Homework 3
Network Dynamics and Learning
Politecnico di Torino

Hajali Bayramov - s288874

January 15, 2022

The goal of this homework is to simulate the spread and control of the 2009 H1N1 pandemic in Sweden using random graph models.

1 Preliminary parts

In the first part, we will introduce the tools that we will use to model the aforementioned pandemic.

1.1 Epidemic on a known graph

One of the simple models used to model disease propagation is the SIR epidemic model in a discrete time scale. This model has two main hyperparameters. The first hyperparameter is the transmission probability β that is in the range $[0, 1]$. This is the probability that an infected node (denoted by the symbol I), transmits the disease to a susceptible node (denoted by the symbol S). The transmission is independent from other infected neighbors of the susceptible node. Therefore, if a susceptible node i has m infected neighbors in time step t, he/she will become infected in time step $t+1$ with probability $1 - (1 - \beta)^m$. The second hyperparameter is the recovery rate ρ that is again in range $[0, 1]$ and is the probability that an infected node i will become recovered (denoted by symbol R) during one time step. We may formalize these two dynamics with the following transition probabilities:

$$P(X_i(t+1) = I | X_i(t) = S, \Sigma_{j \in \mathcal{V}} W_{ij} \delta^I_{X_j(t)} = m) = 1 - (1 - \beta)^m \tag{1}$$

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

where $\Sigma_{j \in \mathcal{V}} W_{ij} \delta^I_{X_j(t)} = m$ is the number of infected neighbors for node i.

We can simulate the SIR dynamics on a k-regular graph. A k-regular graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a graph with node indices $\mathcal{V} = 1, ..., n$ where every node is directly connected to its k neighbor nodes. For example, the graph in Figure 1 shows a 4-regular graph with $n = 8$ nodes.
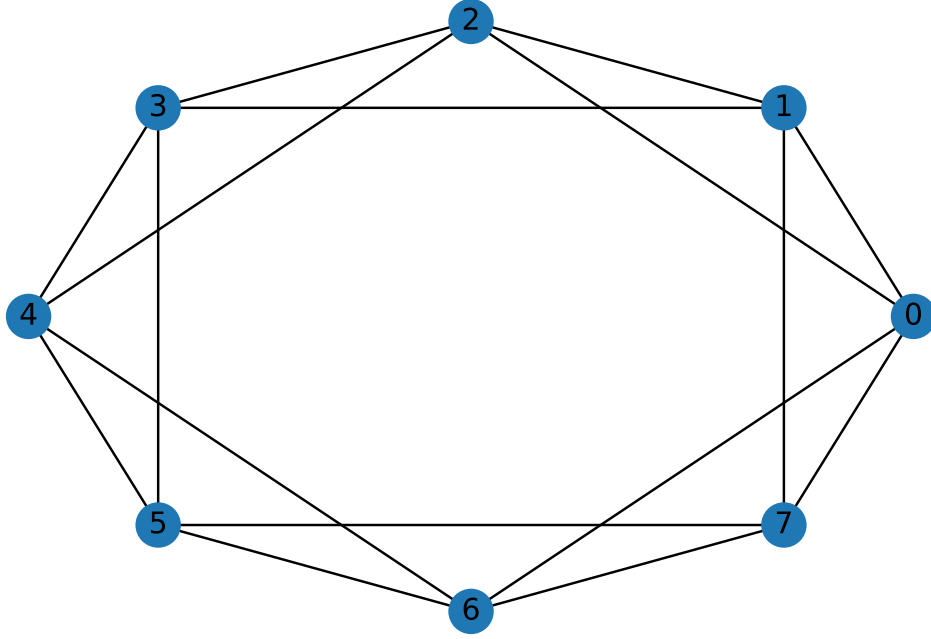
Figure 1: A k-regular graph with 8 nodes and $k=4$

Now, we will simulate the SIR epidemic on a 4-regular graph with 500 nodes with 10 initial infected nodes. For the first experiment, we choose the initial infected nodes at random, and run the SIR dynamics for 15 weeks. For statistical significance, we run the experiment 100 times and average the results. Figure 2 shows the average number of newly infected individuals each week during the 15-week period. Furthermore, Figure 3 shows the average total number of infected and recovered individuals, and Figure 4 shows the average total number of susceptible individuals over the weeks.
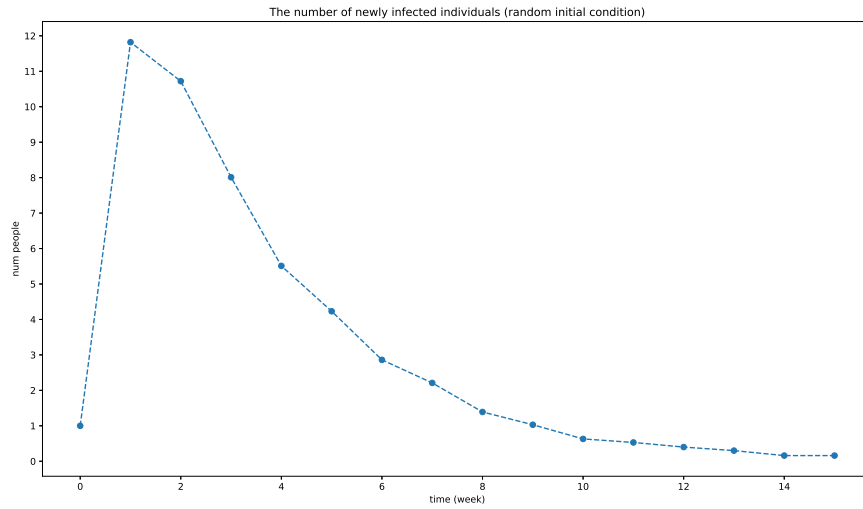

