Simple Epidemic Network Model for Highly Heterogeneous Populations

María del Valle Rafo, Juan Pablo Aparicio

PII: S0022-5193(19)30425-4
DOI: https://doi.org/10.1016/j.jtbi.2019.110056
Reference: YJTBI 110056

To appear in: *Journal of Theoretical Biology*

Received date: 3 July 2019
Revised date: 18 October 2019
Accepted date: 21 October 2019

Please cite this article as: María del Valle Rafo, Juan Pablo Aparicio , Simple Epidemic Network Model for Highly Heterogeneous Populations, *Journal of Theoretical Biology* (2019), doi: https://doi.org/10.1016/j.jtbi.2019.110056

Highlights

- Highly variable networks are complex networks with long-tailed degree distributions.
- We show that an extrem simplification which considers only two degrees captures main features of disease transmission in scale-free networks.

Simple Epidemic Network Model for Highly Heterogeneous Populations

María del Valle Rafo[a], Juan Pablo Aparicio[a,b,c]

[a]*Instituto de Investigaciones en Energía no Convencional (INENCO), Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Universidad Nacional de Salta, Av. Bolivia 5100, 4400 Salta, Argentina.*
[b]*Simon A. Levin Mathematical, Computational and Modeling Sciences Center, Arizona State University, PO Box 871904 Tempe, AZ 85287-1904, USA*
[c]*Corresponding author: juan.p.aparicio@gmail.com*

Abstract

Network models for disease transmission and dynamics are popular because they are among the simplest agent-based models. Highly heterogeneous populations (in the number of contacts) may be modeled by networks with long-tailed degree distributions for which the variance is much greater than the mean degree. An example is given by scale-free networks where the degree distribution follows a power law. In these type of networks there is not a typical degree. Some nodes may have low representation in the population but are key to drive disease transmission. Coarse graining may be used to simplify these complex networks. In this work we present a simple model consisting in of a network where nodes have only two possible degrees, a low degree close to the mean degree and a high degree about ten times the mean degree. We show that in spite of this extreme simplification, main features of disease dynamics in scale-free networks are well captured by our model.

Keywords: scale-free networks, core-group model, disease dynamics

1. Introduction

A networks is a set of nodes connected by edges among them. The number of edges connecting a node, known as the degree of the node is, in general, a random variable K which takes non negative values ($k = 0, 1, 2...$) and its distribution is called the degree distribution of the network.

A random Poisson network, like the Erdös-Rényi (1959) network, has a low degree of heterogeneity. In this class of networks, the degree distribution is Poisson and, therefore, the variance is equal to the mean. We will consider these networks, with low heterogeneity in the distribution of contacts, as a model for a prototypical 'homogeneous' population where the mean degree is representative of most of the nodes degree.

A scale-free network has an asymptotic degree distribution which follows a power law, $f(k) \sim k^{-\beta}$, and therefore these are networks with a high level of variability where the variance of the degree distribution is much greater than its mean. In this case the mean is not representative (at all) of the degree distribution.

In this work we will consider epidemic spread in static networks with a fixed number of nodes and fixed connections among them. Nodes can be in one of three mutually exclusive states: susceptible, infectious and recovered. Intensity of the transmission is the same for every edge and the probability of infection per edge will be denoted by ρ.

Epidemics dynamics is quite different in scale-free and random poisson networks. In both cases the basic reproduction number is given by $R_0 = n_e\rho$ (see 5.2.1 below) where the effective mean number of contacts or excess degree is (see, for example, Brauer 2008, Lindquist et al. 2011)

$$n_e \doteq \frac{\langle K(K-1)\rangle}{\langle K\rangle} = \langle K\rangle - 1 + \frac{var}{\langle K\rangle} \tag{1}$$

where $\langle K\rangle$ and var are the mean and the variance of the degree distribution.

For the same mean degree, the excess degree of a scale-free network is greater than the excess degree of a Poisson network. If we consider that in both networks the basic reproduction number is the same, then we should use different values for the transmission probability per edge ρ. In such case, epidemics in the Poisson networks are significantly larger than the corresponding epidemics in a scale-free network (Bansal et al. 2007, see also Fig. 1). On the other hand if we use the same probability of transmission ρ, epidemics in the Poisson network are much smaller than epidemics in scale-free networks (Fig. 1).

Dynamics in scale-free networks cannot be described by using a simple random network with low heterogeneity. In this work we will present a simple network model which captures the main features of disease dynamics in scale-free networks.

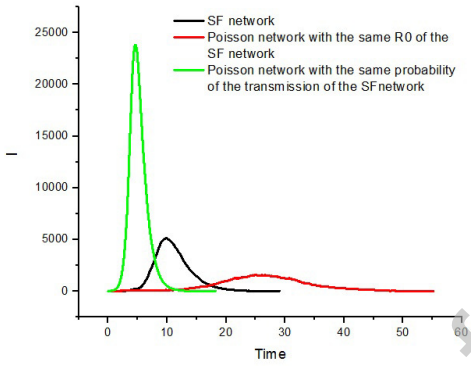

Figure 1: A typical realization of an epidemic in a scale-free network (black line) and two realizations obtained with random poisson networks. When we consider the same R_0 for both networks, epidemics in the Poisson network are significantly greater (green line), while when we consider the same probability of transmission, epidemics are significantly smaller (red line).

