Università degli Studi di Padova

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Physics of Data

EX 2 STATISTICAL MECHANICS OF COMPLEX SYSTEM

OF

MARCO AGNOLON, TOMMASO TABARELLI

Anno accademico 2018-2019

Task 1

We are provided two matrices of contact between individuals, which represent two different contact networks. We want to study the spreading of epidemics in the networks. The epidemic dynamic is the SIS one. This model is made of two type of individuals: the susceptible (S) and the infected/infectious (I). The names explain themselves the characteristics of these sub-populations; after contracting the disease, becoming infectious and recovering one may become again infected. The only way for the disease to spread is exploiting the contact between an infected individual and a susceptible one; this occurs with rate λ. On the other hand when one node has been infected, it recovers with rate μ: this is a spontaneous process that occurs independently with respect to the possible contacts one has with other individuals.
From these information, we want to obtain the critical thresholds λ_C of an epidemic outbreak, using different types of mean-field approximations.

Homogeneous The first is the homogeneous one that considers all individuals equivalent and with the same number of contact $< k >$ (neglecting spatial structure of the network). This approximation leads to a critical threshold of:

$$\lambda_C = \frac{\mu}{< k >}.$$

Heterogeneous In this approximation we relax the fixed number of contact constraint in favor of a weaker one that fixes the distribution of the nodes degrees of the network. This returns us:

$$\lambda_C = \frac{\mu < k >}{< k^2 >}.$$

Quenched The last approximation takes into account the structure of the network and it gives us:

$$\lambda_C = \frac{\mu}{\Lambda_{max}(A)},$$

where $\Lambda_{max}(A)$ is the largest eigenvalue of the adjacency matrix A.

The results are shown in the Table 1.

Table 1: Models results

λ_C	Net 1	Net 2
Homogeneous	0.097	0.139
Heterogeneous	0.082	0.076
Quenched	0.080	0.066

Task 2

Here we want to simulate the behavior of the disease in the networks. Firstly, we set the couple of parameters λ and μ and we simulate the spreading for several λ parameters starting from different initial conditions: a single infected individual, 10% of population infected and 50% of population infected. The simulation is based on a Gillespie-like algorithm. At each iteration of it one reaction occurs; the length of the intervals between two occurrence times are exponentially distributed with total rate equal to the so called "total propensity rate", which is given by $a_0 = a_\mu + a_\lambda$, that in its turns is evaluated summing the two different propensity rates for each reaction. So we sampled the interval times from an exponential distribution, in order to be sure to have an event at each step.
Let us now explain how to evaluate the two considered propensity rates. The recovering reaction is spontaneous and so its propensity rate depends only on the number of possible individuals that can get heal and on the healing rate (referred to a single individual). Thus the result is: $a_\mu = \mu I$.
The propensity rate of infecting reaction is closely related to the possibility of a susceptible individual to come across to an infected one. This is in turn strongly linked to the shape of the network and specially to its connections. Indeed, we calculate it basing on the edges: we count all edges that link a susceptible individual to an infected one. Therefore this is the number of possible times that this reaction may occur; multiplying it by the rate λ we obtain the propensity rate for the infection process a_λ.
To select which reaction takes place, we split the interval $[0, a_0]$ in sections proportionals to the propensity rates (for example, in our case the interval was divided into 2 parts: $[0, a_\lambda]$ and $[a_\lambda, a_0]$). Then we draw a unifororm random number in $[0, a_0]$ and basing on where it was in the aforementioned interval we selected the respective reaction (for example, if it was less then a_λ then we selected the infection process).
We repeat this procedure till the fraction of infected individuals reached a stationary number (see an example in figure 1). To be sure that this happens, we verified 2 different conditions:

1. at each step we check if the difference between the averages of the last 2000 iterations before and after a single step is lower than a fixed (and arbitrary) threshold (stationary of the first order); if so, we decided to save the average fraction of infected individuals in the last 100 events.

2. we stop the procedure after an arbitrary number of 5000 iteration and keep the average of the last 100 iterations.

We started using the first condition but we realized the algorithm was a kind of slow, so we decided to have a direct look to a plot *I vs t* to see how many steps it would took to converge. We saw they were about 2000, so we decided to let the simulations go untill 5000 iterations and then take as before the average on the last 100 individuals. In this way the algorithm worked much faster.

