Network Analysis: Assignment 3

Simone Campisi s4341240

Jacopo Dapueto s4345255

In this assignment we implemented the **SIR epidemic model** and simulated on different networks: *The Karate Club Graph* and *Facebook dataset* already used in the last assignment. The SIR model simplify the mathematical modeling of infectious diseases by identifying 3 different states that a node:

- **Susceptible (S)**. The healthy individuals, who have not yet contacted the virus.
- **Infectious (I)**. Contagious individuals who have contacted the virus and hence they can infects other individuals.
- **Recovered (R)**. Individuals who have recovered from the disease, not more infectious.

In this model there are several parameters:

- The **disease transmission probability p**, which defines the probability of an individual (in state *S*) being infected if it is neighbors of an node in state I.
- T_i represents the minimum number of time steps that an individual in the state *I* (*infectious)*. And after an Individual can move from the state *I* to the state *R (recovered)* with a certain probability q. The probability q is also a parameter.
- The number of individual infected at the beginning of the simulation, i_0 .

All the nodes are initialized to *S* and the first i_0 nodes are infected. the simulation is executed until there is at least one infected node. For each healty neighbor of an infected node a number is sampled randomically and if the result is less than *p*, the contagion occurs and the neighbor is moved from the state *S* to *I*. If the time T_i for an infected nodehas elapsed, it is moved from the state *I* to *R* with probability *q* in the same way a node move from *S* to *I*.

Hence, we have plotted the model with the two types of network, testing different parameters, in order to observe the different behaviors of the model. Therefore, we plotted a graph showing the evolution of the relative number of nodes in the three states S,I,R, for each time step. For the smaller network (Karate Club), we have build also a gif, which shows the evolution of the spread epidemic.

Since we are not domain experts we decided to perform three experiments for each network according to the value of **R** (expected number of nodes infected at each wave by an infected):

- *R < 1*, less than one node will be infected at each wave and so the disease dies out after a finite number of waves.
- *R = 1* At each wave one infected node will infect exactly one Susceptible node.
- *R > 1* the disease persists by infecting at least one person in each wave.

The value of R * is given by * k x p where k is the number of expected edges that a node have and p is the probability of being infected. The value of R remain constant during the simulation.

We decide to infect only one person at the time *zero* ($i_0 = 0$). As minimum number of waves that a node should stay in the compartment *I* is always *15*, this value is greater enough that a node should infect a reasonable number of the neighbors.

In the following linecharts and animations, the three states are mapped into 3 different colors:

- State *S* is *blue*
- State *I* is *red*
- State *R* is *green*

Karate Club Graph

As already mentioned, the first network used to test the SIR model is the Karate Club graph. The model has been created with the disease transmission probability *p=0.1*, a minimum amount of time steps for the infected individuals $T_i = 10$, the transition probability (from I to R) *q=0.2*, and only 1 infected individual at the beginning, i_0.

The graph below shows the evolution of the fraction of nodes in S, I, R. As can be seen, the number of susceptible individuals decreases rapidly, and then there is an equally rapid growth in the number of infected individuals. The curve of infected individuals, therefore, grows until it reaches a maximum peak, after which the number of infected individuals begins to decrease until it disappears. The moment this decrease begins, the number begins to increase rapidly until it reaches the total number of individuals.

