Statistical Mechanics of Complex Systems

Determining the Spread of Epidemics

Alberto Chimenti - 1210360

1 Network Overview

The data provided reports the nodes couples corresponding to the edges of two different networks.

The data have been loaded using `loadtxt` function, therefore the data was mainly handled using `numpy` arrays.

We can see in the following figure the ajacency matrix corresponding to each unweighted and undirected network.

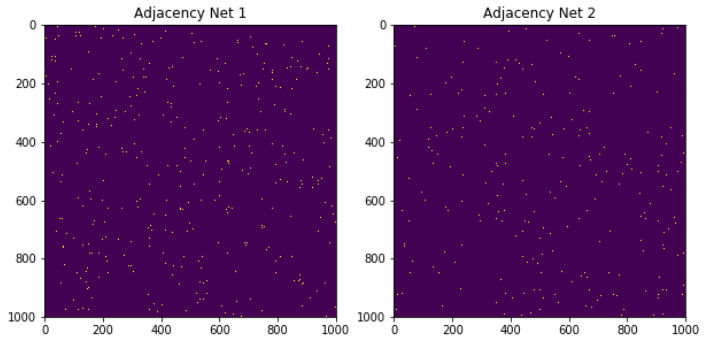

Figure 1: Adjacency matrices (Left: Network 1; Right: Network 2)

The adjacency matrices shown were used to build connected graph models to investigate epidemics dynamics.

2 SIS Models

2.1 Homogeneous SIS

In order to build a rather simple analitically solvable SIS model, one has to impose the assumption that the "populations" are *Well-Mixed* (which is equivalent to a *Mean-Field* approximation). The latter can be expressed with the following requirements:

- All individuals have the same number of contacts $\langle k \rangle$
- All individuals are equivalent
- Closed population (no migration, births or deaths).

The SIS model can be schematized like: Therefore, the model

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

has to take into account these two "events", where μ and λ are the *recovery rate* and the *infection rate*, respectively.

Therefore, the **Homogeneus SIS** model can be written as:

$$\frac{di}{dt} = \lambda \langle k \rangle si - \mu i \quad \Rightarrow \quad i(t) \simeq i(0) e^{\left(\lambda \langle k \rangle - \mu\right)t}$$

In the previous solution one has to impose that $s = \frac{S}{N} \simeq 1$. And the threshold value for the infection rate is:

$$\lambda_c = \frac{\mu}{\langle k \rangle} \qquad \textit{homogeneous SIS} \tag{1}$$

Obviously one has to take into account that in the *outbreak case*, for a closed system, the fraction of the ill individuals i has to saturate even if not predicted by the solution found.

2.2 Heterogenous SIS

Now we drop the first assumption and consider a system in which all the individuals with the same number of links have the same properties. We obtain in this case a stationary solution:

$$\frac{di_k}{dt} = \lambda k (1 - i_k)\Theta_k(t) - \mu i_k \quad \Rightarrow \quad i_k = \frac{k\lambda\Theta_k}{\mu + k\lambda\Theta_k}$$

$\Theta_k = \sum_{k'} P(k'|k) i_{k'}$ is the probability that a node of degree k has an infected neighbor.

In case of *uncorrelated networks* one finds the following non-trivial solution:

$$\Theta = f(\Theta) = \frac{1}{\langle k \rangle} \sum_k \frac{k^2 P(k) \lambda \Theta}{\mu + k\lambda\Theta} \quad \Rightarrow \quad \frac{df}{d\Theta}\Big|_{\Theta=0} = \frac{\lambda}{\mu} \frac{\langle k^2 \rangle}{\langle k \rangle}$$

And we have a non trivial solution for $\frac{df}{d\Theta} = 1$, obtaining:

$$\lambda_c = \frac{\mu \langle k^2 \rangle}{\langle k \rangle} \qquad \textit{heterogeneous SIS} \tag{2}$$

Note that for *well-mixed networks* ($\langle k^2 \rangle = \langle k \rangle^2$) therefore we obtain the solution in (1).