2. Scale-free networks and the preferential attachment algorithm

A scale-free network has an asymptotically potential degree distribution $f(k) \sim k^{\beta}$. For finite networks it is convenient to consider the complementary cumulative distribution $1 - F(k)$ (see Li et al 2005). For a scale-free network this distribution also presents an asymptotically potential distribution $1 - F(k) \sim k^{-\alpha}$ where the cumulative distribution is

$$F(k) = \sum_{j=1}^{k} f(j)$$

2.1. The preferential attachment algorithm

In this work we will consider (approximately) scale-free networks built using a preferential attachment algorithm (Barabassi & Albert, 1999). In the rest of this work the mean degree $\langle K \rangle$ will be denoted by n.

A network with mean degree n has $nN/2$ edges. This number of edges were allocated among the N nodes as follow. First, we selected N_0 equivalent nodes and we assigned one contact (chosen at random) to each one. Then, we assigned one contact to every other node with probability proportional to the degree of the contact. Note that in this part of the process only nodes with degree greater than zero can be chosen. Finally we selected the rest of the $N\,n/2 - N$ pairs (with no repetition) of nodes and connect them in the following way. First we chose a

node at random, then we chose another node but with a probability proportional to the degree of the node. This algorithm produces networks with an approximate potential degree distribution (Fig. 2) where nodes with low degree are most likely connected to high degree nodes. The variability in the degree distribution and the maximum degree obtained in this process depends on N_0.

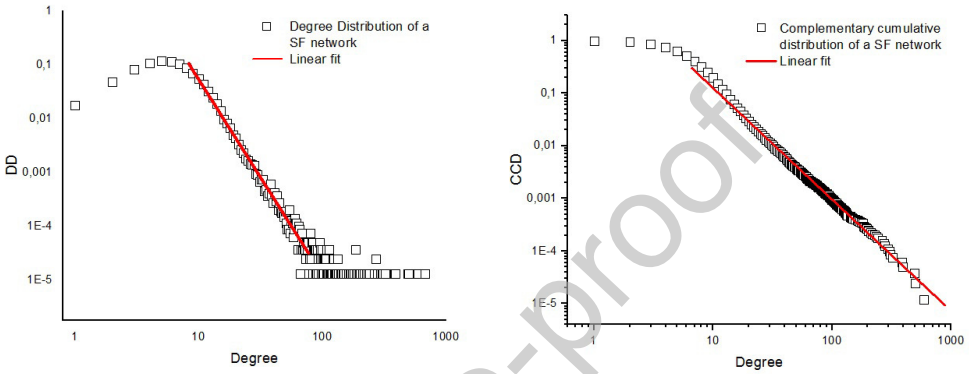

Figure 2: Degree distribution and complementary cumulative distribution for a typical network used in this work. Exponents obtained from linear fit are $\beta = -3.64(0.07)$ and $\alpha = -2.10(0.011)$.

The networks created with this algorithm will be called SF networks. We considered two cases, SF networks with mean degree $n = 8$ and $n = 16$. In each case we created 100 different networks. In the former, maximum degree ranged between 339 and 1475, while the variance of the degree distribution varied between 66 and 106 (mean value 78). For $n = 16$, maximum degree varied between 459 and 2060, while the variance ranged between 142 and 221 with a mean value of 168. In both cases the variance is about ten times the mean of the degree distribution, and therefore the variance is a significant contributor to R_0.

3. A simple model for scale-free networks

In scale free networks most of the nodes have a low degree and are most likely connected to high degree nodes. But there are few nodes with degree much higher than the mean degree and therefore they are super-spreaders. The most simple model consists of a population with only two homogeneous sub-populations: a large population of nodes with low connectivity and a small population of super-spreaders, that is, a core-group model (Yorke & Hethcote 1978; see also Hethcote & Yorke 2014).

Our network model is, therefore, composed by a large population of size N_l of nodes with low degree n_l, and a small population of size N_h of nodes with high degree n_h. These networks will be called two-degrees networks or just 2D networks.

Such simple network is specified by the values of N_j and n_j as well as a rule for connecting the nodes. As we are interested in a model of a specific SF network we will impose some constraints. First, both networks should have the same size and, therefore, only one sub-population size is independent (the other is obtained from the constraint $N = N_l + N_h$). In other words, we have to choose only three independent values: N_l, n_l and n_h. Those (integer) values are selected to match the first three moments of the specific scale-free (or other complex) network being modeled. The preferential attachment algorithm we implemented produced a variety of scale-free networks with a prescribed mean degree n. Realizations of the process of network construction produce networks differing in the degree of variability. Each network presents a frequency distribution of degrees which may be considered as a sample of the implicit degree distribution. Sample variance and (not normalized) skewness are

$$var = \frac{1}{N}\sum_{j=1}^{N}(n_j - n)^2$$

$$skw = \frac{1}{N}\sum_{j=1}^{N}(n_j - n)^3$$

where n_j is the degree of the node j and n is the mean degree.