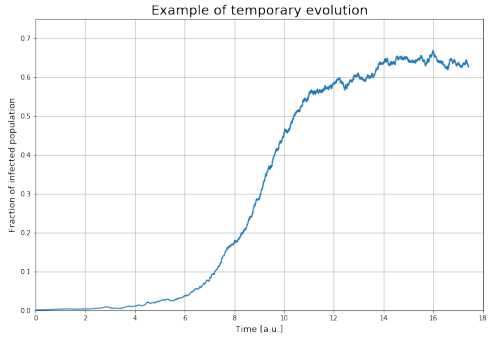

Figure 1: Example of time evolution of the system G1 with 1 starting infected and $\lambda = 0.3$.

Moreover, if the fraction of infected became 0 in the first 5000 steps we decided to save '0' instead of the average value of last 100 steps for 2 reasons:

- first because in the phase space plot it has to be plotted the fraction of infected individuals in the stationary limit, so $t \to \infty$; if the disease goes to extinction there would be no infected for large t;
- second because if we kept a number different from 0 in those calculations then averaging over all 100 simulations for a single λ often resulted in a value different from zero but quite close to it ($O(10^1)$), making harder the evaluation of λ_C.

We made an arbitrary grid for the λs to try to visit a meaningful part of the phase space.
We also zoom the interesting part of each graph near the critical value of λ. We run 100 simulations for every λ and we make the averages over it. Then we repeat it for three different initial conditions (1 infected, 10% of initial population infected and 50% of initial population infected) and for both networks. We also plot the standard deviation for each λ.

One can see that changing the initials conditions affects the stability of the algorithm: increasing the number of infected individuals at the beginning returns a better defined curve, because there are less simulations that go to zero (because of the randomness of the process some infections that starts from 1 single individual can get cured before they start to spread, so they can extinguish at the very beginning of the process). Indeed also the error bars are much smaller as initial infected population becomes bigger. This allows us to predict better the value of λ_C, using the zoomed phase space, that is around 0.09 for both networks, and it quite agrees the theoretical expectations. Nonetheless we expected that it should have matched the quenched approximation.

Task 3

In this section we try to understand the results in light of the connectivity properties of the networks.
In each histogram we use the Freedman-Diaconis's Rule

