Exercise 2 - Determining the spreading of epidemic

Anna Braghetto, Davide Maniscalco

Task 1

The two given datasets contain two different undirected mobility networks among individuals that are analysed in term of epidemic spreading.

The two networks are made up of 1000 nodes each; their adjacency matrices are shown in Fig.1 .

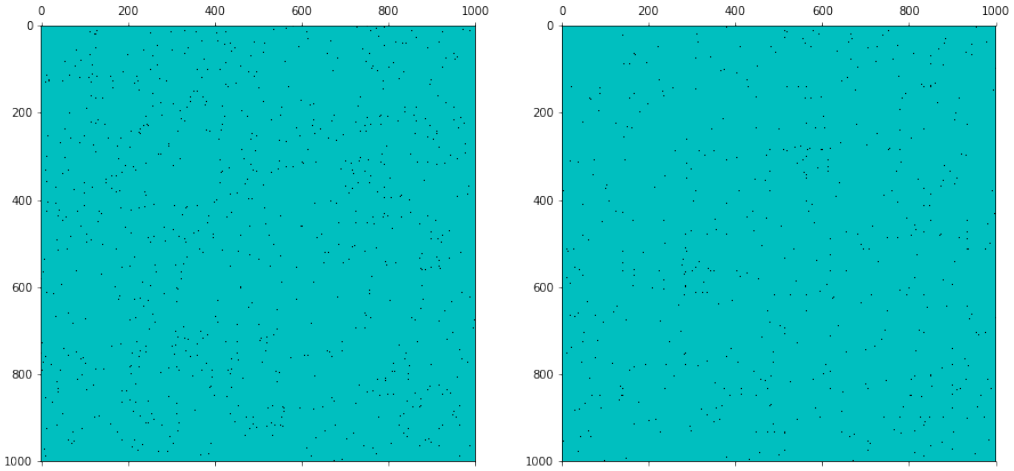

Figure 1: Adjacency matrix of Net1 (on the left) and Net2 (on the right). Dark blue points indicate connected nodes ($a_{ij} = 1$) while the greens refer to unconnected nodes ($a_{ij} = 0$).

At first, we analysed the behavoiur of the networks in the SIS model, which is ruled by two processes, the recovery (with rate μ) and the infection (with rate λ):

$$I \xrightarrow{\mu} S \qquad\qquad S + I \xrightarrow{\lambda} I + I$$

where I and S are an infected and a susceptible node respectevely. In order to study the spread of epidemic we compute the theoretical values for the epidemic thresholds in the SIS dynamics in the homogenous, heterogenous and quenched mean-field approximation.

$$\lambda_c^{MF} = \frac{\mu}{\langle k \rangle} \qquad \lambda_c^{HMF} = \frac{\mu \langle k \rangle}{\langle k^2 \rangle} \qquad \lambda_c^{QMF} = \frac{\mu}{\Lambda_{max}(A)} \, .$$

The rate μ is set to 0.5 for both network in all the approximations; the results are shown in Tab.1.

For Net1 the three approximations show similar values, while for Net2 the mean field result differs a bit, probably it is due to its degree distributions, which has a higher variance ($\simeq 10.768$), making mean field inadequate.

	MF	HMF	QMF
Network 1	0.097163	0.081605	0.079566
Network 2	0.139198	0.075877	0.066400

Table 1: Critical lambdas λ_C for the two networks in three different approximations

Task 2

The theoretical values are then compared with the results obtained through a stochastic process; by using the Gillespie algorithm we simulate the spread of the epidemic and we build the phase diagram, finding the critical lambdas λ_C as the point at which there is the outbreak of the epidemic. To build the phase diagram, we let λ vary in the interval [0.04, 1] and we fix μ to 0.5. For each λ, we intialized the system selecting at random 50 infected sites, and making the system evolve with 1000 iterations of the Gillespie algorithm. At each iteration, in order to select the reaction ($S \to I$ or $S + I \to I + I$), we compute the propensity rates as

$$a_1(I) = \mu I \qquad\qquad a_2(S, I) = \lambda \frac{SI}{N} c$$

where I is the number of infected individuals, S the number of supsceptible individuals, N the total population, and $c = \langle k \rangle / (N - 1)$ is the connectivity of the network. To avoid fluctuations, at the end of the 1000 iterations the simulation is repeated; the fraction of infected is computed as the average over 100 values for each λ.