For a given scale-free network, we computed those moments and then we obtained the values of N_i, n_i $(i = l, h)$ to match the mean, variance and skewness of the SF network. Because we also have the constraint that these parameters are integers, the equality is, in general, not satisfied. We proceeded, therefore, in two steps. First we use the four constraints

$$
\begin{aligned}
N_l + N_h &= N \\
n_l \frac{N_l}{N} + n_h \frac{N_h}{N} &= n \\
(n_l - n)^2 \frac{N_l}{N} + (n_h - n)^2 \frac{N_h}{N} &= var \\
(n_l - n)^3 \frac{N_l}{N} + (n_h - n)^3 \frac{N_h}{N} &= skw
\end{aligned}
\tag{2}
$$

to find a first approximation to n_l, n_h, N_l, and N_h. Then we chose the closest integer value for n_l for which $n_l < n$ (otherwise the mean of the two-degree network would result greater than the mean degree n). Finally we use only the equations for the mean and variance (together with $N = N_l + N_h$) to find the new values for N_l, N_h and n_h. Networks created in this form will be called two-degrees networks (or just 2D networks).

Here we show an example of this process. All the networks considered in this work has $N = 80000$ nodes. One of the scale-free network used has a degree distribution with mean degree $n = 8$, variance $var = 79$ and skewness $skw = 16071$. By solving system 2 we obtain $n_l = 7.6124$, $N_l = 79848.2$, $n_h = 211.818$, $N_h = 151.847$. Closest integer for n_l is 8, but in this case mean degree will result greater than $n = 8$, and therefore, we consider $n_l = 7$. Now we use the first three equations of the system 2,

$$
\begin{aligned}
N_l + N_h &= N \\
n_l \frac{N_l}{N} + n_h \frac{N_h}{N} &= n \\
(n_l - n)^2 \frac{N_l}{N} + (n_h - n)^2 \frac{N_h}{N} &= var
\end{aligned}
$$

to obtain the rest of the values. In this example these values are $N_l = 79000$, $n_h = 87$, N_h=1000, for which the mean and the variance are 8.00 and 79.00 identical (up to 3 decimal places) to the values of the original scale-free network.

3.1. 2D network model connectivity

Given the population parameters N_j and n_j, the nodes may be connected among them in different topological ways. Because preferential attachment produces networks where low degree nodes are most likely connected to high degree nodes

than among themselves, consequently, we build the simple two-populations network model by connecting the small population of super-spreaders at random. Because $N_h \ll N_l$ most of this connections would result between high degree nodes and low degree nodes. After all of the available connections of the super-spreaders were assigned, we randomly connected the rest of the nodes in the N_l population among them.

4. Disease transmission and Basic Reproduction Numbers for Networks

4.0.1. Simulating disease transmission

Each node can be in one of three mutually exclusive states: Susceptible, Infected (and infectious), and Recovered. We assume that the infectious period is exponentially distributed with parameter γ (the recovery rate). Probability of transmission per contact and per unit of time is assumed to be constant and we denote it by λ. Therefore, the probability of disease transmission to a susceptible node in contact with one infected node during a period of time δt is given by

$$P(S \to I, s) = 1 - e^{-\lambda \delta t}$$

In the case of fixed infectious period (of value T) the probability of transmission during the entire infectious period is therefore

$$\rho = 1 - e^{-\lambda T}$$

while for exponentially distributed infectious periods (with mean $1/\gamma$) it is

$$\rho = \frac{\lambda}{\lambda + \gamma}$$

In the simulations we considered a fixed time step δt and two-layers to avoid spurious correlations. Individuals were enumerated and we checked the state of each of them in a sequential way.

In a time step δt recovered nodes may lose immunity (at the rate δ) becoming susceptible with probability

$$p(R \to S, \delta t) = 1 - e^{-\delta \delta t}$$

while infected individual may recover with probability

$$p(I \to R, \delta t) = 1 - e^{-\gamma \delta t}$$

4.1. Basic Reproduction number

In any network, the first infectious case is exceptional because all of its contacts are susceptible. Further cases have, at least, one contact not susceptible, the node from where it got the infection. For an infinite network the basic reproduction number is defined as the mean number of cases produced per infectious individual. It is computed as the ratio between all the cases produced by any generation of infectious individuals except the first generation (see for example Diekmann and Heesterbeek 2000, Brauer 2008). For a finite network we define the basic reproduction number as the expected number of cases produced by an average infectious individual of the second generation (see Aparicio and Pascual 2007).

In this work we considered networks with negligible clustering coefficients. In this case any node of degree k is in contact with k nodes which are not in contact among them (this configuration is known as a *star*). If the central node is infectious and i of its k contacts are susceptible, then the number of infections produced during the whole infectious period is a random variable that we call r_0.

