Analysis of epidemic dynamics on networks

Andrea Perin, 1206078, Carlo Polato, 1205688, and Vincenzo Maria Schimmenti, 1204565

(Dated: 23 April 2019)

In this exercise, an analysis of the basic SIS model is carried out on two different networks, highlighting the role of the structure in the breakout of the epidemic itself. Additional analysis is aimed at the resilience of the two networks with respect to different attacking strategies, and then a basic SIR model stochastic simulation is carried out, allowing the measurement of the *basic reproduction number*, R_0.

I. PRELIMINARIES

The given data consists of three data files: two of them are the link lists of two undirected networks, and they will be the subject of the first part of the analysis. The last file represents the evolution of a chicken pox outbreak, and will serve as the basis of the SIR model simulation performed in the last point.

II. TASK 1: EPIDEMIC THRESHOLD ESTIMATES

In this initial task, we employ three different approximations of epidemic spreading in order to estimate the critical *epidemic threshold* λ_C. The approximations are:

- *homogeneous case*: we assume that the model is well-mixed. This results in the following expression for the epidemic threshold:

$$\lambda_C = \frac{\mu}{\langle k \rangle}$$

- *heterogeneous case*: we now suppose statistical equivalence amongst degrees; this results in

$$\lambda_C = \mu \frac{\langle k \rangle}{\langle k^2 \rangle}$$

- *quenched mean-field*: given the adjacency matrix $\mathbf{A}$ and its largest eigenvalue $\Lambda_{max}(\mathbf{A})$, the epidemic threshold is

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

For all three cases, the *healing rate* was set to be $\mu = 0.5$. In Table I, the results for the three procedures are reported.

Approximation	Network 1	Network 2
Homogeneous	0.097	0.139
Heterogeneous	0.082	0.076
QMF	0.079	0.066

TABLE I. The resulting estimates for the three employed approximations.

III. TASK 2: SIS MODEL SIMULATIONS

The two networks were then used in order to simulate the evolution of a disease in different combinations of the parameters μ and λ, which once again denote respectively the healing rate and the infection rate. In all the simulations, $\mu = 0.5$, while λ ranged from 0.01 to 1. The total number of individuals was set to be $N = 1000$, and the initial number of infected was set to be $I_0 = 10$. Two different algorithms were implemented:

- a Gillespie-like implementation, taking the network structures into account;
- a probabilistic approach, which uses the rates in order to define some time-step dependent probabilities, on the basis of which to describe the evolution of the infected population.

Gillespie-like algorithm

This algorithm is based on the assumption that each process of the same type is equally likely; as such, the next event taking place is chosen on the basis of a kind of weighted average. Let there be $I(t)$ infected individuals at iteration t; then, the possible reactions are divided in two categories: remission and infection. There are a total of I possible remission events, characterized by the remission rate μ, while the possible infections can take place between infected-susceptible pairs (provided such pairs are actually linked), with rate λ. The equiprobability assumption states that, at any iteration t, the probability of an infected node experiencing remission is

$$P_{healing}(t) = \frac{\mu I(t)}{\mu I(t) + \lambda\, e_{active}(t)}$$

where $e_{active}(t)$ denotes the number of active two-individuals edges at time t. On the basis of this reasoning, at each iteration one reaction is chosen (in Gillespie fashion) and the susceptible-infected state of the node is set accordingly. We discard the time information (i.e. we don't record how much time passes between two reactions) since we are only interested in the long time behaviour of the system.

Probabilistic approach

This approach is based on the assumption that both λ and μ are pure probabilities (when multiplied by a unitary time step); λ is the probability that a susceptible individual coming into contact with an infected one is infected by the latter; conversely, μ denotes the probability that an infected individual achieves remission. Therefore, at time t, a susceptible individual with $\tilde{k}_i(t)$ infected neighbors has a probability $P(\tilde{k}_i(t)) = 1 - (1-\lambda)^{\tilde{k}_i(t)}$ of being infected. An infected individual, on the other hand, has a probability $P = \mu$ of healing. At each unitary time step, then, each individual experiences the appropriate reaction with the aforementioned probabilities; the update of the system state is done in a parallel fashion, i.e. the state of each individual is updated only after all the individual's new state is determined.

