On the dynamics of infectious diseases in non-homogeneous populations

by

Lilia Leticia Ramírez Ramírez

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2008

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The principal motivations for studying epidemics and their dynamics are understanding the biological characteristics of the epidemic agents and reducing the economical and social costs originating from epidemic outbreaks.

The most commonly used epidemic models have important assumptions such as the law of mass action, and the latent and infectious periods being exponentially distributed with fixed parameters. Under this kind of suppositions the models are analyzed with well known algorithms as the Euler and Euler-Maruyama, and methodology and results from the theory of Markovian processes. However, these assumptions are selected largely for their analytic convenience and in many cases are far from describing the agent's transmissibility attributes in the population and its biological characteristics in a host.

The epidemic models studied here relax two important epidemic assumptions. The first to be relaxed is the one that susceptible individuals are equally likely to acquire the disease. A structure for the kind of individual contacts that can result in the infection transmission is incorporated in the population. This contact structure can be non-homogeneous and it is modeled as a random graph whose edges describe the contacts between individuals.

The second assumption that is generalized, is the distribution of the latent and infectious period in the host individuals. This research work allows the latent and infectious period to have a distribution other than the exponential and hence the epidemic process is more general than a Markovian process.

As in most stochastic models, the infectious contact is modeled as a random variable with Poisson distribution. However, to introduce the individual variations, the transmission rate is assumed to be a non negative random variable.

This work extends the epidemic models suggested by Newman (2002) in two directions. The first, studies the hierarchical networks that have a more complex network structure, involving the interaction of populations. The second direction examines the evolution in times for outbreaks in networks. In this work, results for discrete and continuous time are obtained.

The results for the continuous time model considers the infectious process to be a bivariate Markovian process. However, the results for the final outbreaks size and the developed simulation program include the general case were the latent and infectious period can have a distribution other than exponential.

This research work also analyze the effect of four control measures in the contact structure, and using the simulation program and Monte Carlo-likelihood methodology, it estimates the parameters for measles and influenza.

The results here obtained can be directly applied to study the dynamics of other kind of "agents" such as information and ideas. For example, the dynamics can involve the spread of computer viruses, rumors, eating habits and personal positions regarding a fact or idea.

Acknowledgements

To my grandmother Linda.
To my family: Beto, Carlos, Emma, Mary, Claudy and Paty.

To Juan.

The most important person who has supported not only the development of this research work, but also my growth as a researcher, is definitely is my advisor Dr. Mary Thompson. I want to express my gratitude for her encouragements, guidance, patience and excellent example.

I am also deeply indebted to Dr. Mary E. Thompson for the financial support she provided to me.

I wish to thank to my thesis committee members: Dr. Don McLeish, Dr. Yulia Gel, and Dr. Ping Yan, for their general advice, for the time they spent reviewing previous version of my thesis, and for all their helpful comments and feedbacks.

I am thankful to the Consejo Nacional de Ciencia y Tecnología (CONACyT) for the financial support granted to me. In particular I wish to thank the personnel at the Management of Grants for Students in Foreign Countries for the close contact we kept during the first four years of my PhD program.

I am grateful for the laughs, advice and support of my friends in Waterloo: Biel, Dasha, Itzel, Rosy, Wanda, Esther, Alberto, Pierre, Audrey, Oana, Alma, Gabriel and Ivan.

My thanks also go to my dearest friends in Mexico: Cris, Martha, Claudia, Rafa, Luz, Mars and Johan.

Contents

List of tables

List of figures

Chapter 1

Introduction

The reduction of the economic and social cost in a community experiencing any epidemic activity are the principal motivations that have led scientists and government to pursue the improvement of both the public health and the financial administration of epidemic control strategies.

Very early in the study of epidemics scientists understood that mathematical models can be utilized to achieve those goals, since the mathematical formulation of the infection process allows us to measure the interaction among the variables included in the model. The description of theses interaction help us not only to understand some of the underlying mechanisms which dictate the spread of a disease, but in many cases, pointing to more effective control strategies.

Deterministic and stochastic infective models for directly transmissible diseases, such as influenza, measles, HIV and SARS, have been intensively studied since the early 20th century. The most important are based on hypotheses such as that the infectious period has an exponential distribution, and that susceptible individuals within a population have the same probability of acquiring the disease when one or more infective individuals are introduced in that population.

1.1 Problem description

Since some directly transmissible infections require specific kind of contacts between individuals to propagate, some epidemics are heavily affected by the population connectivity patterns that characterize the types of contact that can result in infection transmission. For example Liljeros et al. (2001) found that the sexual contacts (sexual partners over 12 months period prior to the survey) can be described by a network in which each individual has a number of sexual partners that is power law distributed.

The effects of non-homogeneous contact rates have been observed by epidemiologists from some disease incidence data. Some monitored epidemics do not evolve uniformly through populations, making evident that the hypothesis of mass action can be very unsuitable to describe the dynamics of some infections.

In order to obtain more realistic inference about the parameters that describe an infection process, it is important to incorporate the social behavior that affects the evolution of an outbreak.

Another hypothesis that can be relaxed in deterministic and stochastic epidemic models is the one that assumes that the infected individuals are latent and infectious for a time that is exponentially distributed. This premise allows us to obtain a system of ordinary partial differential equations and Markovian processes to describe the epidemic dynamic. Then the analysis incorporates numerical simulation that can include the Euler and the Euler-Maruyama algorithms, or the use of general results for Markovian processes. Although this assumptions can lead to elegant results, it may not obey the biological characteristics of the specific agent.

1.2 Main contributions

This work generalizes some important epidemic models, relaxing the hypotheses that the population is homogeneously mixed, the infection rate is constant between any two individuals and the latent and infectious period have length that are exponentially distributed.

This research incorporates the social contact network first used by (Newman, 2002; Ancel-Meyers et al., 2003). The network describes the type of contacts that can result in the disease transmission, and it is defined in the terms of the distribution of the number of contacts that each susceptible individual has with other individuals in the community.

Newman (2002) studied the mean for the total number of affected individuals (*final outbreak size*) in large communities with simple and bipartite network structures. We migrate these results to the standard statistical notation and complement them providing the explicit expression for the variances of the final outbreak size.

We also obtain the mean and variance for the final outbreak size in outbreaks evolving in a more general hierarchical network of contacts. This structure describes a network with two different populations interacting between and within them. For this network, the simple and bipartite networks are two particular cases.

We incorporate the outbreak evolution in discrete time in each of the three different kinds of networks, obtaining the expression for the mean and variance of the number of infected individuals at time t, $t \in 1, 2, \ldots$. In the case of the hierarchical networks, the covariance of the number of infected at time t, in populations A and B is also derived.

The use of both the mean and the variance for the final outbreak size (and for the number of infected individuals at time t) allow us to define the most likely scenarios for the evolution of the outbreak. With these scenarios then the government and/or institutions can prepare and administrate the resources that are needed to handle the contingencies.

We obtain the probability generating function (p.g.f.) for continuous-time-SIR outbreaks that transmit in the networks and with infectious agent having infectious period that is exponentially distributed.

In order to include the study of SIR or SEIR outbreaks that evolve in a small contact network, (with latent and infectious period exponentially, Gaussian or Log normal distributed), we developed a computer program to simulate the transmission of an infectious agent introduced in the population.

The simulation program can include the random network structures that we study in

Chapters 3 and 4. It runs the outbreak in continuous time and extends the model in Section 4.2, to include agents with random latent and infectious period with distribution other than exponential.

The simulation algorithm was written in R (R Development Core Team, 2007) and an earlier version was conformed as the R package InfNet V.0.1.

Using the simulation program, and data sets on measles and influenza outbreaks, we obtain the Monte Carlo-likelihood point estimates for the parameters that describe the agents' infectious characteristics. In both cases we consider that the infection process occurring in the population is not Markovian, since the infectious period is assumed to be Gaussian(μ_L, σ_L).

Finally, four control measures are studied for infectious outbreaks that transmit according to a network of contacts.

1.3 Organization of thesis

This thesis is organized as follows. Chapter 2 introduces some of the most relevant epidemic terminology to be used in the rest of the work, while describing some of the most important early epidemic models. This chapter also presents important definitions in network theory and describes the most relevant networks due to the fact that they have been used to approximate several systems, including the World-Wide Web and sexual networks.

Chapter 3 contains the results obtained by Newman (2002), the variances for the final outbreak size and the mean and variance for the hierarchical networks.

In Chapter 4 we study the evolution of the outbreaks in networks of contact over time.

Chapter 5 studies the effect of three different vaccination strategies and isolation as control measures.

Finally, in Chapter 6 we analyze two different agents. They correspond to measles and influenza, and using information from several outbreaks and simulations we estimate their model parameters.

Appendix A contains the description of the utilized simulating program and Appendix B presents the complete code written in R (R Development Core Team, 2007).

Chapter 2

Background

In this chapter we introduce epidemic terminology, while describing some important early epidemic models and results. Section 2.1 presents some of the most relevant models and results obtained since the first half of the 19th century.

Section 2.2 describes the most important characteristics of direct transmissible diseases to consider in this work. They are presented in terms of the stages any individual susceptible to the infection develops with respect to transmissibility and symptoms.

Based on the infection stages, Section 2.3 presents in more detail three classical epidemic models, and some of their results that will be compared with the models obtained in the following sections.

Finally, Section 2.4 introduces some basic definitions in random graphs that will be used in the rest of this work, and also describes some special social networks that have been studied and applied in diverse practical situations.

2.1 The early history of mathematical epidemic models

As specified by Anderson and May (1991), Daniel Bernoulli appears to have initiated the mathematical study of infectious diseases. With the purpose of influencing public health policy, he developed a method to evaluate the effectiveness of the techniques of *variolation*

against smallpox (Bernoulli, 1760).

After this work there was a long period without significant progress in the study of infectious diseases. The next relevant works arose at the second part of the nineteenth and beginning of the twentieth centuries.

After mapping both the reported deaths from cholera in London during 1854, and the location of water pumps, John Snow realized that the cholera epidemic was being spread through the water supply. Following a similar analysis, in 1873 William Budd established that the typhoid fever in North Tawton (UK) was being transmitted through the sewage systems. The infected cases visualization helped Snow and Budd to detect a relationship between the cases and a common infection source. The basic idea that the closer individuals are to each other (or to an infectious element), the more they affect each other (or the more they are affected by the source), is one of the premises in spatial analysis.

In 1840 William Farr analyzed data on deaths from the England smallpox epidemic of 1837-1839. He observed the periodicity of disease epidemics and with the aim to describe this pattern mathematically, developed a prediction technique based on the second ratios of the incidence data. Farr smoothed the observations with the moving average of quarterly deaths and calculated the second ratios of death in successive quarters. Since he observed that these quantities were approximately equal, he could obtain the expected series of deaths assuming the second ratios were constant.

The method to predict the end of an epidemic proposed by Farr is equivalent to fitting a Gaussian distribution to the smoothed data (Brownlee, 1906), but Farr's approach was inconsistent for real epidemic data. In his analysis of the cattle plague in England of 1865, he utilized the third ratios instead of the second ratios as in 1840. Also Evans (1875) tried to use Farr's method on the English smallpox outbreak of 1871-1872 with little success (Ross, 1916).

In 1906 John Brownlee published *Statistical studies in immunity: the theory of an epidemic* in which he fitted several Pearson curves (mostly Type IV) to different diseases' epidemic data. Some of the data used by Brownlee were the number of new cases and deaths

for miliary fever, cholera, influenza, smallpox and measles, and the number of observations varied between 40 (enteric fever in Coventry, 1900) to 55,000 (deaths in London plague, 1665). In further works (from 1909 to 1918) Brownlee concluded that a Pearson Type IV frequency curve would fit most of his observations.

The curves of the epidemics Brownlee studied were typically either symmetrical or slightly positively skewed, but according to Hamer (1906), apparently Brownlee was familiar with the expectation that epidemics are negatively skewed, a consequence of the accepted hypothesis that the progress of an epidemic is regulated by the decreasing number of susceptibles and the decreasing rate of contact between the infected and non infected individuals. Brownlee then attributed the lack of negative skewness to biological infectivity change in the agent.

Brownlee's biological assumption was criticized by Sir Ronald Ross, who doubted that the agent changed its infectivity power. Nevertheless Ross conceded the possibility of change in the infection due to combinations of variation in environmental conditions, habits of human and vector populations.

In contrast to the descriptive approaches of Farr and Brownlee, who expected to deduce some underlying biological causes or laws which would lead to the regular recurrence of the epidemic, Ross constructed a mathematical model based on accepted biological/epidemiological relationships.

In 1911 Ross published in *The Prevention of Malaria* a difference equation model for malaria based on the proportion of individuals only affected with malaria (infected stage), the proportion of individuals affected and also infective (infective stage), and the rate of recovery among the affected. For the vector (*Anopheline mosquito*) Ross included the number of local mosquitoes capable of carrying malaria, the proportion of these which would succeed in biting an infected person, the proportion of the latter which would succeed in maturing gametes, and of these, the proportion which would succeed in biting a uninfected person.

In this work, Ross also introduced the still important concept of *reproduction ratio* or

reproduction number (denoted as R_0), that is, the (expected) number of secondary cases produced by an infected individual in a large susceptible population. The obtained model determined that there was a threshold density of man and mosquitoes below which malaria would become extinct. Although Ross used the idea of chance or probability in formulating his basic equations, they are completely deterministic.

Some other early advances were made by Hamer (1906), Kermack and McKendrick (1927-1933 and 1937-1939) and Soper (1929).

Hamer (1906) postulated that the evolution of an epidemic depends on the constant rate of interactions among individuals, that dictates the rate of contact between the susceptible and the infectious populations. This notion is still one of the most important concepts in mathematical epidemiology (Anderson and May, 1991) and is referred as the *law of mass action* or *homogeneous mixing*.

The most important result obtained by Soper was the discovery that the basic assumptions on the contact rate, recovery, etc. lead to periodic but damped oscillations in the number of infected individuals. (Bailey, 1975).

Observed data in epidemics do show a marked oscillation in incidence but they do not show a tendency to damp (Wilson and Worcester, 1945). Since deterministic models failed to describe this persistent behavior, other researchers explored incorporating stochastic concepts, that led to stochastic version of the deterministic models.

Another important reason why probabilistic models became increasingly necessary is that given a set of initial conditions, a deterministic model describes only one infectious evolutionary process of the epidemic. Since there are several real life variations that can affect the evolution and final result of this process, such as variation within the population structure, environmental characteristics, etc., it is evident that in the deterministic model the possibility of observing deviations was neglected (Serfling, 1952).

According to Bailey (1975), McKendrick (1926) was apparently the first to obtain a genuinely stochastic treatment of an epidemic process in continuous time, but the lack of satisfactory methods for handling such models had much to do with the fact this approach

did not attract much attention.

Greenwood (1931) proposed a model in which the period of infectiousness was short and the latent and incubation periods were approximately constant. This model can be applied to diseases such as mumps that has an incubation period of 14-18 days (range, 14-25 days), and maximum transmission rate between the period extending from 1 to 2 day before the onset of symptoms to 5 days after (PHAC). This disease's characteristics result in generations of new cases evolving from a single case in a closed group.

Then under certain conditions and given the number of susceptible and infectious individuals present at a previous stage, and the chance of infectious contact, the generations followed a chain of binomial distributions. In 1928 Lowell J. Reed and Wade Hampton Frost were already using the same kind of ideas related to epidemics by generation as well, in lectures and discussions, but they did not think their models were worth publishing. The method nowadays known as the Reed-Frost model was described in 1951 by Helen Abbey. See Section 2.3.1.

As did some other authors in the first half of the twentieth century, Abbey continued extending and re-examining the previous epidemic models. She analyzed the epidemics of measles in families and explained the discrepancy between the calculated results from the Reed-Frost model and the observed results, in terms of differences in within-family contact rates.

In 1946 and 1949 Maurice Bartlett revisited the idea of *continuous-infection* introduced by McKendrick 20 years earlier. He obtained a partial differential equation for the probability generating function of the transition probabilities. The continuous time transitions considered two variables, the numbers of susceptible and infectious individuals. He basically developed a bivariate Markov SIR model based on the Kermack and McKendrick models (see Section 2.3.2) and observed that for large numbers of susceptibles the deterministic mean would approximate the curve of the stochastic mean (Serfling, 1952).

The developments of deterministic and stochastic models for epidemics since then have been evolving very closely. According to Isham (2004), almost all stochastic epidemic

models have a deterministic counterpart obtained by setting the deterministic population variations to be the expected values of the conditional variations in the stochastic model.

Bailey (1975) made the point that the deterministic treatments would normally be regarded as approximately valid formulations under certain circumstances (for example, an outbreak occurring in a population with a large size) and recommended examining the deterministic model consequences before passing to more precise stochastic formulations.

In 1950 Norman Bailey drew an analogy between the stochastic process and the fact that in a population the epidemic curve is the sum of a large number of epidemics in small groups of the population. Using the same deterministic model as Bartlett, he calculated the expected course of an epidemic according to the deterministic equation. The deterministic curve was symmetrical, whereas the stochastic mean was positively skewed in accordance with Brownlee's observations.

Another important advance was the obtaining of a stochastic *threshold theorem* (Whittle, 1955) in which a set of stochastic statements replaced the rigidly specific alternatives described by Kermack and McKendrick in the first version of this result.

Bartlett in 1953, 1956 and 1957 studied the deterministic models proposed by Hamer and Soper in which they obtained unrealistic damping waves. He showed that by adopting a stochastic model the recurrent epidemics may be modeled much more realistically. In sufficiently small communities complete fade-out of infection may occur if new cases are not introduced, whereas in communities larger than a certain critical size, the infection reaches a low level for a time before building up for a new outbreak. This conclusions are in strong agreement with observed and simulated data (Bailey, 1975).