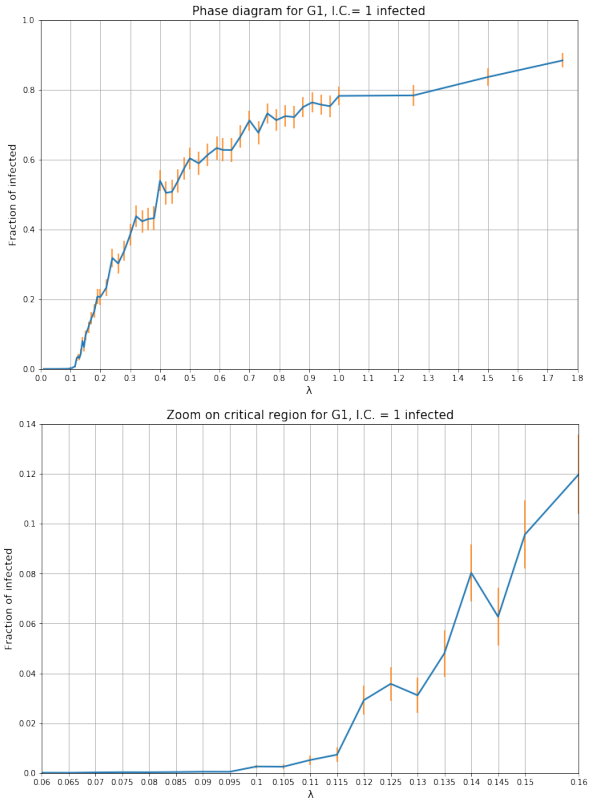

Figure 2: Network 1 phase space diagrams with 1 infected at the beginning

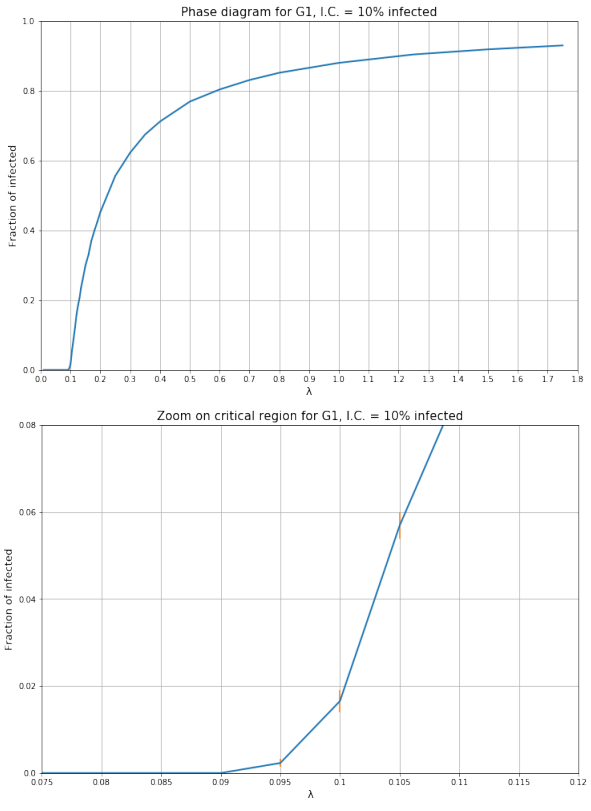

Figure 3: Network 1 phase space diagrams with 10% infected at the beginning

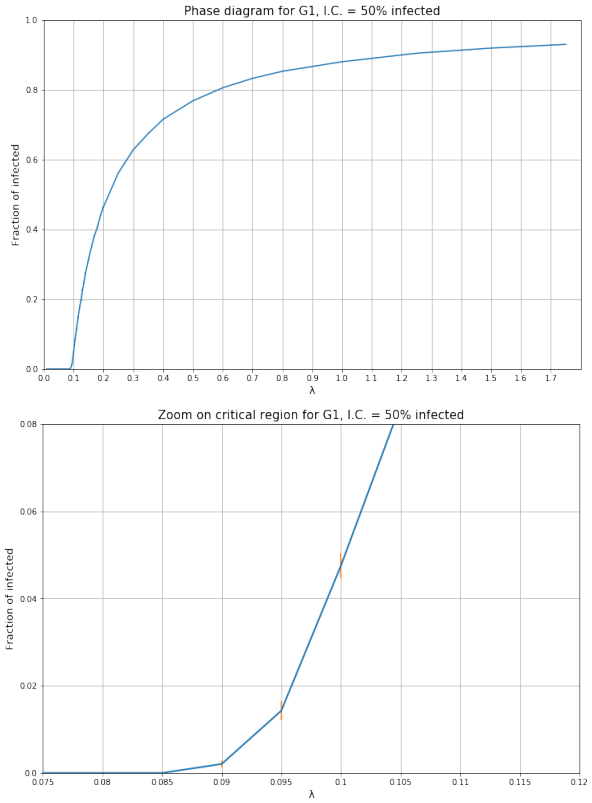

Figure 4: Network 1 phase space diagrams with 50% *infected at the beginning*

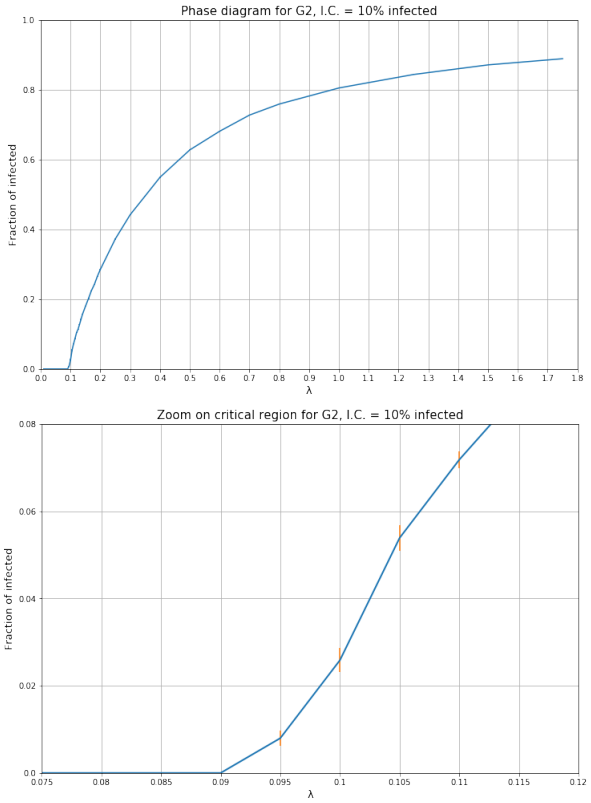

Figure 6: Network 2 phase space diagrams with 10% *infected at the beginning*