The basic reproduction number is computed as the expected value of r_0 for an average infectious node of the second generation and in this case is given by (Aparicio and Pascual 2007)

$$R_0 = \left[(n-1) + \frac{var}{n}\right]\rho \tag{3}$$

where var is the variance of the degree distribution, n is the mean degree, and $\left[(n-1) + \frac{var}{n}\right]$ is the effective number of contacts or excess degree 1.

For homogeneous networks, every node has the same degree; the variance of the degree distribution is zero and therefore $R_0 = (n-1)\rho$. That is, at the beginning of the epidemics an average infectious individual will produce $(n-1)\rho$ infections because one of the contacts is already infected, the contact from where the node caught the infection.

For a Poisson random network we have a relatively low heterogeneity. Degree distribution is Poisson and therefore, the mean is equal to the variance and $R_0 = n\rho$. In this case no a significant difference is observed with respect to the homogeneous case. However for highly heterogeneous networks, as the ones considered in this work, the variance is much greater than the mean and the contribution of the variability of the degree distribution plays a significant role.

5. Testing for the goodness of the approximation

5.1. Topology

Scale-free networks and the two-degrees networks have very different degree distributions. In our case, mean and variance are practically the same, by construction, but the distributions differ in the other moments.

The clustering coefficient and mean path length are other common topological features that characterize network structure. The (global) clustering coefficient is a measure of the probability that two contacts of a node are also in contact between them. Mean path length is the average minimum distance between two nodes chosen at random. All the networks considered in this work have low clustering coefficients and short mean path lengths. In table 1 we display the values obtained for the networks used in this work.

	Clustering Coeff.		Mean Path Length	
Mean Degree	SF	2D	SF	2D
8	0.000439	0.000300	3.430107	4.769741
16	0.000486	0.000438	2.903585	3.958496

Table 1: Clustering coefficients and mean path lengths for some of the networks used in this work.

5.2. Disease dynamics

5.2.1. Basic reproduction numbers

Because the two types of networks considered in this work are random networks with negligible clustering coefficients the basic reproduction number, defined as the number of cases produced by an infectious individual of the second generation, is given by 3. Empirical estimations of R_0 for each network were obtained as follow. We set the immunity-loss rate equal to zero and selected I_1 index cases at random. If the number of secondary cases I_2 was greater than zero we computed the ratio of tertiary cases (I_3) over the number of secondary cases, I_3/I_2. This process was repeated a thousand times and averages and standard errors were computed. According to equation (3) it should be a straight line as a function of ρ. However this result was deduced for an infinite size network. Finite size effects are apparent for large values of ρ lowering the observed value with respect to values predicted by expression 3 (but this effect is not shown in the figure 3).

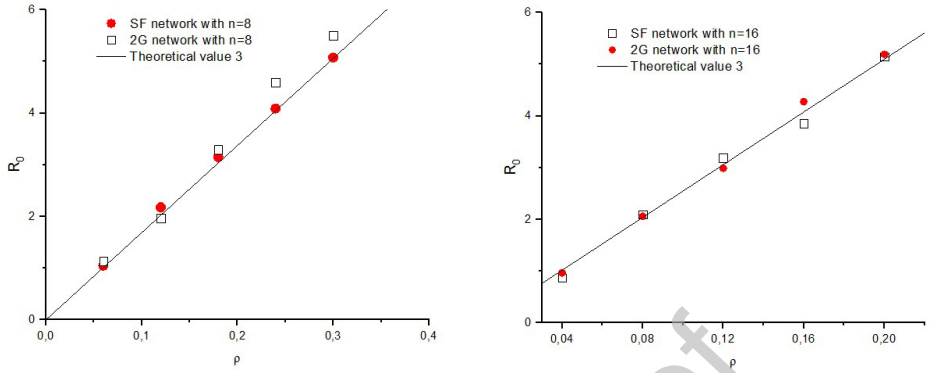

Figure 3: Empirical values obtained for R_0 for the scale-free network and the two-degree model. In both cases, a thousand simulations were performed (given $I_2 > 0$) and averages and standard errors were computed for the ratio I_2/I_3. Continuous line represent the theoretical value 3

Despite of the significative differences in structure, basic reproduction numbers for the two types of networks are the same since the only significant features are a clustering coefficient close to zero, and the mean and variance of the degree distribution.

5.2.2. *Disease invasions*

Probability of an epidemic. In an homogeneously mixed population the probability of an epidemic is a function of R_0. However, in more complex networks, the topology plays a significant role. In our case, the two types of networks have quite different topology so differences in the probability of an epidemics is expected although both networks have the same values for the basic reproductive number. The simulations were started with one infected node, randomly chosen, with all the other nodes in the susceptible state. We considered that an epidemic took place if more than one percent of the population resulted infected while if more than 10 cases were observed (but less than 1% of the total population) we considered that a small outbreak took place. The process was repeated 1000 times and the fraction of simulation for which we observed an epidemic was used as an estimator for the probability of disease invasion. In tables 2 and 3 we show the results obtained for scale-free and two-degrees networks.