Figure 2: The average number of newly infected individuals in the 4-regular graph SIR simulation
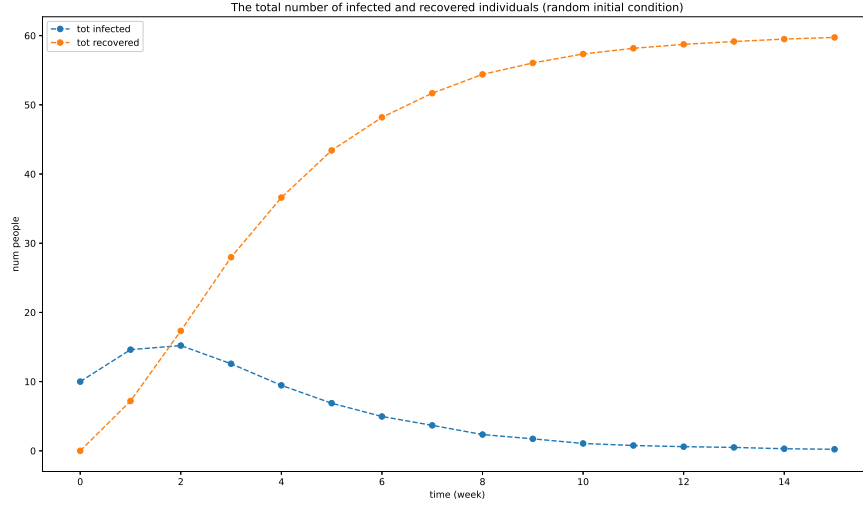
Figure 3: The average number of total infected and recovered individuals in the 4-regular graph SIR simulation
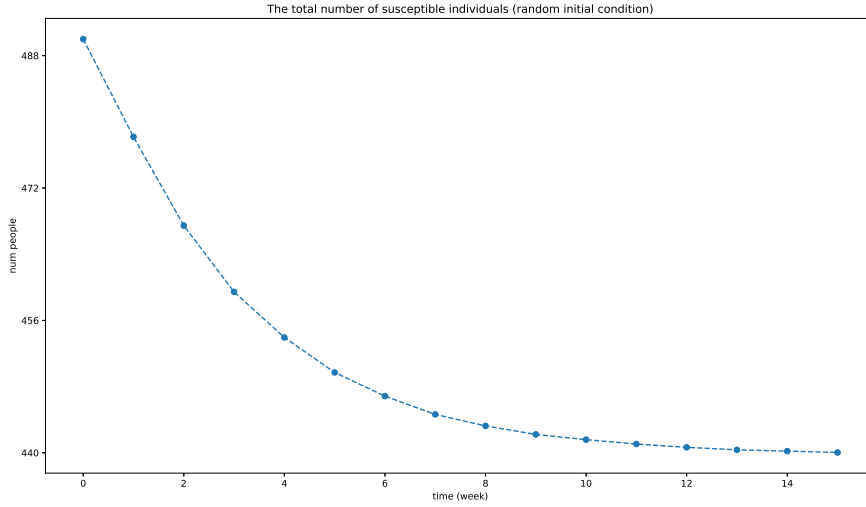

Figure 4: The average number of total susceptible individuals in the 4-regular graph SIR simulation

We can repeat this experiment with a more realistic initial infected population. To do so, we choose the first eight nodes in $\mathcal{V}$ as the initial infected nodes. Note that because of the structure of the graph (k-regular with $k = 4$), these nodes can be thought as eight people in a social circle with multiple ties among each other, and the pandemic might start with one of them infecting the whole circle because of physical proximity to each other, and then based on the SIR dynamics, it could spread throughout the network. This situation is analogous to Covid-19 clusters in the current pandemic, such as the White House Cluster that emerged after the ceremony held in the Rose Garden [1]. With this initial condition, we again run the SIR dynamics for 100 times and plot the average results that can be seen in Figures 5 to 7.
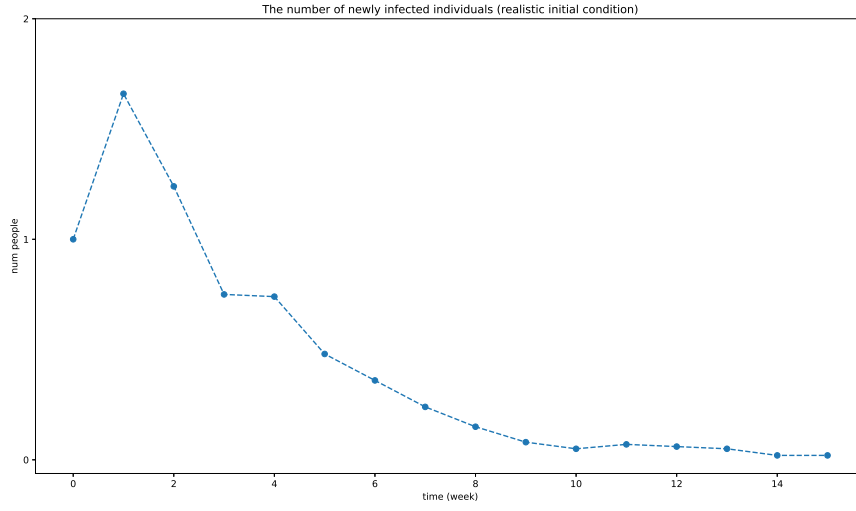
Figure 5: The average number of newly infected individuals on a 4-regular graph with a realistic initial infected nodes
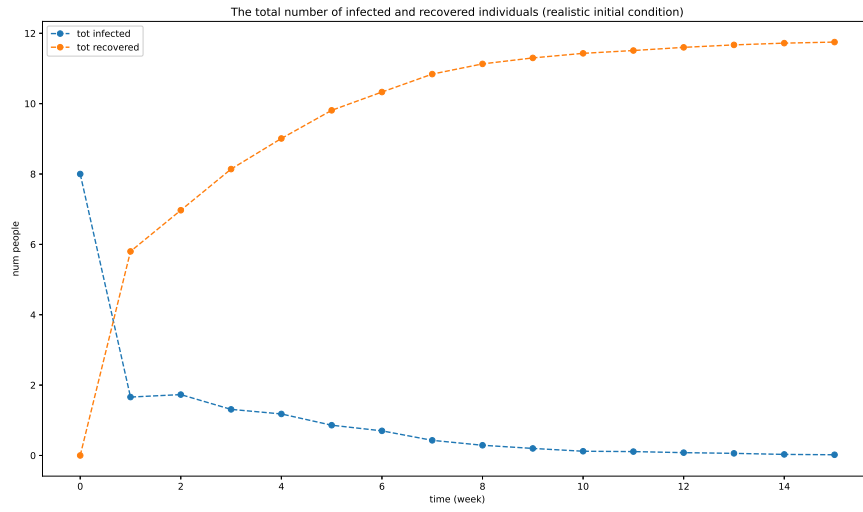