2.2 Infectious diseases

As noticed by Ross, it is important to understand and incorporate into the model the biologic course of an infection and the disease the agent induces. In this section we present the infectious and disease stages that most infectious diseases develop.

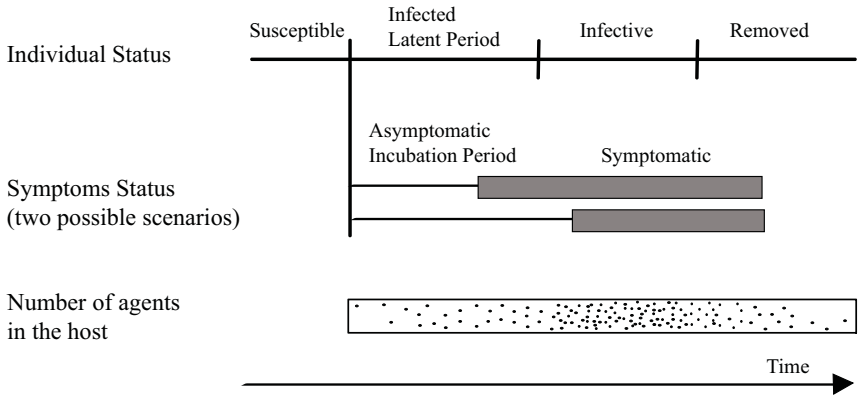


Figure 2.1: *Infection-disease evolution.*

It is usual to divide the infection course into a series of stages that start with a susceptible state. See Figure 2.1. After the host becomes infected, the agent replicates inside, so that the host becomes able to transmit the infection to others. The period between being infected and being infectious is known as the *latent period*.

Close to the evolution of the infection is the disease evolution. The period between being infected and having any agent-related disease symptoms is referred to as the *incubation period* and should not be confused with the latent period since the host can become infective before or after having any symptoms.

An *infectious contact* or transmission occurs when given a contact between a susceptible and an infective, the agent transmits through that contact to the susceptible individual. The definition of the term 'contact' is according to the transmission mode for the specific agent under study.

The removal stage represents death caused by the infection or the acquisition of natural immunity that prevents reinfection. For some diseases the removal rate does not exist since the individual reenters the susceptible stage after recovery.

In general the diseases transmitted by viral agents, such as influenza, measles, rubella and chicken pox confer immunity against reinfection, while the diseases transmitted by bacteria, such as tuberculosis, meningitis and gonorrhea, confer no immunity against rein-

fection.

2.3 Basic mathematical models for epidemic

In this section we present three classical epidemic models and some of their results. A comprehensive comparison of the most important models, the variables they consider, and their relevant results are presented in Yan (2007).

Most of the stochastic models for epidemics still refer to the models obtained in the second half of the nineteenth and first half of the twentieth century. According to the infectious disease discrete stages included in the model, the most important models (usually called compartmental models) are:

1. SIR Models: Consider three different subpopulations at time t: X_t, Y_t and Z_t, are respectively the numbers of susceptible, infected and removed individuals. The infected individuals are capable of transmitting the infection and they are referred to as *infectives*. The *removed* population is that part which recovers and becomes immune to the infection or dies, or for some other reason is not important to the transmission of the infection process.

2. SEIR Models: These are similar to SIR models but they include an extra stage or population: individuals that are infected or exposed but do not transmit the infection (latent period).

3. SIS Models: In these models the infectives return to the susceptible state after recovery.

2.3.1 Reed-Frost model

This model might be classified as the simplest SEIR stochastic model for epidemics. The model considers a closed population with the susceptible, infective and removed subpopulation sizes in discrete times $t = 1, 2, 3, \ldots$. The infection is passed from one infective to one

susceptible in a relatively short time interval $(t, t+\epsilon)$ at the beginning of the period. The newly infected individuals Y_{t+1} will themselves become infectious in $(t+1, t+1+\epsilon)$, while the current infectives Y_t will be removed. In discrete time models like this, it is natural to think of the infectious period as being short and preceded by a longer and almost constant latent period.

Each susceptible is assumed to have the same probability $0 \leq q \leq 1$ of not having an infectious contact with some infective. Then the probability that a susceptible at time t remains susceptible in the next period is q^{Y_t}.

The new infections occur in generations that are separated by the latent period as the discrete time unit. Since the event probabilities in a given generation depend only on the state of the epidemic in the previous generation, the model describes a Markov process with binomial transition probabilities

$$P(Y_{j+1} = y_{j+1} | X_j = x_j, Y_j = y_j) = \binom{x_j}{y_{j+1}} (1 - q^{y_j})^{y_{j+1}} (q^{y_j})^{x_j - y_{j+1}},$$

and $X_{j+1} = X_j - Y_{j+1}$. Of course, the infectious transitions occur until $\phi_1 = \inf\{t | Y_t = 0\}$ or $\phi_2 = \inf\{t | X_t = 0\}$, and the *final outbreak size*, is $\sum_1^{\phi_0 \wedge \phi_2} Y_t$.

Given the initial state $X_0 = n$ and $Y_0 = m$ the probability distribution of the complete chain $y_1, y_2, \ldots$, can be obtained by conditioning sequentially and using the Markov property of the chain. This can be very useful for a small population, but finding these distributions becomes very complicated to compute even for moderate seized populations.

2.3.2 Kermack-McKendrick model

The basic deterministic time-continuous compartmental models to describe the transmission of communicable diseases are contained in a sequence of three papers by Kermack and McKendrick in 1927, 1932 and 1933. The first of these papers describes the general epidemic model, for which the so called Kermack-McKendrick epidemic model is actually a special case.

The model known as the Kermack-McKendrick epidemic model characterizes the epidemic with the number of susceptibles X_t, infectives Y_t and immunes or removed Z_t at time $t \geq 0$. This model supposes the time scale of the disease to be much faster than the time scale of birth and deaths, so that demographic effects on the population may be ignored and the total population $N_t = X_t + Y_t + Z_t$ remains constant ($N_t = N$).

If we assume that each individual in a population mixes homogeneously, the rate of interaction between two different subsets of the population is proportional to the product of the numbers in each of the subsets concerned. Then under the law of mass action the evolution of the epidemic is defined by the equations

$$\frac{dX_t}{dt} = -\frac{\beta}{N} X_t Y_t, \tag{2.3.1}$$

$$\frac{dY_t}{dt} = \frac{\beta}{N} X_t Y_t - \gamma Y_t, \tag{2.3.2}$$

$$\frac{dZ_t}{dt} = \gamma Y_t, \tag{2.3.3}$$

subject to the initial conditions $X_0 = x_0$, $Y_0 = y_0$ and $Z_0 = 0$. The parameter β is called the standard incidence rate and represents the infection rate per unit of time, describing the number of infectious contacts made by one infectious individual. When N is fixed β can be replaced by $\alpha = N\beta$, that is called the mass action incidence rate.

The parameter γ is the removal rate that describes the speed at which infectives become immune. In cases where death or isolation may occur, Z_t represents all removals from the population, including individuals that become immune, die or are isolated.

The law of mass action can then be seen as the result of superposing all possible contributions of the individual components to the interactions; the individuals are regarded as being equally likely to interact with each other in a given (small) interval of time.

Other similar models based on the idealistic law of mass action can be built and analyzed, but not even the simplest case presented can be analytically solved. Some results derived from this kind of model involve qualitative analysis. The two most important are presented next.

Threshold theorem

Kermarck and McKendrick obtained two basic results, referred to as their threshold theorem. The first is a critical statement, and comes from equation (2.3.2). Writing this equation as:

$$\frac{dY_t}{dt} = \frac{\beta}{N} Y_t (X_t - N\rho), \tag{2.3.4}$$

where ρ is the relative removal rate γ/β, shows that the epidemic can grow (Y_t can increase), if and only if $X_t > N\rho$ or $X_t/(N\rho) > 1$.

The quantity $\mathcal{R} = X_0/(N\rho)$ is known as the *effective reproduction number* or *replacement number* and represents the total number of secondary infections produced by a typical infective.

Then, based on the initial conditions it is possible to know if the outbreak will grow to become an epidemic or not. If $\mathcal{R} > 1$, Y increases to a maximum attained when $X = N\rho$ and then decreases to zero (epidemic scenario), and if $\mathcal{R} < 1$, Y decreases to zero and the infection dies out before reaching a substantial fraction of the population.

The term $\mathcal{R}_0 = 1/\rho$ is the average number of secondary infections that occur when one infective is introduced into a completely susceptible population. Then $\mathcal{R} = \mathcal{R}_0 X_0/N$ and when Y_0 is very small $X_0 \approx N$ and the basic reproduction number $\mathcal{R}_0$ is a good approximation to $\mathcal{R}$.

It is natural to think that the immunization of a single individual not only protects the individual but also indirectly protects others against the possibility of disease transmission from this individual. The protection of an entire population via the immunity of a fraction of the population is known as *herd immunity*.

Thanks to a result derived from (2.3.4) it is possible to obtain an estimation of the immunization proportion required to achieve the herd immunity. The minimum size of a susceptible population necessary for an epidemic to occur is defined by the threshold $\mathcal{R} = 1$, reached when

$$X_0 = N\frac{\gamma}{\beta} =: X_{\text{th}}.$$

Then a pathogen will go extinct if the size of the susceptible population is less than this threshold X_{th}. If a fraction p of the population is immunized, $\mathcal{R}$ is reduced to

$$\mathcal{R}^* = (1-p)\mathcal{R}, \tag{2.3.5}$$

and according to this model, if $\mathcal{R} > 1$, the herd immunity is attained when $p > 1 - 1/\mathcal{R}$.

Extension of this basic model have been used to predict the minimum immunizaion coverage necessary to drive specific diseases to extinction (Ancel-Meyers, 2007). For example, according to Anderson and May (1991), the respective required coverages are thought to be 90-95% for measles and whooping cough; 85-90% for chicken pox and mumps; 82-97% for polio and scarlet fever; and 70-80% for smallpox.

Uninfected susceptible population

The second result from the Kermarck-McKendrik model is obtained from (2.3.1) and (2.3.2):

$$\frac{dY}{dX} = -1 + \frac{N\rho}{X},$$

and integrating the last expression with respect to X, we can obtain the function

$$Y(X) = -X + N\rho \ln(X) + c,$$

that can be expressed as the orbits function

$$g(X,Y) = X + Y - N\rho \ln(X)$$

where c, the constant of integration determined by the initial values, is $c = g(X_0, Y_0)$.

The maximum value of Y_t on each of these orbits defined by $g(X,Y)$ is attained when $X = N\rho$, and none of the orbits reaches the Y-axis, for any t. In particular, $\lim_{t\to\infty} X_t = X_\infty > 0$. This implies that part of the population always escapes infection. At the time this was a significant result and still nowadays, together with the threshold theorem, is a typical result to compare with when using new mathematical epidemic models.

2.3.3 Stochastic model based on the Kermack-McKendrick model

When the number of individuals is very large, it was customary to represent the infection process deterministically, but it was already known that these models were very unsuitable for small populations.

Since the spread of a disease is naturally a random phenomenon, a stochastic model is a more appropriate model for any epidemic dynamic. In this section we consider the stochastic model obtained by McKendrick (1926) and further developed by Bartlett in 1949 and 1955.

The immediate extension to the Kermack-McKendrick model is its stochastic version obtained by setting the population increments equal to the expected values of the conditional increments in the stochastic model. For example we write

$$E(Y_{t+dt} - Y_t | (X,Y)_t = (x,y)) = \left(\frac{\beta}{N}x - \gamma\right) y dt + o(dt).$$

The solution of the deterministic equations is not simply the mean of the stochastic process since the expected value of the population increments can include covariance terms the determinist model does not include. For example, from the last equation we have

$$\begin{aligned} \frac{dE(Y_t)}{dt} &= \lim_{dt\to 0} E(E(Y_{t+dt} - Y_t | (X,Y)_t = (x,y)))/dt \\ &= \left(\frac{\beta}{N}E(X_t) - \gamma\right) E(Y_t) + \frac{\beta}{N}\mathrm{Cov}(X_t, Y_t). \end{aligned}$$

When the rates in the SIR deterministic model (2.3.1)-(2.3.3) are interpreted as the infinitesimal transition probabilities

$$\begin{aligned} &\Pr\{(X,Y)_{t+dt} = (x-1, y+1) | (X,Y)_t = (x,y)\} = \tfrac{\beta}{N} xy dt + o(dt), \\ &\Pr\{(X,Y)_{t+dt} = (x, y-1) | (X,Y)_t = (x,y)\} \quad = \gamma y dt + o(dt), \\ &\Pr\{(X,Y)_{t+dt} = (x,y) | (X,Y)_t = (x,y)\} \quad = 1 - \left(\tfrac{\beta}{N}x + \gamma\right) y dt - o(dt). \end{aligned}$$

we have a bivariate Markov processes that assumes the mass action and in which the number of individuals in each state are the random variables (X_t, Y_t, Z_t) with initial values $(X_0, Y_0, Z_0 = 0)$.

In this model each susceptible individual makes βy infectious contacts per unit of time and the infectives leave that state at mean rate γy per unit of time. This last assumption is the equivalent to supposing that the "cohort" of members who were all infective at one time stay in that new state for a period that is distributed exponentially with mean $1/\gamma$.

Looking to the expression $\mathcal{R}_0 = \beta/\gamma$ it is clear that in the context of this stochastic model the basic reproductive number is equal to the mean number of infectious contacts during the mean time of being infective. As in the deterministic model, the quantity $1/\rho$ coincides with the basic reproduction number $\mathcal{R}_0$ that is close to $\mathcal{R}$ when $X_0 \sim N$ and Y_0 is small.

The model can be regarded as a simple birth and death process in which being born is equivalent to becoming infected. At the beginning of an epidemic, when the infective and removed populations have very few members, then the probability that the epidemic ends quickly (no epidemic) can be approximated by the simple birth and death process result:

$$\Pr(\text{no epidemic}) \;=\; \begin{cases} 1, & \mathcal{R}_0 \leq 1 \\ (1/\mathcal{R}_0)^{Y_0}, & \mathcal{R}_0 > 1. \end{cases} \tag{2.3.6}$$

Kendall (1956) explained Bailey's findings about the distribution of the total number of secondary cases and the distribution of the final outbreak size. Bailey found two different kinds of distributions that generalize the Kermack-McKendrick threshold theorem. If $X_0 \leq \rho$ the probability distribution for the final size is decreasing (*J-shaped*) and if $X_0 > \rho$ the distribution is bimodal with one peak at 0 and the other corresponding to a major outbreak (*U-shaped*).

Unlike in the deterministic model, the threshold theorem of its stochastic version indicates whether the probability of having an epidemic is zero or not, so it does not predict an epidemic with certainty if $\mathcal{R}_0 > 1$.

2.4 Networks

In this section we introduce the basic definitions in random graphs theory and define the small world and scale-free networks which have been fitted in diverse practical situations.

2.4.1 Random graphs

A *graph* G consists of an ordered pair (V, E) of vertices (points or nodes) $V = \{v_1, \ldots, v_n\}$ and edges (lines or links) that connect pairs of vertices, $E \subset V^2$. The number n is called the *order* of G and the *size* of G refers to the number q of edges. We say that a graph G is labeled if the vertices have fixed identities.

Since several structures can be represented as graphs (electric circuits, communication paths, etc.), the properties of graphs have been extensively studied since the mid nineteenth century. The graphs can be classified as directed (or digraph), weighted, etc.

The mathematicians Paul Erdős and Alfréd Rényi (1959; 1960; 1961) took a particular approach to the study of communication networks. Together they introduced the theory of random graphs.

A *random graph* is the name given to a graph in which the network nodes are randomly connected by edges.

In the original random graphs proposed by Erdős and Rényi, the n nodes are joined by edges which are placed between pairs of vertices chosen uniformly at random, and they provided a number of versions of their model. The most commonly studied graphs $G_{n,p}$ were those in which each possible edge between two vertices is present independently with probability p, and absent with probability $1 - p$.

The *degree of a node* ($\mathrm{d}_G(x)$, $x \in V$) is the number of edges connected to that particular vertex, and the graph is called *regular* if each of its vertices has the same degree.

Random graphs are not regular, but they can be characterized by a probabilistic distribution that describes their nodes' degrees. This distribution gives the probability that a randomly selected node has exactly k edges. That is

$$\Pr(K = k) := \Pr(\mathrm{d}_G(x) = k | x \text{ selected at random}).$$

In the random graphs $G_{n,p}$ the majority of nodes have approximately the same degree, close to the *average degree* $E(K) = (n-1)p \approx np$ and distributed approximately according to the Poisson distribution with parameter $E(K)$.

Following the finding of Erdős and Rényi, random graphs were extended according to their degree distribution. Then random graphs can be modeled and generated using degree distributions in any family of probability distributions. There are several contributors to the topic of degree distribution, including Erdős, Rényi, Ivčhenko, Bollobás (1985) and Newman et al. (2001).

Random graphs have been employed as models of real-world networks of various types, like social, manufacturing, telecommunication and computer networks. Two important dynamics studied considering network characteristics are information access in the World-Wide Web and the evolution of an epidemic.

An important characteristic of a graph that is closely related with the degree distribution is called a *component*. A component in a graph is a subset of vertices each of which is reachable from the others by some path through the network. For small values of $E(K)$, when there are few edges in the graph, most of the vertices are disconnected from one other and the components that can be built have small sizes. However, there is a critical value of $E(K)$ above which, in an asymptotic sense, the one largest component in the graph contains a positive fraction S of the total number of vertices. This largest component is called a *giant component* (Solomonoff and Rapoport, 1951; Erdős and Rényi, 1960). In general there will be other components in addition to the giant component, but these are still small and with fraction size $S_0 \to 0$ as $n \to \infty$.