	Epidemic freq.		Small outbreak freq.	
R_0	SF	2D	SF	2D
1.5	0.036	0.06	0.055	0.05
2.5	0.278	0.277	0.02	0.03
5	0.644	0.685	0.0	0.0

Table 2: Observed frequency of epidemics and small outbreak for a scale free network and its two degrees approximation obtained with of 1000 simulations. In all cases I_0=1. The two-degrees network were built with n_l=7, n_h=87, N_l=79000, N_h=1000.

	Epidemic freq.		Small outbreak freq.	
R_0	SF	2D	SF	2D
1.5	0.1350	0.10	0.060	0.069
2.5	0.417	0.374	0.13	0.22
5	0.718	0.734	0.0	0.0

Table 3: Observed frequency of epidemics and small outbreak for a scale free network and its two degrees approximation obtained from 1000 simulations. In all cases I_0=1. The two degrees network were built with n_l=15, n_h=184, N_l=79528, N_h=472.

5.2.3. *Comparing single epidemics disease dynamics: epidemic size, peak, time-to-peak, average curves*

Epidemic curves are quite similar in both, the scale-free and the corresponding two-degrees networks. Most differences are noted for low values of the basic reproductive numbers where stochasticity amplifies the differences in network topology and connectivity. In all cases we found significant statistical differences between the statistics characterizing disease dynamics in both types of networks (see tables 4 and 5), however typical, or averaged epidemic realizations are close to each other for the different networks. The two-degrees model captures the disease dynamics of the scale-free networks surprisingly well with differences less than 15%, which are usually below observational error (see Figures 4 and 5).

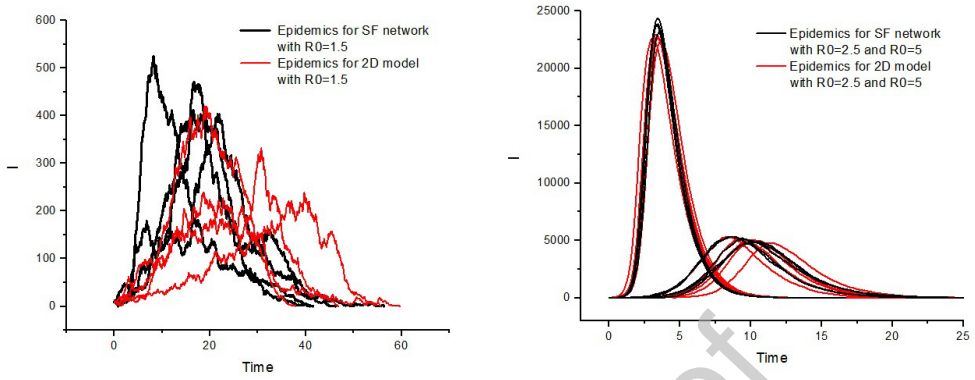

Figure 4: Epidemics obtained with a scale-free network (Black lines) and its two-degrees approximation (red lines) for R_0=1.5 (left panel), R_0=2.5 and R_0=5, (right panel). Mean degree n = 8

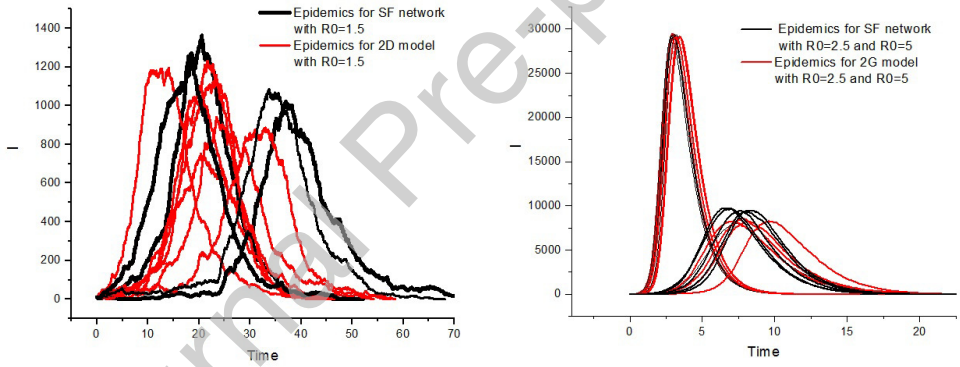

Figure 5: Epidemics obtained with a scale-free network (Black lines) and its two-degrees approximation (red lines) for R_0=1.5 (left panel), R_0=2.5 and R_0=5, (right panel). Mean degree n = 16

In the tables 4 and 5 we compare epidemic size, peak of the epidemic curve, and time at which the peak occur for both types of networks for different values of ρ.

R_0	*Peak*		t_{peak}		FS	
	SF	2D	SF	2D	SF	2D
1.5	570(14)	368(12)	20.1(0.8)	29(1)	0.101(0.001)	0.08(0.001)
2.5	5273(24)	5106(23)	9.2(0.17)	8.97(0.15)	$0.42(7\times10^{-4})$	$0.368(9\times10^{-4})$
5	23900(27)	22870(24)	3.23(0.03)	3.3(0.02)	$0.82(2\times10^{-4}$	$0.87(2\times10^{-4}$

Table 4: Epidemic peak, time of the peak and epidemic final size for a scale free network and its two degrees approximation. Averages and standard errors (in parenthesis) obtained with of 100 simulations. Two degrees network with n_l=7, n_h=87, N_l=79000, N_h=1000.