To choose the number of iterations, we check the convergence by running the simulation at increasing number of iterations at $\lambda = 0.04$, at which the fraction of infected goes to 0 quickly; the results for the two networks are shown in Fig.2 and Fig.3. To be sure that no fluctations occur, we set the number of iterations to 1000, which is higher than the minima needed to go to 0.

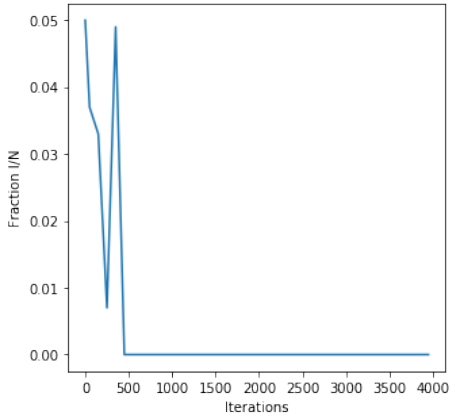

Figure 2: Fraction of infected as a function of the number of iterations of the algorithm for network 1, at $\lambda = 0.04$

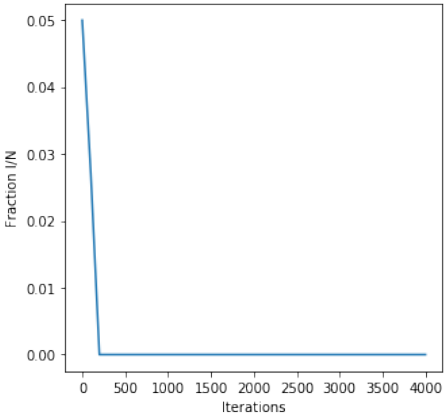

Figure 3: Fraction of infected as a function of the number of iterations of the algorithm for network 2, at $\lambda = 0.04$

The phase diagrams resulting from the simulations are shown in Fig.4 and in Fig.5.

In order to have a better estimation of the critical point, we make a further simulation with 10 equidistant λ between the values at which the number of infected passes from $I = 0$ to $I > 0$, for each network. The phase diagrams around the critical point are shown in Fig.6 and Fig.7.

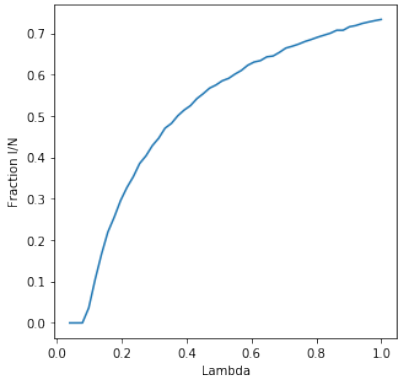

Figure 4: Phase diagram for network 1

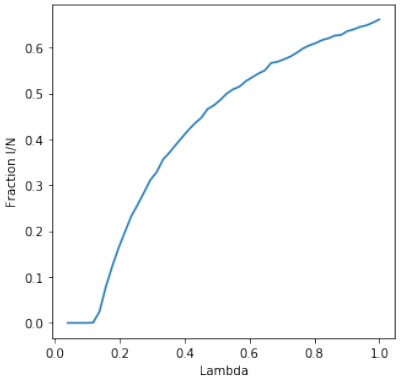

Figure 5: Phase diagram for network 2

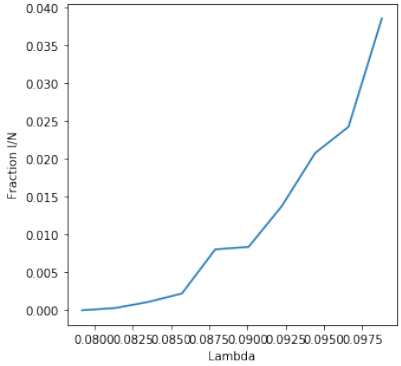

Figure 6: Phase diagram around critical point for network 1.

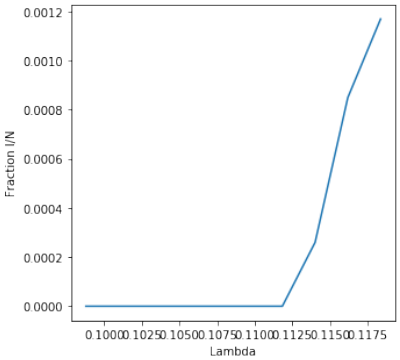

Figure 7: Phase diagram around critical point for network 2.