Results

In the following, the two approaches are compared. For both algorithms, a total of 100 values of λ were tried, ranging from 0.01 up to 1, while $\mu = 0.5$.
For the Gillespie-like method, each value of λ was tested 100 times, and each run lasted 10000 iterations (provided that the infected population did not go extinct before reaching that number, in which case the simulation was cut short). These 100 runs (understood to be representative of a long time behaviour, an estimate of the otherwise unattainable asymptotic behaviour) were used in order to compute the average fraction of infected individuals and its standard deviation. The resulting points were then plotted against the respective values of λ; the results for the two networks are shown in Figure 1 and 2.
For the probabilistic method, each value of λ was tested

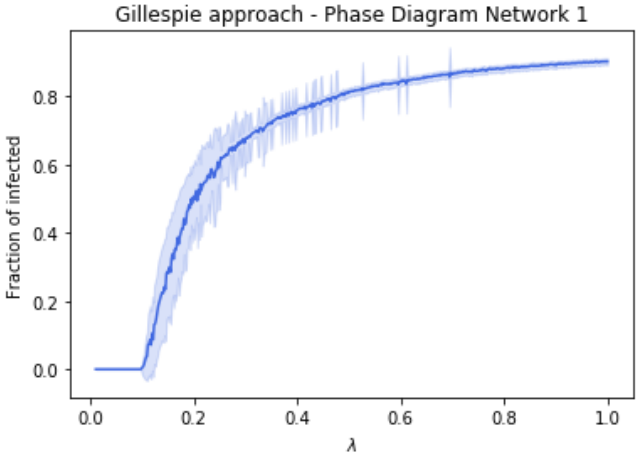

FIG. 1. The plot of the infected fraction vs. the values of λ, obtained via the Gillespie-like algorithm for the network 1.

100 times, and each run lasted 100 iterations (since in these case each iteration updates the whole configuration, not just one individual as before). These 100 runs were used to compute the average long time limit fraction of infected individuals. The resulting points were then plotted against the respective values of λ; the result is shown in Figures 3 and 4.

Both plots display the expected behavior: once the epidemic threshold is reached, the fraction of infected individuals at the end of the simulation quickly increases. The estimation of the actual value of λ_C was performed by analyzing the points where the phase diagram curve was close to zero (and extracting the value of λ_C directly or using a fitting procedure). The results are collected in Table II.

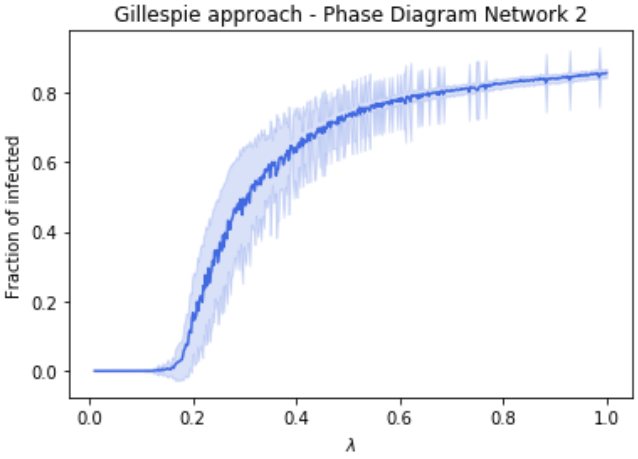

FIG. 2. The plot of the infected fraction vs. the values of λ, obtained via the Gillespie-like algorithm for the network 2.

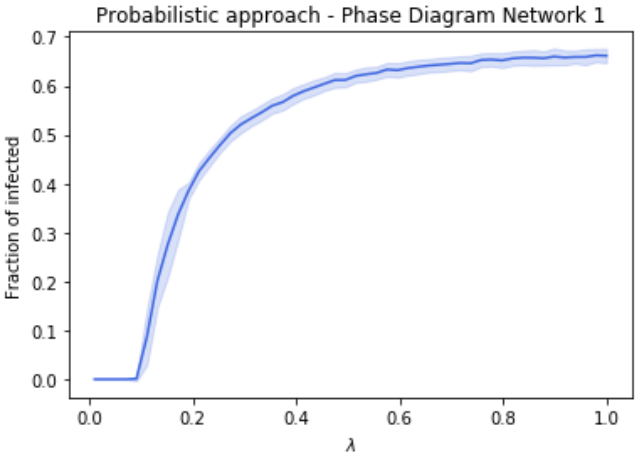

FIG. 3. The plot of the infected fraction vs. the values of λ, obtained via the probabilistic algorithm for the network 1.

