Report Homework III

Network Dynamics and Learning

Benfenati Luca
Politecnico di Torino
Torino, Italy
s286582@studenti.polito.it

De Cristofaro Carmine
Politecnico di Torino
Torino, Italy
s291129@studenti.polito.it

Scarano Nicola
Politecnico di Torino
Torino, Italy
s287908@studenti.polito.it

Vacca Alessio
Politecnico di Torino
Torino, Italy
s288240@studenti.polito.it

***Abstract*—The following report addresses, firstly, three different problems on epidemic models and pairwise network systems. Then, all the previous exercises are combined in order to estimate the social structure of the Swedish population and the disease-spread parameters during the H1N1 pandemic in 2009. The aim of this work is to propose solutions to specific problems, dealing with both the theoretical and the practical side of the subject, presenting an approach which remains as general and broad as possible.**

INTRODUCTION

First of all, in order to deal with epidemic model simulation, we need to characterize what a *pairwise interacting network system* is. Specifically, in such system we have a finite population of agents, identified as nodes in a graph $G = (V, E, W)$, that update their state as a result of the superposition of independent mutations and pairwise interactions with their neighbors. We define A as the finite space of the states of the nodes and, consequently, we define $X = A^V$ as the *configuration space*. Then, the pairwise interacting network systems are continuous-time Markov chains $X(t)$ with state space X. The entries $X_i(t)$ of the system configuration $X(t)$ represent the states of the single nodes. Independent state mutations are described in terms of mutation kernels, as in:

$$\psi^{(i)} \in \mathbb{R}^{AxA}, \qquad i \in V \tag{1}$$

whose entries $\psi_{a,b}$ represent the conditional probability that, when a node i gets activated, it spontaneously mutates its state into b given that its current state is a. The pairwise interactions among nodes are described by pairwise interaction kernels, as in:

$$\varphi^{(i,j)}(c) \in \mathbb{R}^{AxA}, \qquad (i,j) \in E, c \in A \tag{2}$$

whose entries $\varphi^{(i,j)}(c)$ represent the conditional probability that, when a link (i, j) gets activated, node i changes its state into b given that it is currently in state a and node j in state j.

For our problems, we consider the SIR epidemics model, where population is divided into groups:

- S: susceptible people that are currently healthy an that can be infected;
- I: infected people that can transmit the infection;
- R: removed people that had the infection but that can no longer infect (because immune, or isolated or dead).

The mechanism that describes how contagion takes place is the pairwise interaction previously described. Pairwise interactions associated to links (i, j) occur at the ticking of independent Poisson clocks of rate βW_{ij}, where $\beta > 0$ is the interaction frequency parameter: if link (i, j)'s clock ticks at some time $t \geq 0$, and node j's current state is $X_j(t) = c$, node i's state changes from $X_i(t^-) = a$ to $X_i(t) = b$ with conditional probability $\varphi^{(i,j)}_{a,b}(c)$.

In the second part of the homework, we consider an additional state, V, which stands for *vaccinated*, in order to take into account an action to slow down the epidemic. Once a person is vaccinated it cannot be infected. Furthermore, the vaccination is assumed to take effect immediately once given, i.e. as soon as person a is vaccinated, then a is no longer susceptible, and can therefore not infect any other individual.

I. PRELIMINARY PARTS

A. Epidemic on a known graph

In this first exercise we are asked to simulate an epidemic following the simplified version of the SIR model over a known graph. Specifically, it is a k-regular undirected graph $G = (V, E)$, with $|V| = 500$ nodes and an average degree of 4 (i.e every node is directly connected with the four closest nodes). In particular, in SIR model are eligible three possible states, Susceptible (S), Infected (I) and Recovered (R) and the probability of an agent to change state depends on its own state and its influenced by the number of infected neighbours textitm. The transition matrix reported below summarizes these probabilities, which depend on the two parameters $\beta \in [0, 1]$ and $\rho \in [0, 1]$.

$$\begin{array}{c} \\ \text{S} \\ \text{I} \\ \text{R} \end{array} \begin{array}{c} \begin{array}{ccc} \text{S} & \text{I} & \text{R} \end{array} \\ \begin{bmatrix} (1-\beta)^m & 1-(1-\beta)^m & 0 \\ 0 & 1-\rho & \rho \\ 0 & 0 & 1 \end{bmatrix} \end{array}$$

β represents the probability that an infection is spread from an infected agent to a susceptible one during one step and is $\beta = 0.3$; ρ, instead, is the probability that an infected node will recover during one step and is $\rho = 0.7$. The factor m in the transition matrix represents the number of infected neighbors of given node and it is used to properly model the probabilities.