Figure 6: The average number of total infected and recovered individuals on a 4-regular graph with a realistic initial infected nodes
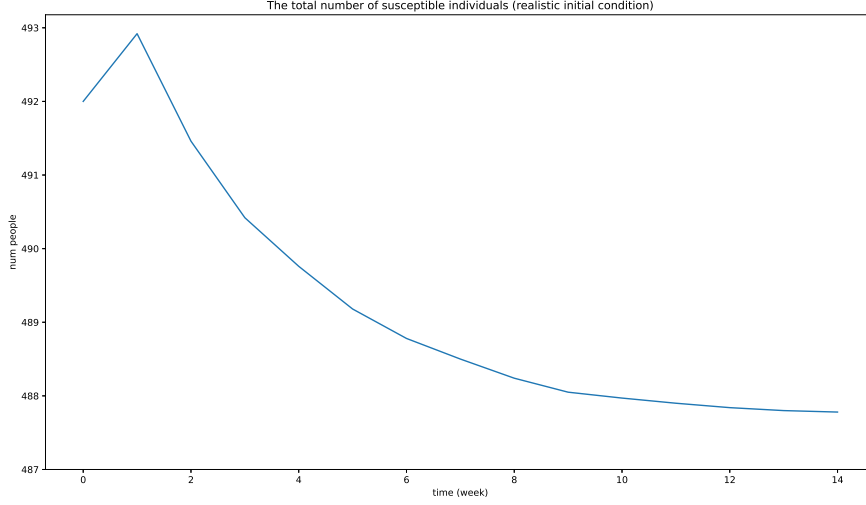
Figure 7: The average number of total susceptible individuals on a 4-regular graph with a realistic initial infected nodes

With this realistic initial condition, we can observe that the number of infected individuals does not increase as expected in a pandemic. A plausible explanation for this phenomenon is that the k-regular graph is not suitable for modeling the connections in a real society (for example, it is obvious that the pandemic will die out soon because of extreme local connectivity of the network and the absence of the small-world property). In the next part, we will work with another graph model that may better capture the real connections in a society.

1.2 Generating a random graph

A feasible way to generate the population with a connection pattern more suitable for the spread of an epidemic is to generate a random graph with the preferential attachment model. The goal is to have a connected graph with average node degree of k. To achieve this goal, at $t = 1$ we start with an initial graph $\mathcal{G}_1$ that is complete with $k + 1$ nodes. Then, at every time step $t \geq 2$, we create $\mathcal{G}_t$ by adding a new node to $\mathcal{G}_{t+1}$ and adding $_t(t) = c = k/2$ undirected edges from the existing nodes to this node. The nodes to be linked to the new node are chosen based on a preferential "the rich get richer" scheme and according to the following probability distribution (at most one link can be added for each node):

$$P(W_{n_t,i}(t) = W_{i,n_t}(t) = 1|\mathcal{G}_{t-1} = (\mathcal{V}_{t-1}, \mathcal{E}_{t-1})) = \frac{\omega_i(t-1)}{\Sigma_{j \in \mathcal{V}_{t-1}} \omega_j(t-1)}, i \in \mathcal{V}_{t-1} \tag{3}$$

where $W(t)$ represents the adjacency matrix at time step t, and $\omega_i(t-1)$ is the degree of node i prior to adding the new node.

This model, contrary to the k-regular graph, does not create a symmetric and locally connected pattern between the nodes. It also satisfies the small-world property, and at the same time maintains the average degree of nodes in the graph. It is therefore more similar to the actual connections in a social network. We can experimentally validate these statements statement by implementing the preferential attachment model algorithm and creating a random graph. Creating a graph with $n = 500$ nodes and $k = 6$ results in the following degree histogram and with a average degree of 6 over all nodes. Furthermore, we observed that the longest shortest path in the graph was 4, which shows that the graph satisfies the small-world property.
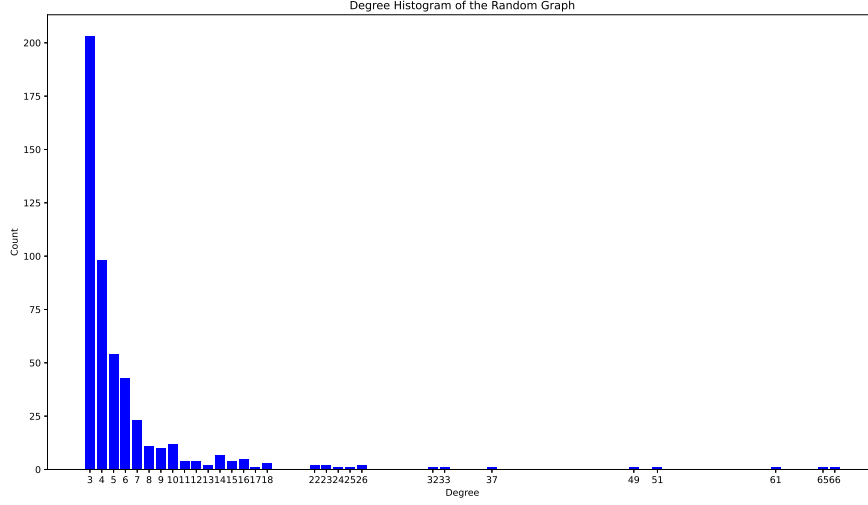
Figure 8: The degree histogram of the preferential attachment random graph model with n=500 and k=6