2.3 Quenched SIS

We drop the second assumption too this time, considering individual probabilities for each specific node $p_i(t)$ ="*probability that node i is infected at time t*". So we end up with the equation:

$$p_i(t+1) = p_i(t)(1-\mu) + (1 - p_i(t)) q_i(t)$$

with

$$q_i(t) = 1 - \prod_{j=1}^{N} [1 - \lambda A_{ij} p_j(t)]$$

Supposing that the system reaches a stable regime where $lim_{t\to\infty} p_i(t) = \epsilon_i^* << 1$ we have:

$$\epsilon_i^* = \frac{\lambda}{\mu} \sum_{j=1}^{N} A_{ij} \epsilon_i^*$$

We can now observe that the latter equation tells us that $\frac{\mu}{\lambda}$ is an eigenvalue of the adjacency matrix A.

In the end we can infer that:

$$\lambda_c = \frac{\mu}{\Lambda_{max}(A)} \qquad \textit{quenched SIS} \tag{3}$$

2.4 Results for our networks

The previous models were used to compute some reference values for the critical rate in our networks (see `code`).

	Homogeneous	Heterogeneous	Quenched
Network 1	0.0971628	0.0816048	0.0795659
Network 2	0.1391982	0.0758766	0.0663996

3 SIS Model: Simulation and Phase Diagram

A simulation of the epidemics spread was performed using Gillespie algorithm in order to enlighten the critical behaviour of the system. The *healing rate* $\mu = 0.5$ was chosen to be equal to the previous model predictions in order to confront the results.

3.1 Gillespie Algorithm

The implementation of the algorithm was performed as follows. The system was initialized with $I = 10$ ill individuals chosen randomly among $N = 1000$ graph nodes.

The transition rates are defined as:

$$a_1 = \mu I$$

$$a_2 = \lambda I(N - I) \sum_{i \neq j}^{N} a_{ij}$$

where

$$a_{ij} = \begin{cases} 1 & if \;\; i \iff j \;\; and \;\; i_{state} \neq j_{state} \\ 0 & otherwise \end{cases}$$

It consists in the sum of the number of links between the ill and susceptible individuals.

The computation of the latter definition for a_2 is very expensive and for this reason it was approximated with:

$$a_2 = \lambda I(N - I)c \qquad with \quad c = \frac{\langle k \rangle}{N - 1}$$

c is the connectivity of the graph.

Using the previously defined transition rates, a random number r was sampled from a uniform distribution leading to the following alternatives:

$$\begin{cases} if \quad r(a_1 + a_2) < a_1 \quad \Rightarrow \quad ill \; recovery \\ otherwise \quad \Rightarrow \quad transmit \; to \; susceptible \; neighbor \end{cases}$$

3.2 Phase Diagram and Critical Value Determination

The simulation was run for several intervals of λ to have see the behaviour of both the system and the algorithm.

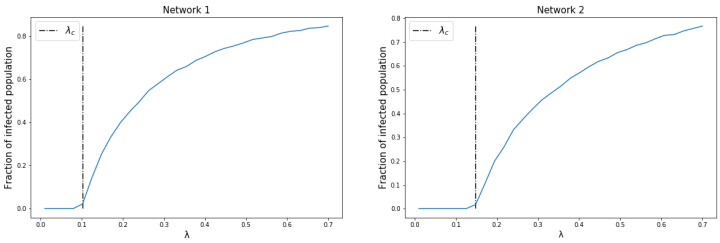

Figure 2: Phase diagrams for the two networks resulting from Gillespie sampling

For each value of λ several samples were taken in order to obtain a good statistic over the results. The python implementation of the Gillespie algorithm proved itself to be computationally expensive, and some trade-off values for the simulation parameters were chosen.

The phase diagrams in Figure (2) were plotted using $\simeq 30$ values of λ with 50 samples each with 2000 iterations for every run of the Gillespie algorithm. This simulation was removed from the code since it was too time consuming and aiming towards system visualization more than actual parameters estimation.

Similar runs were performed but using smaller intervals to speed up the processing time.