The epidemic is run for a total of 15 weeks, considering a week being one unit of time and an initial configuration with 10 infected nodes selected at random from the node set V. We are required to repeat the simulation 100 times and plot some graphs which exhibit the dynamic of the pandemic. Figure 1 shows the average number of newly infected each week, Figure 2 shows the average total number of susceptible, infected and recovered individual at each week.

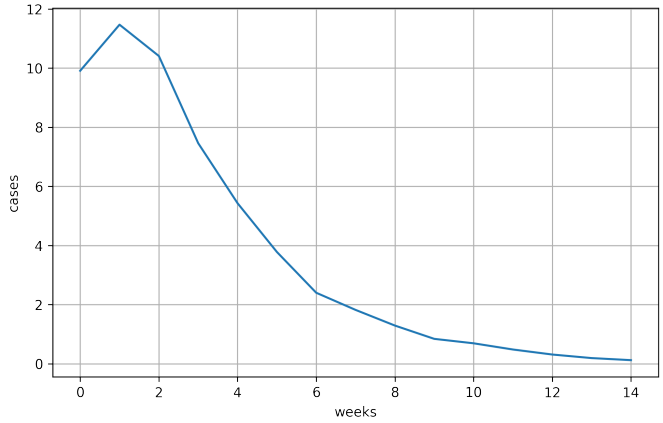

Fig. 1. Average number of newly infected individuals each week

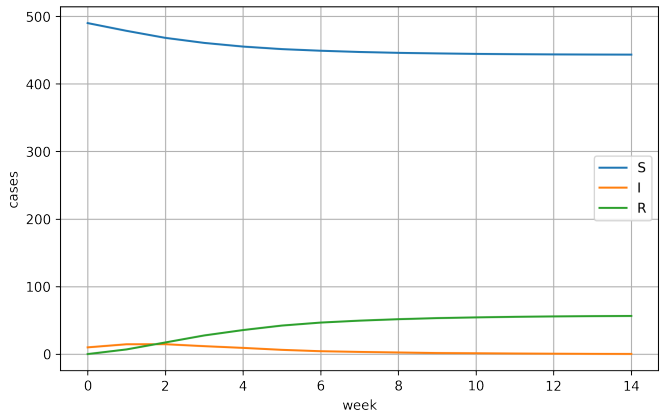

Fig. 2. Average total number of S, I, and R individuals at each week

Looking at these two figures, we can extrapolate that the infection rate is not high enough compared to the recovery one to support a spread of the pandemic. As a consequence, the curve of the newly infected individuals, after a small increase up to 11 cases in the first week, rapidly decreases to 0 cases at the end of the 14^{th} week. Even in the second graph, we can see that the state of each individual slightly changes in the first 3 weeks and almost reaches an equilibrium value from the 6^{th} week. This validate, once again, the low impact of this epidemic model, due to a not very-well connected graph ($k = 4$) and a dominant recovery rate with respect to the infection one.

B. Generate a random graph

The second part asks us to generate a random graph of a large size with average degree $k \in \mathbb{Z}^+$, by using preferential attachment. This means that at every time-step $t \geq 2$, every new node added at time t will have a degree $w_t(t) = c = \frac{k}{2}$. Hence, it should add c undirected links to the existing graph G_{t-1}, and decide to which of the nodes it should connect to, based on some probability that is proportional to the current degree of the node it is currently connected to. For the sake of simplicity, Figure 3 shows a graph with 12 nodes and an average degree $k = 4$ created with preferential attachment method.