R_0	*Peak*		t_{peak}		FS	
	SF	2D	SF	2D	SF	2D
1.5	1122(14)	1010(16)	20.0(0.45)	21.5(0.6)	0.200(0.001)	0.17(0.001)
2.5	9471(25)	8322(27)	7.90(0.11)	7.96(0.10)	$0.62(7\times5^{-4})$	$0.60(9\times7^{-4})$
5	29272(23)	29323(26)	3.00(0.03)	3.20(0.03)	$0.91(1.4\times10^{-4}$	$0.95(2\times1.1^{-4}$

Table 5: Epidemic peak, time of the peak and epidemic final size for a scale free network and its two degrees approximation. Averages and standard errors (in parenthesis) obtained with of 100 simulations. Two degrees network with n_l=15, n_h=184, N_l=79528, N_h=472.

5.2.4. *Immunity loss and endemic equilibria.*

A simple way to obtain an endemic equilibria is by considering that recovered nodes may become susceptible again at constant rate δ. We considered that the system is in an endemic equilibrium if the infected population does not die out during a long period of time (see figures 6, 7 for example). For a weak homogeneous-mixing population, as in a random Poisson network, the group of contacts of any node represents a random sample of the population and therefore the expected susceptible proportion in this sample is S/N, the total susceptible proportion. In this case it is expected that at the endemic equilibrium $R_0 S/N = 1$, or $R_0 = N/S$. For non-homogeneously mixed population, as in networks with high clustering coefficient, this relation usually does not hold. In our case we are considering random networks with negligible clustering coefficients but which are not homogeneously mixed. Numerical simulations show that at the endemic equilibrium $R_0 \neq N/S$. In Fig. 6 we show the evolution of the susceptible proportions for both types of networks. The steady state fluctuates around a value significatively higher than the value expected according to the homogeneous mixing assumption.

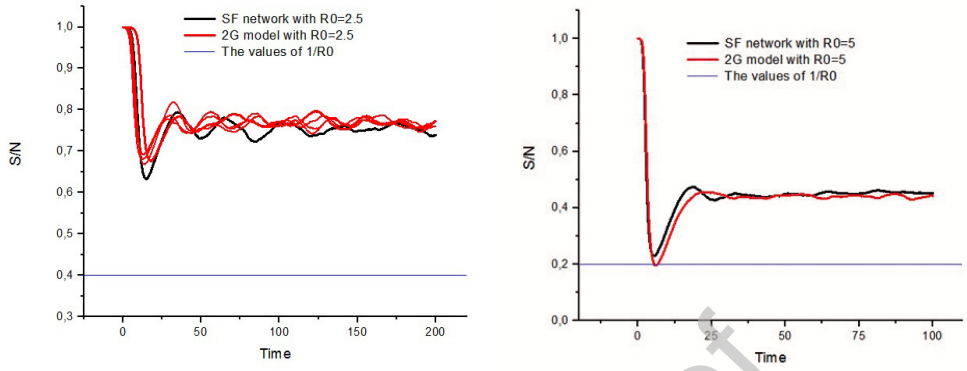

Figure 6: Evolution of the susceptible proportion for the SF network (black line) and the 2D model (red lines) for R_0=2.5 (left) and $R_0 = 5$ (right). Horizontal lines represent the values of $1/R_0$.

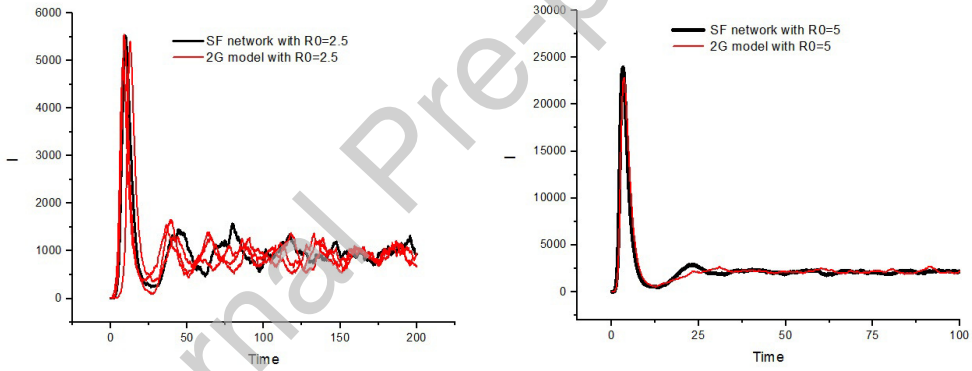

Figure 7: Evolution of the Infected populations for the SF network (black line) and the 2D model (red lines) for R_0=2.5 (left) and $R_0 = 5$ (right).

6. Super-spreaders and Disease dynamics

We applied the simple two-degrees model to gain insight into the disease dynamics in complex networks like the scale free networks. As we show many of the main characteristics of scale free networks may be captured considering only two sub-populations: a large population of nodes with low degree, close to the mean degree, and a small population of super-spreaders, nodes with a much higher degree. For each group we kept track of the populations of susceptible, infected and recovered nodes: S_i, I_i and R_i $(i = l, h)$.