It is worth noting how the two algorithms, while producing similar results in the estimation of the epidemic threshold, are not easily comparable. For instance, while in the Gillespie-like algorithm the parameters λ and μ

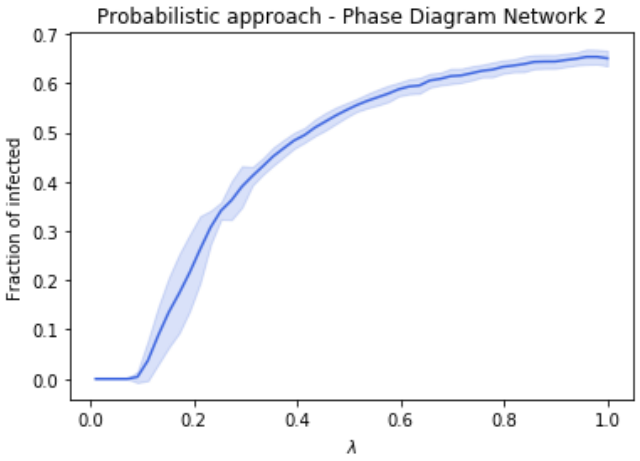

FIG. 4. The plot of the infected fraction vs. the values of λ, obtained via the probabilistic algorithm for the network 2.

Algorithm	Network 1	Network 2
Gillespie-like	0.103	0.165
Probabilistic	0.090	0.088

TABLE II. The values of λ_C estimated from the plots.

are meant as rates, and so can have values higher than 1, the same can not be said for the probabilistic approach, where the same parameters denote a pure probability, which is bounded to be at most 1.
Apart from these considerations, it seems that the Gillespie-like approach validates the homogeneous approximation, while the probabilistic one seems to be less linked to either approximation.
Of course, when analysing these results, one has to also take the finite-size effect into account, which partially affects the abruptness of the transitions, making the choice of the actual value of λ_C more of an arbitrary procedure (especially for Network 2, cfr. Figure 2).

IV. TASKS 3-4: ANALYSIS OF THE STRUCTURAL PROPERTIES AND SEGREGATION STRATEGIES

The two networks have been characterized by their following structural quantities:

- *degree distribution*
- *betweenness centrality*
- *PageRank*
- *clustering coefficient*

The results are collected and shown in Figure 5 and 6.
The histograms seem to show that Network 2 has a power-law degree distribution, and this power-law behaviour translates also to the other examined quantities.

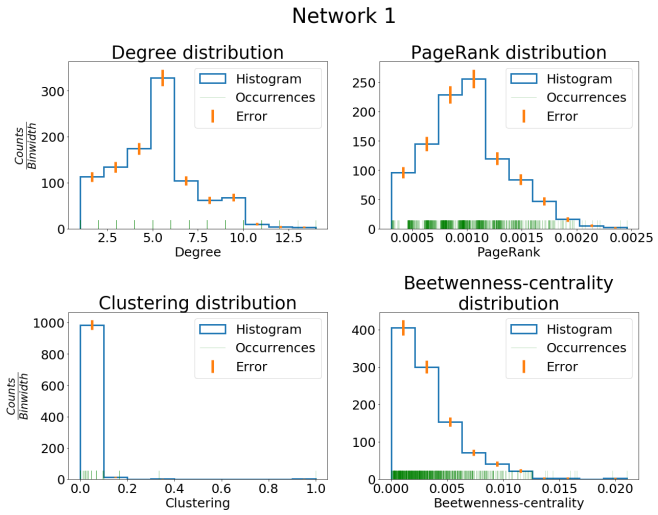

FIG. 5. The plots show the distribution of the quantities of interest for each node of Network 1.

