

Homework 3

Pierluigi Compagnone s288301
Armando La Rocca s279401
Network Dynamics and Learning
2021/2022

January 15, 2022



Collaboration with Giuseppe Desiderio

In this report is briefly explained how the presented Homework 3 has been solved. All the presented charts and simulations are performed by notebooks files delivered with this document. In particular each of the following 5 sections is associated to the a notebook with the same section number in the name. An utility script containing common functions in the different notebooks is available in the directory as *utility.py*.

1 Preliminary parts

The first task faced consists in the simulation of a SIR epidemic on a graph $G = (V, \epsilon)$ with $|V| = 500$ and number of closest neighbors $k = 4$. This problem can be modeled as a discrete time Markov chain, where each step corresponds to a week in which the nodes can change their states. In particular three possible states are eligible: Susceptible (S), Infected (I) and Recovered (R). The probability of a node to change state depends by its current state and is influenced by the number of infected neighbours m . These probabilities are summarized in the following transition rate matrix.

$$\begin{matrix} & S & I & R \\ \begin{matrix} S \\ I \\ R \end{matrix} & \begin{pmatrix} (1 - \beta)^m & 1 - (1 - \beta)^m & 0 \\ 0 & 1 - \rho & \rho \\ 0 & 0 & 1 \end{pmatrix} \end{matrix}$$

The parameters β and ρ are respectively the infection and recovery probabilities and they are set to $\beta = 0.3$ and $\rho = 0.7$ accordingly to the assignment. To obtain an overall trend the simulation is performed 100 times.

As specified in the assignment, 10 nodes at the beginning are initialized to be infected, and two different approaches are considered to choose these nodes. The first one consists in randomly select the infected nodes on the graph. While for the second policy all the infected nodes are neighbours, assuming that a pandemic spread over a population from a specific starting point representing a specific place or a specific population.

Here, some graphs are reported with the aim to understand the dynamic of the pandemics over the time. In particular, Figure 3 shows the number of new infected individuals per week and Figure 6 plots the number of nodes in the different states per week. Each chart is computed for both the starting policies of infection, in order to make a comparison of the results.

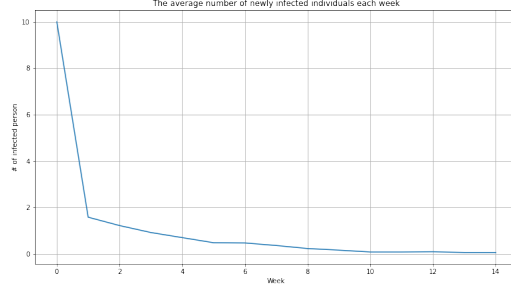


Figure 1: Neighbor

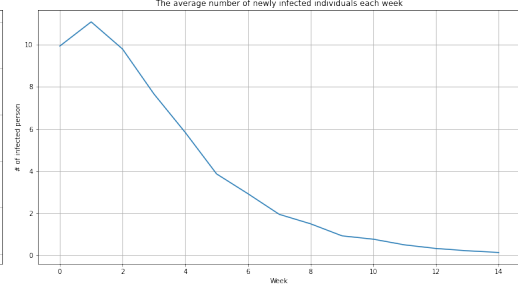


Figure 2: Random

Figure 3: Average newly infected per week considering two initial infected policies

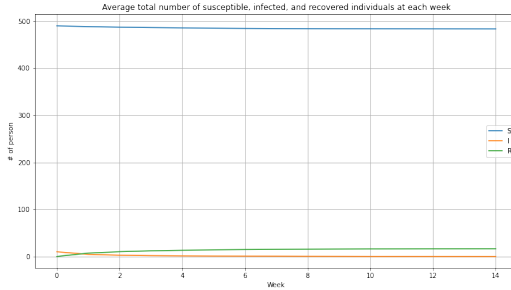


Figure 4: Neighbor

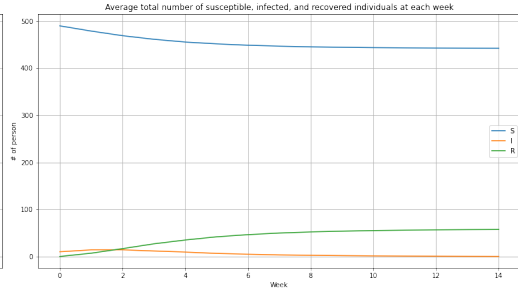


Figure 5: Random

Figure 6: Average number of infected, susceptible and recovered per week considering two initial infected policies

The Graphs clearly point out how the epidemic reach an equilibrium before the end of the 15th week in both cases, probably because the recovery rate is dominant respect to the infection one. From the two policies can be evinced that the random one encourage the infection over the population more than the other. This result can be explained looking at the structure of the graphs, since if the infected people are restrained in a specific group of the population, is less likely that the infection will spread over all the nodes.