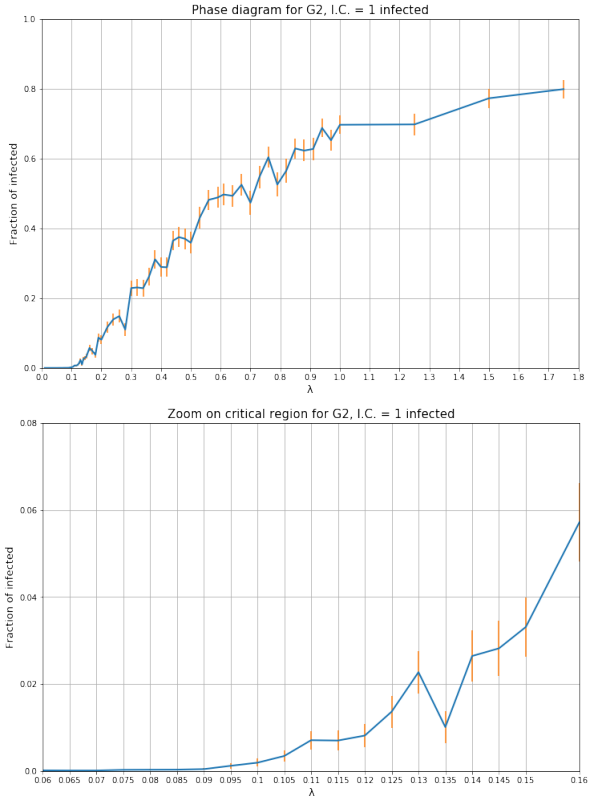

Figure 5: Network 2 phase space diagrams with 1 infected at the beginning

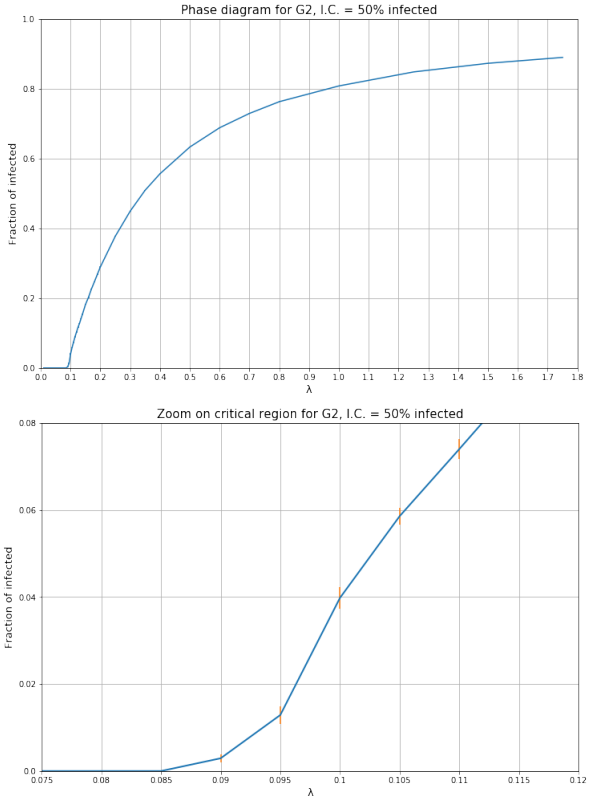

Figure 7: Network 1 phase space diagrams with 50% *infected at the beginning*

to choose the number of bins.
We start comparing the degree distribution of the two (see Figure 8):

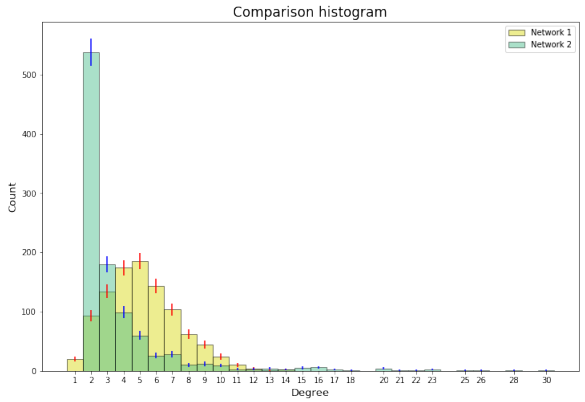

Figure 8: Degree distributions of the two networks

The second quantity we compare is the Page-rank distribution (see Figure 9):