The two-degrees model has N_l nodes with low degree n_l and the population of super-spreaders of size N_h and high degree n_h. In figure 8 we show the relative contributions of I_l and I_h in a typical epidemic. At the beginning of the outbreak the super-spreaders play a significant role in the dynamics as the proportion of high-degree nodes infected I_h/N is greater than the representation of such nodes in the population, N_h/N. In figure 8 we display the normalized proportions $\frac{I_j/I}{N_j/N}$ ($j = l, h$). When this fraction is greater than one the corresponding infected population has a greater representation than expected. For mean degree $n = 8$ and $R_0 = 1.5$, at the beginning of the epidemic we observe that the population of super-spreaders is up to 40 times the representation of that population (N_h/N). In the steady, endemic equilibrium this proportion fluctuates between 2 and 10. On the other hand the normalized proportion for low degree infected nodes fluctuates (almost imperceptibly in the figures scale) around 1. For other cases the situation is similar (see Fig 8).

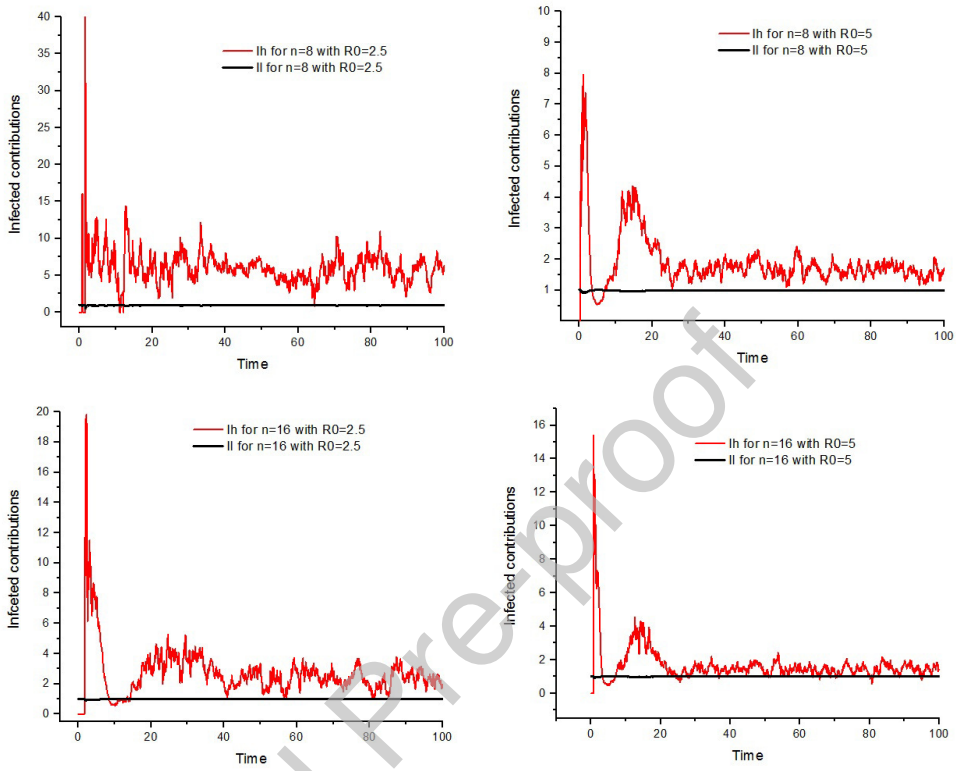

Figure 8: Relative contributions of the I_l (black lines) and I_h (red lines) populations for $R_0 = 2.5$ (left figures) and $R_0 = 5$ (right figures) . Top panel for mean degree $n = 8$, bottom panel for $n = 16$.

Another way to show that the super-spreaders are key to understanding disease dynamics is computing the epidemic final sizes. If at least one super-spreader node resulted infected, the epidemic final size in each population, obtained with 100 simulations, are displayed in table 6. For example, when mean degree is $n = 8$ and $R_0 = 1.5$, average epidemic final size in the low degree population is about 7% while in the population of super-spreaders is greater than 40%.

R_0	FS		FS_L		FS_H	
	n=8	n=16	n=8	n=16	n=8	n=16
1.5	0.074(0.002)	0.169(0.001)	0.0698(0.002)	0.165(0.001)	0.43(0.01)	0.832(0.003)
2.5	0.369(0.007)	0.605(0.006)	0.362(0.007)	0.602(0.006)	0.989(0.003)	1(0)
5	0.872(0.003)	0.9548(0.0001)	0.870(0.03)	0.9545(0.0001)	1(0)	1(0)

Table 6: Epidemic final sizes (given at least one super-spreader node was infected) for different 2D networks. Averages and standard errors (in parenthesis) obtained with of 100 simulations. Total epidemic final size FS, final size for the population of low degree nodes FS_L and for the population of super-spreaders FS_H