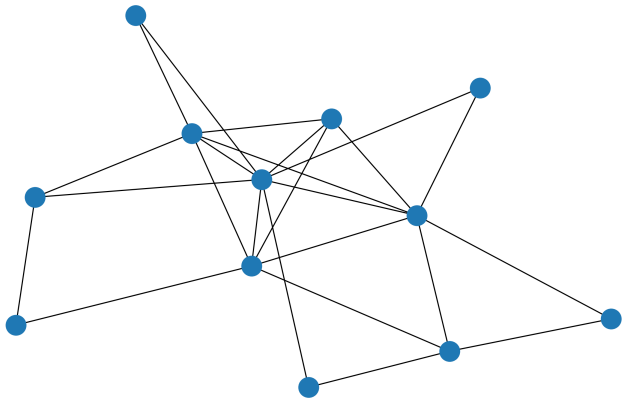

Fig. 3. Graph created with preferential attachment

II. Simulate a pandemic without vaccination

In this exercise we use the graph generate in Section I-B in order to simulate the epidemic on it. We need to generate a preferential attachment random graph $G = (V, E)$, with $|V| = 500$ nodes and average degree $k = 6$, with $\beta = 0.3$ and $\rho = 0.7$. We are asked to simulate the epidemic for 15 weeks, starting with an initial configuration of 10 infected nodes, randomly selected.

Figure 4 shows the average number of newly infected individuals each week (i.e. how many individuals become infected each week). With respect to the previous case (Section I-A), we can see that the epidemic spreads with much more impact on the number of cases, reaching its peak at the 4^{th} week with over 100 casualties (in Section I-A we have just 11 cases). This increasing growth in the number of cases per week can be justified considering that we have a graph on which the epidemic spreads far more quickly: the higher average degree ($k = 6$) can be, indeed, interpreted as an higher number of interactions among people, which results in an easier spread of the infection. It is worth also noticing that, even if the *random* initial infected policy has a slight impact on the number of cases per week, this contribute is far less than relevant with respect to the Section I-A case, and will further decrease in Sections IV and V. Figure 5 shows the average total number of susceptible, infected, and recovered individuals at each week (i.e. many individuals in total that