This is because the infected individuals gradually recover from the virus, and thus will no longer be infected. The algorithm ends when all individuals are in the R state, and thus the virus has been eradicated. The graphic and the animation below show how the state of the nodes changes during the execution of the SIR model.
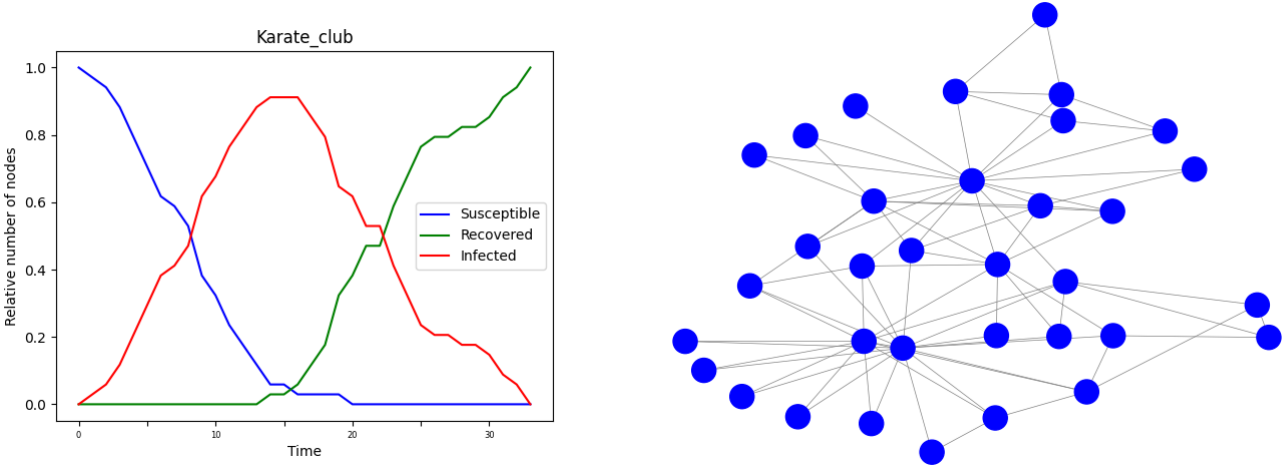

As said before, we simulate the model considering the 3 different values of R, the parameters of the experiments are the following:

R	< K >	p	q	i0	T0

R	< K >	p	q	i0	T0
0.45	4.58	0.015	0.2	1	15
1	4.58	1 / < K >	0.2	1	15
1.376	4.58	0.05	0.2	1	15