Another important characteristic of graphs is the set of *distances* $d(i, j)$ between pairs of vertices i and j. The distance $d(i, j)$ is defined as the minimum number of edges that must be traversed in order to reach vertex j from vertex i. That is, the distance is the shortest path length between i and j, (Watts, 1999).

The *diameter* of a graph G, $\text{diam}(G)$, is the maximal distance between pairs of vertices of G. A graph that is disconnected is defined to have a diameter equal to ∞.

The *neighborhood* of a vertex v_0 is the set of first order neighbors or the set of nodes that have an edge to the vertex v_0.

Another graph characteristic is the *cluster coefficient* introduced in 1998 by Watts and

Strogatz to quantify the *clustering* or *transitivity*. The clustering is a measure of the probability that two vertices will be connected directly if they have a common neighbor.

First, the *clustering coefficient for a vertex* v_0 is the proportion of links between the vertices within its neighborhood divided by the number of links that could possibly exist between them.

Then the *clustering coefficient* for the graph is defined as the average of the clustering coefficient for each vertex in the graph. Newman et al. (2001) and Newman (2003) noticed that the cluster coefficient defined this way does not correspond to the mean probability that two nodes with a common neighbor are neighbors, but the correct average probability is

$$\begin{aligned} C &= \frac{3 \times (\text{number of triangles on a graph})}{\text{number of connected triples of vertices}} \\ &= \frac{6 \times (\text{number of triangles on a graph})}{\text{number of paths of length 2}} \end{aligned} \tag{2.4.7}$$

where a triangle means three vertices that are each connected to both of the others, connected triple means a trio of vertices in which at least one is connected to both the others and path of length 2 is any three distinct vertices v_1, v_2 and v_3 for which there are edges $\{(v_1, v_2), (v_2, v_3)\}$ in the network.

The value of the clustering coefficient in a fully connected graph (i.e. where every pair of nodes are neighbors) is $C = 1$. Since on random graphs, the probability that any two vertices are connected is $E(K)/(n-1)$, where K is the degree distribution and n is the order of the graph, then the cluster coefficient is also $E(K)/(n-1)$.

2.4.2 Social, small world, and scale-free networks

A *social network* is a representation of relevant relationships between individuals. This map consists of a graph where the nodes represent the individual units (persons, institutions, etc.) and an edge between two nodes stands for the existence of the relationship of interest.

According to Wasserman et al. (2005), the increasing interest in the study of social networks and its methodology experienced after 1990 can be attributed to the realization

of the impact of the "social context" in behavioral sciences.

The definitions previously given for random graphs then can be translated into the language of social networks. For example, the distance between two individuals is the degree of separation or number of acquaintances away from each other, and the expression (2.4.7) for the clustering coefficient (also known as the fraction of transitive triples) measures the extent to which two people are more likely to be acquainted with one other if they have another common acquaintance.

Watts and Strogatz (1998) raised the possibility of constructing random graphs that have some of the important properties of 'real-world' networks. The real-world networks they studied included neural networks, the power grid of the western United States and the collaboration graph of film actors.

Watts and Strogatz noticed that these networks had diameters considerably smaller than those of regularly constructed graphs, such as lattices or grid graphs. More precisely, Watts and Strogatz found that real-world networks tend to be highly clustered ($C >> C_{\text{R.G.}} = E(K)/n$), like regular lattices, but have small diameters, like random graphs.

That large social networks have rather small diameters had been noticed considerably earlier, by psychologist Stanley Milgram (Milgram, 1967) and this phenomenon is called *the small-world effect*.

Watts and Strogatz defined a network to be a *small-world network* if it shows both properties: the diameter $\text{diam}(G)$ is comparable with that of a random graph (that is $\text{diam}(G)/\text{diam}_{\text{R.G.}}(G) \approx 1$), and the clustering coefficient C is much greater than that for the random graph ($C/C_{\text{R.G.}} >> 1$).

It is important to notice that the term *small-world network* has been used inconsistently by other authors. Some authors use it to mean specifically networks taking the form of the Watts-Strogatz model and some others use it to mean networks that show the small-world effect (Newman et al., 2006).

Amaral et al. (2000) suggested a general classification for the small-world networks (in the sense given by Watts and Strogatz, 1998). Using the empirical analysis of several

networks, the authors classifed this kind of networks as *single-scale, broad-scale* or *scale-free* (SF), depending on their degree distribution.

The single-scale networks are characterized by a connectivity distribution with a fast decaying tail, such as exponential or Gaussian, while the broad-scale networks (or truncated scale-free networks) are described by a connectivity distribution that has a power law regime followed by a sharp cutoff, like an exponential or Gaussian decay of the tail.

Finally, the scale-free networks are characterized by a connectivity distribution with a tail that decays as a power law. It is the last kind of network that has found more interest since it describes several kind of interactions and they emerge in the context of a growing network.

The work of Watts and Strogatz inspired the study of large-scale networks by random graphs and for a large number of networks the degree distribution is well described with a discrete *power law distribution*

$$\Pr(K = k) \propto k^{-\delta}; \qquad \delta > 1, \quad k \in \{1, 2, \ldots\}.$$

This distribution has been used to approximate several systems, including the World-Wide Web (Albert et al., 1999), Internet (Faloutsos et al., 1999), metabolic and protein networks (Jeong et al., 2000), language (Ferrer i Cancho and Solé, 2001) or sexual networks (Liljeros et al., 2001).

Barabási and Albert (1999) named the networks with power-law degree distribution as *scale-free* (SF) networks since their degree kernel distribution function $p_k := \Pr(K = k)$ remains unchanged when scaling k with any constant a. That is

$$p_{ak} \propto a^{-\delta} k^{-\delta} \propto p_k.$$

The identification of networks with power-law degree distribution has allowed the development of new studies of the dynamics that are naturally associated to network structures. Barabási and Albert (1999) also pointed out that this kind of network can potentially model generic properties of many real networks, and they proposed that the properties of these

networks can be explained using a model in which a network grows dynamically, rather than being a static graph.

The network growing process that Barabási and Albert (1999) outlined gives as a result a SF network with parameter 2.9 ± 0.1 and average connectivity $2m$. Starting with a small number (m_0) of connected nodes, in each step a new node with degree $m \leq m_0$ is introduced in the network and attached independently to m existing nodes. The new node is connected to node i that has degree k_i, with probability $k_i / \sum_j k_j$. This preferential attachment means that the most connected nodes are more likely to increase, even more, their degree at every step.

There are major topological differences between random graphs $G_{n,p}$ and SF networks. For the former most nodes have approximately the same number of links $E(K)$ since the exponential decay of the distribution p_k guarantees the absence of nodes with appreciably more links than $E(K)$. In contrast, the power-law distribution implies the existence of numerous nodes with only few links, and few nodes with a very large number of links. Thus SF networks are extremely heterogenous.

Chapter 3

Final outbreak size and probability of epidemic

On the basis of epidemic data, epidemiologists realized that epidemics do not evolve uniformly through populations, making evident that the law of mass action can be very unsuitable to describe the dynamic of some infections.

Since some infections require very specific contact between individuals to propagate, their epidemics are heavily affected by the population connectivity patterns that characterize the type of contact that can result in infection transmission (Bailey, 1975; Diekmann and Heesterbeek, 2000; Anderson and May, 1991).

Some epidemic models map this contact pattern in terms of random graphs or networks. The nodes in the graph represent individuals or units susceptible to becoming infected and transmitting the illness (such as wards, facility's units, cities, etc.). The links between them represent the kind of contacts that can lead to a transmission of the infection between two individuals (Newman, 2002).

This chapter presents some definitions used in the context of random graphs and some results that describe the infectious outbreak evolution in a population with a network contact structure among individuals, such as the probability of an outbreak to develop into a big outbreak (epidemic), and variance and mean for the size of the total affected population.

Sections 3.1 and 3.2 present the two network structures studied by Newman (2002). The author obtained the mean number of infected individuals during an outbreak and the probability that the outbreak evolves into an epidemic. This derivation is presented in statistical notation, and we obtain the expression for the variance of the final outbreak size.

Newman (2002) also showed that the degrees of individuals who are infected have a larger mean than the mean degree distribution for the nodes in the graph. In Section 3.1 we present an alternative proof of this fact.

In Section 3.3 we extend the results for the simple and the bipartite network to a hierarchical network with two populations.

3.1 Simple random graph

We call *simple random graphs* those graphs that are completely characterized by their order and degree distribution. In these graphs a node can be connected to any other node in the graph and the nodes' degrees have the same distribution.

3.1.1 Homogeneous and independent transmission rates and infectious periods

Consider a pair of individuals who are connected, one of whom i is infective and the other j is susceptible. Suppose that the rate of infectious contacts between them is a random variable R_{ij}, and that the infective individual remains infective for a time I_i (infectious period). Then the conditional probability that the individual i transmits the infection to j during the entire period of infectiousness is

$$\Pr\left(\text{disease is transmitted from } i \text{ to } j | R_{ij} = r, I_i = l\right) = 1 - \mathrm{e}^{-rl}.$$

As a first case, assume that $\{R_{ij}\}$ and $\{I_i\}$ are two independent series of i.i.d. random variables with distributions F_R and F_I. Then the probability that an infective transmits

the disease to a connected susceptible (*transmissibility*) is

$$\pi := P(\text{disease is transmitted}) = \int_0^\infty \int_0^\infty \left(1 - \mathrm{e}^{-rl}\right) dF_R(r) dF_I(l) = E_{(R,I)}\left(1 - \mathrm{e}^{-RI}\right).$$

Newman (2002) commented that π lies between 0 and 1, and showed this using some simulations; however he omitted remarking that π is actually a probability.

An *occupied edge* is an edge of the network for which the disease is transmitted. Then an edge is occupied with probability π and the size of the outbreak would be the size of the cluster formed with occupied edges.

Let K be the degree for the nodes in the graph and $\{p_k\}$ the degree distribution of a randomly chosen vertex. Its probability generating function (p.g.f.) is then

$$G_K(s) := \sum_{k=0}^{\infty} p_k s^k.$$

If instead of choosing a vertex, we randomly choose an edge, we can derive the p.g.f of the degree of one of the vertices at the end of the edge. Since the probability of thus selecting a vertex of degree k is equal to $kp_k/E(K)$, then the p.g.f. of that distribution is equal to

$$G_{K_e}(s) = \sum_{k=0}^{\infty} \frac{kp_k}{E(K)} s^k = \frac{s}{E(K)} \frac{d}{ds} G_K(s) = \frac{sG'_K(s)}{E(K)}. \tag{3.1.1}$$

If we exclude the edge we arrived along, then the *excess degree* K_1 has p.g.f.

$$G_{K_1}(s) := \sum_{k=1}^{\infty} \frac{kp_k}{E(K)} s^{k-1} = \frac{G'_K(s)}{E(K)}. \tag{3.1.2}$$

Similarly, we can define the p.g.f. of the nodes' degree of occupied edges K_T (*occupied degree*) and excess degree of occupied edges K_{T1} (*occupied excess degree*) at the end of the outbreak.

To illustrate these two concepts consider Figure 3.1. Here an outbreak started with individual i_1 (patient zero). Then this individual was able to transmit the infection using all its edges, and at the end of its infectious period individual i_1 had infected nodes i_2, i_4, i_5 and i_6.

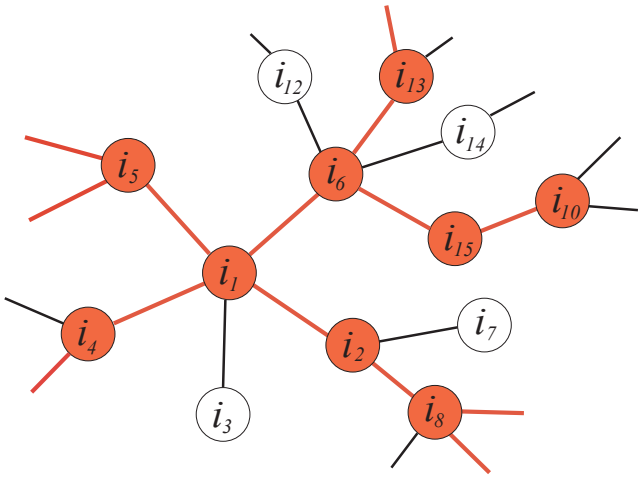


Figure 3.1: *Transmission in network.*

Since one of the edges of i_6 and i_2 were already used to infect them, then these nodes were capable to transmit the illness through one of the edges of their excess degree. By the end of their infectious period node i_6 infected i_{13} and i_{15}, and i_2 infected i_8.

Then the occupied degree of i_1 is four and the occupied excess degrees of i_6 and i_2 are two and one, respectively.

It is important to notice that all the random graphs we consider in this work are undirected, but the next definitions and expressions are obtained considering the flow of the infectious transmissions during an outbreak.

If T_j is the final status (1=infected, 0=not infected) of a single vertex connected to the patient zero i then $T_j|(R=r, I=l) \sim \text{Bernoulli}(1-\mathrm{e}^{-rl})$ and hence the total number of occupied edges connected to i given (R, I) is $\sum_1^K T_j|(R=r, I=l)$ where K is the degree of the occupied node i.

Since we consider that $\{T_1, \ldots, T_K\}$ are independent of K the p.g.f. of the occupied excess degree K_T, $G_{K_T}(s)$ is

$$G_{K_T}(s) = E_{(R,I)}\left(G_K(G_T(s))\right) = E_{(R,I)}\left(G_K\left(s+(1-s)\mathrm{e}^{-RI}\right)\right). \tag{3.1.3}$$

For simplicity we have assumed that $G_{K_T}(s)$ depends only on the degree distribution and the transmission probability. This assumption holds if we have a contact network that does not have any loop, so that any node can be reached by no more than one other node.

Since it is desirable that the component of occupied edges grows entirely treelike without adding any restriction to the network other than being a random graph, in the results obtained and showed in this section, it is assumed the networks have a large number of nodes.

If we randomly select an occupied edge, then the occupied excess degree of the node that was infected through that edge, K_{T1}, has p.g.f. equal to

$$G_{K_{T1}}(s) = E_{(R,I)}\left(G_{K_1}(G_T(s))\right) = E_{(R,I)}\left(G_{K_1}(s+(1-s)\mathrm{e}^{-RI})\right) \tag{3.1.4}$$

If $R \equiv r_0$ and $I \equiv l_0$ ($r_0, l_0 \in \mathbf{R}^+$) then (3.1.3) and (3.1.4) reduce to

$$\begin{aligned} G_{K_T}(s) &= G_K(s+(1-s)\mathrm{e}^{-r_0 l_0}) = G_K\left(1+(s-1)(1-\mathrm{e}^{-r_0 l_0})\right) \\ &= G_K\left(1+(s-1)\pi\right), \quad \text{and} \\ G_{K_{T1}}(s) &= G_{K_1}(s+(1-s)\mathrm{e}^{-r_0 l_0}) = G_{K_1}\left(1+(s-1)(1-\mathrm{e}^{-r_0 l_0})\right) \\ &= G_{K_1}\left(1+(s-1)\pi\right). \end{aligned}$$

Final outbreak size distribution

Let Z be the final size of the outbreak with transmissibility π and let the p.g.f. of Z be

$$H_0(s) := \sum_{z=0}^{\infty} \Pr\left(Z=z\right) s^z.$$

Similarly, we define the p.g.f. $H_1(s)$ for the cluster of occupied vertices we reach by following a randomly chosen occupied edge in the same direction as the one induced by the outbreak evolution.

Since the cluster size reached by following an occupied edge is equal to the sum of the cluster sizes reached by following the other occupied edges of the vertex we first arrive at, plus one, then

$$H_1(s) = G_{\text{r.v.}\equiv 1}(s) \cdot G_{K_{T1}}(H_1(s)) = s\, G_{K_{T1}}(H_1(s)), \tag{3.1.5}$$

where $G_{\text{r.v.}\equiv 1}(s)$ is the p.g.f. of the degenerated random variable at 1.

Under the same reasoning, the size of the occupied cluster reachable, following the infection transmission direction, from a randomly chosen vertex can be expressed as

$$H_0(s) = s\, G_{K_T}(H_1(s)). \tag{3.1.6}$$

Using (3.1.5) and (3.1.6) we can compute H_1 and H_0, and then recover the probability density function of the final outbreak size using the consecutive derivatives of H_0 evaluated at zero. Since it is not always possible to find a result in closed form, we can turn to numerical approximations (Newman, 2002).

Final outbreak size mean and variance

Using (3.1.5) and (3.1.6) and numeric approximations it is possible to obtain the final outbreak size distribution, but based on these two equations it is also possible to derive a closed form expression for the first two moments of the final outbreak size. These moments are easier to obtain and they provide useful information about the distribution of the final outbreak size.

Using (3.1.5) and the properties of the p.g.f. we have

$$H_1'(1) = 1 + G'_{K_{T1}}(1)H_1'(1). \tag{3.1.7}$$

Then

$$H_1'(1) = \frac{1}{1 - G'_{K_{T1}}(1)}. \tag{3.1.8}$$

Now, based on (3.1.6) we have

$$E(Z) = H_0'(1) = 1 + G'_{K_T}(1)H_1'(1) \tag{3.1.9}$$

and hence

$$E(Z) = 1 + \frac{G'_{K_T}(1)}{1 - G'_{K_{T1}}(1)}.$$

Since from (3.1.3) and (3.1.4) we have that the expected number of occupied edges and expected excess number of occupied edges are

$$G'_{K_T}(1) = \pi G'_K(1) \qquad \text{and} \qquad G'_{K_{T1}}(1) = \pi G'_{K_1}(1), \tag{3.1.10}$$

then the expected final size can be written as

$$E(Z) = 1 + \frac{\pi G'_K(1)}{1 - \pi G'_{K_1}(1)}, \tag{3.1.11}$$

which in terms of the first two moments of the degree distribution of the network, is

$$E(Z) = 1 + \frac{\pi E(K)}{1 - \pi E(K_1)} = 1 + \frac{\pi E(K)}{1 - \pi \left(\frac{E(K^2)}{E(K)} - 1\right)}. \tag{3.1.12}$$

The equation (3.1.11) diverges when $\pi E(K_1)$ approaches to 1. This quantity coincides with the replacement number $\mathcal{R}$ representing the expected excess number of occupied edges of a typical infected node. Thus we have that the outbreak has a finite mean size when $\mathcal{R} < \infty$.