The second part of the exercise consists in generating a random graph according with the preferential attachment policy, in order to have an average

degree close to k . Figure 8 is an example of the graph generated in this way with 9 nodes and an average degree $k = 3$, while in Figure 7 is reported the nodes degree and the average one considering the 900 nodes and $k = 4$.

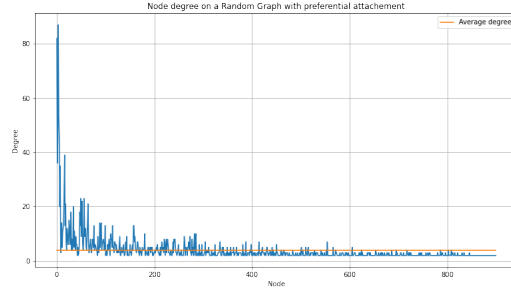


Figure 7: Node degrees

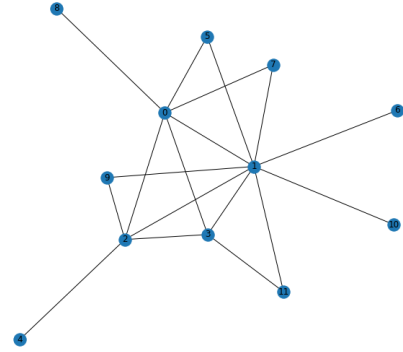


Figure 8: Random Graph

2 Simulate a pandemic without vaccination

The goal of this exercise is to simulate the SIR epidemics as explained in the previous point, but using a random graph with preferential attachment $G = (V, \epsilon)$. The Graph G has $|V| = 500$ and $k = 6$ and the probabilities that a node changes its state are the ones described in the previous section. Also in this case the number of initial infected is equal to 10 and the two different initialization policies are considered. Figure 11 and 14 show some plots in order to better understand the evolution of the pandemics, considering the two proposed initial infected policies. As in the previous case the results are the average calculated on 100 simulations.

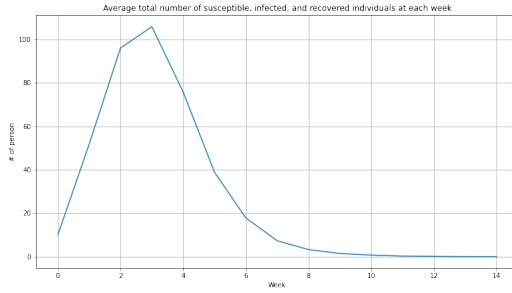


Figure 9: Neighbors

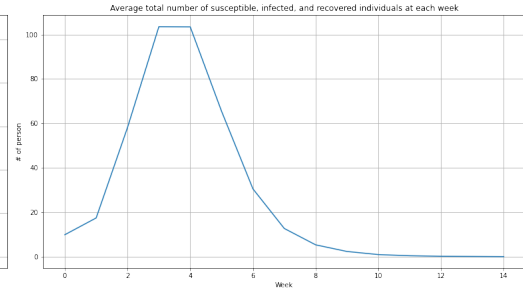


Figure 10: Random

Figure 11: Average newly infected per week considering two initial infected policies

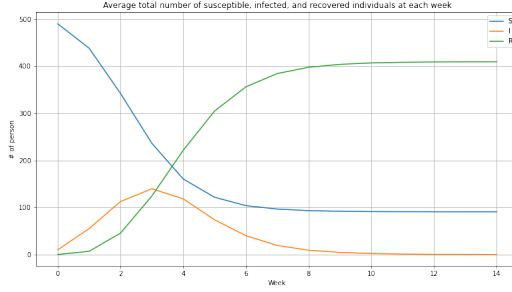


Figure 12: Neighbor

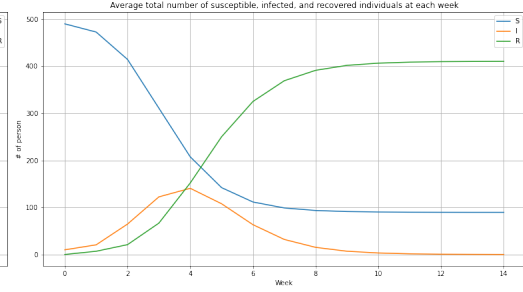


Figure 13: Random

Figure 14: Average number of infected, susceptible and recovered per week considering two initial infected policies