By "zooming" the plot, it was clear that the mean value of the fraction of ill individuals is subject to strong fluctuations in the neighborhood of the transition value. This introduced some difficulties in the determination of the correct value for λ_c.

The non-zero values of the mean ill population vector were extracted along with their standard deviation. Subsequently, all of those were checked for compatibility with the zero and the index of the last value compatible with zero was designated to be the index of the critical value.

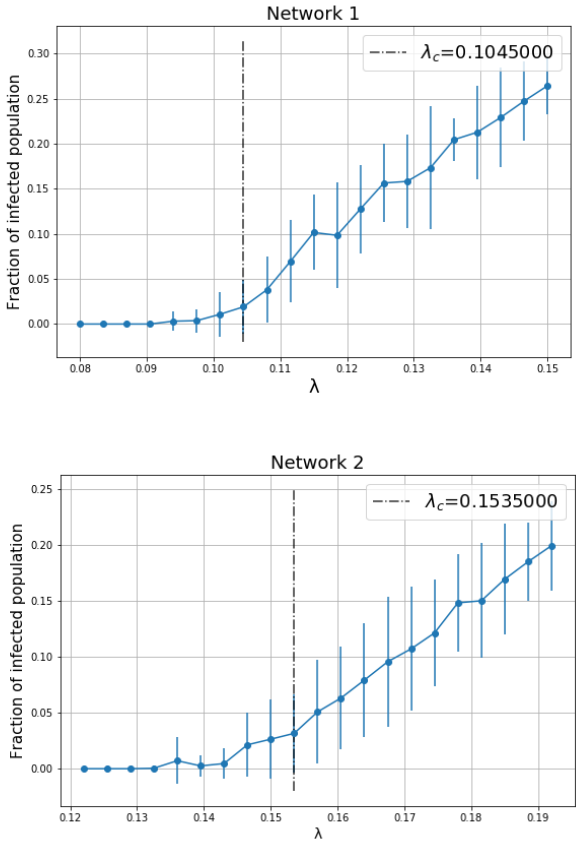

Figure 3: "Zoomed" simulation results (obtained with 20 values of λ in the interval)

By confronting the results obtained with the ones extracted theoretically, we can tell that there is an overshooting or undershooting in one of the two models.

One has to consider the possibility that our critical value estimation can be wrong and therefore improved even if it seems to be statistically correct. We can observe, following the thesis just proposed, that our Gillespie algorithm uses similar transition probabilities to the homogeneous model. Hence it seems reasonable to confront the obtained values with the theoretical homogeneous ones. By looking at Figure (3), we can see that the first non-zero points obtained from the simulation are close to match almost perfectly the previously cited theoretical values.

However, statistically speaking, it is not very feasible to just choose the first non-zero points as predicted critical values for our simulation.

4 Network Characterization

Typical descriptive graph quantities distribution were plotted in order to understand the structure of the two networks provided.