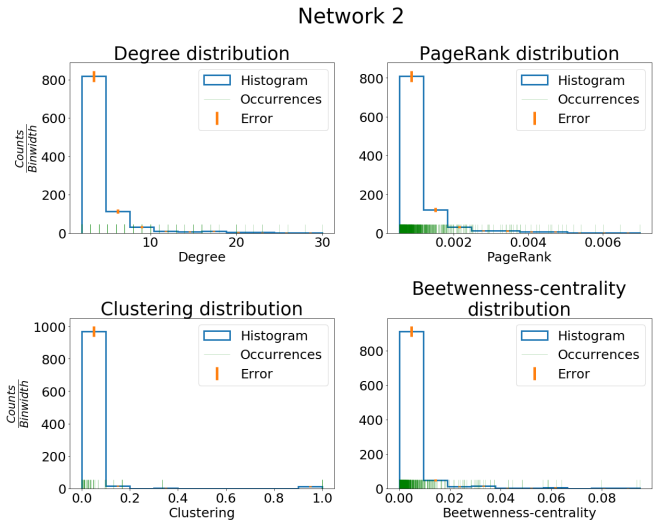

FIG. 6. The plots show the distribution of the quantities of interest for each node of Network 2.

Network 1, on the other hand, does not display the same distribution for all the quantities.
On the basis of this data, some segregation strategies were performed. The elimination was limited to 1% of the overall nodes, corresponding to 10 nodes. Each of the four aforementioned quantities have been used in order to choose the nodes to eliminate. In particular, the 10 nodes which exhibited the highest values of degree, beetwenness-centrality, PageRank and clustering coefficient were eliminated. The pruned networks were then once again tested in order to compare the effects of such segregation strategies on the epidemic threshold.
The values of λ_C for the segregated networks are basically the same as the original networks' ones, except for the degree case of Network 2. A visualization of these results is provided in Figures 7 and 8. The values are

obtained by means of the previously discussed Gillespie-like algorithm.

The reasoning behind the addition of beetwenness-beetwenness-centrality seems like an effective way to halt (or, at least, slow down) the percolation process.

Strategy	Network 1	Network 2
Degree	0.109	0.186
PageRank	0.105	0.169
Clustering	0.105	0.169
Beetwenness centrality	0.105	0.169

TABLE III. The values of the epidemic threshold λ_C after the four different segregation strategies.

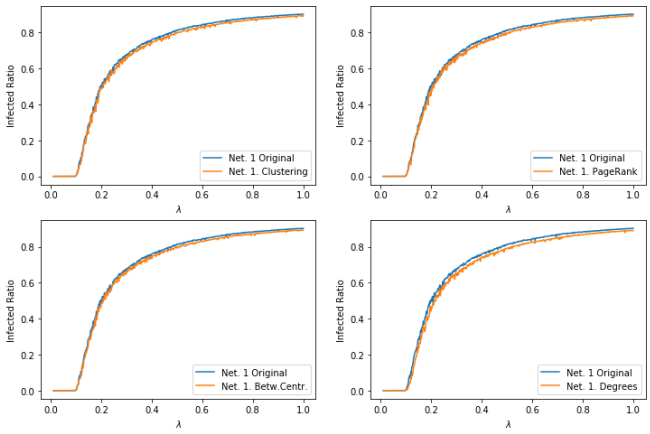

FIG. 7. A comparison between the phase diagram of the original Network 1 and its pruned counterparts.

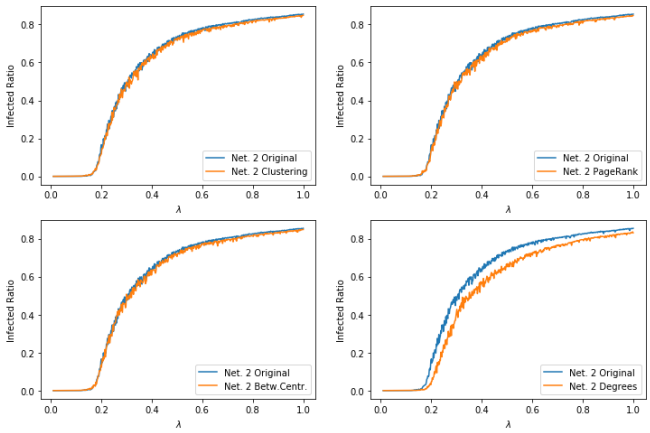

FIG. 8. A comparison between the phase diagram of the original Network 2 and its pruned counterparts.

centrality as a meaningful metric by which to perform segregation is that it measures how much a certain node is connecting any two parts of the overall network. As the outbreak of an epidemic in a network can be thought of as a percolation, targeting the nodes with the highest

Prior knowledge about the properties of networks with power-law distributed degrees suggests a particularly low resilience of Network 2 with respect to highest-degree based attacks. Network 1, instead, is not ascribable to the same network class, so that this reasoning does not apply to it.