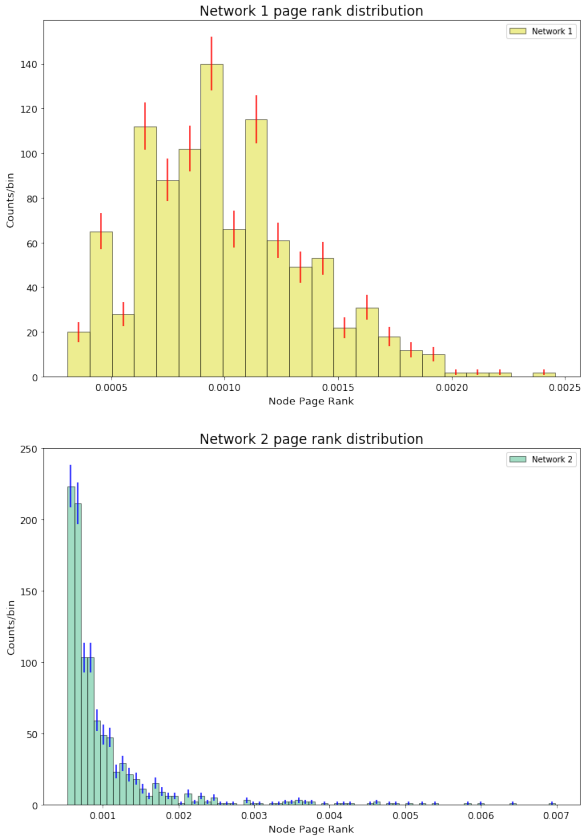

Figure 9: Page-rank distributions of the two networks

And finally we compare the clustering coefficients:

- Clustering coefficient network 1: 0.039
- Clustering coefficient network 2: 0.011

As one can recognize from the graphs the two networks are quite different in term of connectivity. The first one presents a greater number of nodes with higher degree, with respect to the second network where the majority of the nodes have a small degree and just few of them (hubs)

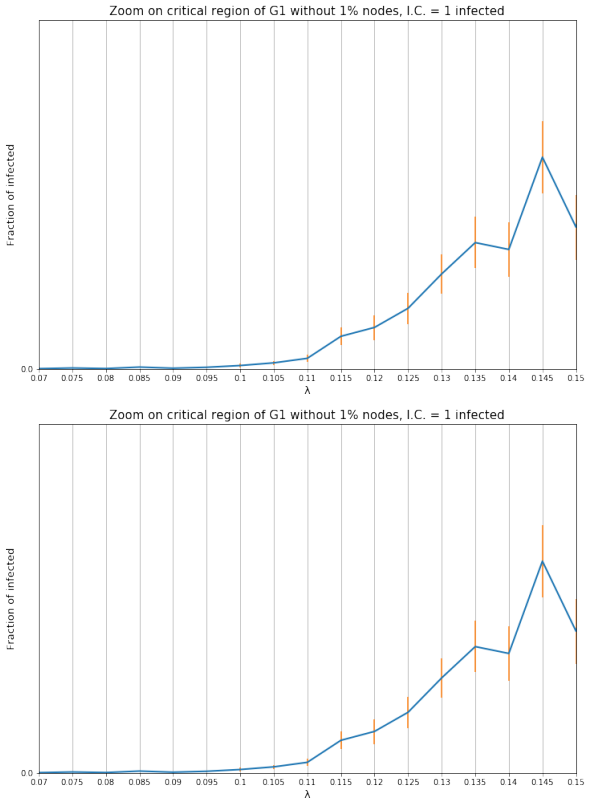

Figure 10: Network 1 phase space diagrams with 1 infected at the beginning and without 1% of nodes

present a high one. This is also reflected by the Page-rank distribution which histogram presents a similar shape. Also the clustering coefficient are quite different between each other. It seems that the first network shows a random distribution, instead the second seems to be a scale-free one.

