, 2010 of the H1N1 epidemic in Sweden, and the average evolution of total susceptibles, infected, recovered and vaccinated on a graph randomly generated following the Small World model with 934 nodes and parameters ($k^* = 6$, $\beta^* = 0.2$, $\rho^* = 1$, $p^* = 0.005$).

parameters using Small World random graphs each time and we obtain an RMSE $\simeq 6.04290$ and the graphs reported in Figure (6). While the RMSE is very close to the best ones previously obtained using preferential attachment random graphs, the shape of the newly infected curve looks closer to the true newly infected one and is able to almost reach its same peak in the same week, even though the newly infected line is slower to rise and descend than the real one.

6. Conclusions

We could consider other different random graph models that could maybe estimate better the social structure of the Swedish population and also find better and more precise algorithms for parameter inference but probably it would also beneficial to consider a more complex SIR model that works in continuous time and does not consider transitions between different states to happen all at once at the start of the week. Also it would be interesting to make the parameter β evolve in time during the different weeks, mimicking how the interactions between people change in the different weeks following maybe stay at home suggestions or mandates from the government or a choice of a more cautious lifestyle of most people after the wide spread of the virus. Of course, accounting for all these different variations would mean an increase complexity of the model and thus more computational power would be needed in order to perform all the simulations we reported before with our simpler SIR model.