- $\mathrm{R} < 1$. This is the best case for an epidemiological point of view: the disease dies out faster because each infected node infects less than one node at each wave. The linecharts below shows that the epidemic reaches the peak (more or less the 90% of the population) after 20 time steps the curve of the infected nodes starts immediately to decrease untill no one is sick anymore, this happens at the time steps 45 because the nodes remains infected for at least 15 steps and also the probability q is very low. In this regime the epidemic is contained.
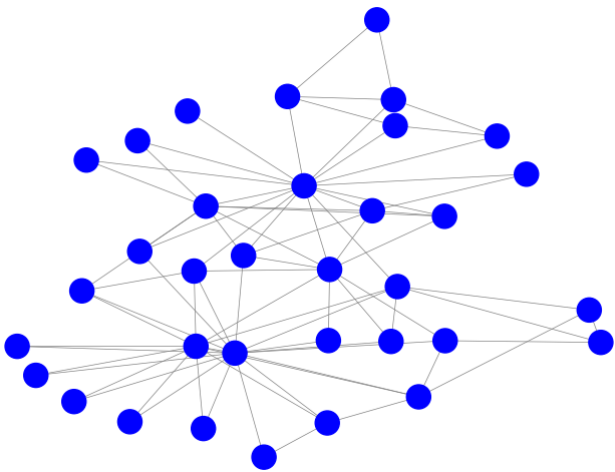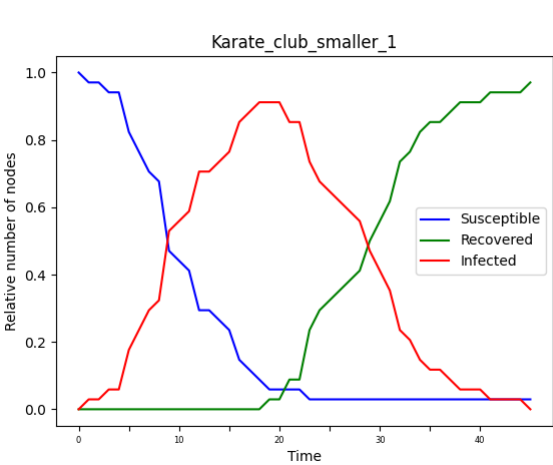
- $\mathrm{R} = 1$. This is the *endemic* regime, where the number of infections encreases faster untill the curve reaches a plateau where all the nodes are infected. With respect to the previous case, the number of total infections remains stable for some time steps and but it requires less steps (35 time steps) to eradicate the virus.
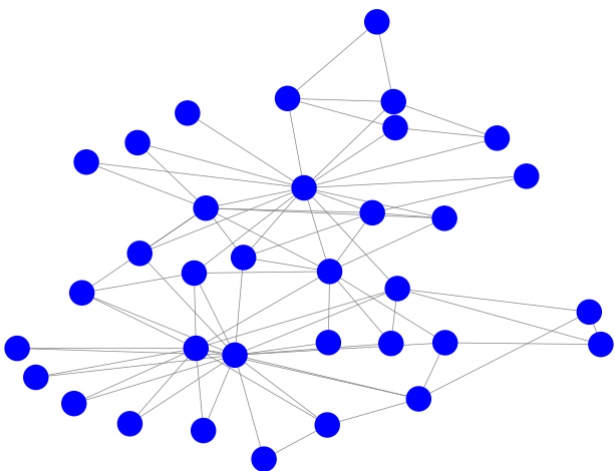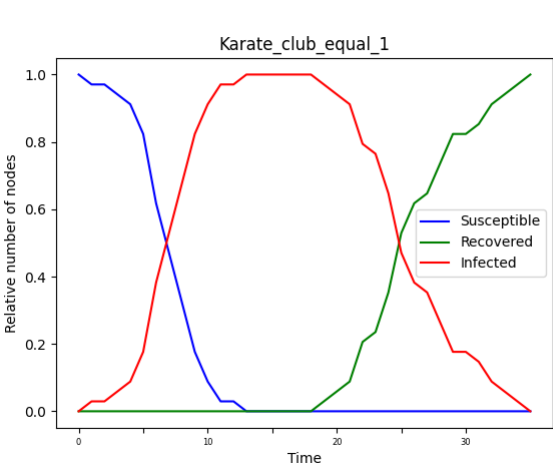
- $\mathrm{R} > 1$. This is the *epidemic* regime, and it is the worst case. Here every one infects more than one people at aech wave. For this reason, we have a exponential growth, rapidly leading to the entire population becoming infected. All the population remain infected for a large amount of time steps, then the curve starts the degrowth rapidly. The simulation ends in 33 time steps.
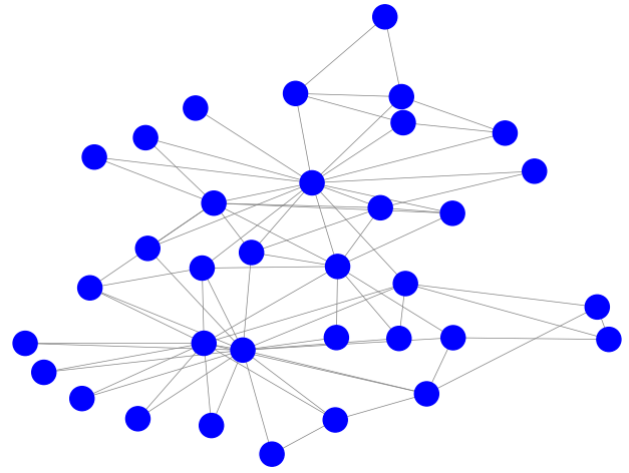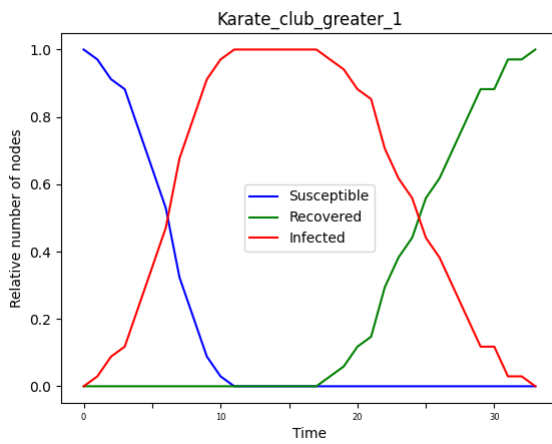
Looking at the three previous charts, the first experiment requires more steps to end the simulation. It is the slowest case, in fact the other cases R = 1 and R > 1, require lesser time steps to reach the convergence criterion, mainly because all the individuals get in the state *R* in less time. So the larger the value of R is, the faster the curve of contagions grows. But this is a very small graph even though it displays many interesting features. In the last assignment we discover that the Karate club graph is disassortative so if one hubs get infected also all its neighbors get to the state *R* because of $\mathrm{T_i}$ is quite big.