6.1. Disease control

In the networks used in this work the effective degree $n_e = (n-1) + var/n$ has two contributions of comparable sizes. For example, the scale free networks considered have mean degrees $n = 8$ and n =16 while the variances are about 80 and 170 respectively. Using the two-degrees models is straightforward to understand the importance of the super-spreaders in disease dynamics. In all the cases considered this small sub-population is between 0.5 to 1.25% of the total population. However removing this small number of nodes from the transmission process (through vaccination for example) has dramatic effects in disease dynamics. In such a case we may estimate the reduction in the basic reproduction number just by disregarding the variance contribution, that is $R_{0vacc} \sim (n-1)\rho$ which is, as discussed above, approximately $R_0/2$. Therefore for cases where the basic reproduction number is less than two, targeted vaccination of about 1% of the population is enough to drive the system below the epidemic threshold.

While for $R_0 < 1$ the probability of a major outbreak is zero, in the practice, a value for R_0 slightly greater than one is enough to reduce significantly the occurrence of an epidemic. For example, with the two-degrees model, with mean degree 8 and variance 79, $n_e \sim 17$. Vaccination of the super-spreader population reduces the effective number of contacts to 7, almost a 60% reduction. If the basic reproduction number is 3 (no vaccination) then $R_{0vacc} \sim 1.24$ and major outbreaks are very unlikely. From 100 realizations we obtained a 0% of major outbreaks and only a 17% of small outbreaks. For $R_0 = 4$ frequency of major outbreaks is only 33% while for small outbreaks this frequency decreases to 7%. But even in this case where major outbreaks may take place with relatively high probability, the final size of the epidemics is greatly reduced. Without vaccination, epidemic final size is about 0.74 while with vaccination this value drops to 0.42.

7. Discussion and Conclusions

In a strong homogeneous mixing population all individuals are in contact among them and a complete graph is the natural network model. A random Poisson network is a more realistic model for an homogeneously mixed population. In this case, the number of contacts of each node, that is its degree, is a random variable with Poisson distribution and the group of contacts of any node represents a random sample of the population. Deterministic, homogeneous-mixing, mass-action, models capture the epidemics accurately (Aparicio & Pascual 2007). These networks provide some degree of heterogeneity (in the number of contacts of a random node) but the variance of the degree distribution is the same as the mean degree and therefore the excess degree is the same as the mean degree ($n_e = (n-1) + var/n = n$). In this sense we consider that a Poisson random network represents an (almost) homogeneous population where the mean degree is representative of the degree distribution.

Scale-free networks, on the other hand, present a high variability in the degree distribution and therefore the mean degree is not representative of the degree distribution. Simple deterministic homogeneous mixing models are not adequate to describe disease dynamics in these type of networks. We show, however, that a simple two-degrees network model may capture the main features of the dynamics.

Our work highlights that highly heterogeneous population may be modeled by a simple model with two homogeneous sub-populations where most of the individuals (or nodes), about 99% of the total population have the (same) number of contacts and it is close to the mean number of contacts ($n-1$ and n respectively). The other sub-population is composed of super-spreaders with a much higher number of contacts (n_h is about 10 times n_l).

Further refinements are expected to improve the approximations.

Acknowledgements

This work was partially supported by grants CIUNSA 2467 and PICT 2014-2476. JPA is a member of the CONICET. MR holds a postdoctoral fellowship from CONICET.

References

[1] Aparicio, J. P., & Pascual, M. (2007). Building epidemiological models from R_0: an implicit treatment of transmission in networks. Proceedings of the Royal Society B: Biological Sciences, 274(1609), 505-512.

[2] Bansal, S., Grenfell, B. T., & Meyers, L. A. (2007). When individual behaviour matters: homogeneous and network models in epidemiology. Journal of the Royal Society Interface, 4(16), 879-891.

[3] Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509-512.

[4] Brauer, F. (2008). An introduction to networks in epidemic modeling. In Mathematical epidemiology (pp. 133-146). Springer, Berlin, Heidelberg.

[5] Diekmann, O., & Heesterbeek, J. A. P. (2000). Mathematical epidemiology of infectious diseases: model building, analysis and interpretation (Vol. 5). John Wiley & Sons.

[6] Erdös, P., & Rényi, A. (1959). On random graphs, I. Publicationes Mathematicae (Debrecen), 6, 290-297.

[7] Hethcote, H. W., & Yorke, J. A. (2014). Gonorrhea transmission dynamics and control (Vol. 56). Springer.

[8] Li, L., Alderson, D., Doyle, J. C., & Willinger, W. (2005). Towards a theory of scale-free graphs: Definition, properties, and implications. Internet Mathematics, 2(4), 431-523.

[9] Lindquist, J., Ma, J., Van den Driessche, P. and Willeboordse, F.H., (2011). Effective degree network disease models. Journal of mathematical biology, 62(2), pp.143-164.

[10] Watts, D. J. & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, **393**, 440-442.

[11] Yorke, J. A., Hethcote, H. W., & Nold, A. (1978). Dynamics and control of the transmission of gonorrhea. Sex Transm Dis, 5(2), 51-56.