Task 4

The results of the task 3 may lead us to think what will happen if we remove some nodes from the two networks. The most simpler choice is removing the hubs of the network, i.e. that nodes which governs the flow of the disease. We choose to remove those 1% with highest Page-rank distribution, because for all our purposes what we will consider and what we have already considered is the stationary case, and the Page-rank will give us the probability to end up in a node in this situation. Moreover it allows random jumping in order to escape from possible loops that may take place in the networks, hence it gives us a more truthful results.

After having removed all necessary nodes we notice a different behavior among the two networks. The first one seems not to be affected by removing of nodes, because it shows the same critical threshold. The second one, instead, behaves completely different with respect what it used to. The critical threshold is shifted very far, near 0.25. This great changing denotes a different robustness of the two graphs after node removing. This is reflected by the

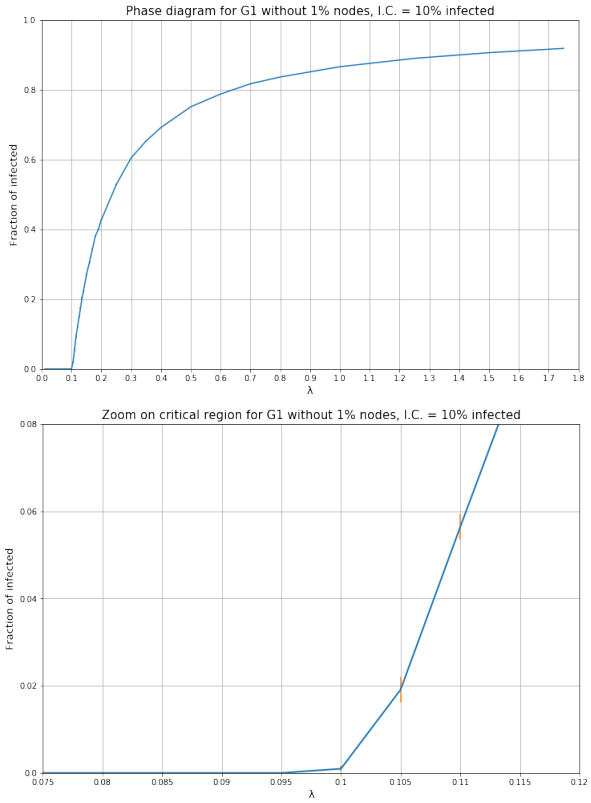

Figure 11: Network 1 phase space diagrams with 10% *infected at the beginning and without 1% of nodes*

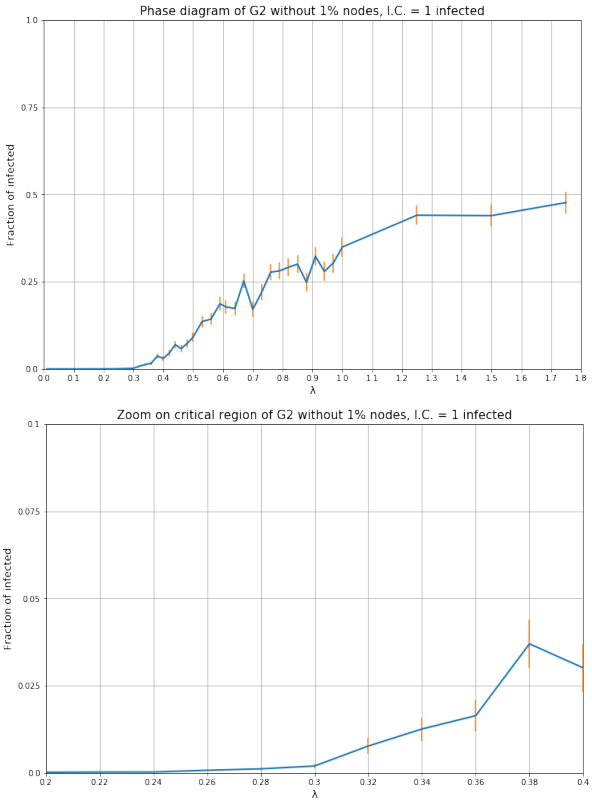

Figure 13: Network 2 phase space diagrams with 1 infected at the beginning and without 1% of nodes