From the graphs can be seen that, respect to the previous case, the epidemics spread in a more serious way. This can be explained by the structure of the graph and by an higher value of average degree, that is interpreted as a higher number of interactions among people and consequently an easier spread of the infection. Indeed also the differences between the presented policies are reduced, obtaining very similar trends and values.

3 Simulate a pandemic with vaccination

Vaccination is a fundamental resource to reduce the spreading of the pandemic. In this section the vaccination is integrated in the previously presented SIR model, so to understand the benefits it can bring. In particular, at each week a percentage of the population is chosen accordingly to the data proposed in the assignment. For the vaccination, people are randomly chosen between all the states except from the vaccinated itself. The new condition is identified with a new state (V) that is basically very similar to the Recovered one since a vaccinated node cannot change its state. However in this analysis it is considered as a new state so this epidemic model can be identified as SIRV. In plots 17 and 20 the results are showed considering the parameters used in the previous sections.

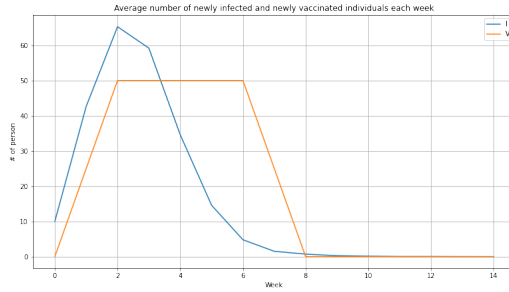


Figure 15: Neighbors

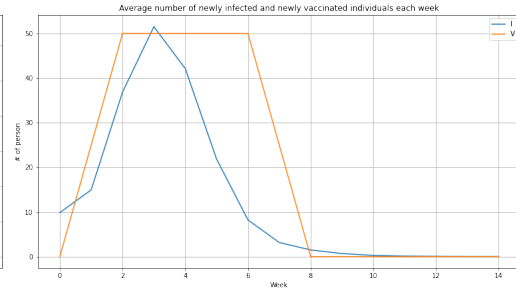


Figure 16: Random

Figure 17: Average newly infected and vaccinated per week considering two initial infected policies

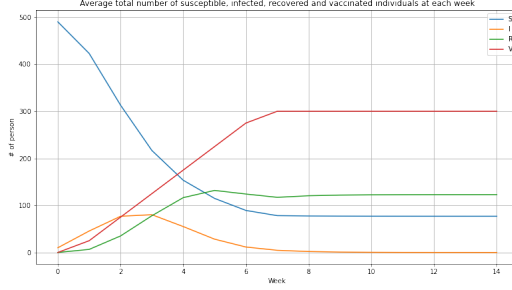


Figure 18: Neighbors

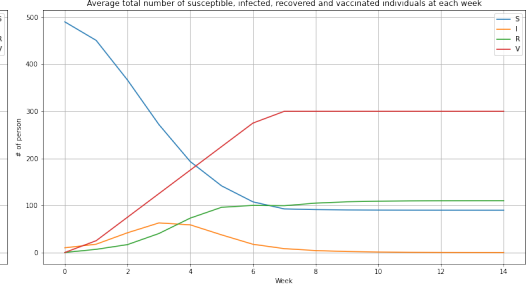


Figure 19: Random

Figure 20: Average number of infected, susceptible, recovered and vaccinated per week considering two initial infected policies

Comparing the dynamics with and without the vaccination it is clear that produces an important decrease on the contagion. Regarding the two different policies the vaccination performs better taking in account random initial infection. However, considering that in neighbors approach the infection is initially related to a specific part of the network, the contagion dynamic can be less confined by the randomly vaccination, at least in the first weeks. In this case a targeted vaccination could be a more effective strategy to bound the infection.

4 The H1N1 pandemic in Sweden 2009

Now, starting from data about the H1N1 pandemic in Sweden of 2009 an algorithm is used to estimate the structure of the graph that better represents Swedish society and the disease-spread parameters. The data provided concern the fraction of population vaccinated and the number of new infected each week, and cover a sixteen week range. In order to speed up the simulations the population is scaled down of a factor of 10^4 , so it can be represented with 934 nodes.