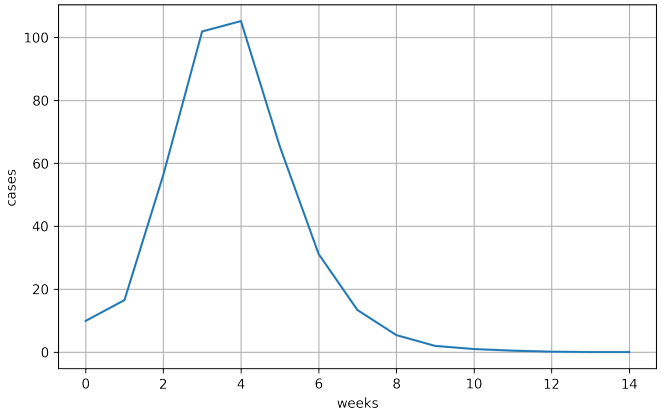

Fig. 4. Average number of newly infected individuals each week

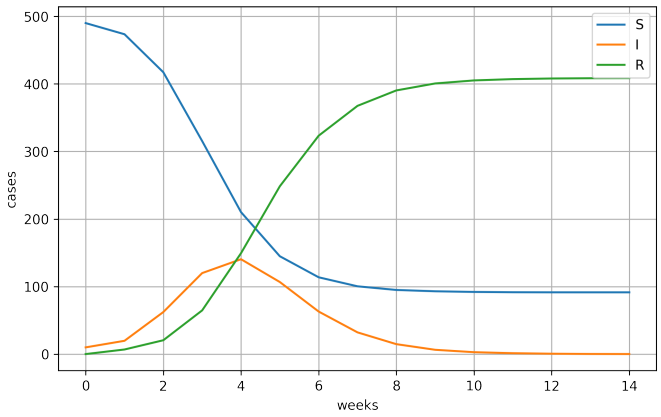

Fig. 5. Average total number of S, I, and R individuals at each week

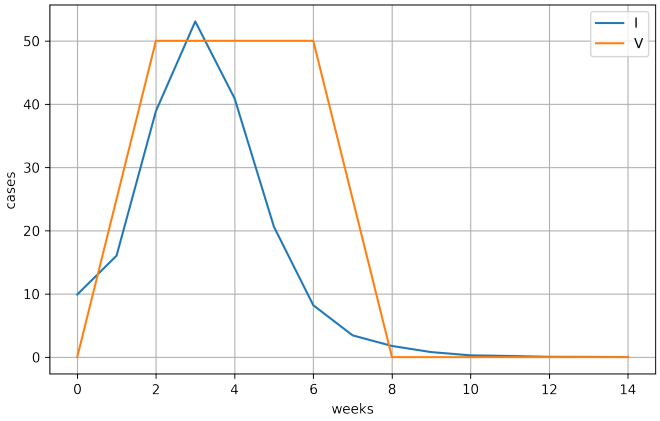

Fig. 6. Average number of newly infected and vaccinated per week

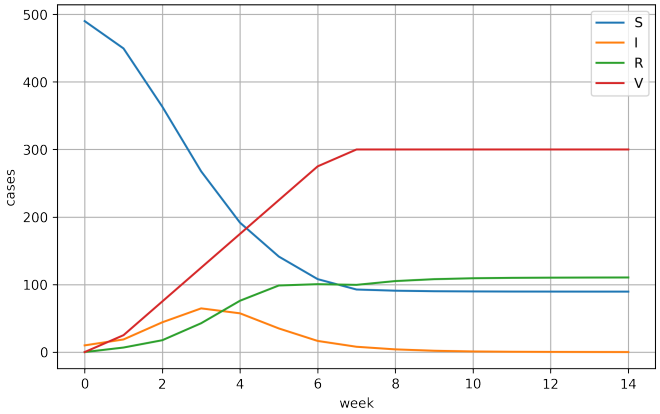

Fig. 7. Average total number of S, I, R and V individuals at each week

are susceptible/infected/recovered at each week). Once again, this graph validates the observations made for the previous one and confirms the behavior of the epidemic already discussed. Differently from the previous case in Section I, as the epidemic spreads and reaches a greater share of the total population, the number of the susceptible individuals decreases together with the increase of the number of recovered ones. At the end of the epidemic, almost all people that were susceptible took the virus (i.e. become recovered): in the absence of an effective counter to the spreading of a possible fatal virus, this could represent a possible disaster in terms of cases and deaths. For this reason, in the Section III, we take into account the role of a counter-action as a vaccine.

III. Simulate a pandemic with in vaccination

In this part we want to simulate the pandemic changing the state paradigm. Indeed, until now we have analyzed the pandemic according to three states, which were susceptible, infected and recovered. Our aim is to slow down the epidemic by means of vaccination. For vaccination, we mean that, at each unit of time, some parts of the population will receive vaccination. Vaccinated (**V**) is the new state in the SIR paradigm, thus becoming SIVR. The vaccination schema follows the cumulative percentages given by the assignment.

We focus on the benefits that the new state can bring. Taking into account Figure 6, we can highlight how for the first two weeks both curves are increasing and this is understandable since the epidemic is spreading as before. However, when the number of vaccinated individuals becomes constant, we experience a critical decrease of the infected people. The impact of vaccination on epidemic spread is evident since we assumed, by assignment, that a vaccinated individual cannot be infected anymore. The last part of the graph is clarifying the eradication of the epidemic: newly infected remains zero for the rest of our timeline.

Figure 7 shows the dynamics of SIRV paradigm at each week. With respect to Figure 5 (dynamics of infections without vaccination) we can further investigate the number of recovered and infected people. For what concerns Infected (I), Figure 7 corroborates the previous statements, whereas, regarding Recovered individuals (R), their count does not overcome the 200 cases during all the timeline. That is explained by the increase in the number of vaccinated that allows a reduction of people that get sick lowering by implication recovered cases.

IV. The H1N1 pandemic in Sweden 2009

In this fourth exercise, we exploit all the previous Sections in order to estimate the social structure of the Swedish population and the disease-spread parameters during the H1N1 pandemic. We will simulate the pandemic between week 42, 2009 and week 5, 2010. During these weeks, once again, the vaccination schema follows the cumulative percentages given in the text assignment. To reduce simulation time, we scale down the population of Sweden of a factor of 10^4, thus considering a graph with nodes $n = |V| = 934$. For the scaled version, the number of newly infected individuals each week in the period between week 42 and week 5 was:

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

Our objective is to develop an algorithm that leverages 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. The pseudo-code of our algorithm is showed in Algorithm 1.