From branching processes theory we have that if $\mathcal{R} = G'_{K_{T1}}(1) < 1$ then the probability that the infectious agent becomes extinct is one, while if $\mathcal{R} = G'_{K_{T1}}(1) > 1$ the extinction probability is less than one.

In the terminology of random graphs the event of an outbreak evolving into an epidemic is the same as the cluster of occupied nodes forming a giant component.

The expression (3.1.12) can be derived as well considering a discrete outbreak evolution. If there exists only one patient zero at time 0, then the p.g.f of the number of infected at time 1 is $G_{K_T}(s)$, at time 2 the p.g.f. is $G_{K_T} \circ G_{K_{T1}}(s)$, at time 3 is $G_{K_T} \circ G_{K_{T1}} \circ G_{K_{T1}}(s)$, etc. Then the mean number of individuals infected at time 1, 2, 3 are $E(K_T)$, $E(K_T)E(K_{T1})$ and $E(K_T)[E(K_{T1})]^2$, respectively. Hence the mean of the total number of individuals that have been infected up to time n is

$$1 + \sum_{m=0}^{n-1} E(K_T)[E(K_{T1})]^m$$

and this converges to (3.1.12) as $n \to \infty$.

From (3.1.12) we have that the critical transmissibility threshold for the outbreak to become an epidemic, given the degree distribution, is

$$\pi_c := \frac{1}{G'_{K_1}(1)} = \frac{E(K)}{E(K^2) - E(K)}.$$

Then if $\pi < \pi_c$ the outbreak remains small and $E(K) < \infty$.

Like Pastor-Satorras and Vespignani (2001) (also Pastor-Satorras and Vespignani, 2003), we can see that the transmission threshold approaches to zero as $E(K^2)$ increases. Then, in a network with a large degree variance, any infectious disease outbreak will have the potential to turn into an epidemic.

Since the p.g.f. for degree of the nodes in the component of occupied nodes is (3.1.1) and their p.g.f. for their occupied degree is $E_{(R,I)}\left(G_{K_e}(s + (1-s)\mathrm{e}^{-RI})\right)$, then the clustering coefficient for this component is equal to $\pi E(K^2)/(E(Z)E(K))$ when $\pi E(K_1) < 1$, and it is always greater than $\pi E(K^2)/(nE(K))$.

For some specific cases Newman (2002) and Ancel-Meyers et al. (2003) obtained the mean final outbreak size based on the epidemic model and they compared their results using the average final outbreak size observed in a large number of computational simulations. However the authors did not explicitly describe the behavior of the deviations from the mean. In order to complement their results, next we derive the variance of the final outbreak size, in the cases $\mathcal{R} < 1$.

The variance of the final outbreak size is

$$\begin{aligned} \mathrm{Var}(Z) &= \mathrm{Var}(E(Z|K_T)) + E(\mathrm{Var}(Z|K_T)) \\ &= \mathrm{Var}(K_T E(Z_1) + 1) + E(K_T \mathrm{Var}(Z_1)) \\ &= E(Z_1)^2 \mathrm{Var}(K_T) + \mathrm{Var}(Z_1) E(K_T) \end{aligned} \tag{3.1.13}$$

where Z_1 is the size of the component at the end of a randomly chosen occupied edge.

Similarly

$$\begin{aligned}\mathrm{Var}(Z_1) &= \mathrm{Var}(E(Z_1|K_{T1})) + E(\mathrm{Var}(Z_1|K_{T1})) \\ &= \mathrm{Var}(K_{T1}E(Z_1)+1) + E(K_{T1}\mathrm{Var}(Z_1)) \\ &= E(Z_1)^2\mathrm{Var}(K_{T1}) + \mathrm{Var}(Z_1)E(K_{T1}).\end{aligned}$$

Then

$$\mathrm{Var}(Z_1) = \frac{E(Z_1)^2\mathrm{Var}(K_{T1})}{1-E(K_{T1})} = E(Z_1)^3\mathrm{Var}(K_{T1}),$$

using

$$E(Z_1) = \frac{1}{1-\pi E(K_1)},$$

from (3.1.8). Hence

$$\begin{aligned}\mathrm{Var}(Z) &= E(Z_1)^2\left[\mathrm{Var}(K_T) + E(Z_1)\mathrm{Var}(K_{T1})E(K_T)\right] \\ &= E(Z_1)^2\left[\mathrm{Var}(K_T) + E(Z_1)\mathrm{Var}(K_{T1})\pi E(K)\right]. \qquad (3.1.14)\end{aligned}$$

Now

$$\begin{aligned}\mathrm{Var}(K_T) &= E(K)^2\mathrm{Var}_{(R,I)}(1-\mathrm{e}^{-RI}) + \pi_2[\mathrm{Var}(K)-E(K)] + \pi E(K) \\ &= E(K)^2\mathrm{Var}_{(R,I)}(\mathrm{e}^{-RI}) + \pi_2[\mathrm{Var}(K)-E(K)] + \pi E(K), \qquad (3.1.15) \\ \mathrm{Var}(K_{T1}) &= E(K_1)^2\mathrm{Var}_{(R,I)}(1-\mathrm{e}^{-RI}) + \pi_2[\mathrm{Var}(K_1)-E(K_1)] + \pi E(K_1) \\ &= E(K_1)^2\mathrm{Var}_{(R,I)}(\mathrm{e}^{-RI}) + \pi_2[\mathrm{Var}(K_1)-E(K_1)] + \pi E(K_1). \qquad (3.1.16)\end{aligned}$$

where

$$\begin{aligned}\pi_2 &= E_{(R,I)}\left((1-\mathrm{e}^{-RI})^2\right) \\ E(K_1) &= \frac{E(K^2)}{E(K)} - 1, \\ \mathrm{Var}(K_1) &= \frac{1}{E(K)}\left(E(K^3) - \frac{E(K^2)^2}{E(K)}\right).\end{aligned}$$

Then the $\mathrm{Var}(Z)$ can be computed using the expressions (3.1.14), (3.1.15) and (3.1.16).

From (3.1.14) we have that the variance for the final outbreak size depends on the mean and variance of transmission $1-\mathrm{e}^{-RI}$ and the first three moments of the connectivity. Hence fixing the distributions of R and I the final outbreak results can be completely different the more heterogeneous the connectivity pattern is (skewness of the degree distribution).

Newman (2002) exemplifies its results using a network with Polylogarithmic (or Gutenber-Richter law) degree distribution

$$p_k = \frac{k^{-\delta}\mathrm{e}^{k/\lambda}}{\mathrm{Li}_\delta(\mathrm{e}^{-1/\lambda})},$$

which tends to the power law distribution as $\delta \to \infty$, is skewed, and all its moments are finite for any parametric values $\delta \in [1, \infty)$ and $\lambda \in (0, \infty)$.

The author verifies the agreement between the theoretical results for the mean final outbreak size and the probability of epidemic, with those obtained from computer simulations.

3.1.2 Degree of infected individuals

From the properties of the p.g.f. and results derived for branching processes, we know that if $G'_{K_{T1}}(1) > 1$ the probability that an outbreak does not evolve into an epidemic is the smallest root of

$$G_{K_{T1}}(u) = u.$$

Then the probability that a vertex does not belong to a giant component of occupied nodes via one of its edges, given that its degree is k, is

$$\begin{aligned} &\Pr\left(\text{node is not in giant component} \mid \text{degree is } k\right) = \\ &= \sum_{i=0}^{k} \binom{k}{i} q^{k-i} \left[(1-q)E_{(R,I)}\left(\mathrm{e}^{-RI}\right)\right]^i \\ &= \left[q + (1-q)E_{(R,I)}\left(\mathrm{e}^{-RI}\right)\right]^k \end{aligned} \tag{3.1.17}$$

where q is the probability that a selected node by randomly choosing one edge (from which our node can be reached) is not in a giant component.

Now, since

$$\Pr\left(\text{selected node by randomly choosing one edge is not in g.c. } | \text{ degree is } k_{\mathrm{e}}\right) =$$

$$= u + (1-u)\left[q + (1-q)E_{(R,I)}\left(\mathrm{e}^{-RI}\right)\right]^{k\mathrm{e}},$$

then q is the solution of

$$q = u + (1-u)G_{K\mathrm{e}}\left(q + (1-q)(1-\pi)\right) \tag{3.1.18}$$

where K_{e} has the p.f. $kp_k/E(K)$.

Similarly to (3.1.17) we have

$$\Pr\left(\text{node in giant component } | \text{ degree is } k\right) = \left(1 - \left[q + (1-q)E_{(R,I)}\left(\mathrm{e}^{-RI}\right)\right]^k\right).$$

Hence the degree distribution of a node given that it *was not* and *was* infected during the epidemic are

$$\Pr\left(\text{degree is } k| \text{ not in giant component}\right) = \frac{\left[q + (1-q)(1-\pi)\right]^k p_k}{G_K\left(q + (1-q)(1-\pi)\right)}$$

and

$$\Pr\left(\text{degree is } k| \text{ in giant component}\right) = \frac{\left(1 - \left[q + (1-q)(1-\pi)\right]^k\right) p_k}{1 - G_K\left(q + (1-q)(1-\pi)\right)}.$$

Based on the conditional degree distributions we can easily calculate the respective conditional degree means. They are

$$E(K|\text{not in g.c.}) = \frac{\left[q + (1-q)(1-\pi)\right] G_K'\left(q + (1-q)(1-\pi)\right)}{G_K\left(q + (1-q)(1-\pi)\right)}$$

and

$$E(K|\text{in g.c.}) = \frac{E(K) - \left[q + (1-q)(1-\pi)\right] G_K'\left(q + (1-q)(1-\pi)\right)}{1 - G_K\left(q + (1-q)(1-\pi)\right)}.$$

Since the infection transmits through the network edges, it is natural to think that the larger the degree of a node is, the more likely the node would be affected by the epidemic. Next we will prove that the mean degree of affected nodes is larger than the mean degree of the nodes that were unaffected.

To prove that $E(K|\text{in g.c.}) > E(K|\text{not in g.c.})$ we have to show

$$E(K)G_K\left(q+(1-q)(1-\pi)\right) - \left[q+(1-q)(1-\pi)\right]G'_K\left(q+(1-q)(1-\pi)\right) > 0. \tag{3.1.19}$$

Now, since we are interested in the cases when $E(K) = G'_K(1) > 1$ then there exist $w < 1$ such that $G_K(w) = w$.

Let $z = q + (1-q)(1-\pi)$. If $z < w$ then $z < G_k(z)$ and $G'_K(z) < G'_K(1) = E(Z)$, the inequality (3.1.19) immediately follows.

If $z > w$ but $z < r$, where r is such that $rG'_K(r) = G_K(r)$ then $zG'(z) < G_K(z)$ and hence we have (3.1.19).

For the cases $z > r$ refer to Figure 3.2. In this case $G_K(z) < zG'_K(z)$ but $w < G(z)$ and then $zG'_K(z) = wG'_K(1) < G(z)G'_K(1)$.

In corroboration, it may be noted that inequality (3.1.19) holds for values of z close to 1 due to the fact that at 1 the function $G'_K(1)G_K(z) - zG'_K(z)$ is equal to zero and the derivative evaluated at 1 is negative, describing that the last expression was positive for values at the left of 1.

3.1.3 Individual transmission rates

There are several infections in which is important to consider that transmissions are not i.i.d. r.v.'s, that is, the probability of transmission from a given individual i to another j could be drawn from different distributions for different individuals. In the most general case if the infectious contact rate and infectious period $\{R_{ij}\}$ and $\{I_i\}$ are two series of independent random variables with distributions $\{F_{R_{ij}}\}$ and $\{F_{I_i}\}$, the probability of

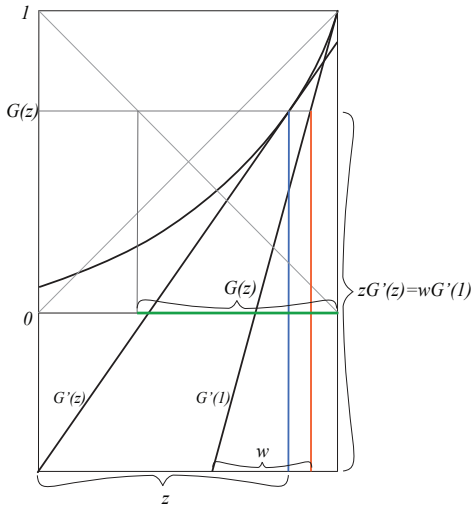


Figure 3.2: *Probability generating function.*

transmission is

$$\pi_{ij} := \Pr\left(i \text{ transmit to } j\right) = \int_0^\infty \int_0^\infty \left(1 - \mathrm{e}^{-rl}\right) dF_{R_{ij}}(r) dF_{I_i}(l).$$

In contrast to the model presented in Section 3.1.1 suppose that R varies with respect to the infective individual. Assume that the transmission rates from an infective i to each of the k_i others to whom it is connected are drawn from a distribution $F_{R_i}(r)$, which can be different for each individual i. Then

$$\pi_i := \Pr\left(i \text{ transmit to } j\right) = \int_0^\infty \int_0^\infty \left(1 - \mathrm{e}^{-rl}\right) dF_{R_i}(r) dF_I(l).$$

Similarly we can assume that the distribution of the infectious period I_i is different for each individual i. In this case the set of individual transmission probabilities is still express as $\{\pi_i\}$.

Now, let N be the number of nodes in the graph, where N is large. Since for a randomly

selected node we have that the occupied degree distribution is equal to

$$\begin{aligned}
&\Pr\left(m \text{ occupied edges proceeding from node}\right) = \\
&= \sum_{i=1}^{N} E_{(R_i,I_i)} \left(\sum_{k=m}^{\infty} \Pr\left(m \text{ occupied edges } | \text{select node } i, \text{ node has degree } k, R_i, I_i\right) \right) \times \\
&\quad \Pr\left(\text{degree } k \mid \text{ select node } i \text{ }\right) \times \Pr\left(\text{select node } i\right) \\
&= E_{(R_i,I_i)} \left(\sum_{i=1}^{N} \sum_{k=m}^{\infty} \binom{k}{m} (1-\mathrm{e}^{-R_i I_i})^m (\mathrm{e}^{-R_i I_i})^{k-m} p_k \frac{1}{N} \right),
\end{aligned}$$

then we have that the p.g.f. for the occupied degree distribution of patient zero is

$$\begin{aligned}
G_{K_T}(s) &= \sum_{m=0}^{\infty} s^m E_{(R_i,I_i)} \left(\sum_{i=1}^{N} \sum_{k=m}^{\infty} \binom{k}{m} (1-\mathrm{e}^{-R_i I_i})^m (\mathrm{e}^{-R_i I_i})^{k-m} p_k \frac{1}{N} \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(\sum_{k=0}^{\infty} p_k \left(s(1-\mathrm{e}^{-R_i I_i}) + \mathrm{e}^{-R_i I_i} \right)^k \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(G_K \left(s + (1-s)\mathrm{e}^{-R_i I_i} \right) \right). \qquad (3.1.20)
\end{aligned}$$

If we have a secondary case, then the distribution of the number of nodes this node infects is

$$\begin{aligned}
&\Pr\left(m \text{ excess occupied node} \mid \text{node was infected}\right) = \\
&= E_{(R_i,I_i)} \left(\sum_{i=1}^{N} \sum_{k=m+1}^{\infty} \Pr\left(m \text{ exc. occ. degree} \mid \text{degree } k, \, R_i, \, I_i\right) \Pr\left(\text{degree } k \mid i \text{ }\right) \Pr\left(i\right) \right) \\
&= E_{(R_i,I_i)} \left(\sum_{i=1}^{N} \sum_{k=m+1}^{\infty} \binom{k-1}{m} (1-\mathrm{e}^{R_i I_i})^m (\mathrm{e}^{-R_i I_i})^{k-1-m} \frac{k p_k}{E(K)} \frac{1}{N} \right).
\end{aligned}$$

Hence the p.g.f. for the occupied excess degree of a vertex that is a secondary case is

$$\begin{aligned}
G_{K_{T1}}(s) &= \sum_{m=1}^{\infty} s^m E_{(R_i,I_i)} \left(\sum_{i=1}^{N} \sum_{k=m+1}^{\infty} \binom{k-1}{m} (1-\mathrm{e}^{-R_i I_i})^m (\mathrm{e}^{-R_i I_i})^{k-1-m} \frac{k p_k}{E(K)} \frac{1}{N} \right) \\
&= \frac{1}{NE(K)} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(\sum_{k=0}^{\infty} k p_k \sum_{m=0}^{k-1} \binom{k-1}{m} (s(1-\mathrm{e}^{-R_i I_i}))^m (\mathrm{e}^{-R_i I_i})^{k-1-m} \right) \\
&= \frac{1}{NE(K)} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(\sum_{k=0}^{\infty} k p_k (s(1-\mathrm{e}^{-R_i I_i}) + \mathrm{e}^{-R_i I_i})^{k-1} \right) \\
&= \frac{1}{NE(K)} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(\sum_{k=0}^{\infty} k p_k (s + (1-s)\mathrm{e}^{-R_i I_i})^{k-1} \right) \\
&= \frac{1}{NE(K)} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(G_k'(s + (1-s)\mathrm{e}^{-R_i I_i}) \right) \\
&= \frac{1}{N} \sum_{i=1}^{N} E_{(R_i,I_i)} \left(G_{K_1}(s + (1-s)\mathrm{e}^{-R_i I_i}) \right). \qquad (3.1.21)
\end{aligned}$$

Newman (2002) also derived the p.g.f. for the occupied degree and occupied excess degree; however expressions obtained here (3.1.20 and 3.1.21) are different. These are the weighted average of the individual p.g.f.'s for the occupied degrees and occupied excess degrees, and it is easy to see that when $\{R_i\}$ and $\{I_i\}$ are both i.i.d. both results reduce to (3.1.3) and (3.1.4).