Results of the segregation strategies are depicted in Figures 7 and 8. The network 1, when attacked, is resilient and does not display any sensible change in the epidemic threshold while the network 2 does this only in the case of degree segregation: this is due the aforementioned power-law structure of network 2.

V. TASK 5: SIR ANALYSIS

The last step in this work was to analyse the file containing information about the outbreak of a chicken pox epidemic in 100 locations during a period of 77 days. The dataset is employing a SIR model, meaning that the whole population is divided into three categories: *susceptible*, *infected* and *recovered*. The situation is depicted in Figure 9, where the evolution of the three categories for each site is shown, together with the average behaviour.

The SIR model can be described by the following set of

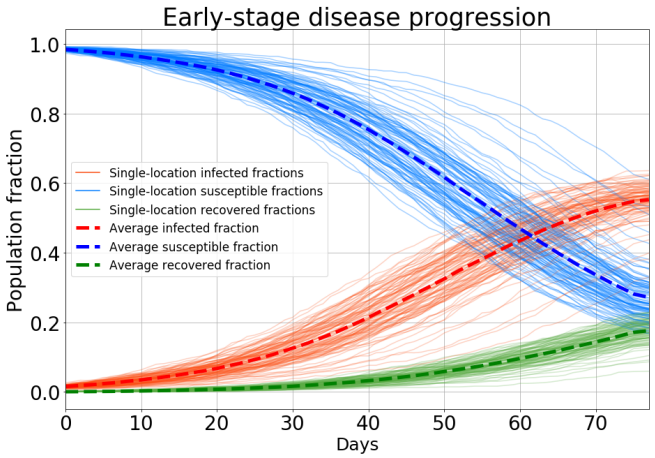

FIG. 9. The early stages of the infection outbreak in the 100 locations. The average behaviours are shown as dashed lines.

differential equations:

$$\begin{cases} \dfrac{di(t)}{dt} &= \lambda\, i(t)s(t) - \mu\, i(t) \\ \dfrac{ds(t)}{dt} &= -\lambda\, i(t)s(t) \\ \dfrac{dr(t)}{dt} &= \mu\, i(t) \end{cases}$$

where $\{i(t), s(t), r(t)\}$ denote respectively the fraction of infected, susceptible and removed individuals at time t,

while λ and μ are respectively the infection rate and the recovery rate.
If one were to invert the equations above as to isolate the two parameters, the following expressions would be found:

$$\lambda = -\frac{s(t) - s(0)}{\int_0^t i(\tau)s(\tau)\, d\tau}, \quad \mu = \frac{r(t)}{\int_0^t i(\tau)\, d\tau} \tag{1}$$

This suggests a way to estimate the values of the two parameters: assuming that the values $s(t), r(t)$ are the last that can be found (either by way of simulation or, as in our case, in a given dataset), then it simply suffices to substitute the appropriate quantities in eq. 1 to get such estimates. This procedure's main interest lies in the possibility of estimating the important value of the *basic reproduction number*, R_0. This number describes the average number of susceptible individuals that each infected one can infect, assuming a *naive* population (that is, a population not knowing about the existence of infected individuals amongst its ranks); as such, it is an effective way of estimating the eventual outbreak of a disease, or the lack of it. For a SIR model, it can be shown that

$$R_0 = \frac{\lambda}{\mu},$$

thus justifying the interest in the values of λ and μ.
For each of the 100 locations provided in the file, the procedure for the calculation of R_0 described above was performed, and the values were then averaged. This lead to the estimate

$$R_0 = 7.77 \pm 0.65$$

An infected individual will then on average infect other 7 individuals before achieving remission. As an $R_0 < 1$ would be required in order to avoid outbreak, it can be concluded that, in the case at hand, the chicken pox will in fact become viral and infect the vast majority of the population.