Algorithm 1 Search algorithm

Initial guess of k_0, β_0, ρ_0 with $\Delta k, \Delta\beta, \Delta\rho$
while $(k_{new}, \beta_{new}, \rho_{new}) \neq (k_0, \beta_0, \rho_0)$ **do**
  for (k, β, ρ) with k, β, ρ in their parameter spaces **do**
    Generate a random graph G using preferential attachment
    while $i \leq 10$ **do**
      Simulate the pandemic for 15 weeks on G
      $i = i + 1$
    end while
    Compute avg of newly infected for each week, $I(t)$
    Compute RMSE between simulation and real pandemic:

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

  end for
  $(k_{new}, \beta_{new}, \rho_{new}) = (k, \beta, \rho)$ with lowest RMSE

The initial guess of parameters is given by the text assignment, specifically we have $(k_0, \beta_0, \rho_0) = (10, 0.3, 0.6)$ and $(\Delta k, \Delta\beta, \Delta\rho) = (1, 0.1, 0.1)$. Basically, at each step we explore the neighbourhood of the parameter configuration (k, β, ρ), defined as:

- neighbourhood of $k = [k - \Delta k, k, k + \Delta k]$
- neighbourhood of $\beta = [\beta - \Delta\beta, \beta, \beta + \Delta\beta]$
- neighbourhood of $\rho = [\rho - \Delta\rho, \rho, \rho + \Delta\rho]$

considering all the possible combinations of the parameters. Our simulation found that the best configuration is $(k, \beta, \rho) = (9, 0.214, 0.555)$ with the lowest achieved RMSE $= (3.91)$.

Figures 8 and 9 show the obtained results for the H1N1 pandemic simulation from week 42 (2009) to week 5 (2010) on a smaller sample of the Swedish population. We can immediately notice that the predicted behavior of the epidemic is sufficiently similar to the actual one, with some exceptions. First of all, the peak in our simulated epidemic is reached 1 week before the actual one, as we can see from the shift to the left of the predicted curve. Secondly, also the magnitude of the impact on number of cases is slightly over-estimated with respect to the actual one: the maximum number of cases reached in the peak is higher by some units than the original one. Finally, Figure 9 shows a similar pattern with respect to the one already discussed in Section III. We can notice just two differences:

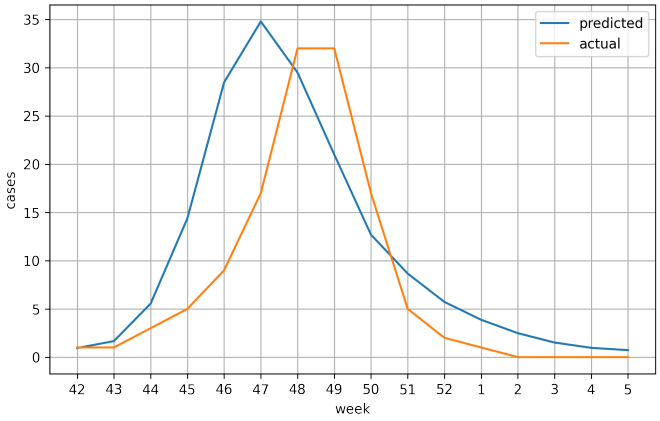

Fig. 8. Average number of newly infected and vaccinated per week

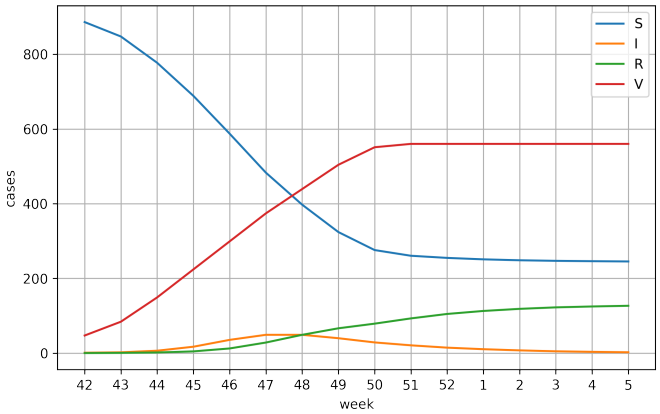

Fig. 9. Average total number of S, I, R and V individuals at each week

1) the plot in the section III shows an higher peak of infected individuals compared to the one during H1N1 pandemic, which remain more flat.
2) the number of recovered individuals in section III outgrows the susceptible one, while in this case the ratio between these two is bigger (thus keeping them more separate).