Using (3.1.20) and (3.1.21) the results from Section 3.1.1 can be immediately used. For example, the mean final outbreaks size is

$$E(Z) = 1 + \frac{\frac{E(K)}{N} \sum \pi_i}{1 - \frac{E(K_1)}{N} \sum \pi_i}.$$

3.2 Bipartite populations

In contrast with the simple random graphs, the bipartite networks or bipartite populations consist of two different populations that are connected to each other with two possible different degree distributions. No connection exists within either population.

Suppose we have two different populations A and B of sizes N and M. Consider that these two subpopulations have contacts between them but that no contact within either subpopulation exists.

Let p_k be the probability that one member in A has k edges that connect it to k different members in B. Analogously we define q_k for the degree of any member in B with respect to connections to A.

Then the p.g.f. for the degree and excess degree distributions of each population are

$$\begin{aligned} G_{K_A}(s) &:= \sum p_k s^k, & G_{K_{A1}}(s) &= G'_{K_A}(s)/G'_{K_A}(1), \\ G_{K_B}(s) &:= \sum q_k s^k, & G_{K_{B1}}(s) &= G'_{K_B}(s)/G'_{K_B}(1). \end{aligned}$$

Since each edge has an end in population A and and end in population B, then the sum of the degrees in populations A and B must be equal. Assuming that N and M are very large, Newman expressed this relation in term of the mean of the degrees, as

$$\frac{G'_{K_A}(1)}{M} = \frac{G'_{K_B}(1)}{N}. \tag{3.2.22}$$

If we consider two distributions for R and I to define two different transmissibilities from infectives in A to susceptibles in B (R_A, I_A), and from infectives in B to susceptibles in A (R_B, I_B), then we have similar results to (3.1.3) and (3.1.4). They are:

$$\begin{aligned} G_{K_{AT}}(s) &= E_{(R_A,I_A)}\left(G_{K_A}(s + (1-s)\mathrm{e}^{-R_A I_A})\right), \\ G_{K_{AT1}}(s) &= E_{(R_A,I_A)}\left(G_{K_{A1}}(s + (1-s)\mathrm{e}^{-R_A I_A})\right), \\ G_{K_{BT}}(s) &= E_{(R_B,I_B)}\left(G_{K_B}(s + (1-s)\mathrm{e}^{-R_B I_B})\right), \\ G_{K_{BT1}}(s) &= E_{(R_B,I_B)}\left(G_{K_{B1}}(s + (1-s)\mathrm{e}^{-R_B I_B})\right). \end{aligned}$$

If an outbreak is started for a single individual, we can obtain the number of new cases in population A. The number of new cases in this group, originated by an individual in the same group, is the number of second neighbors and hence has p.g.f. equal to

$$G_{\underline{\mathbf{AT}}}(s) = G_{K_{AT}}(G_{K_{BT1}}(s)).$$

Since every edge connects two individuals, one of each group, the p.g.f. of the excess degree of a individual in population A is

$$G_{\underline{\mathbf{AT1}}}(s) = G_{K_{AT1}}(G_{K_{BT1}}(s)).$$

In the same way $G_{\underline{\mathbf{BT}}}(s)$ and $G_{\underline{\mathbf{BT1}}}(s)$ are defined.

Using the p.g.f. just defined, we can calculate the p.g.f. H_0 and H_1 for each group. For population A they are

$$H_{A0}(s) = s\, G_{\underline{\mathbf{AT}}}(H_{A1}(s)) \tag{3.2.23}$$

and

$$H_{A1}(s) = s\, G_{\underline{\mathbf{AT1}}}(H_{A1}(s)). \tag{3.2.24}$$

Then the expected number of affected members of population A is

$$\begin{aligned} E(Z_A) &= 1 + \frac{G'_{\underline{\mathbf{AT}}}(1)}{1 - G'_{\underline{\mathbf{AT1}}}(1)} \\ &= 1 + \frac{\pi_A \pi_B G'_{K_A}(1) G'_{K_{B1}}(1)}{1 - \pi_A \pi_B G'_{K_{A1}}(1) G'_{K_{B1}}(1)} \\ &= 1 + \frac{\pi_A \pi_B E(K_A) E(K_{B1})}{1 - \pi_A \pi_B E(K_{A1}) E(K_{B1})}, \end{aligned} \tag{3.2.25}$$

where $\pi_A = E_{(R_A, I_A)}\left(1 - \mathrm{e}^{-R_A I_A}\right)$ and $\pi_B = E_{(R_B, I_B)}\left(1 - \mathrm{e}^{-R_B I_B}\right)$.

Analogously, the expected final outbreak size for the second group $E(Z_B)$ can be obtained. Even though the expected final sizes can be different, the transmission threshold is the same:

$$\pi_A \pi_B = \frac{1}{E(K_{A1}) E(K_{B1})}.$$

The variance of Z_A can be obtained directly from (3.1.13) defining $K_{\underline{\mathbf{A}}}$, $K_{\underline{\mathbf{A1}}}$, $K_{\underline{\mathbf{AT}}}$ and $K_{\underline{\mathbf{AT1}}}$ as the degree, excess degree, occupied degree and occupied excess degree of a node in A with respect to other nodes in A. See Figure 3.3.

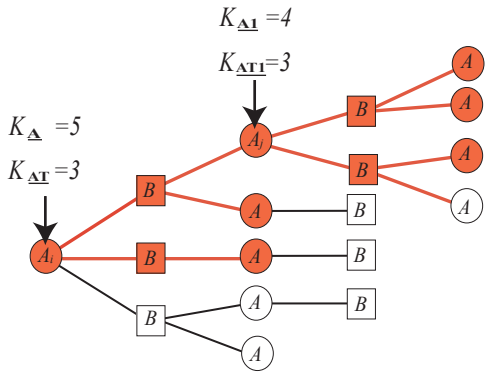


Figure 3.3: *Bipartite graph.*

Then the variance of the final outbreak size in population A is

$$\mathrm{Var}(Z_A) = E(Z_{A1})^2 \mathrm{Var}(K_{\underline{\mathbf{AT}}}) + \mathrm{Var}(Z_{A1}) E(K_{\underline{\mathbf{AT}}})$$

where

$$\mathrm{Var}(Z_{A1}) = E(Z_{A1})^2 \mathrm{Var}(K_{\underline{\mathbf{AT1}}}) + \mathrm{Var}(Z_{A1}) E(K_{\underline{\mathbf{AT1}}}).$$

Then

$$\mathrm{Var}(Z_A) = E(Z_{A1})^2 \left[\mathrm{Var}(K_{\underline{\mathbf{AT}}}) + E(Z_{A1}) \mathrm{Var}(K_{\underline{\mathbf{AT1}}}) E(K_{\underline{\mathbf{AT}}})\right]$$

where

$$E(Z_{A1}) = \frac{1}{1 - G'_{\underline{\mathbf{AT1}}}(1)} = \frac{1}{1 - E_{K_{\underline{\mathbf{AT1}}}(1)}} = \frac{1}{1 - \pi_A \pi_B E(K_{A1}) E(K_{B1})}$$

and

$$E(K_{\underline{\mathbf{AT}}}) = \pi_A \pi_B E(K_A) E(K_{B1}).$$

Now

$$\begin{aligned}
\mathrm{Var}(K_{\underline{\mathbf{AT}}}) &= E(K_A)^2 E(K_{B1})^2 \mathrm{Var}(\mathrm{e}^{-R_A I_A}) \mathrm{Var}(\mathrm{e}^{-R_B I_B}) + \\
&\quad \pi_{2A} \pi_{2B} \left(\mathrm{Var}(K_{\underline{\mathbf{A}}}) - E(K_{\underline{\mathbf{A}}})\right) + \pi_A \pi_B E(K_{\underline{\mathbf{A}}}), \qquad \text{and} \\
\mathrm{Var}(K_{\underline{\mathbf{AT1}}}) &= E(K_{A1})^2 E(K_{B1})^2 \mathrm{Var}(\mathrm{e}^{-R_A I_A}) \mathrm{Var}(\mathrm{e}^{-R_B I_B}) + \\
&\quad \pi_{2A} \pi_{2B} \left(\mathrm{Var}(K_{\underline{\mathbf{A1}}}) - E(K_{\underline{\mathbf{A1}}})\right) + \pi_A \pi_B E(K_{\underline{\mathbf{A1}}}),
\end{aligned}$$

where

$$\begin{aligned}
\pi_{2A} &= E_{(R_A,I_A)}\left((1-\mathrm{e}^{-R_A I_A})^2\right), \\
\pi_{2B} &= E_{(R_B,I_B)}\left((1-\mathrm{e}^{-R_B I_B})^2\right), \\
E(K_{\underline{\mathbf{A}}}) &= E(K_A)E(K_{B1}), \\
\mathrm{Var}(K_{\underline{\mathbf{A}}}) &= \left(E(K_{B1})\right)^2 \mathrm{Var}(K_A) + E(K_A)\mathrm{Var}(K_{B1}), \\
E(K_{\underline{\mathbf{A1}}}) &= E(K_{A1})E(K_{B1}), \quad \text{and} \\
\mathrm{Var}(K_{\underline{\mathbf{A1}}}) &= \left(E(K_{B1})\right)^2 \mathrm{Var}(K_{A1}) + E(K_{A1})\mathrm{Var}(K_{B1}).
\end{aligned}$$

The expression for the variance of Z_A more complex than the one obtained in Section 3.1.1 however it still depends only on the transmissibilities and the first three moments of the degree distributions.

When $K_B \equiv 2$ then the mean and variance final outbreak size in population A reduce to (3.1.11) and (3.1.13) with $K = K_{\underline{\mathbf{A}}}$, $K_1 = K_{\underline{\mathbf{A1}}}$, $\pi = \pi_A\pi_B$ and $\pi_2 = \pi_{2A}\pi_{2B}$.

3.3 Hierarchical networks

The hierarchical networks are the natural extension of the simple and bipartite networks, and consist of two different populations that can have connections between and within them with possibly different degree distributions.

Ancel-Meyers et al. (2003) utilized a bipartite network to model the spread of the *Mycoplasma pneumoniae* in psychiatric institutions. The authors included in their model the interactions between the subpopulations of wards and caregivers.

In their work, Ancel-Meyers et al. considered that any interaction within the population of wards and caregivers was not possible, so that the infection was spread only through the contacts of type ward-caregiver. See Figure 3.4.

A ward was classified as infective if at least one of its residents was infective, and the probability of transmission from a ward to a caregiver, and from a caregiver to a ward were

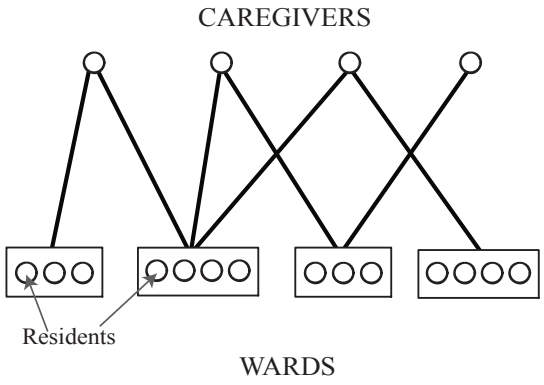


Figure 3.4: *Bipartite model used by Ancel-Meyers et al. (2003).*

assumed constants, regardless of the number of residents in each ward.

As in the bipartite population model, the authors considered that a person or a place (ward) only encounters the infection once. That is, the graph is sufficiently sparse to prevent interconnections among clusters of infections.

Ancel-Meyers et al. compared the final outbreak sizes obtained from applying the results of Newman (2002) and computational simulations for different values of mean degree distribution for the connexions of caregivers to wards ($\mu_{(A,B)}$) given the values of $\pi_A = \pi_B = 0.6$ and $\mu_{(B,A)} = 440/15\,\mu_{(A,B)}$ (that fulfils the relationship from (3.2.22)).

However the data consist of only 15 wards and 440 caregivers, and the low disease transmissibilities compensate for the small number of nodes in the network. This results in a very good agreement between the theoretical results and the simulations for the values of $\pi_A = \pi_B = 0.6$. Assuming that this agreement would hold for any other parameter's values Ancel-Meyers et al. (2003) obtained a range of possible values for the parameter $\mu_{(A,B)}$, π_A, and π_B using the results given in Newman (2002).

In the same work Ancel-Meyers et al. estimated the disease transmission for residents

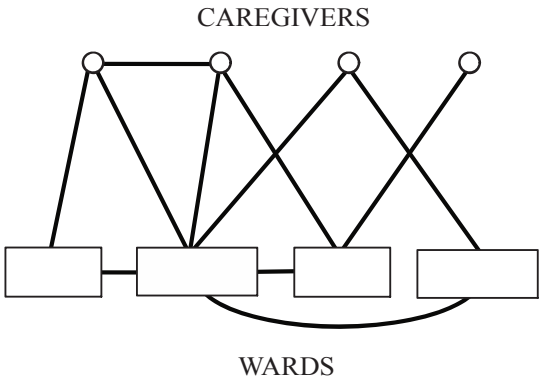


Figure 3.5: *Hierarchical network model.*

in the same ward but they did not explicitly incorporate this information into the model to obtain the mean final outbreak size.

One first generalization we think is important to make to the previous model is allowing interaction within the wards and the set of caregivers. See Figure 3.5.

Then the populations of wards and caregivers become simple subgraphs or local networks that interact connecting their individuals with a specific degree distribution.

Let $K_{(A,A)}$ and $K_{(A,B)}$ be the network connections that a node in population A has with elements in A and B, respectively, with p.g.f. $G_{K_{(A,A)}}(s)$ and $G_{K_{(A,B)}}(s)$. Similarly to the last section, we define the p.g.f. for the occupied degree as $G_{K_{(A,A)T}}(s)$ and $G_{K_{(A,B)T}}(s)$, and for the occupied excess degree $G_{K_{(A,A)T1}}(s)$ and $G_{K_{(A,B)T1}}(s)$.

Analogously we define the p.g.f.'s based on the degree distribution for elements in B.

Let $K_{\underline{\mathbf{ABAT}}}$ be the number of new infections in A that are more than one degree apart. That is, the number of elements in A that are not directly infected by a node i in A, but infected by the the secondary cases of i, all in B. Then $K_{\underline{\mathbf{ABAT}}}$ is the number of occupied nodes that go from the initial case A_0 to a case in B to a case in A ($A_0 \to B \to A$) plus

the number of occupied nodes $A_0 \to B \to B \to A$, plus the number of occupied nodes $A_0 \to B \to B \to A$, and so on. Hence the p.g.f of $K_{\underline{\mathbf{ABAT}}}$ is equal to

$$\begin{aligned} G_{\underline{\mathbf{ABAT}}}(s) &= G_{K_{(A,B)T}} \circ G_{K_{(B,A)T1}}(s) \cdot \\ & \quad G_{K_{(A,B)T}} \circ G_{K_{(B,B)T}} \circ G_{K_{(B,A)T}}(s) \cdot \\ & \quad G_{K_{(A,B)T}} \circ G_{K_{(B,B)T}} \circ G_{K_{(B,B)T1}} \circ G_{K_{(B,A)T}}(s) \cdot \\ & \quad G_{K_{(A,B)T}} \circ G_{K_{(B,B)T}} \circ G^{[2]}_{K_{(B,B)T1}} \circ G_{K_{(B,A)T}}(s) \cdot \\ & \quad G_{K_{(A,B)T}} \circ G_{K_{(B,B)T}} \circ G^{[2]}_{K_{(B,B)T1}} \circ G_{K_{(B,A)T}}(s) \cdot \cdots , \end{aligned} \tag{3.3.26}$$

where $G_{K_{(B,B)T}} \circ G^{[m]}_{K_{(B,B)T1}}(s)$ represents the p.g.f. for the number of elements, in the occupied component, that are the $(m+1)$-th neighbors in population B for a node in B, and with all intermediate neighbors also in B.