2 Simulating a pandemic without vaccination

In this part, we will simulate a pandemic on a random graph generated by the preferential attachment model and using the SIR epidemic dynamics. The graph has 500 nodes with average degree of $k = 6$ and 10 initial infected nodes. The SIR dynamics hyperparameters are $\beta = 0.3$ and $\rho = 0.7$. For the first experiment, we choose the initial infected nodes at random, and run the SIR dynamics for 15 weeks. For statistical significance, we run the experiment 100 times and average the results. Figure 9 shows the average number of newly infected individuals each week during the 15-week period. Furthermore, Figure 10 shows the average total number of infected and recovered individuals, and Figure 11 shows the average total number of susceptible individuals over the weeks.
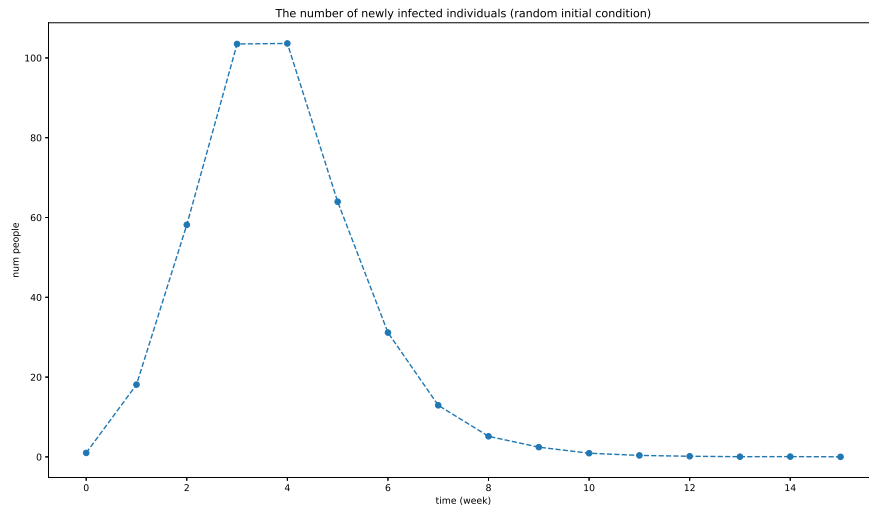
Figure 9: The average number of newly infected individuals with random initial infected nodes and without vaccination
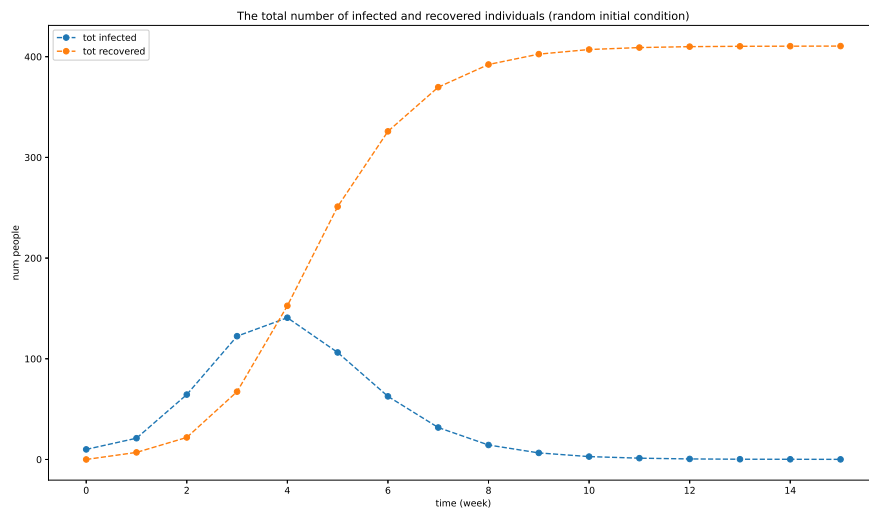

Figure 10: The average number of total infected and recovered individuals with random initial infected nodes and without vaccination
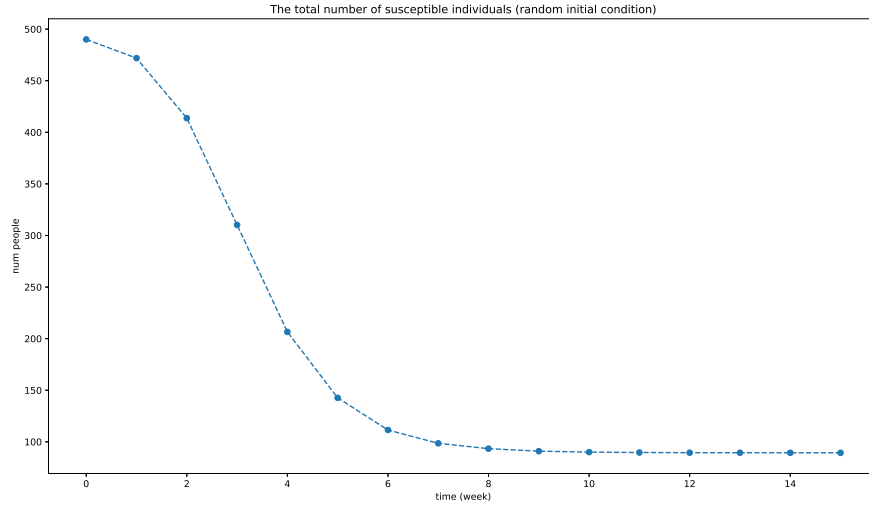
Figure 11: The average number of total susceptible individuals with random initial infected nodes and without vaccination

Again, we can repeat this experiment with a more realistic initial infected population. To do so, we choose a node with at least seven neighbors and create the initial infection cluster and run the SIR dynamics for 100 times and plot the average results that can be seen in Figures 12 to 14.
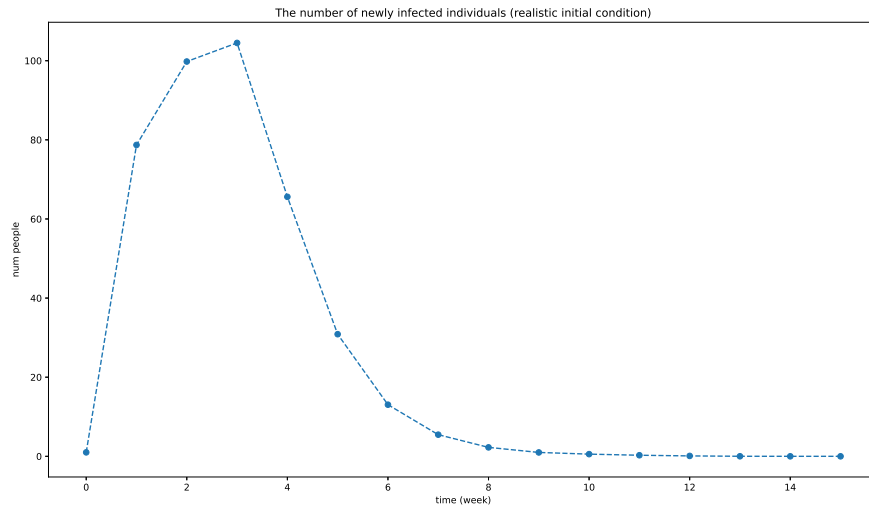