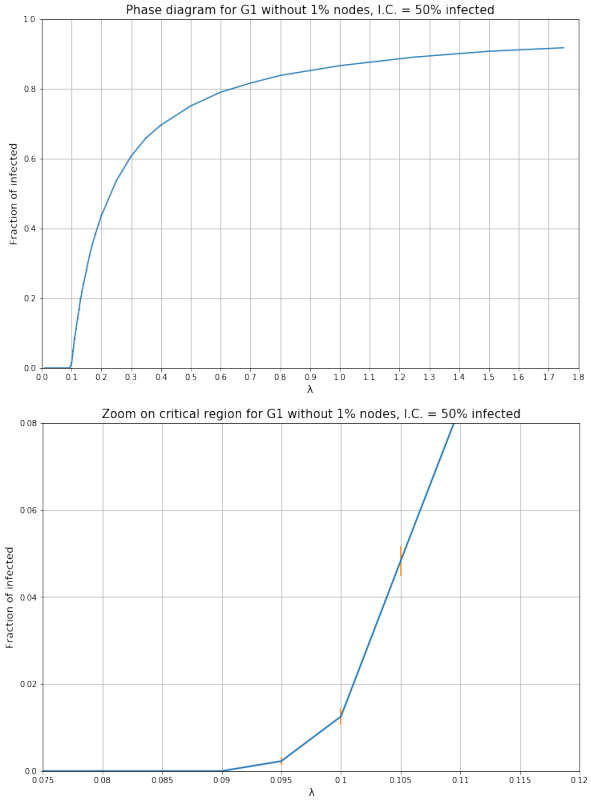

Figure 12: Network 1 phase space diagrams with 50% *infected at the beginning and without 1% of nodes*

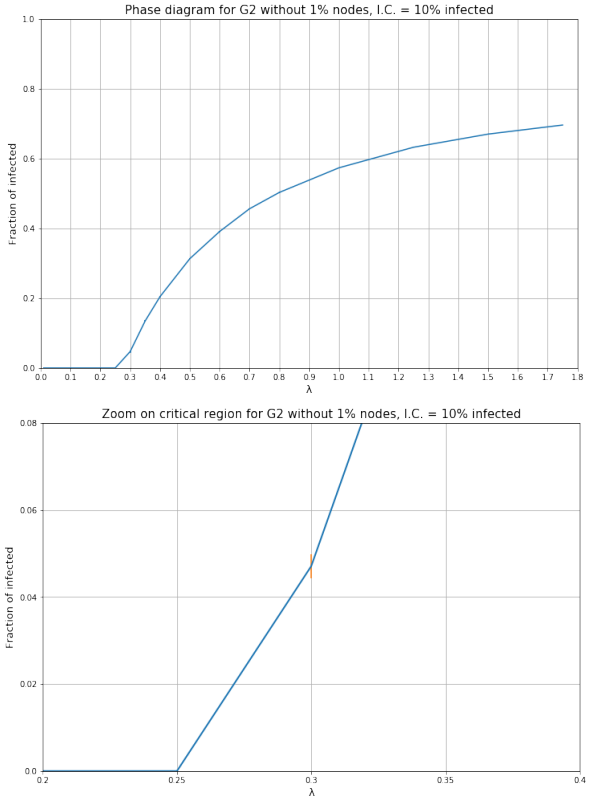

Figure 14: Network 2 phase space diagrams with 10% *infected at the beginning and without 1% of nodes*

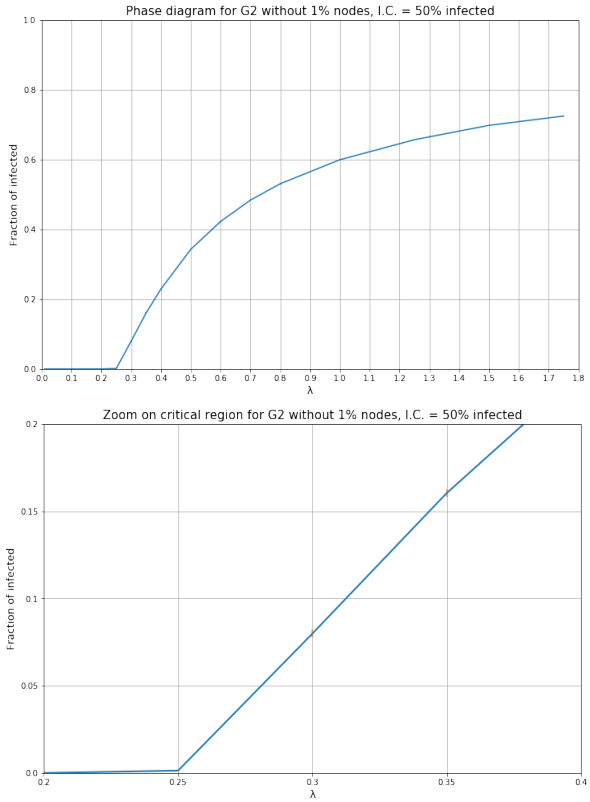

Figure 15: Network 2 phase space diagrams with 50% infected at the beginning and without 1% of nodes