Based on (3.1.11), if the distribution of $K_{(B,B)}$ or the distributions of $R_{(B,B)}$ and $I_{(B,B)}$ make small values more probable (so $\pi_{(B,B)}E(K_{(B,B)}) << 1$) then $G_{K_{(B,B)T}} \circ G^{[m]}_{K_{(B,B)T1}}(s)$ can be approximated by 1 for values of $m > m_0$. In this case

$$\begin{aligned} G_{\underline{\mathbf{ABAT}}}(s) &\sim G_{K_{(A,B)T}} \circ G_{K_{(B,A)T1}}(s) \cdot \\ & \quad G_{K_{(A,B)T}} \circ G_{K_{(B,B)T}} \circ G_{K_{(B,A)T}}(s) \cdot \cdots \\ & \quad G_{K_{(A,B)T}} \circ G_{K_{(B,B)T}} \circ G^{[m_0]}_{K_{(B,B)T1}} \circ G_{K_{(B,A)T}}(s). \end{aligned}$$

Similarly, if we select an occupied node in A that was infected by a node in B, the total number of new infections in A that follow infections in B, $K_{\underline{\mathbf{ABAT1}}}$, has p.g.f equal to

$$\begin{aligned} G_{\underline{\mathbf{ABAT1}}}(s) &= G_{K_{(A,B)T1}} \circ G_{K_{(B,A)T1}}(s) \cdot \\ & \quad G_{K_{(A,B)T1}} \circ G_{K_{(B,B)T}} \circ G_{K_{(B,A)T}}(s) \cdot \\ & \quad G_{K_{(A,B)T1}} \circ G_{K_{(B,B)T}} \circ G_{K_{(B,B)T1}} \circ G_{K_{(B,A)T}}(s) \cdot \\ & \quad G_{K_{(A,B)T1}} \circ G_{K_{(B,B)T}} \circ G^{[2]}_{K_{(B,B)T1}} \circ G_{K_{(B,A)T}}(s) \cdot \cdots . \end{aligned} \tag{3.3.27}$$

Under this new network structure the relations (3.2.23) and (3.2.24) are now

$$H_{A0}(s) = s \cdot G_{K_{(A,A)T}}(H_{A1|(A,A)}(s)) \cdot G_{\underline{\mathbf{ABAT}}}(H_{A1|(B,A)}(s)), \tag{3.3.28}$$

where $H_{A1|(A,A)}(s)$ and $H_{A1|(B,A)}(s)$ refer to the cluster sizes at the ends of occupied edges that go from an element of A to A, and from B to A, respectively. Then

$$H_{A1|(A,A)}(s) = s \cdot G_{K_{(A,A)T1}}(H_{A1|(A,A)}(s)) \cdot G_{\underline{\mathbf{ABAT}}}(H_{A1|(B,A)}(s)) \tag{3.3.29}$$

and similarly for an edge between a node in B and a node in A,

$$H_{A1|(B,A)}(s) = s \cdot G_{K_{(A,A)T}}(H_{A1|(A,A)}(s)) \cdot G_{\underline{\mathbf{ABAT1}}}(H_{A1|(B,A)}(s)). \tag{3.3.30}$$

Solving (3.3.29) and (3.3.30) and using the first derivative of (3.3.28), the expected number of elements in A that are affected by the outbreak is

$$\begin{aligned} E(Z_A) &= 1 + G'_{K_{(A,A)T}}(1)H'_{A1|(A,A)}(1) + G'_{\underline{\mathbf{ABAT}}}(1)H'_{A1|(A,B)}(1) \\ &= \frac{1 + N_{\underline{(\mathbf{A},\mathbf{A})}} + N_{\underline{\mathbf{ABAT}}} + N_{\underline{(\mathbf{A},\mathbf{A})}}N_{\underline{\mathbf{ABAT}}}}{1 - N_{\underline{(\mathbf{A},\mathbf{A})}}N_{\underline{\mathbf{ABAT}}}}, \end{aligned} \tag{3.3.31}$$

where

$$N_{\underline{(\mathbf{A},\mathbf{A})}} = \frac{G'_{K_{(A,A)T}}(1)}{1 - G'_{K_{(A,A)T1}}(1)} = \frac{\pi_{(A,A)}G'_{K_{(A,A)}}(1)}{1 - \pi_{(A,A)}G'_{K_{(A,A)1}}(1)}$$

and

$$N_{\underline{\mathbf{ABAT}}} = \frac{G'_{\underline{\mathbf{ABAT}}}(1)}{1 - G'_{\underline{\mathbf{ABAT1}}}(1)}.$$

If $E(K_{(B,B)T1}) < 1$ it can be shown that

$$N_{\underline{\mathbf{ABAT}}} = \frac{E(K_{(A,B)T})E(K_{\underline{\mathbf{BAT}}})}{1 - E(K_{(A,B)T1})E(K_{\underline{\mathbf{BAT}}})}$$

where

$$E(K_{\underline{\mathbf{BAT}}}) = E(K_{(B,A)T1}) + \frac{E(K_{(B,B)T})E(K_{(B,A)T})}{1 - E(K_{(B,B)T1})}. \tag{3.3.32}$$

When we restrict the model to the simple random model in Section 3.1.1 there are no connections to any node in population B, and (3.3.31) reduces to $E(Z_A) = 1 + N_{\underline{(\mathbf{A},\mathbf{A})}}$, that is equal to (3.1.11).

In the bipartite network model we have that $E(Z_A) = 1 + N_{\underline{\mathbf{ABAT}}}$ and since all the new elements for A are exclusively as second neighbors, then

$$N_{\underline{\mathbf{ABAT}}} = \frac{E(K_{(A,B)T})E(K_{(B,A)T1})}{1 - E(K_{(A,B)T1})E(K_{(B,A)T1})}.$$

Hence $E(Z_A) = 1 + N_{\underline{\mathbf{ABAT}}}$ is equal to the expression in (3.2.25).

The random variable $K_{\underline{\mathbf{BAT}}}$ denotes the number of new nodes in A that an element in B originates after the infection transit through elements in B. This is, the number of cases: $B \to A$, $B \to B \to A$, $B \to B \to B \to A$, etc.

Given the last definition, then (3.3.26) and (3.3.27) can be expressed as

$$G_{\underline{\mathbf{ABAT}}}(s) = G_{K_{(A,B)T}}(G_{K_{\underline{\mathbf{BAT}}}}(s)) \tag{3.3.33}$$

and

$$G_{\underline{\mathbf{ABAT1}}}(s) = G_{K_{(A,B)T1}}(G_{K_{\underline{\mathbf{BAT}}}}(s)). \tag{3.3.34}$$

The variance of Z_A has a complex expression but this can still be computed given the three first moments of the degree, excess degree distribution and $(1 - \exp(1 - RI))$, for the distributions of R and I: $F_{R_{A,A}}, F_{I_{A,A}}, F_{R_{A,B}}, F_{I_{A,B}}, F_{R_{B,A}}, F_{I_{B,A}}, F_{R_{B,B}}$, and $F_{I_{B,B}}$.

The variance for the final outbreak size in population A is

$$\begin{aligned} \mathrm{Var}(Z_A) &= E(Z_{A|(A,A)})^2\mathrm{Var}(K_{(A,A)T}) + E(Z_{A|(B,A)})^2\mathrm{Var}(K_{\underline{\mathbf{ABAT}}}) + \\ &\quad E(K_{(A,A)T})\mathrm{Var}(Z_{A|(A,A)}) + E(K_{\underline{\mathbf{ABAT}}})\mathrm{Var}(Z_{A|(B,A)}), \end{aligned} \tag{3.3.35}$$

where $Z_{A|(A,A)}$ and $Z_{A|(B,A)}$ are the final sizes of the outbreak in population A that are at the ends of occupied edges that connect two elements in A, and an element from B to A, respectively.

The expression for the terms in (3.3.35) are:

1. $\mathbf{E(Z_{A|(A,A)})}$ and $\mathbf{E(Z_{A|(B,A)})}$:

$$E(Z_{A|(B,A)}) = \frac{1 + E(K_{(A,A)T})E(Z_{A|(A,A)})}{1 - E(K_{\underline{\mathbf{ABAT1}}})},$$

where

$$E(Z_{A|(A,A)}) = \frac{1 + N_{\underline{\mathbf{ABAT}}}}{(1 - E(K_{(A,A)T1}))N_{\underline{\mathbf{ABAT}}}N_{\underline{(\mathbf{A},\mathbf{A})}}}.$$

2. $\mathbf{Var}(\mathbf{K}_{(\mathbf{A},\mathbf{A})\mathbf{T}})$:

$$\begin{aligned}\mathrm{Var}(K_{(A,A)T}) &= E\left(K_{(A,A)}\right)^2 \mathrm{Var}(\mathrm{e}^{-R_{(A,A)}I_{R_{(A,A)}}}) + \\ &\quad E\left(\left(1 - \mathrm{e}^{-R_{(A,A)}I_{R_{(A,A)}}}\right)^2\right)\mathrm{Var}(K_{(A,A)}) + \\ &\quad E\left(\left(1 - \mathrm{e}^{-R_{(A,A)}I_{R_{(A,A)}}}\right)e^{-R_{(A,A)}I_{R_{(A,A)}}}\right)E(K_{(A,A)}).\end{aligned}$$

The variance for the occupied excess degree $\mathrm{Var}(K_{(A,A)T1})$ has a similar expression.

3. $\mathbf{Var}(\mathbf{K}_{\underline{\mathbf{ABAT}}})$:

$$\mathrm{Var}(K_{\underline{\mathbf{ABAT}}}) = E(K_{\underline{\mathbf{BAT}}})^2\mathrm{Var}(K_{(A,B)T}) + E(K_{(A,B)T})\mathrm{Var}(K_{\underline{\mathbf{BAT}}}),$$

where

$$\mathrm{Var}(K_{\underline{\mathbf{BAT}}}) = \mathrm{Var}(K_{(B,A)T1}) + E(K_{\underline{\mathbf{BAT1}}})^2\mathrm{Var}(K_{(B,B)T}) + E(K_{(B,B)T})\mathrm{Var}(K_{\underline{\mathbf{BAT1}}})$$

and

$$\mathrm{Var}(K_{\underline{\mathbf{BAT1}}}) = \frac{E(K_{\underline{\mathbf{BAT1}}})^2\mathrm{Var}(K_{(B,B)T1})}{1 - E(K_{(B,B)T1})}.$$

On the other hand, $E(K_{\underline{\mathbf{BAT}}})$ is given in (3.3.32) and

$$E(K_{\underline{\mathbf{BAT1}}}) = \frac{E(K_{(B,A)T})}{1 - E(K_{(B,B)T1})}.$$

4. $\mathbf{Var}(\mathbf{Z}_{\mathbf{A}|(\mathbf{A},\mathbf{A})})$ and $\mathbf{Var}(\mathbf{Z}_{\mathbf{A}|(\mathbf{B},\mathbf{A})})$

$$\begin{aligned}\mathrm{Var}(Z_{A|(B,A)}) &= E(Z_{A|(A,A)})^2\mathrm{Var}(K_{(A,A)T}) + E(Z_{A|(B,A)})^2\mathrm{Var}(K_{\underline{\mathbf{ABAT1}}}) + \\ &\quad E(K_{(A,A)T})\mathrm{Var}(Z_{A|(A,A)}) + E(K_{\underline{\mathbf{ABAT1}}})\mathrm{Var}(Z_{A|(B,A)}),\end{aligned}$$

where

$$\begin{aligned}\mathrm{Var}(Z_{A|(A,A)}) &= \frac{E(Z_{A|(A,A)})^2\mathrm{Var}(K_{(A,A)T1}) + E(Z_{A|(B,A)})^2\mathrm{Var}(K_{\underline{\mathbf{ABAT}}})}{1 - E(K_{(A,A)T1})} + \\ &\quad \frac{E(K_{\underline{\mathbf{ABAT}}})\mathrm{Var}(Z_{A|(B,A)})}{1 - E(K_{(A,A)T1})},\end{aligned}$$

and from relations (3.3.33) and (3.3.34) we have

$$E(K_{\underline{\mathbf{ABAT}}}) = E(K_{(A,B)T})E(K_{\underline{\mathbf{BAT}}})$$

and

$$E(K_{\underline{\mathbf{ABAT1}}}) = E(K_{(A,B)T1})E(K_{\underline{\mathbf{BAT}}}).$$

The expressions involved in the last four sections assume that the number of nodes in the network is large in relation with the distribution of the transmission rate and infectious period. However this supposition can be relaxed for SIS epidemic models since the node degree distribution is less affected by the number of infected-removed and network loops, as the infectious periods are small and the recovered individuals reenter the network of susceptible individuals.

It is important to notice that for SIR and SEIR epidemic models, the last results represent the worst case scenario. Since the degrees and excess degrees of susceptible nodes will be smaller for smaller random graphs we can expect that the means for the final outbreak sizes are smaller in observed data.

Chapter 4

Outbreak evolution in time

In contrast to Chapter 3, in this chapter we are interested in describing the outbreak size at any time t.

Considering that the infection evolution can be approximated by the Reed-Frost model, in Section 4.1 we obtain the general expression for the p.g.f of the size outbreak, the mean and the variance at any time $t \in \{1, 2, \ldots\}$. We obtain these results for infections that happen in simple, bipartite and hierarchical networks.

Section 4.2 presents the p.g.f. for the outbreak evolution assuming that the infection is a Markovian process transmitting through a simple network of contacts.

As in the previous chapter we assume that the graph is very large and then the infection grows treelike.

4.1 Discrete time evolution

In this section we incorporate the Reed-Frost SEIR model in a network of contacts and we obtain the mean affected population size at time t ($t \in 1, 2, \ldots$) for the specific network structures presented in the previous section.

As in the Reed-Frost model we consider the infection passes from one infective node i to a connected susceptible node j in a short time interval $(t-1, t-1+\epsilon)$, and the newly infected individuals become infectious in $(t, t+\epsilon)$, while the current infectives are removed.

Then the infectious cases occur in generations.

Similarly to the models in Chapter 3 we assume the number of nodes in the networks is large with respect to the transmission probabilities. Then the number of loops in the graph is negligible and the infections spread treelike.

4.1.1 Simple random graph

In this section we assume that the infective rate and infectious period are constant. Then $\tau = 1 - \exp(-rl)$ is the probability that a contact between any infectious and infective individuals results in a transmission.

As in Subsection 3.1.1, let $\{p_k\}$ be the degree distribution of the graph. Now let Y_t be the number of infected individuals at time t that become infectious in $(t, t+\epsilon)$. If we have a single initial case ($Y_0 = 1$) then the p.g.f. of Y_t fulfills

$$G_{Y_1}(s) = G_{Y_0}(G_{K_T}(s)) = G_K(1 + (s-1)\tau), \tag{4.1.1}$$

since $G_{Y_0}(s) = s$; and

$$G_{Y_t}(s) = G_{Y_{t-1}}(G_{K_{T1}}(s)) = G_{Y_{t-1}}(G_{K_1}(1 + (s-1)\tau)). \tag{4.1.2}$$

for $t \in \{2, 3, \ldots\}$.

Using (4.1.1) and (4.1.2), the p.g.f of the number of infectives at time t, $G_{Y_t}(s)$, can be computed recursively for any $t \in \{2, 3, \ldots\}$.

Mean of affected individuals at time t

An explicit expression for the mean number of infective individuals in $(t, t+\epsilon)$ can be obtained utilizing the derivatives of (4.1.1) and (4.1.2), or directly as follows:

$$E(Y_t) = E(E(Y_t|Y_{t-1})) = E(Y_{t-1}E(K_{T1})) = E(Y_{t-1})E(K_{T1}) = \tau E(Y_{t-1})G'_{K_1}(1)$$

for $t > 1$.

Now, if $Y_0 = 1$ then $E(Y_1) = E(K_T) = \tau E(K)$, and

$$E(Y_t) = E(K_T)\left[E\left(K_{T1}\right)\right]^{t-1} = \tau E(K)\left[\tau E(K_1)\right]^{t-1}, \tag{4.1.3}$$

for $t \in 1, 2, 3, \ldots$.

Using (4.1.3) we have that the expected number of infected individuals in the long run is

$$\lim_{t\to\infty} E(Y_t) \;=\; \begin{cases} 0, & E(K_{T1}) < 1 \\ \tau E(K), & E(K_{T1}) = 1 \\ \infty, & E(K_{T1}) > 1 \end{cases} \tag{4.1.4}$$

Hence the outbreak dies out if $\mathcal{R} = E(K_{T1}) < 1$, it becomes endemic in the population if $\mathcal{R} = 1$, and it transforms into an epidemic if $\mathcal{R} > 1$. This information coincides with Newman's result presented in Section 3.1.1 and provides us with further information about the endemicity of the disease when $\mathcal{R} = 1$.

However, since any community is finite, we can consider that the endemicity can be reached thanks to the arrival (births, immigration) of new susceptible individuals that connect to individuals in the network with degrees that follow the graph degree distribution.

Variance of affected individuals at time t

In order to study the possible outcomes of an outbreak in a given social network, we obtain the variance of the number of infective individuals in $(t, t+\epsilon)$. The variance can be derived from (4.1.1) and (4.1.2), or as follows.

$$\begin{aligned} \mathrm{Var}(Y_t) &= \mathrm{Var}(E(Y_t|Y_{t-1})) + E(\mathrm{Var}(Y_t|Y_{t-1})) \\ &= \left[E(K_{T1})\right]^2 \mathrm{Var}(Y_{t-1}) + E(Y_{t-1})\mathrm{Var}(K_{T1}) \\ &= \left[E(K_{T1})\right]^2 \mathrm{Var}(Y_{t-1}) + E(Y_{t-1})\left[\tau^2\mathrm{Var}(K_1) + \tau(1-\tau)E(K_1)\right]. \end{aligned} \tag{4.1.5}$$

If Y_0 is known then $\mathrm{Var}(Y_0) = 0$, and

$$\mathrm{Var}(Y_1) = \mathrm{Var}(K_T) = \tau^2\mathrm{Var}(K) + \tau(1-\tau)E(K). \tag{4.1.6}$$

Using the last expression then we have

$$\begin{aligned}\mathrm{Var}(Y_t) &= E(K_{T1})^{2(t-1)}\mathrm{Var}(K_T) + E(K_T)\mathrm{Var}(K_{T1})\sum_{i=0}^{(t-1)} E(K_{T1})^{t-2+i} \\ &= E(K_{T1})^{(t-2)}\left[E(K_{T1})^t\mathrm{Var}(K_T) + E(K_T)\mathrm{Var}(K_{T1})\frac{1-E(K_{T1})^t}{1-E(K_{T1})}\right]\end{aligned} \tag{4.1.7}$$

for $t \in \{3, 6, 7, \ldots\}$, where $\mathrm{Var}(K_T)$ is given by (4.1.6) and

$$\mathrm{Var}(K_{T1}) = \tau^2\mathrm{Var}(K_1) + \tau(1-\tau)E(K_1),$$

as used in (4.1.5).