Figure 12: The average number of newly infected individuals with realistic initial infected nodes and without vaccination
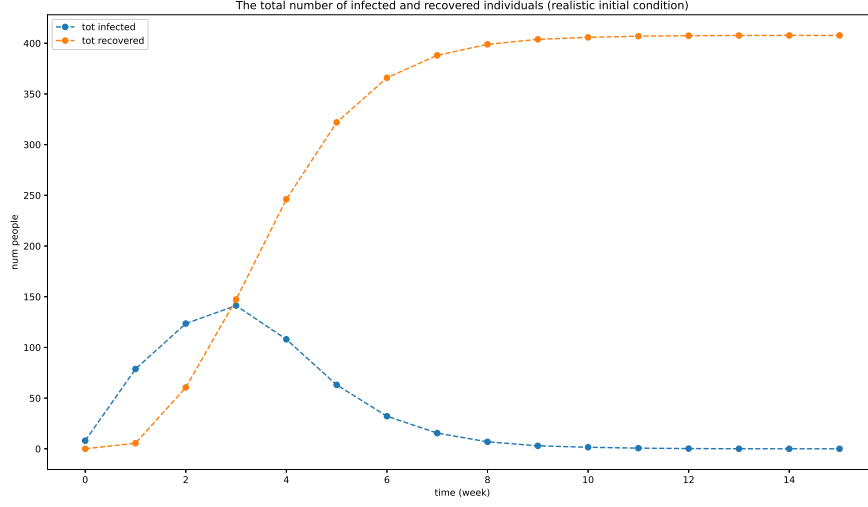
Figure 13: The average number of total infected and recovered individuals with realistic initial infected nodes and without vaccination
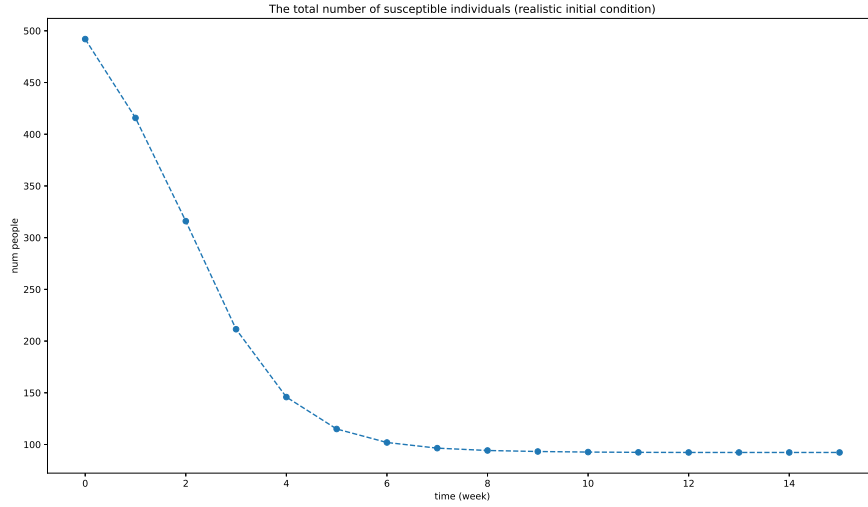

Figure 14: The average number of total susceptible individuals with realistic initial infected nodes and without vaccination

This time, contrary to the k-regular graph, we can observe that the infection trends with a random initialization are similar to those obtained with a realistic initialization. This may be attributed to the small-world property that causes the pandemic to spread faster to the whole graph independently of the initial infected nodes.

3   Simulating a pandemic with vaccination

Now, we introduce vaccination to curb the pandemic and reduce the number of infections. The vaccines take effect immediately, and we assume that a vaccinated person will not become infected

again. The total fraction of the population that have been vaccinated by each week of our 15-week-long simulation is according to:

$$Vacc(t) = [0, 5, 15, 25, 35, 45, 55, 60, 60, 60, 60, 60, 60, 60, 60] \qquad (4)$$

The individuals to be vaccinated each week are chosen randomly from the current unvaccinated population (an infected individual might receive vaccination as well and his/her state will change from infected to vaccinated (denoted by symbol V)). The vaccinated individuals are immune to infection until the end of simulation.

Now, we will run the SIR dynamic with vaccination on a graph with the same characteristics as the previous part and random initial infected nodes. The simulation is done 100 times and the results are averaged for statistical significance. The plots in Figures 15 to 17 show the average number of newly infected and vaccinated individuals, the total infected and recovered individuals, and the total number of susceptible individuals over the weeks.
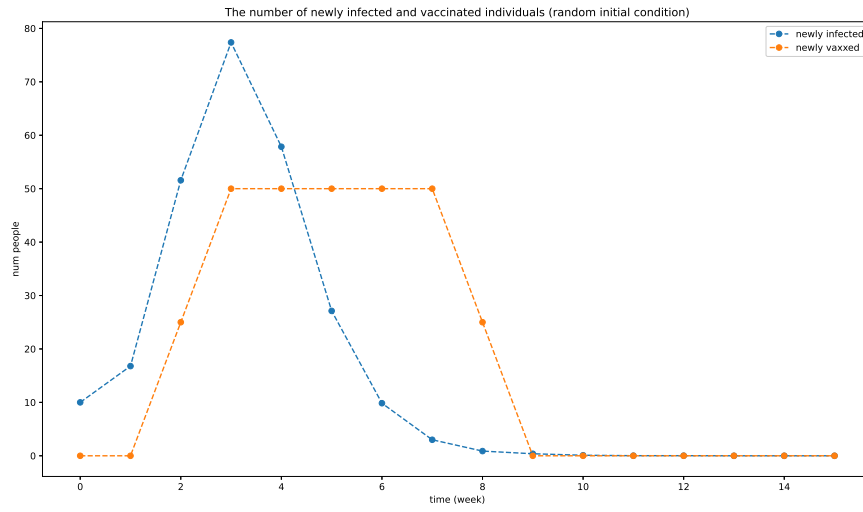