We take the last λ at which $I = 0$ in these simulations as the extimation of the critical points of the two networks:

$$\lambda_1^c \simeq 0.079 \qquad \lambda_2^c \simeq 0.112$$

Task 3

The behavoiur of the networks is related to their structure. The degree, page-rank and clustering distribution are exposed in Fig. 8, 9 and 10 respectevily.

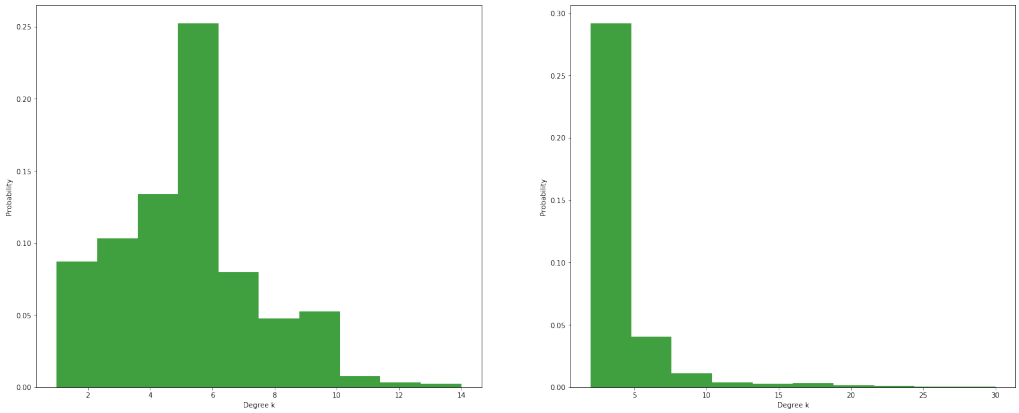

Figure 8: Degree distribution for the first (on the left) and for the second (on the right) network.

The degree of the first network shows a normal distribution with mean $\langle k \rangle = 5.146$, while

in the second network, the degree has a decreasing exponential distribution with mean $\langle k \rangle = 3.592$; Net1 is more connected than Net2, where the average degree is lower.

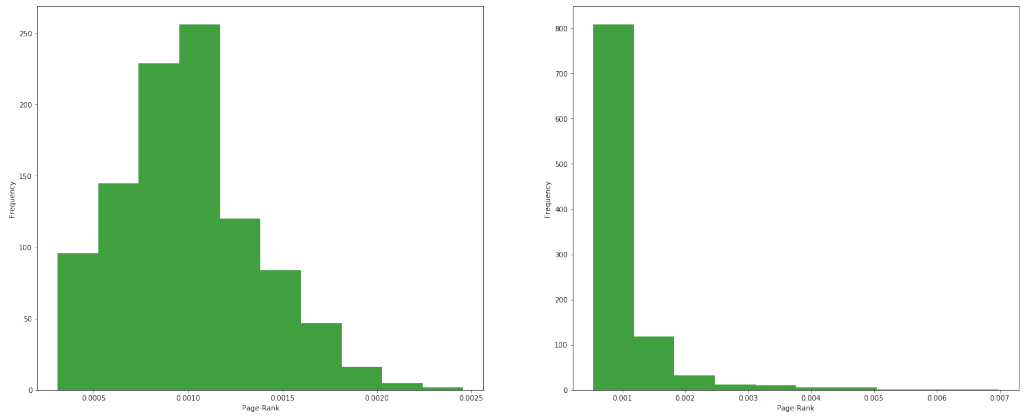

Figure 9: Page-Rank distribution for the first (on the left) and for the second (on the right) network.

The page-rank distribution is quite similar to the degree one. It shows a normal distribution in the first network and a decreasing exponentail distribution in the second network. While the degree distribution is centerd in the different values for different networks, the page-rank is centred in 0.001 for both Net1 and Net2.