4.1.2 Bipartite graphs

As in Section 3.2, we consider a graph with two different populations A and B with degrees K_A, K_B, excess degrees K_{A1} and K_{B1}, occupied degrees and occupied excess degrees K_{AT}, K_{BT}, K_{AT1}, and K_{BT1}.

Let $\tau_A = 1-\exp(-r_A l_A)$ and $\tau_B = 1-\exp(-r_B l_B)$, and let $Y_{t,A}$ and $Y_{t,B}$ be the number of infective individuals in $(t, t+\epsilon)$. Then

$$G_{Y_{t,A}}(s) = G_{Y_{t-1,B}}(G_{K_{BT1}}(s)) = G_{Y_{t-1,B}}(G_{K_{B1}}(1+(s-1)\tau_B))$$

and

$$G_{Y_{t,B}}(s) = G_{Y_{t-1,A}}(G_{K_{AT1}}(s)) = G_{Y_{t-1,A}}(G_{K_{A1}}(1+(s-1)\tau_A)).$$

Then, if the first infectious case occurs within population A, the p.g.f. for the number of infectives in A at time t is

$$G_{Y_{t,A}}(s) = \begin{cases} G_{Y_{t-2,A}}(G_{K_{AT1}}(G_{K_{BT1}}(s))), & \text{if } t \text{ is even} \\ 1, & \text{if } t \text{ is odd} \end{cases} \tag{4.1.8}$$

and an analogous expression is obtained for $G_{Y_{t,B}}(s)$.

If a single first case is observed within population A at the beginning of the outbreak, then $G_{Y_{0,A}}(s) = s$, $G_{Y_{0,B}}(s) = 1$,

$$G_{Y_{1,B}}(s) = G_{Y_{0,A}}(G_{K_A}(1 + (s-1)\tau_A)) = G_{K_A}(1 + (s-1)\tau_A),$$

and

$$G_{Y_{2,A}}(s) = G_{Y_{0,A}}(G_{K_{AT}}(G_{K_{BT1}}(s))) = G_{K_{AT}}(G_{K_{BT1}}(s)). \tag{4.1.9}$$

Mean of affected individuals at time t

Based on either (4.1.8) and (4.1.9), or the expected value given the previous state of population A, we obtain the mean number of affected individuals at time t as

$$\begin{aligned} E(Y_{t,A}) &= E(E(Y_{t,A}|Y_{t-2,A})) \\ &= E(Y_{t-2,A})E(K_{AT1})E(K_{BT1}) \\ &= \tau_A\tau_B E(Y_{t-2,A})E(K_{A1})E(K_{B1}), \end{aligned}$$

where

$$E(Y_{2,A}) = E(K_{AT})E(K_{BT1}) = \tau_A\tau_B E(K_A)E(K_{B1})$$

if $Y_{0,A} = 1$ and $Y_{0,B} = 0$.

In this case then

$$\begin{aligned} E(Y_{t,A}) &= \begin{cases} E(K_{BT1})^{t/2}E(K_{AT1})^{(t/2)-1}E(K_{AT}), & \text{if } t \text{ is even} \\ 0, & \text{if } t \text{ is odd} \end{cases} \\ &= \begin{cases} \tau_A\tau_B E(K_A)E(K_{B1})[\tau_A\tau_B E(K_{A1})E(K_{B1})]^{(t/2)-1}, & \text{if } t \text{ is even} \\ 0, & \text{if } t \text{ is odd.} \end{cases} \end{aligned} \tag{4.1.10}$$

As in the simple network model we can observe that $\tau_A\tau_B E(K_{A1})E(K_{B1})$ is the replacement number for populations A and B ($\mathcal{R}_A = \mathcal{R}_B$). Then from (4.1.10) we have that in the long run the mean number of affected individuals in population A diverges if $\mathcal{R}_A > 1$, is endemic if $\mathcal{R}_A = 1$, and dies out if $\mathcal{R}_A < 1$.

Variance of affected individuals at time t

The recursive expression for the expression for the variance is

$$\begin{aligned}\mathrm{Var}(Y_{t,A}) &= \mathrm{Var}(E(Y_{t,A}|Y_{t-2,A})) + E(\mathrm{Var}(Y_{t,A}|Y_{t-2,A})) \\ &= \left[E(K_{AT1})E(K_{BT1})\right]^2 \mathrm{Var}(Y_{t-2,A}) + \\ &\quad E\left(Y_{t-2,A}\right)\left[E(K_{BT1})^2\mathrm{Var}(K_{AT1}) + \mathrm{Var}(K_{BT1})E(K_{AT1})\right], \quad (4.1.11)\end{aligned}$$

for $t > 3$, where

$$\mathrm{Var}(K_{AT1}) = \tau_A^2\mathrm{Var}(K_{A1}) + \tau_A(1-\tau_A)E(K_{A1})$$

and similarly for $\mathrm{Var}(K_{BT1})$.

From (4.1.11) we can observe that $\mathrm{Var}(Y_{t,A})$ reduces to the simpler case in (4.1.5) if $K_{BT1} \equiv 1$, and it is larger than (4.1.5) if $E(K_{BT1}) > 1$.

If we have that $Y_{0,A} = 1$ and $Y_{0,B} = 0$ then the recursive expression leads to

$$\mathrm{Var}(Y_{t,A}) = \begin{cases} (a^2)^{t/2-1}\mathrm{Var}(Y_{2,A}) + bE(Y_{t-2,A})\sum_{i=0}^{t/2-2} a^{2i}, & \text{if } t \text{ is even} \\ 0, & \text{if } t \text{ is odd,} \end{cases}$$

where

$$\begin{aligned} a &= E(K_{AT1})E(K_{BT1}), \\ b &= E(K_{BT1})^2\mathrm{Var}(K_{AT1}) + \mathrm{Var}(K_{BT1})E(K_{AT1}), \\ \mathrm{Var}(K_{AT1}) &= \tau_A^2\mathrm{Var}(K_{A1}) + \tau_A(1-\tau_A)E(K_{A1}), \\ \mathrm{Var}(K_{BT1}) &= \tau_B^2\mathrm{Var}(K_{B1}) + \tau_B(1-\tau_B)E(K_{B1}), \text{ and} \\ \mathrm{Var}(Y_{2,A}) &= E(K_{BT1})^2\mathrm{Var}(K_{AT}) + \mathrm{Var}(K_{BT1})E(K_{AT}). \end{aligned}$$

4.1.3 Hierarchical networks

As in Section 3.3, we consider a graph with two different populations A and B with degrees $K_{(A,A)}$, $K_{(B,B)}$ and degrees that connect them $K_{(A,B)}$ and $K_{(B,A)}$; excess degrees $K_{(A,A)1}$,

$K_{(B,B)1}$, $K_{(A,B)1}$, $K_{(B,A)1}$; and occupied degrees and occupied excess degrees $K_{(A,A)T}$, $K_{(B,B)T}$, $K_{(A,B)T}$, $K_{(B,A)T}$, $K_{(A,A)T1}$, $K_{(B,B)T1}$, $K_{(A,B)T1}$ and $K_{(B,A)T1}$.

In this model we have

$$G_{Y_{t,A}}(s) = G_{Y_{t-1,A}}(G_{K_{(A,A)T1}}(s)) \cdot G_{Y_{t-1,B}}(G_{K_{(B,A)T1}}(s))$$

and

$$G_{Y_{t,B}}(s) = G_{Y_{t-1,B}}(G_{K_{(B,B)T1}}(s)) \cdot G_{Y_{t-1,A}}(G_{K_{(A,B)T1}}(s))$$

for $t > 1$.

If $G_{Y_{0,A}}(s) = s$ and $G_{Y_{t,B}}(s) = 1$, then

$$G_{Y_{1,A}}(s) = G_{K_{(A,A)T}}(s) \quad \text{and} \quad G_{Y_{1,B}}(s) = G_{K_{(A,B)T}}(s).$$

Mean of affected individuals at time t

Based on the last expression for the p.g.f. of the number of infective individuals in each population at time t, we have

$$\begin{aligned} E(Y_{t,A}) &= E(Y_{t-1,A})E(K_{(A,A)T1}) + E(Y_{t-1,B})E(K_{(B,A)T1}) \\ &= \tau_{AA}E(Y_{t-1,A})E(K_{(A,A)1}) + \tau_{BA}E(Y_{t-1,B})E(K_{(B,A)1}), \end{aligned}$$

and similarly,

$$E(Y_{t,B}) = \tau_{BB}E(Y_{t-1,B})E(K_{(B,B)1}) + \tau_{AB}E(Y_{t-1,A})E(K_{(A,B)1})$$

for $t > 1$.

If the initial infectious case occurs in population A, then

$$E(Y_{1,A}) = E(K_{(A,A)T}) = \tau_{AA}E(K_{(A,A)})$$

and

$$E(Y_{1,B}) = E(K_{(A,B)T}) = \tau_{AB}E(K_{(A,B)}).$$

Now let $\mathbf{E_{K_{T1}}}$ be the matrix of occupied expected excess degree among the populations. This is

$$\mathbf{E_{K_{T1}}} = \begin{pmatrix} E(K_{(A,A)T1}) & E(K_{(B,A)T1}) \\ E(K_{(A,B)T1}) & E(K_{(B,B)T1}) \end{pmatrix}. \tag{4.1.12}$$

Let $\mathbf{E(Y_t)} = (E(Y_{t,A}), E(Y_{t,B}))^t$; then the expected number of infectious individuals at time t is

$$\mathbf{E(Y_t)} = \mathbf{E_{K_{T1}}E(Y_{t-1})} = (\mathbf{E_{K_{T1}}})^k\mathbf{E(Y_{t-k})}$$

for $k \leq t-1$.

Since patient zero can transmit the infection using all of his/her connections, then

$$\mathbf{E(Y_1)} = (E(K_{\tau_{AA}}), E(K_{\tau_{AB}}))^t$$

and hence $\mathbf{E(Y_t)}$ can be expressed in terms of the populations initial conditions as

$$\mathbf{E(Y_t)} = (\mathbf{E_{K_{T1}}})^{t-1}\mathbf{E(Y_1)}.$$

Provided that $E(K_{(B,A)T1}) > 0$ and $\mathbf{E_{K_{T1}}}$ has full rank, the last expression is easier to compute using $(\mathbf{E_{K_{T1}}})^{t-1}$ equal to

$$\frac{E(K_{(B,A)T1})}{e_2 - e_1} \begin{pmatrix} 1 & 1 \\ \frac{e_1 - E(K_{(A,A)T1})}{E(K_{(B,A)T1})} & \frac{e_2 - E(K_{(A,A)T1})}{E(K_{(B,A)T1})} \end{pmatrix} \begin{pmatrix} e_1^{t-1} & 0 \\ 0 & e_2^{t-1} \end{pmatrix} \begin{pmatrix} \frac{e_2 - E(K_{(A,A)T1})}{E(K_{(B,A)T1})} & -1 \\ -\frac{e_1 - E(K_{(A,A)T1})}{E(K_{(B,A)T1})} & 1 \end{pmatrix},$$

where e_1 and e_2 are the eigenvalues associated to $\mathbf{E_{K_{T1}}}$ and such that $e_1 \leq e_2$.

Variance of affected individuals at time t

Let $\mathbf{E^*_{K_{T1}}}$ and $\mathbf{V_K}$ be the matrices of the mean occupied degrees square and occupied degree variances, respectively. That is,

$$\mathbf{E^*_{K_{T1}}} = \begin{pmatrix} E(K_{(A,A)T1})^2 & E(K_{(B,A)T1})^2 \\ E(K_{(A,B)T1})^2 & E(K_{(B,B)T1})^2 \end{pmatrix}$$

and

$$\mathbf{V_K} = \begin{pmatrix} \mathrm{Var}(K_{(A,A)T1}) & \mathrm{Var}(K_{(B,A)T1}) \\ \mathrm{Var}(K_{(A,B)T1}) & \mathrm{Var}(K_{(B,B)T1}) \end{pmatrix}.$$

Then $\mathbf{V_{Y_t}} = \text{diag}(\text{Var}(\mathbf{Y_t})) = (\text{Var}(Y_{t,A}), \text{Var}(Y_{t,B}))^t$ fulfils

$$\mathbf{V_{Y_t}} = \mathbf{E^*_{K_{T1}}} \mathbf{V_{Y_{t-1}}} + \mathbf{V_K E(Y_{t-1})}$$

for $t > 1$.

Hence the variance can be expressed as

$$\mathbf{V_{Y_t}} = \left(\mathbf{E^*_{K_{T1}}}\right)^{t-1} \mathbf{V_{Y_1}} + \sum_{i=0}^{t-2} \left(\mathbf{E^*_{K_{T1}}}\right)^i \mathbf{V_K E(Y_{t-(i+1)})}$$

where

$$\mathbf{V_{Y_1}} = \left(\text{Var}(K_{(A,A)T}), \text{Var}(Y_{(A,B)T})\right).$$

On the other hand the covariance between the number of infectives in each population at time t is

$$\begin{aligned}
&\text{Cov}(Y_{t,A}, Y_{t,B}) = \\
&= E\left(Y_{t,A} Y_{t,B}\right) - E(Y_{t,A}) E(Y_{t,B}) \\
&= E\left(E\left(Y_{t,A} Y_{t,B} | Y_{t-1,A}, Y_{t-1,B}\right)\right) - E(Y_{t,A}) E(Y_{t,B}) \\
&= \left[E(K_{(A,A)T1}) E(K_{(B,B)T1}) + E(K_{(B,A)T1}) E(K_{(A,B)T1})\right] \text{Cov}(Y_{t-1,A}, Y_{t-1,B}) + \\
&\quad E(K_{(A,A)T1}) E(K_{(A,B)T1}) \text{Var}(Y_{t-1,A}) + E(K_{(B,A)T1}) E(K_{(B,B)T1} \text{Var}(Y_{t-1,B}).
\end{aligned}$$

4.2 Continuous time evolution

Suppose we have a SIR process in a simple network. If Y_t and Z_t are the sizes of the infective and removed populations at time t, then we can describe a bivariate Markovian process in the network with the transitional probabilities conditioned on the number of susceptibles connected to the cluster of infectives at time t (for short, referred as the cluster's excess degree and denoted by K_{c_t}).

As in the previous section we assume that the graph is very large. Then the infection grows treelike, and the node degree distribution of the susceptible population is hardly affected by the increments of the infective and removed population over time.

If we assume that the rate of infection is β and the infectious period has exponential distribution with parameter γ then the conditional transitional probabilities for the infective and removed populations are

$$\Pr\{(Y_{t+\Delta}, Z_{t+\Delta}) = (y+m, z+n) | (Y_t, Z_t) = (y,z), K_{c_t} = k_{c_t}\} = \\ = \begin{cases} \beta k_{c_t}\Delta + o(\Delta), & \text{if } m=1, n=0 \\ \gamma y \Delta + o(\Delta), & \text{if } m=-1, n=1 \\ 1 - \beta k_{c_t}\Delta - \gamma y \Delta + o(\Delta), & \text{if } m=0, n=0 \\ o(\Delta), & \text{otherwise.} \end{cases}$$

If we denote $\mathrm{P}_{(y,z)}(t) = \Pr\{(Y_t, Z_t) = (y,z)\}$, then the transition probability satisfies

$$\begin{aligned} &\mathrm{P}_{(y,z)}(t+\Delta) = \\ &= \Pr\{(Y_{t+\Delta}, Z_{t+\Delta}) = (y,z)\} \\ &= \sum_{k_{c_t}} \{\beta k_{c_t}\Delta \Pr(k_{c_t}|(y-1,z))\,\mathrm{P}_{(y-1,z)}(t) + \gamma(y+1)\Delta \Pr(k_{c_t}|(y+1,z-1))\,\mathrm{P}_{(y+1,z-1)}(t) + \\ &\quad + (1 - \beta k_{c_t}\Delta - \gamma y \Delta)\Pr(k_{c_t}|(y,z))\,\mathrm{P}_{(y,z)}(t)\} + o(\Delta) \end{aligned}$$

with boundary

$$\mathrm{P}_{(0,z)}(t+\Delta) = \sum_{k_c} \{\gamma\Delta \Pr(k_{c_t}|(1,z-1))\,\mathrm{P}_{(1,z-1)}(t) + \mathrm{P}_{(0,z)}(t)\} + o(\Delta).$$

From the last two expression we obtain the forward equations of the process:

$$\begin{aligned}\frac{d}{dt}\mathrm{P}_{(y,z)}(t) &= \beta\mathrm{P}_{(y-1,z)}(t)E(K_{c_t}|y_t=y-1,z_t=z)+\gamma(y+1)\mathrm{P}_{(y+1,z-1)}(t)+\\ &\quad+\left[\beta E(K_{c_t}|y)+\gamma y\right]\mathrm{P}_{(y,z)}(t),\end{aligned}$$

and

$$\frac{d}{dt}\mathrm{P}_{(0,z)}(t)=\gamma\mathrm{P}_{(1,z-1)}(t).$$

If the process grows treelike then distribution of K_{c_t} depends only on the number of infective individuals. In this case $E(K_{c_t}|y_t=y,z_t=z)=E(K_c|y)$.