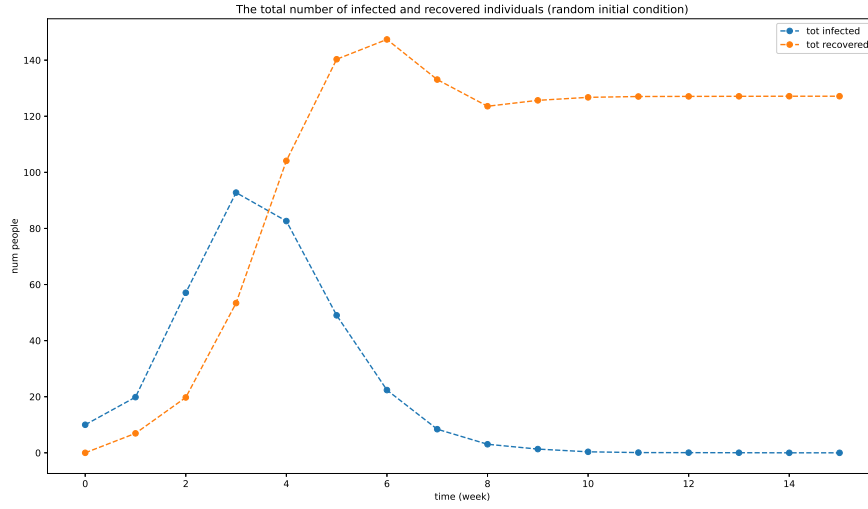Figure 15: The average number of newly infected and vaccinated individuals

Figure 16: The average number of total infected and recovered individuals with vaccination
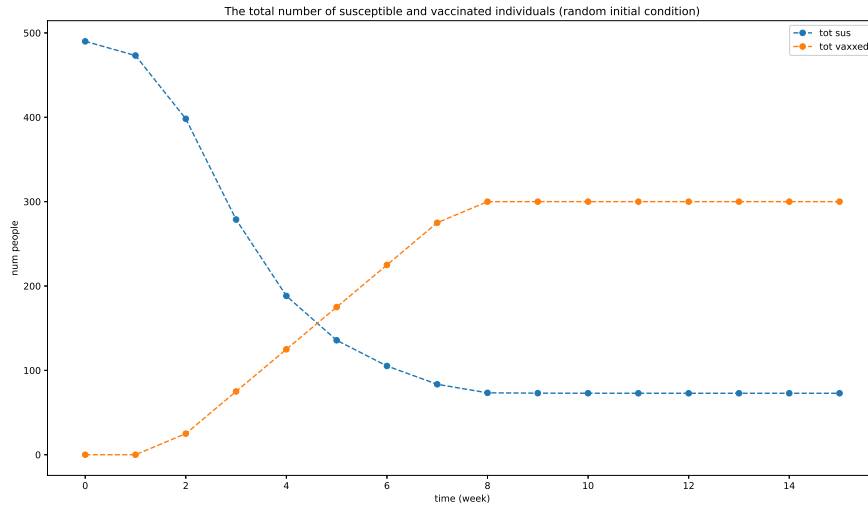

Figure 17: The average number of total susceptible individuals with vaccination

We can observe that vaccination has had a significant effect on reducing both the new weekly infections (compare Figures 9 and 15) and the total number of infected people in each week (compare Figures 10 and 16).

4   The 2009 H1N1 Pandemic in Sweden

Using the previous parts, we now try to simulate and estimate the parameters of the 2009 H1N1 pandemic in Sweden. 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. We will simulate the pandemic between week 42, 2009 and week 5, 2010. During these weeks, the fraction of population that received vaccination was:

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

To reduce the simulation time, we scale down the Sweden population, and the number of infections and vaccinations by a factor of 10^4. Therefore, the number of nodes in the graph will be $n = |\mathcal{V}| = 934$, and the number of newly infected individuals during the aforementioned 15-week period will be:

$$I_0(t) = [1, 1, 3, 5, 9, 17, 32, 32, 17, 5, 2, 1, 0, 0, 0, 0] \tag{6}$$

Now, we try to find a set of hyperparameters for the SIR dynamics and the preferential attachment model to simulate the H1N1 pandemic in Sweden. The hyperparameters to tune are k for the graph model, and β and ρ for the SIR dynamics. Our evaluation criterion for a set of hyperparameters will be the root-mean-square error (RMSE) (defined in (7)) between the estimated newly infected individuals over the weeks (averaged over 10 independent runs) and the real values from (6). The number of independent runs could be increased for a better estimation if we had access to more computational power.

$$RMSE = \sqrt{\frac{1}{n} \times \Sigma_{t=1}^{n}(I(t) - I_0(t))^2} \tag{7}$$

For finding the optimal hyperparameters, we utilize a discrete combinatorial gradient-based search, i.e., at each iteration we test every possible combination of the three hyperparameters in $\{a - \Delta a, a, a + \Delta a\}$ for each hyperparameter a. We start with $k_0 = 10$, $\beta_0 = 0.3$, and $\rho_0 = 0.6$, with steps $\Delta k = 1$, $\Delta\beta = 0.1$, $\Delta\rho = 0.1$. In order to fine-grain our search when we do not find a better solution, we employ a halving mechanism for the float hyperparameters β and ρ. In this method, we reduce the step by half if the specific parameter does not change. We stop the algorithm when there is no improvement in the RMSE for two consecutive iterations. The final optimal hyperparameters were found as below:

$$\{k, \beta, \rho\} = \{9, 0.2, 0.5\} \tag{8}$$

After running the simulation with these hyperparameters for 100 times and averaging the results, we obtain the plots in Figures 18 and 19 for the total number of infected and recovered individuals, and the total number of susceptible individuals over the weeks.
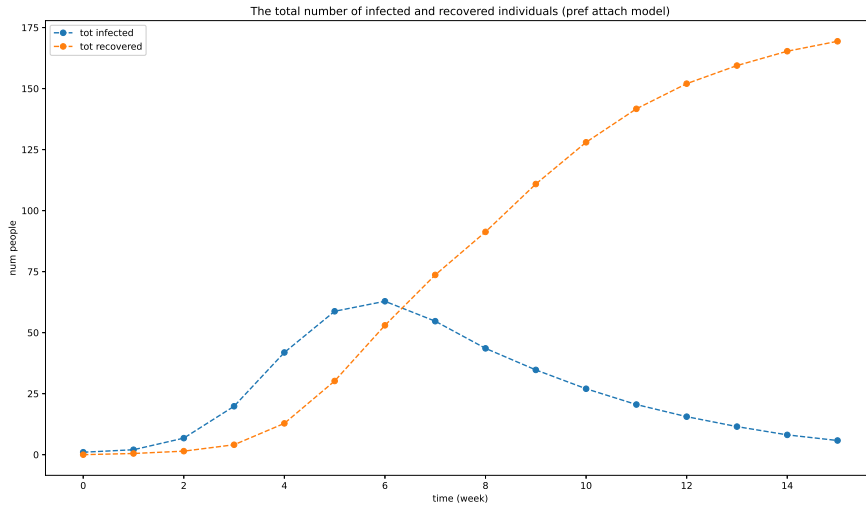