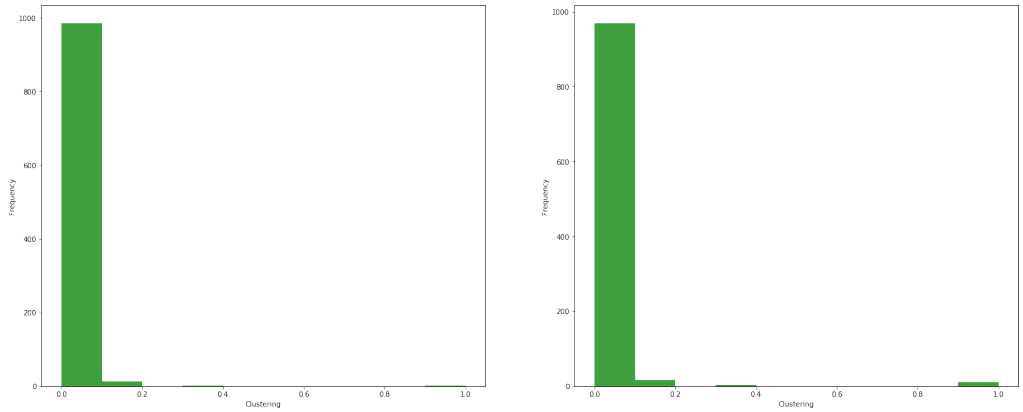

Figure 10: Clustering distribution for the first (on the left) and for the second (on the right) network.

The clustering distribution has a similar shape for both networks but different mean; for the first network $\langle C \rangle = 0.005$ while for the second $\langle C \rangle = 0.02$.

The different structure of the networks, such as the number of connections, are reflected on the fact that the epidemic spread faster in network 1 and λ_C^1 is greater than λ_C^2. In fact, if the a network is well connected, we expect that the epidemics is more likely to happen and, if the outbreak occurs, it spreads faster.

Task 4

While the recovery process doesn't involve interaction between individuals, infection does, and individuals with higher degree make the epidemic spread easier. For this reason, we decided to remove from both networks the 10 individuals with the highest degree; the adjacency matrices for the new networks are shown in Fig.11

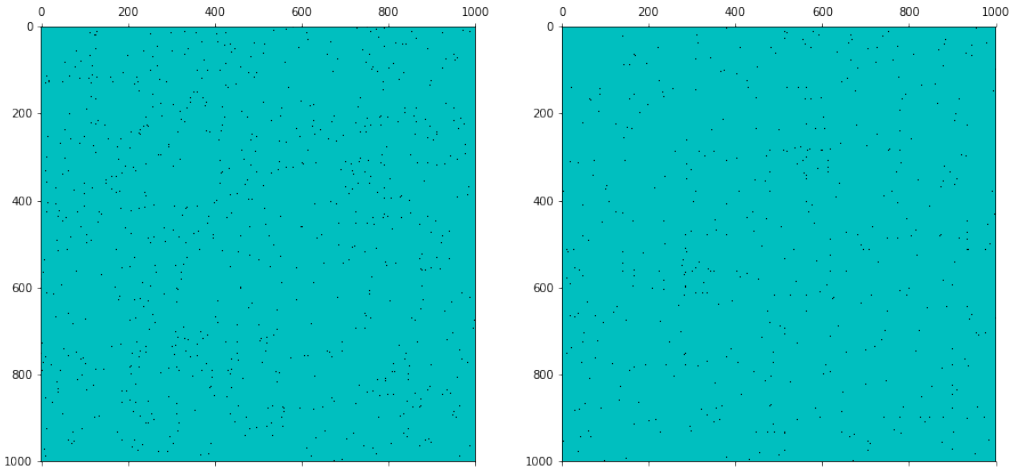

Figure 11: Adjacency matrix for the first (on the left) and for the second (on the right) network after the removal of the 10 highest-degree individuals.

We run again the simulation for these new networks, obtaining the phase diagrams shown in Fig.12 and in Fig.13.

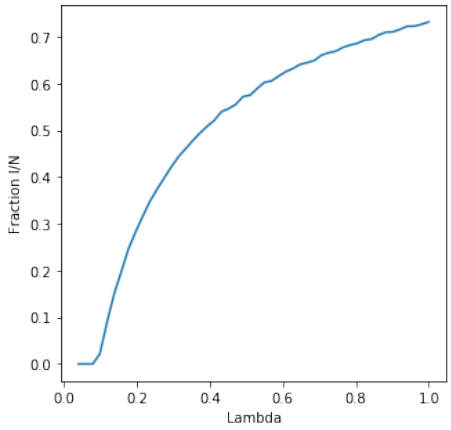

Figure 12: Phase diagram of network 1'