The algorithm is built following the instructions reported in the assignment and it performs a gradient based search over the parameters k, β, ρ .

At each step of the algorithm, starting from an initial configuration $\{k, \beta, \rho\}$, its neighborhood, defined by adding or subtracting $\{\Delta k, \Delta \beta, \Delta \rho\}$, is examined simulating the epidemic with all the possible combinations of parameters. At the end, the simulation with the lowest *root-mean-square error* respect to the real trend of the pandemic is selected and the cycle restarts examining its neighborhood. When there is no improvement in the best solution the Δ are halved and the process is completely reiterated and finally stopped when there is no improvement.

For the simulation the values are initialized as suggested in assignment, and a random seed is set for the graph creation because since it is an operation repeated at each cycle, it could affect a lot the parameters choice and on the RMSE.

The parameters that best fit the pandemic are $\{k = 9, \beta = 0.2140625, \rho = 0.5546875\}$, with **RMSE** = 3.83371. Figure 21 compares the trend of the number of newly infected individuals each week obtained in the simulation with the real trend.

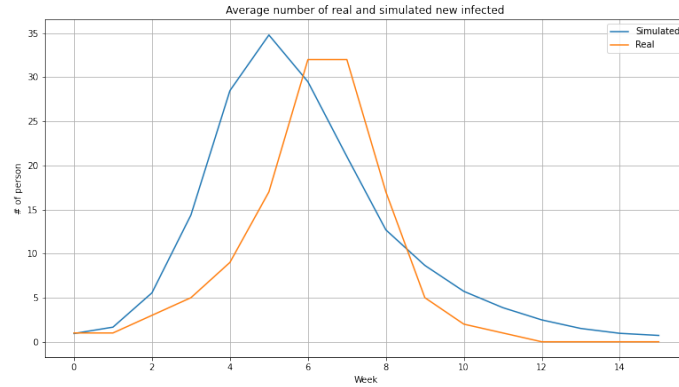


Figure 21: Comparison between the real and simulated with the best model newly infected

It shows how the shape of the pandemic is relatively faithful to reality. The simulation is affected by an overestimation, indeed at the maximum point the simulated number of newly infected are about ten units more than real data. Moreover the shape of the simulated trends is quite shifted on the left, indeed the peak is reached 1/2 weeks in advance.

The following is the plot of the total number of susceptible, infected, recovered and vaccinated individuals at each week, as estimated by the model.

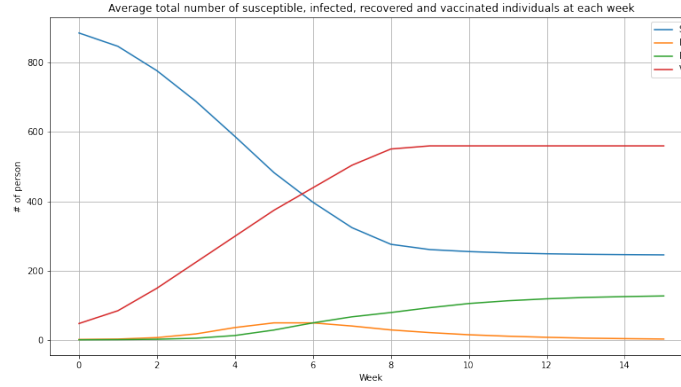


Figure 22: Average number of infected, susceptible, recovered and vaccinated per week considering two initial infected policies

The infections increase in the first weeks reaching the peak in week 5, in this range the number of vaccination increase and consequently the number of susceptible observe the opposite trend. At the 8th week almost 60% of the population is vaccinated and the dynamic starts to converge, from now on the new infected each week are very few and the main changes are in the data about recovery.

5 Challenge

Now the aim is to find a graph and an algorithm that potentially could approximate the H1N1 epidemic better than those used in Section 4. The solution inspected are respectively the *Small World* graph and the *Taboo Search* metaheuristic algorithm.

The Small World graph is built starting from a k -regular Graph, where each node has probability p to be connected with another node randomly chosen. In this case p represents another parameter that has an impact on the approximation of the epidemic. The new link can be seen as the contact that a person have with people that are not considered neighbors or as less strictly related contacts. This network is implemented using the function `newman_watts_strogatz_graph(...)` by `networkx` library.