This may be because we are considering that the 5% of the population already received a vaccine in the first week, whereas in the previous case we start to vaccinate from the second week. This early start of the process has the effect of damp the spread of the pandemic.

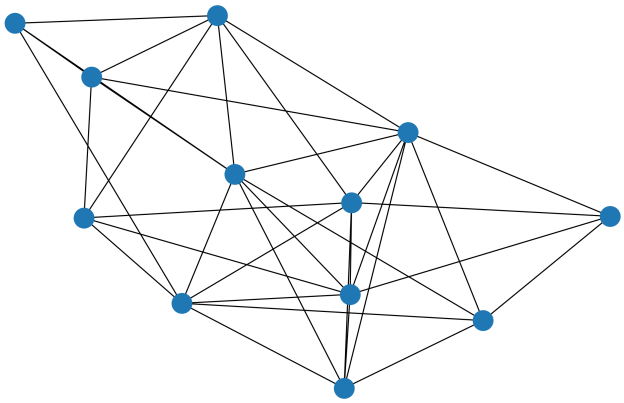

Fig. 10. Graph created using small world

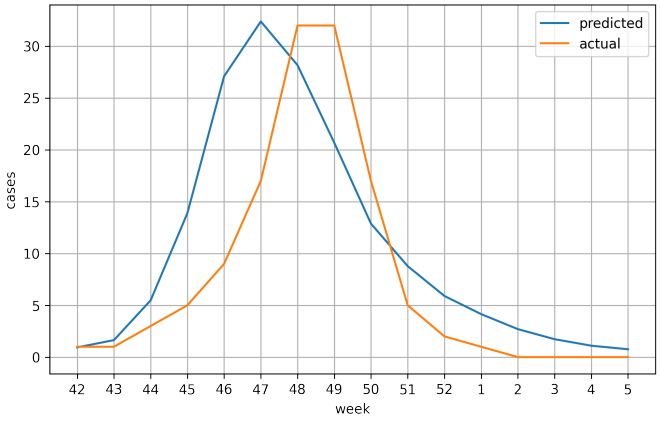

Fig. 11. Predicted vs actual avg # newly infected with random graph

V. Challenge[1]

In the fifth point of the assignment, we are asked to implement a better random graph and a better fine-tuning parameters algorithm with respect to Section IV. As the previous point, we want to approximate the H1N1 epidemic. The adopted solutions are the *Small World graph* and the *Tabu Search* meta-heuristic algorithm.

The Small World graph is a type of mathematical network in which most nodes are not neighbors of one another, but neighbors of any given node are more likely to be neighbors of each other. This means that most nodes can be reached from every other node with a small number of steps. Specifically, a small-world network is defined to be a network where the typical distance between two randomly chosen nodes grows proportionally to the logarithm of the number of nodes in the network. The graph is built, through the `networkx` built-in function *newman_watts_strogatz_graph(...)*, starting from a *k-regular* Graph, where each node has the 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. For completeness, Figure 12 shows a small graph $(n = 12, k = 4, p = 0.5)$.

The *Tabu search algorithm* provides a way to improve the search of the k, β and ρ parameters. It is a meta-heuristic search method employing local search methods used for mathematical optimization. Local searches take a potential solution to a problem and check its immediate neighbors in the hope of finding an improved solution. Classic local search methods tend to be stuck in sub-optimal regions or on plateaus where many solutions are equally fit. Tabu search, instead, enhances the performance of local search by relaxing its basic rule. First, at each step worsening moves can be accepted if no improving move is available. In addition, prohibitions are introduced to discourage the search from coming back to previously-visited solutions. The algorithm implementation uses memory structures that describe the visited solution or user-provided sets of rules. If a potential solution has been previously visited within a certain short-term period or if it has violated a rule, it is marked as '***tabu***' so that the algorithm does not consider that possibility repeatedly.