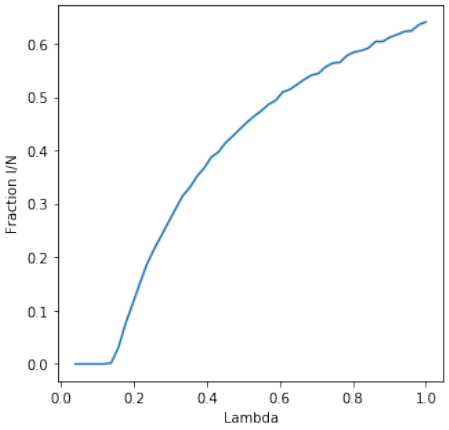

Figure 13: Phase diagram of network 2'

As before, for each network we make a further simulation with ten trial λ around the critical points found in the phase diagrams shown in Fig.12 and in Fig.13.

We then take the last λ at which $I = 0$ as the extimation of the critical point for the two networks:

$$\lambda^c_{1'} \simeq 0.081 \qquad\qquad \lambda^c_{2'} \simeq 0.118$$

As expected, new critical points are bigger than previous ones, but the difference is very small: there is not a qualitative change. The networks are homogeneous enough and not fragile to a target attack.

We also simulated the processes in the two networks obtained by removing the nodes with higher bewteenness centrality, but the results are very similar to the ones computed by removing the nodes with higher degree.

Task 5

The dataset contains the early stages of an outbreak of Chicken Pox in 100 different locations in the SIR model, ruled by

$$S + I \xrightarrow{\beta} I + I \qquad\qquad I \xrightarrow{\gamma} R$$

where β and γ are the infection and recovery rate respectively. The fraction of infected, susceptible and recovered individuals is given; in Fig.14, these values are shown as function of time.

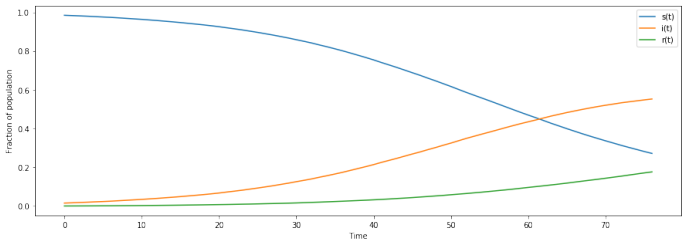

Figure 14: Fraction of infected, subsceptible, and recovered individuals.

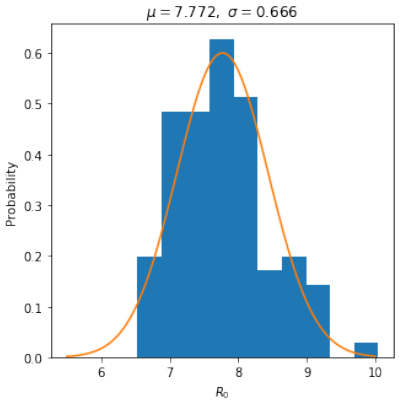

Figure 15: Histogram and gaussian fit for the R_0

The two rates β and γ are computed from the equations of the SIR model, as follows:

$$\beta = -\frac{s(t) - s(0)}{\int_0^t s(\tau)\, i(\tau) d\tau} \qquad\qquad \gamma = \frac{r(t)}{\int_0^t i(\tau) d\tau}$$

where s, r, i are respectively the fraction of supsceptible, recovered and infected individuals. For each of the 100 locations, we can estimate these parameters from the data with the following discrete approximations:

$$\beta = -\frac{s(T) - s(0)}{\sum_{t=0}^{76} s(t) i(t)} \qquad\qquad \gamma = \frac{r(T)}{\sum_{t=0}^{76} i(t)}\,.$$

where T is the final time set to 76 days and the sums run over all the days that we have.

For each location we calculate the basic reproduction number $R_0 = \beta/\gamma$, that indicates if the outbreak becomes viral ($R_0 > 1$) or not ($R_0 < 1$). In Fig.15 is shown the distribution of the basic reproduction number and its gaussian fit, with the parameters $\mu = 7.772$ and $\sigma = 0.666$ calculated from the data.

Since $R_0 = 7.772 \pm 0.666 > 1$ we predict that the outbreak will become viral.