Facebook Circles

Here the model has been experimented on the Facebook Circles graph. We performed similar experiments testing the model in the 3 different regimes (R < 1, R = 1, R > 1). The parameters chosen are similar to those used for the Karate Club, only for the parameter *p* is adjusted because of *k* being much bigger. The parameters are reported in the following table:

R	< K >	p	q	i0	T0
0.655	43.69	0.015	0.2	1	15
1	43.69	1 / < K >	0.2	1	15
2.184	43.69	0.05	0.2	1	15

- $\mathrm{R} < 1$.In this case we have a not too fast growth of infections were only the 60% of population is infected. The GIF shows the peripheral communities not getting any case of infections because they are more isolated wrt the core of the network. The epidemic spread is contained but it requires more than 100 steps to dies out.

Facebook_R_smaller_1
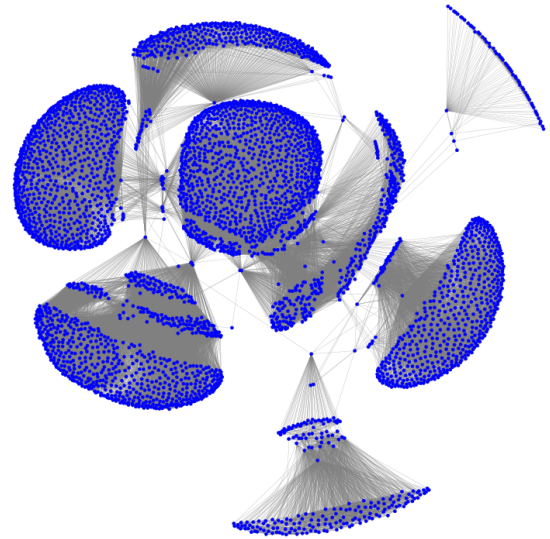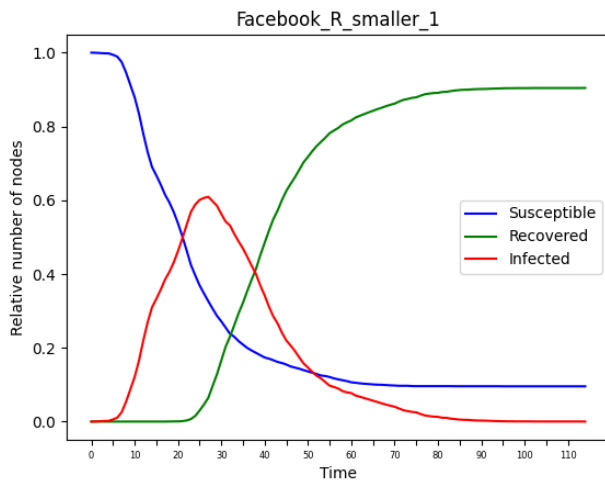
- $\mathbf{R = 1}$ The growth of the contagion curve is a little faster than in the previous case, but still al lot of nodes remain in the state *S* in the end. Only 65% of the population get infected but it requires less steps to stop the infection.