First of all, we test the presented algorithm on the Random Graph with preferential attachment as proposed in Section IV. The best configuration that we found for our parameters is $(k, \beta, \rho) = (9, 0.21, 0.55)$ with a lowest achieved RMSE $= 3.91$. The obtained results are comparable with the ones of the algorithm implemented in the previous section, but *Tabu Search* provides an improvement in computations time reducing the number of needed iterations. The comparison between predicted and real newly infected is reported in Figure 11.

For the second step, we apply the Tabu Search over a graph created using the Small-world technique. We choose a value of 0.5 for p (the probability of adding a new edge for each edge). With this configuration the set of best parameters found is $k = 12$, $\beta = 0.19$, $\rho = 0.86$ yielding a lowest RMSE of 4.617. Figure 12 shows the comparison between predicted and actual newly infected individuals per week. With respect to the case using preferential attachment, the shape of the predicted infected is shifted to the right fitting more accurately the real shape of the actual epidemic, slightly overestimating the number newly infected during the analysed weeks.

Conclusions

We addressed three different problems on epidemic models and pairwise network systems. Then, we combined all the previous exercises in order to estimate the social structure and the spreading of an epidemic in a real-world scenario: the H1N1 pandemic in Sweden in the final weeks of 2009 and the first weeks of 2010. Initially, we modeled how a SIR epidemic spreads over a known graph. After that, we explored its behavior on an unknown random graph: firstly, we considered the SIR model as before, then we added the contribution of a counter-action as the vaccine, in order to mitigate the spread of the epidemic, thus adding a state V (vaccinated) to the SIR model. We further discussed the impact that the vaccination

[1]In collaboration with Pierluigi Compagnone (s288301)

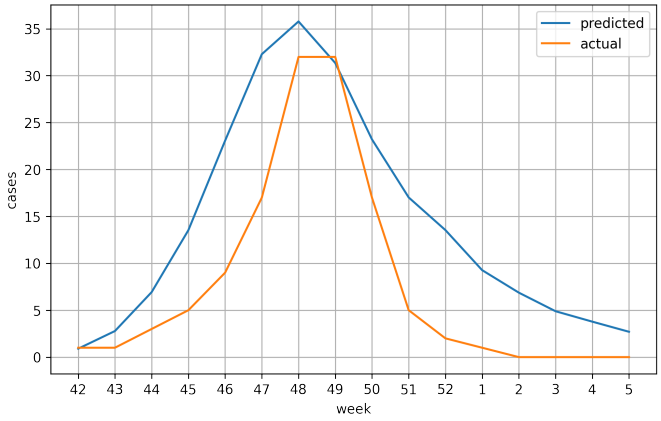

Fig. 12. Predicted vs actual avg # newly infected with small world graph

had on the spread of the epidemic, underlying the importance of it.

Our last contribution was to explore the Small-World random graph generator, a different algorithm to generate random graph without exploiting the preferential attachment used in the first simulation. We also deployed Tabu search, a different algorithm for the search of the best configuration of the parameters, parameters that were then used to generate the graph and to simulate the epidemic. Once again, we showed and discussed the results obtained.

Future works could take into account the implementation of other meta-heuristic and heuristic algorithms for the parameter optimizations, such as *Ant Colony Optimization* and *Genetic algorithm*, that may provide better results. With the same principle, other techniques can be explored to generate random graphs, such as *Erdős–Rényi model*.

REFERENCES

[1] G. Como and F. Fagnani, "Lecture on Network Dynamics"
[2] Networkx manual v. 2.6.2 - https://networkx.org/
[3] L. Benfenati, "Homework 3 - ND&L, Jupyter Notebooks (.ipynb)"
[4] C. De Cristofaro , "Homework 3 - ND&L, Jupyter Notebooks (.ipynb)"
[5] N. Scarano, "Homework 3 - ND&L, Jupyter Notebooks (.ipynb)"
[6] A. Vacca, "Homework 3 - ND&L, Jupyter Notebooks (.ipynb)"
[7] Networkx manual v. 2.6.2 - https://networkx.org/