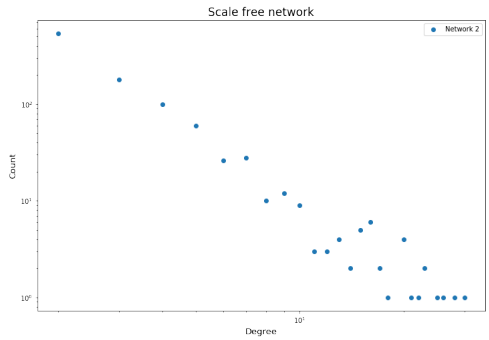

Figure 16: log-distribution of node degrees

fact that the first network has a random graph-like degree distribution, instead the second one seems to present a typical scale free behavior (as we confirm in Figure 16) that is known to have less robustness against target attack.

Task 5

We want to analyze a SIR infection, i.e. a disease where after being infected from another individual one could die or recover. Anyhow it enters in a state where it is no longer involved in the disease spreading. We are going to plot the fraction of susceptible, infected and recovered (or died) individuals. We just plot the early stages of the spreading for 100 locations, here we show only on example of those (see example for location 43 of 100 in the left part of Figure 17).
We treat all the 100 configurations as different samples in order to obtain a relevant statistic for the early stages. Our task is to find the value of the basic reproduction number R_0: the average number of secondary infections caused by an infected host for this disease. We know that if this value is greater than 1 the disease will have an outbreak.
We can find it in each plot solving the differential equation and finding β and γ, where $s(t)$, $x(t)$ and $r(t)$ are respectively the number of susceptible, infected and recovered individuals at time t:

$$\beta = -\frac{s(t)-s(0)}{\int_0^t s(\tau)-x(\tau)d\tau} \quad \gamma = \frac{r(t)}{\int_0^t x(\tau)d\tau}$$

the one easily get:

$$R_0 = \frac{\beta}{\gamma}$$

So we handle this for each location and we plot the evolution of R_0 over time (see right part of Figure 17 for location 43).

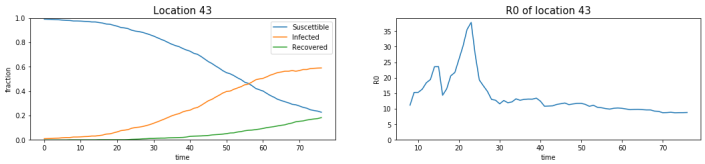

Figure 17: Early stage for 43th location

In the example shown we may see that the value of R_0 after some initial oscillation (transient part), it start to converge to its proper value. We are interested in find it for each location. The reader could find the results for all the other locations in the notebook.
Finally we make an histogram (Figure 18) where we show the distribution of the R_0 values.

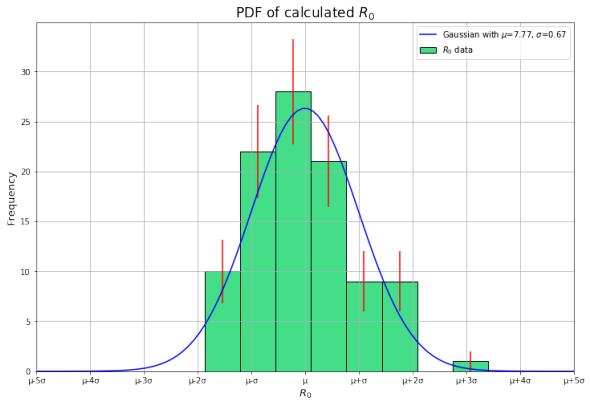

Figure 18: Distribution of R_0

The distribution follows approximately a gaussian with average $\mu = 7.77$ and standard deviation $\sigma = 0.67$.
Therefore, this give us a estimation for the R_0 value that is pretty high, and we can say that the outbreak almost certainly occurs.