Facebook_R_equal_1
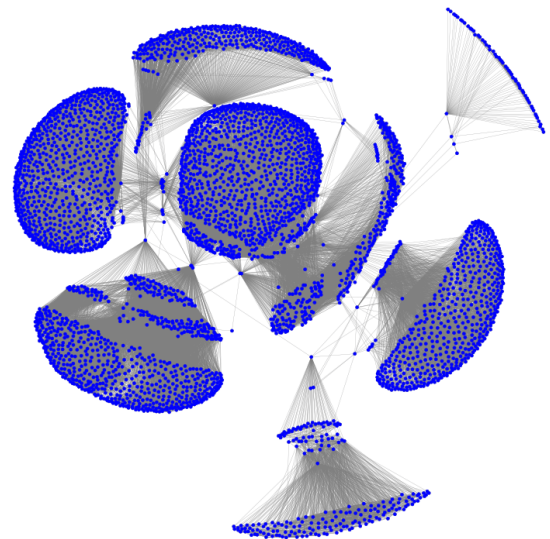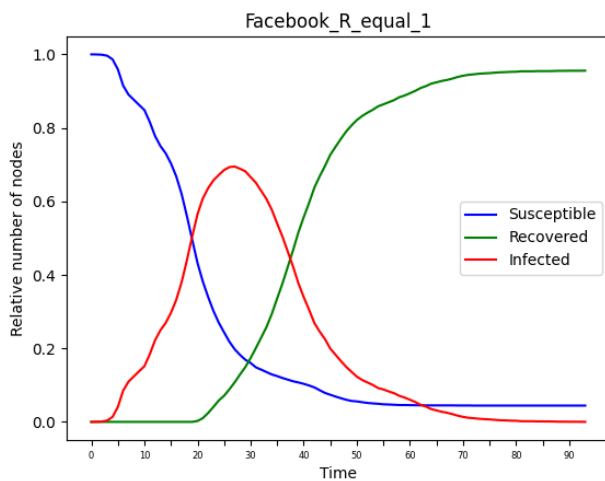
- $\mathbf{R > 1}$ Also here can be observed an exponential growth of the infected individuals, reaching in few time steps, the 80% of the population. But in this case the transition between the state *I* to *R* seems to be as faster as the transition from *S* to *I*. At the end there are still few nodes that are not infected.
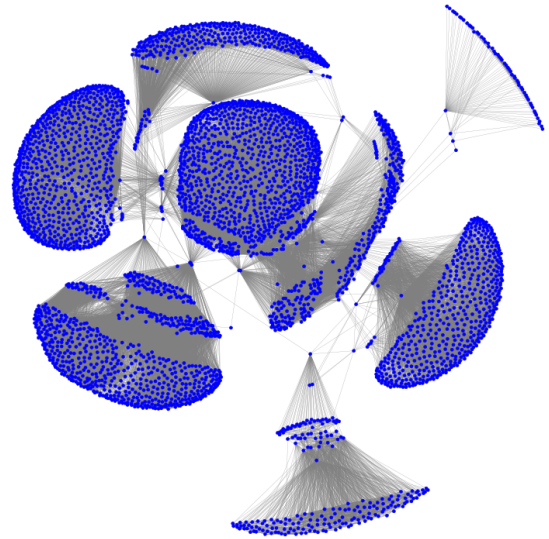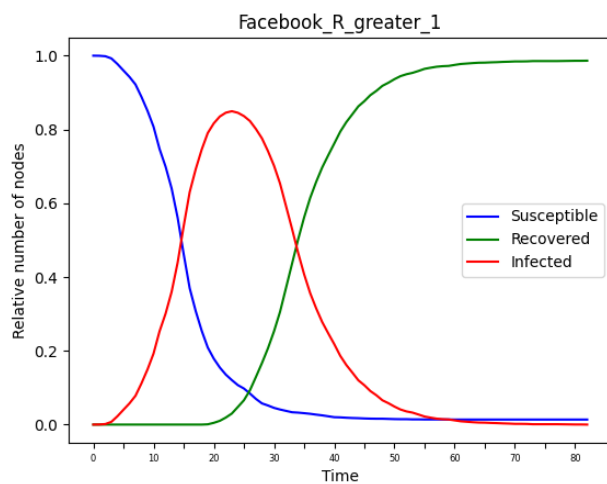
Comparing these experiments with those of Karate club there are major differences that must be highlighted:

- First the curve of the infections are completely different. The model on Karate Club easely the majority (or all) of the nodes in all the cases, and on the other hand the model doesn't infect all the nodes in the Facebook dataset in any case.
- On Facebook in each experiment the infection curves are characterized by a long tail mainly because of the small probability q. But such behaviour is not present on those of Karate club where the simulation ends after few steps after the initial decreases. Another reason for that is the kind of structure of the network, the Facebook dataset seems to be less conductive in certain regions between the communites.
- The experiments on Facebook requires way more times steps to end up, also because of the number of nodes.