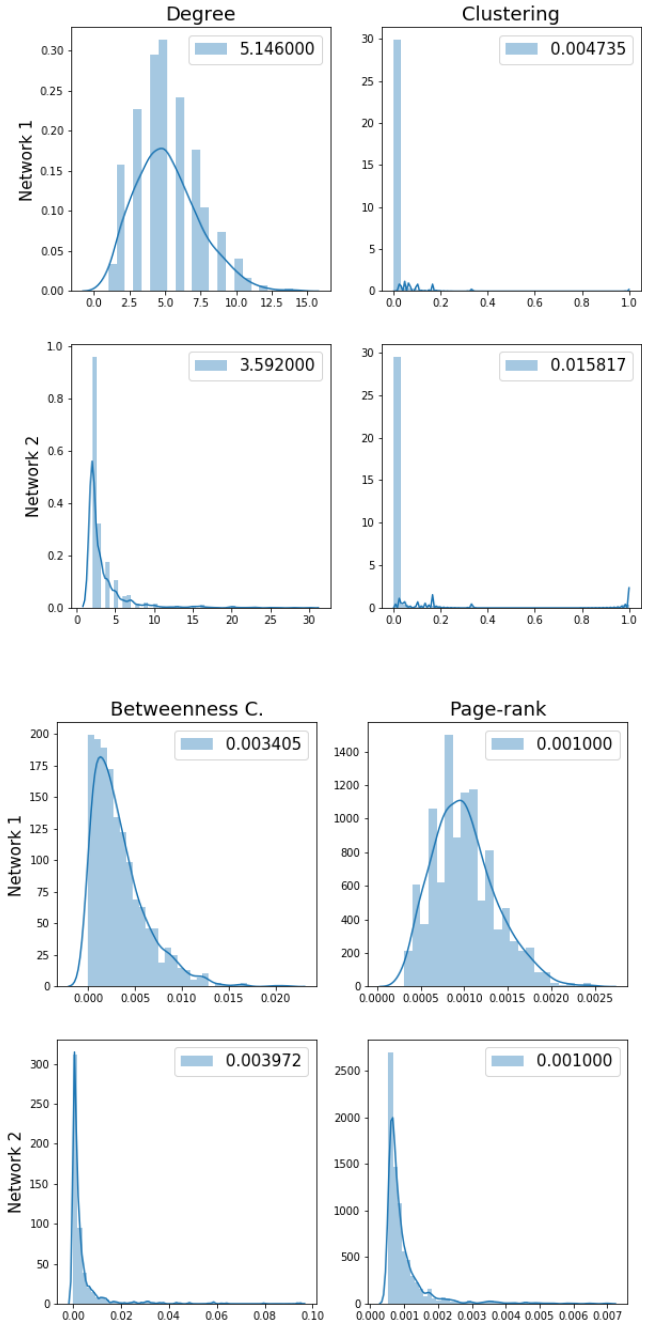

Figure 4: Network properties distributions (mean values are reported in legends)

By looking at Figure (4) we can clearly deduce that the network 2 is closer to an homogeneous configuration while the first one is more heterogeneous.

Since the graphs are undirected, the Page-Rank doesn't give much information becoming very similar to the degree distribution in the "homogeneous" case.

In particular we can see that the degree distribution of network 2 looks like a power law hence a *scale-free* network. This last consideration gives us a good discussion point over the network different structures and intrinsic susceptibility to epidemic outbreaks.

Basically speaking, what gives the disease the possibility to spread over the network is the degree of the infected nodes; for this reason, the *betweenness centrality* doesn't give any structural information of interest in this particular problem.

5 Segregation

What gives the disease the possibility to spread over the network is the degree of the infected nodes; for this reason, the *betweenness centrality* doesn't provide any structural information of interest in this particular problem as well as the Page-rank.

Furthermore, the low clustering values of both the networks don't represent useful parameters for segregation choice.

Therefore, the segregation procedure was performed over 1% of the maximum degree nodes for both networks. We expect that the "aimed attack" to particular nodes, would perform better on the scale free network. We obtained the following results:

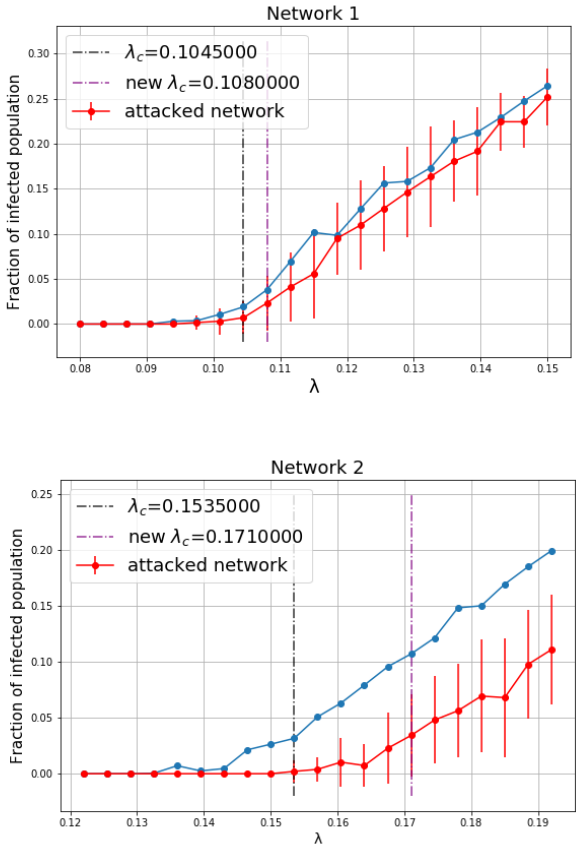

Figure 5: Segregation results

By looking at the Figure (5) we can tell that it had a noticeable effect on the scale free network (as expected) while leaving the first network's phase diagram almost identical.

6 SIR: Chicken Pox

Lastly, data about a chicken pox outbreak were loaded and analyzed. The simulation data provided reports data of the SIR model population fractions per time unit divided in 100 different locations. The SIR model can be summarized by the following time evolution differential equations:

$$\begin{cases} \frac{di}{dt} = \beta s(t) i(t) - \gamma i(t) \\ \frac{ds}{dt} = -\beta s(t) i(t) \\ \frac{dr}{dt} = \gamma i(t) \end{cases}$$

where i, s and r are respectively the fractions of *infected*, *susceptible* and *recovered* population and γ, β are the *recovery rate* and the *infection rate*.

The early stages of the population fraction evolution are plotted in the following Figure.

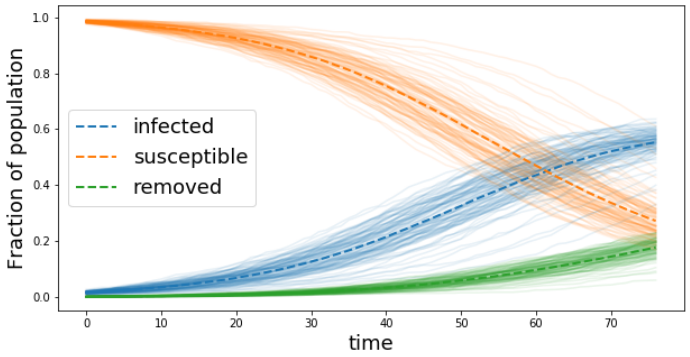

Figure 6: Population fraction time evolution

As we can see we cannot forecast what is gonna happen after the end of the data record just by looking at its evolution.

It is important to define a quantity called *basic reproduction number* $R_0 = \frac{\beta}{\gamma}$ which represents the average number of secondary infections caused by an infected host.

By looking at the R_0 distribution we can tell whether the infection is spreading faster or slower than the healing rate.

Therefore we can find the β and γ values from the population fractions as:

$$\beta = -\frac{s(T) - s(0)}{\int_0^T s(t) * i(t) dt} \qquad \gamma = \frac{r(T)}{\int_0^T i(t) dt}$$

Therefore the R_0 distribution was plotted:

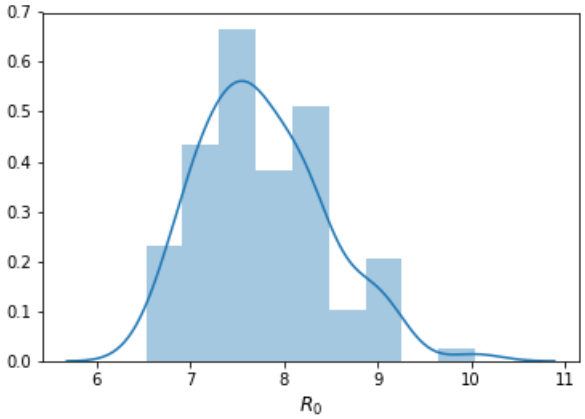

Figure 7: R_0 distribution

Obtaining the following mean values:

$$\beta = 0.0751 \pm 0.003$$
$$\gamma = 0.0107 \pm 0.0008$$
$$R_0 = 7.771 \pm 0.666$$

Since the value of $R_0 > 1$ we can tell that the Chicken Pox will spread causing an outbreak over the population.