In Figure 23 is represented an example of this kind of graph considering 12 nodes and a number of initial nodes $k = 2$

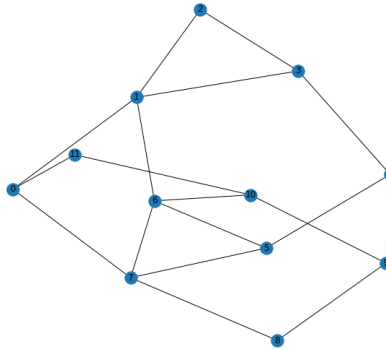


Figure 23: Small World Graph example

In order to improve the parameters search an implementation of the taboo search is used. The taboo search is a metaheuristic algorithm that at each step analyze the neighborhood of a specific solution trying to find a better one. Instead of stopping when no better solution is found it can consider a worst solution as new starting point adding it on a specific list called taboo list. The configurations inside the taboo list are forbidden for a certain amount of steps in order to avoid some fatal loops in the algorithm. This strategy is performed to avoid that search stuck in a local minimum. However the efficiency of this algorithm is highly influenced by the size of the neighborhood and by the chosen number of iterations.

The first step of this analysis have been the implementation of the taboo search that is tested initially on the Random Graph with preferential attachment as proposed in Section 4, considering also the same seed, such as to compare it with previous algorithms. The best configuration found is $k = 9$, $\beta = 0.21$, $\rho = 0.55$, with $RMSE = 3.91$. Given the granularity proposed in the neighborhood generation, the results are very similar to the one obtained in the previous section but reached after only 2 iterations. In Figure 24 is reported a graph that compares the real newly infected with the ones obtained by the simulation considering the presented parameters.

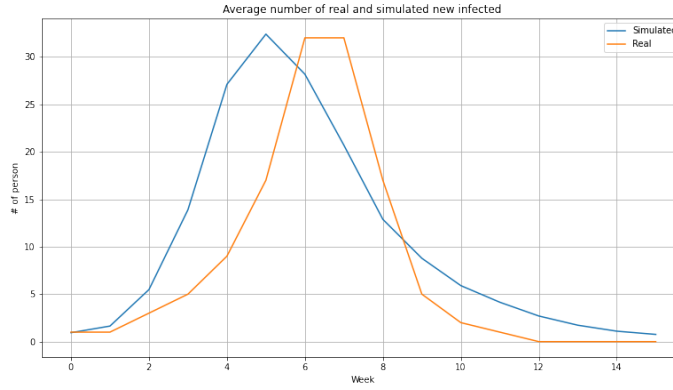


Figure 24: Comparison between the real and simulated with the best model newly infected considering a Graph with preferential attachment

Now the taboo search is applied on the Small World Graph. For the presented simulation a value of $p = 0.5$ is used and the best set of parameters found is $k = 12$, $\beta = 0.19$, $\rho = 0.86$ with an $RMSE = 4.621$. Also for this tuning a seed is set in the Graph generation to reduce the variability in the experiments. In Figure 25 is reported the comparison between real and simulated new infected.

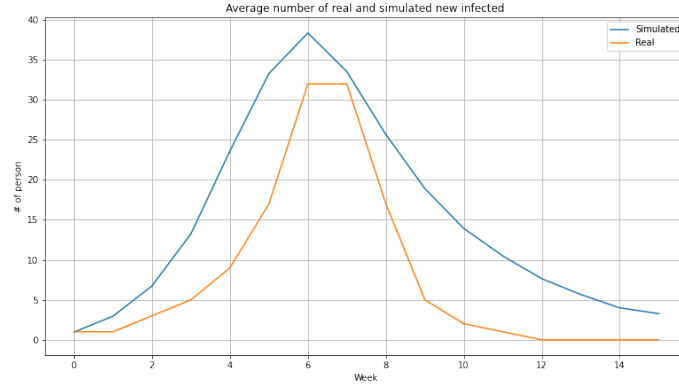


Figure 25: Comparison between the real and simulated with the best model newly infected considering a Small World Graph

As can be seen from the previous figure, the simulated pandemic is in the same weeks of the real ones. However the simulation overestimate the newly infected for all the analysed period. A possible solution can be a better tuning of the parameter p that can lead to a better approximation of the real number of connections in the Graph and consequently improve the estimation of the new infected.