Figure 18: The average number of total infected and recovered individuals (optimal model for the H1N1 pandemic)

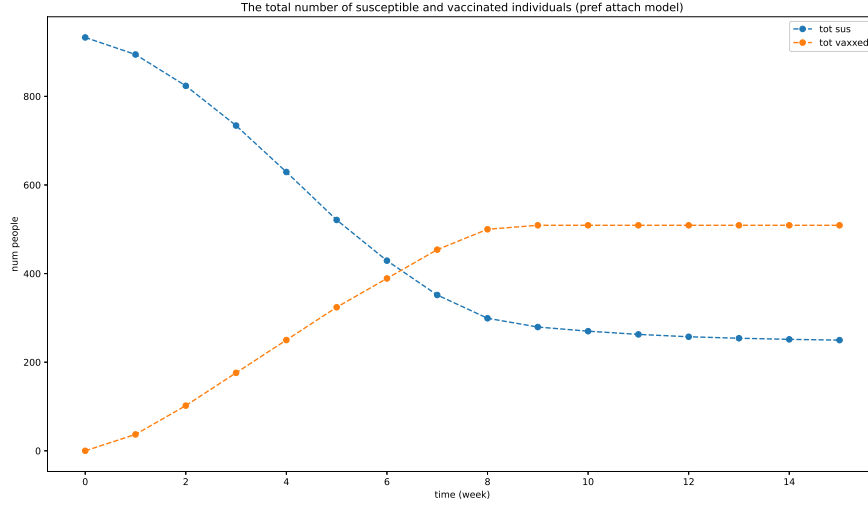The total number of susceptible and vaccinated individuals (pref attach model)

Figure 19: The average number of total susceptible individuals (optimal model for the H1N1 pandemic)

We can also compare the estimated newly infected individuals with the real data in (6). The graph in Figure 20 visualizes this comparison. It can be seen that the model has given a fairly good estimation of the newly infected individuals over the weeks.

The comparison of the real and estimated newly infected indivuals using the preferential attachment model
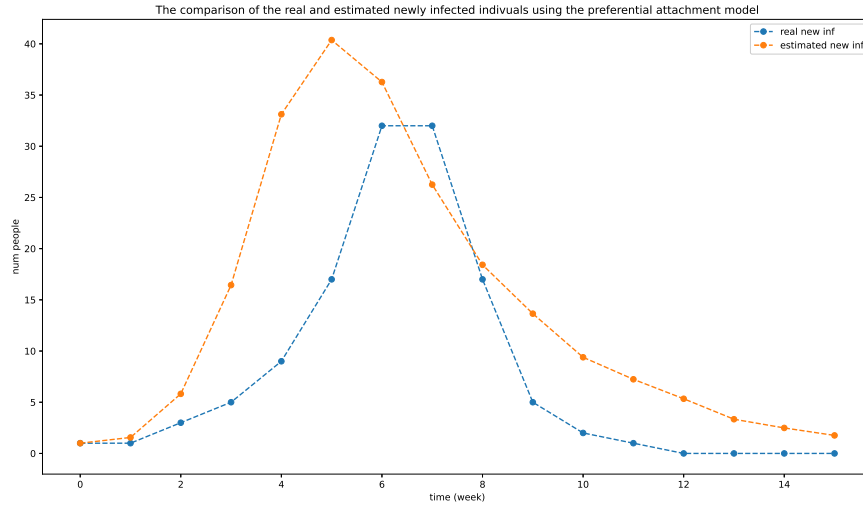
Figure 20: The comparison between the estimated and real number of newly infected individuals in the H1N1 pandemic using the optimal preferential attachment model

5 Watts-Strogatz (WS) small-world network to model the pandemic (challenge)

In this part, we use a better random graph and a slightly modified optimization technique to model the society and the pandemic. The graph model we use is the Watts-Strogats small-world model

[2]. This model gives us the ability to distinguish between contacts within clustered cliques and unclustered, random ones in the population which may be helpful in making and implementing policies to sustain an epidemic or pandemic. It has also been used to model the Covid-19 pandemic in the last two years [3, 4]. In this graph model, we first generate a k-regular graph, and then with a probability p rewire each edge to model the random connections between the nodes. This probability can decrease if the government imposes lockdowns and restrictions. When we do not have any restrictions, a reasonable choice for p is 0.5 [4] which we will use here. Again, we have three hyperparameters to optimize which are k in the WS graph model, and ρ and β in the SIR dynamics. We also use an enhanced gradient-based optimization technique with a restart mechanism. In this method, we run the same gradient-based algorithm as the previous part until convergence. After convergence, we restart the algorithm with the converged parameters as the initial condition but we reduce the steps of the float hyperparameters ρ and β by half. The motivation behind this is to fine-grain the search and get as close as possible to the optima. We restart the algorithm twice, and report the final hyperparameters:

$$\{k, \beta, \rho\} = \{10, 0.275, 0.85\} \tag{9}$$

After running the simulation with these hyperparameters for 100 times and averaging the results, we obtain the plots in Figures 21 and 22 for the total number of infected and recovered individuals, and the total number of susceptible individuals over the weeks.
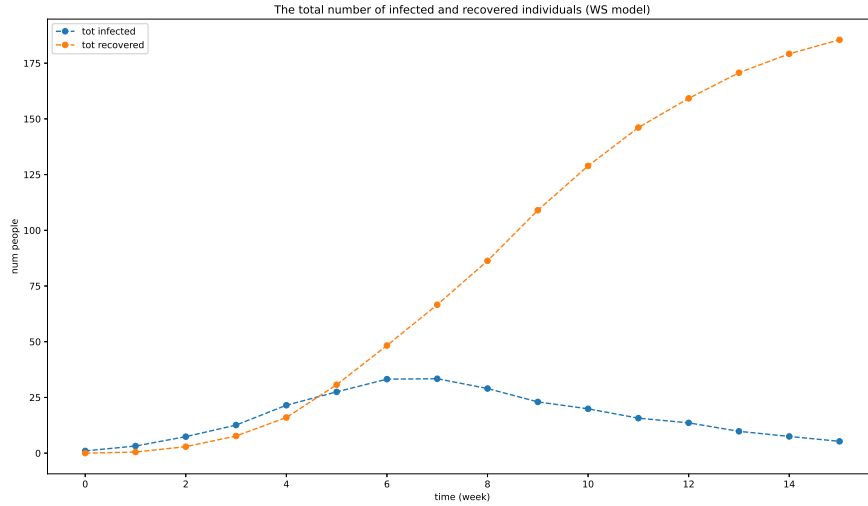

Figure 21: The average number of total infected and recovered individuals (optimal Watts-Strogatz model for the H1N1 pandemic)
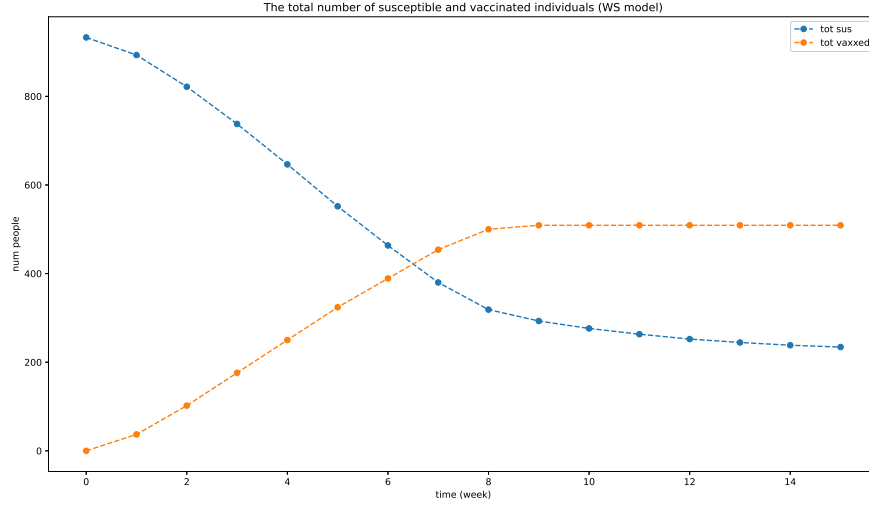
Figure 22: The average number of total susceptible individuals (optimal Watts-Strogatz model for the H1N1 pandemic)

We can also compare the estimated newly infected individuals with the real data in (6) and also the data achieved using the preferential attachment model. The graph in Figure23 visualizes this comparison.
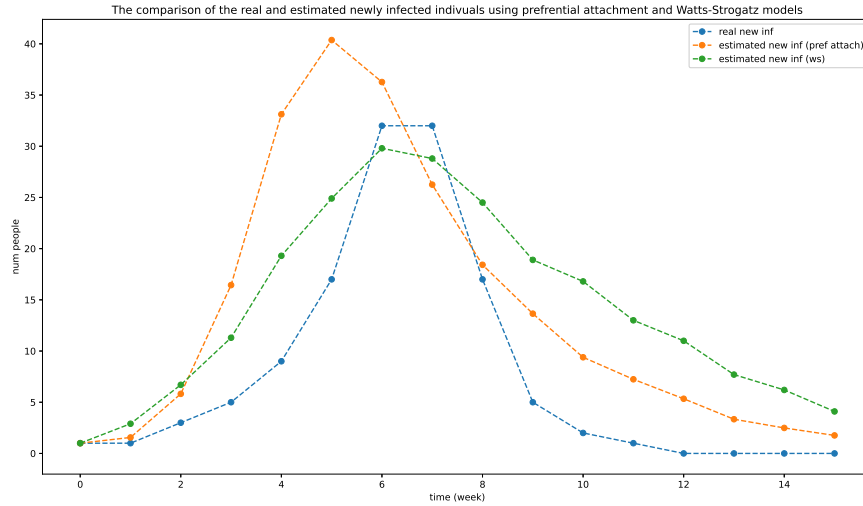

Figure 23: The comparison between the estimated and real number of newly infected individuals in the H1N1 pandemic using the optimal preferential attachment and Watts-Strogats models

It can be seen that the stronger Watts-Strogtz model alongside with the improved optimization technique have resulted in a better estimation of the newly infected individuals compared to the preferential attachment model.

Notes

For most of the graph operations in this assignment, the Python package NetworkX [5] was used.
I have collaborated with Hafez Ghaemi in solving the exercises and writing the report.

References

[1] Wikipedia, "White House COVID-19 outbreak — Wikipedia, the free encyclopedia," http://en.wikipedia.org/w/index.php?title=White\%20House\%20COVID-19\%20outbreak&oldid=1065343165, 2022, [Online; accessed 14-January-2022].

[2] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world'networks," *nature*, vol. 393, no. 6684, pp. 440–442, 1998.

[3] L. Xue, S. Jing, J. C. Miller, W. Sun, H. Li, J. G. Estrada-Franco, J. M. Hyman, and H. Zhu, "A data-driven network model for the emerging covid-19 epidemics in wuhan, toronto and italy," *Mathematical Biosciences*, vol. 326, p. 108391, 2020.

[4] S. Syga, D. David-Rus, Y. Schälte, M. Meyer-Hermann, H. Hatzikirou, and A. Deutsch, "Inferring the effect of interventions on covid-19 transmission networks," *arXiv preprint arXiv:2012.03846*, 2020.

[5] A. Hagberg, P. Swart, and D. S Chult, "Exploring network structure, dynamics, and function using networkx," Los Alamos National Lab.(LANL), Los Alamos, NM (United States), Tech. Rep., 2008.