Now, multiplying $\frac{d}{dt}\mathrm{P}_{(y,z)}(t)$ by $s^y r^z$ and adding over the possible values of y and z we obtain

$$\begin{aligned}\frac{\partial}{\partial t}G(s,r;t) &= \beta\sum_{z\geq 0}\sum_{y\geq 1}E(K_c|y-1)s^y r^z\mathrm{P}_{(y-1,z)}(t)+\gamma\sum_{z\geq 0}\sum_{y\geq 0}(y+1)s^y r^z\mathrm{P}_{(y+1,z-1)}(t)-\\ &\quad-\beta\sum_{z\geq 0}\sum_{y\geq 1}E(K_c|y)s^y r^z\mathrm{P}_{(y,z)}(t)-\gamma\sum_{z\geq 0}\sum_{y\geq 0}ys^y r^z\mathrm{P}_{(y,z)}(t).\end{aligned}$$

When $t=0$, $E(K_c|y)=yE(K)$ and once the patient zeros are removed, then

$$y(E(K_1)-1)\leq E(K_c|y)\leq yE(K_1).$$

If the infectious period is small the new infected individuals are likely not to be infectives at the same time that the individual that transmitted the infection to them. Then $E(K_c|y)\sim yE(K_1)$ and

$$\frac{\partial}{\partial t}G(s,r;t) = (\beta E(K_1)s(s-1)-\gamma(s-r))\frac{\partial G(s,r;t)}{\partial s} \tag{4.2.13}$$

with boundary condition

$$G(s,r,0)=s^{Y_0}.$$

If we evaluate (4.2.13) at $r=1$ then we obtain the differential equation for the p.g.f. of a simple birth and death process with birth and death rates $\beta E(K_1)$ and γ, respectively.

The solution to equation (4.2.13), found by the method of characteristics, is given by

$$G(s,r;t) = \begin{cases} \left(\frac{1-(\gamma t-1)(s-1)}{1-\gamma t(s-1)}\right)^{Y_0}, & \text{if } \beta E(K_1) = \gamma \text{ and } r = 1 \\ \left(\frac{a(s-b)-b(s-a)\exp\{-\beta E(K_1)t(b-a)\}}{(s-b)-(s-a)\exp\{-\beta E(K_1)t(b-a)\}}\right)^{Y_0}, & \text{otherwise} \end{cases}$$

where a and b are the functions of r:

$$\begin{aligned} a(r) &= \frac{\beta E(K_1)+\gamma+\sqrt{(\beta E(K_1)+\gamma)^2-4\beta E(K_1)\gamma r}}{2\beta E(K_1)}, \text{ and} \\ b(r) &= \frac{r\gamma}{a(r)\beta E(K_1)}. \end{aligned}$$

The last result can easily be extended to bipartite networks substituting $E(K_1)$ by

$$G'_{\underline{\mathbf{A1}}}(s) = G'_{K_{A1}}(1)G'_{K_{B1}}(1) = E(K_{A1})E(K_{B1}).$$

The function $G(s,r;t)$ can be numerically solved to find the terms in its series expansion and hence the outbreak states distributions.

In this research work we do not further analyze this continuous model since it is able to relax the law of mass action assumption but it is restricted to the SIR model and infectious periods with exponential distribution.

Chapter 5

Control measures

In this section we study four outbreak control measures. The first two are vaccination actions taken before the outbreak starts and the last two are implemented during the outbreak and based on the infective individual's identification.

The principal goal of the control measures is to minimize the final outbreak size. Since information about the total number of infected individuals is in part determined by the replacement number, then the control measures attempt to modify the replacement number to reach its threshold value one, to assure (a.s.) that the outbreaks do not evolve into epidemics.

One manner to evaluate the control measures then is comparing the necessary efforts in each control strategy to reduce the replacement number to one. The efforts are measured in terms of the control measures' parameters.

A more practical approach to evaluate the control strategies is measuring the number of resources needed to contain the outbreak. Here we examine the mean number of vaccines and number of individuals to isolate, to reduce the replacement number to the value one.

5.1 Mass vaccination

Mass vaccination is the simplest immunization procedure. We assume that a fraction ν of susceptibles is randomly selected and vaccinated prior to an outbreak.

In this work we assume that the vaccination is always effective in providing immunity against the agent. Then the proportion of susceptible individuals at the beginning of an outbreak is $1-\nu$.

In the case that the probability that the vaccine is effective in any individual is $p<1$ then the following results hold for a fraction of susceptibles that is $1-\nu p$.

Since the susceptibility status after vaccination is independent of any node's characteristic such as the degree distribution, then the p.g.f. of the number of susceptible nodes that are connected to a single susceptible vertex is

$$G_J(s) = G_K(s+(1-s)\nu) = G_K(1+(s-1)(1-\nu)).$$

Similarly, the excess degree distribution becomes

$$G_{J_1}(s) = G_{K_1}(s+(1-s)\nu) = G_{K_1}(1+(s-1)(1-\nu))$$

and the p.g.f for the occupied degree and occupied excess degree are

$$G_{J_T}(s) = E_{(R,I)}\left(G_K(1+(s-1)(1-\nu)(1-e^{-RI}))\right)$$

and

$$G_{J_{T1}}(s) = E_{(R,I)}\left(G_{K_1}(1+(s-1)(1-\nu)(1-e^{-RI}))\right). \tag{5.1.1}$$

Hence the mean final outbreaks size (3.1.11) transforms into

$$E(Z) = 1+\frac{(1-\nu)\pi E(K)}{1-(1-\nu)\pi E(K_1)},$$

from which can be observed that the original basic reproductive number $\mathcal{R}_0$ and replacement number $\mathcal{R}$ are multiplied by $(1-\nu)$, with the result that the epidemic threshold with respect to the transmissibility π is increased to $1/[(1-\nu)E(K_1)]$.

When the probability of having an epidemic is greater than zero, the probability that the outbreak does not evolve into an epidemic is the smallest root of

$$u = E_{(R,I)}\left(G_{K_1}\left(1+(u-1)(1-\nu)\left(1-e^{-RI}\right)\right)\right).$$

In the bipartite population the modification can consider two different vaccination coverage proportions for populations A and B: ν_A and ν_B.

The transformation of the p.g.f for the degree, excess degree, occupied degree and occupied excess degree are similar to those shown. For example, the p.g.f.s for the degrees in population A and B are

$$G_{J_A}(s) = G_{K_A}(1 + (s-1)(1-\nu_B)) \qquad \text{and} \qquad G_{J_B}(s) = G_{K_B}(1 + (s-1)(1-\nu_A)),$$

resulting in a replacement number for population A:

$$\mathcal{R}_{A_\nu} = (1-\nu_A)(1-\nu_B)\pi_A\pi_B E(K_A)E(K_B) = (1-\nu_A)(1-\nu_B)\mathcal{R}_{\mathcal{A}}.$$

In the case of the hierarchical networks the transformation for the p.g.f's is done accordingly. The expressions for the mean and variance are more complex but still obtainable. For example,

$$\begin{aligned} &G'_{J_{\underline{\mathbf{ABAT1}}}}(1) = \\ &= (1-\nu_B)(1-\nu_A)G'_{K_{(A,B)T1}}(1)G'_{K_{(B,A)T1}}(1) + \\ &\quad +(1-\nu_B)^2(1-\nu_A)G'_{K_{(A,B)T1}}(1)G'_{K_{(B,B)T}}(1)G'_{K_{(B,A)T}}(1)\frac{1}{1-(1-\nu_B)G'_{K_{(B,B)T1}}(1)}. \end{aligned}$$

and the basic reproductive number transforms into

$$\mathcal{R}_{0,(\nu_A,\nu_B)} = (1-\nu_A)\pi_{(A,A)}E(K_{(A,A)}) + G'_{J_{\underline{\mathbf{ABAT1}}}}(1).$$

5.2 Acquaintance vaccination

In the acquaintance vaccination strategy there are two step in the selection of the nodes that are immunized. First, a fraction ω of susceptibles are randomly chosen. In contrast with the mass vaccination the nodes are randomly selected with replacement and they are not vaccinated but used to identify those to be vaccinated. In the second step a neighbor (acquaintance) of each selected node is randomly chosen and vaccinated.

It has been observed that for some diseases immunization of randomly selected individuals requires treating a very large fraction of the population in order to arrest epidemics (Zanette and Kuperman, 2002; Pastor-Satorras and Vespignani, 2001; Cohen et al., 2000; Callaway et al., 2000; Anderson and May, 1991). This is specially observed when the random network is very heterogeneous.

Some heterogeneous random networks such as the Scale Free networks present noticeable resilience to random connection failures (immunization, in our context); however this kind of network is strongly affected by selective modification. Only a few removals of the most highly connected nodes are enough to prevent any outbreak from developing into an epidemic.

Although targeting the nodes with highest degree for vaccination is more efficient than mass vaccination in any network, this strategy is impractical in most cases since it requires global information about the network.

If the strategy is changed to randomly selecting with replacement a fraction ω of susceptible nodes and randomly choosing one of their acquaintances to immunize, then as pointed out by Cohen et al. (2003) this vaccination design is local and only needs the knowledge of the selected nodes' neighborhoods.

Since the susceptibles nodes are sampled with replacement, the fraction ω may be larger than 1, while the fraction of nodes that are immunized ρ is always less than or equal to 1.

If under this plan of action we select the acquaintance to vaccinate by randomly choosing one of its edges (the edge where one end point is the randomly selected node), then the nodes that are vaccinated have degree distribution $kp_k/E(K)$. Hence without having all the structure information network, this selection process tends to include for vaccination the nodes with higher degrees.

Cohen et al. (2003) studied the immunization threshold for this strategy and showed that this is dramatically reduced in all the cases the authors studied.

Next we explicitly obtain the p.g.f. related to the susceptible nodes after the vaccination process.

First we compute the probability that a randomly selected node v is susceptible given that it is connected to another node v_a with degree k .

$$\begin{aligned}\eta_k &= \sum_{k_v=1}^{\infty} \Pr\left(v \text{ is susceptible } |k_v, k, v_a\right) \Pr\left(k_v|k, v_a\right) \\ &= \sum_{k_v=1}^{\infty} \frac{k_v p_{k_v}}{E(K)} \frac{1}{E_{K_e}\left(e^{-\omega/k}\right)} e^{-\omega/k_v} e^{-\omega/k} \left(E_{K_e}\left(e^{-\omega/k}\right)\right)^{k_v-1} \\ &= e^{-\omega/k} \sum_{k_v=1}^{\infty} \frac{k_v p_{k_v}}{E(K)} e^{-\omega/k_v} \left(E_{K_e}\left(e^{-\omega/k}\right)\right)^{k_v-2}.\end{aligned}$$

As in Cohen et al. (2003) the term $\exp(-\omega/k)$ is the approximation to the probability $(1 - 1/(Nk))^{N\omega}$, that corresponds to the event that for a specific pair of connected nodes (w_1, w_2) with w_1 of degree k, w_2 is not vaccinated after randomly selecting w_1. $E_{K_e}(\exp(-\omega/k))$ is the same but unconditional probability, and K_e is the degree distribution of a node selected by selecting one of its edges. Then K_e has p.f. equal to $kp_k/E(K)$.

Hence the probability that a randomly selected node is connected to k_s susceptibles given that its degree is k is

$$\binom{k}{k_s} \eta_k^{k_s} (1 - \eta_k)^{k-k_s}.$$

Then the p.g.f. for the degree to remaining susceptible nodes for any randomly selected node is

$$G_J(s) = G_K(1 + (s - 1)\eta_K),$$

the mean number of susceptible neighbors is

$$G'_J(1) = E(K_s) = E_K(K\eta_K),$$

and the basic reproductive number transforms into $\mathcal{R}_0 = \pi E(K\eta_K)$.

It is easy to observe when the vaccination rate is zero ($\nu = 0$) we have $\eta_K \equiv 1$ and then the p.g.f and basic reproductive number reduce to those presented in Section 3.1.1.

The p.g.f. for the excess degree to unvaccinated nodes is

$$\begin{aligned}G_{J_{T1}}(s) &= E_{(R,I)}\left(G_{J_1}(1 + (s - 1)(1 - e^{-RI}))\right) \\ &= \frac{1}{E(K_s)} E_{(R,I)}\left(G'_J(1 + (s - 1)(1 - e^{-RI}))\right).\end{aligned}$$

Based on the stable distribution for the excess degree distribution Cohen et al. (2003) derived the following expression for the replacement number

$$\mathcal{R} = \pi \sum_{k=2}^{\infty} \frac{k(k-1)p_k}{E(K)} \left(E_{K_e}\left(e^{-\omega/k}\right)\right)^{k-2} e^{-2\omega/k}. \tag{5.2.2}$$

The fraction of immunized individuals is estimated from the probability that a randomly sampled node was not vaccinated,

$$f_\omega = 1 - \sum_{k} \left(E_{K_e}\left(e^{-\omega/k}\right)\right)^{k} p_k. \tag{5.2.3}$$

5.3 Ring vaccination

The ring vaccination, like isolation, is a control measure that is implemented during an outbreak. We assume that during the outbreak all infective individuals are identified but their neighbors are vaccinated with probability q after m units of time since the node has become infective.

Since the individuals are usually identified after they become symptomatic and this occurs, for many infectious diseases, after they have been infectious, we introduce the constant m to represent the time between start being infective and identification.

If some of the neighbor nodes are already infected or infectives when vaccinated, we consider that the vaccine does not have any effect on their disease and transmissibility evolution.

The case when $m = 0$ has been studied, in combination with the mass vaccination, by Takeuchi and Yamamoto (2006).

When $m = 0$ the p.g.f's for the degree and excess degree that connect a just infective node to those who remain susceptible after ring vaccination is

$$G_J(s) = G_K(1 + (s-1)(1-q)) \quad \text{and} \quad G_{J_1}(s) = G_{K_1}(1 + (s-1)(1-q)).$$

Then the corresponding p.g.f. for the occupied degrees are

$$G_{J_T}(s) = E_{(R,I)}\left(G_K(1 + (s-1)(1-q)(1-e^{-RI}))\right)$$

and

$$G_{J_{T1}}(s) = E_{(R,I)}\left(G_{K_1}(1+(s-1)(1-q)(1-e^{-RI}))\right). \tag{5.3.4}$$

When $m > 0$ then

$$\begin{aligned} G_{J_T}(s) &= \int_0^m \int_0^\infty G_K(1+(s-1)(1-e^{-rl}))dF_R(r)dF_I(l) + \\ &\quad + \int_m^\infty \int_0^\infty G_{J_m}(1+(s-1)(1-e^{-rl}))dF_R(r)dF_I(l) \end{aligned}$$

where

$$G_{J_m}(s) = \int_0^m \int_0^\infty G_K(1+(s-1)(1-q)e^{-r_0 l_0})dF_R(r_0)dF_I(l_0).$$

and

$$G_{J_m}(1+(s-1)(1-e^{rl})) = \int_0^m \int_0^\infty G_K(1+(s-1)(1-q)(1-e^{-rl})e^{-r_0 l_0})dF_R(r_0)dF_I(l_0).$$

If we define

$$\pi_m = \int_0^m \int_0^\infty (1-e^{-rl})dF_R(r)dF_I(l) \quad \text{and} \quad \pi_{m+} = \int_m^\infty \int_0^\infty (1-e^{-rl})dF_R(r)dF_I(l),$$

then

$$\mathcal{R}_0 = (\pi_m + (1-q)(1-\pi_m)\pi_{m+})E(K) \tag{5.3.5}$$

and

$$\mathcal{R} = (\pi_m + (1-q)(1-\pi_m)\pi_{m+})E(K_1), \tag{5.3.6}$$

that is, the mean number of neighbors (excess neighbors) of an infective node that are infected up to time m plus the mean number of individuals not vaccinated, not infected in $[0, m]$, and infected in (m, ∞).

5.4 Isolation

As in the case of the vaccination, there are a lot of different criteria to implement the isolation of an individual during an outbreak. For example, quarantine isolates the individuals that have been in contact with infective individuals for at least the infection's incubation period.

In this work we consider the isolation of individuals that become infective and as in the case of the ring vaccination, we incorporate the fact that infected individuals are detected after m units of time of becoming infectives. In contrast to the ring vaccination we assume that the isolation is 100% effective and only a fraction γ of all the infective cases are detected.

Then the isolation is equivalent to the ring vaccination when $q = 1$ and $\gamma = 1$.

An isolated individual is prevented from any contact with other individuals. In the network setting this is equivalent to erasing the edges around the node. However, we can incorporate this control measure into the model modifying the infectious period I.

When the infective individuals are detected immediately after becoming infective ($m = 0$) we have that the occupied degree and occupied excess degree's p.g.f. are

$$\begin{aligned} G_{J_T}(s) &= (1-\gamma)G_{K_T}(s) + \gamma \\ G_{J_{T1}}(s) &= (1-\gamma)G_{K_{T1}}(s) + \gamma \end{aligned} \tag{5.4.7}$$

and the basic reproductive and replacement numbers then are

$$\mathcal{R}_0 = \pi(1-\gamma)E(K) \quad \text{and} \quad \mathcal{R} = \pi(1-\gamma)E(K_1). \tag{5.4.8}$$

When $m > 0$ the expressions in (5.4.7) transform to

$$\begin{aligned} G_{J_T}(s) &= (1-\gamma)G_{K_T}(s) + \gamma \int_0^m \int_0^\infty G_K(1+(s-1)(1-e^{-rl}))dF_R(r)dF_I(l) \\ G_{J_{T1}}(s) &= (1-\gamma)G_{K_{T1}}(s) + \gamma \int_0^m \int_0^\infty G_{K_1}(1+(s-1)(1-e^{-rl}))dF_R(r)dF_I(l). \end{aligned}$$

Then the basic reproductive and replacement numbers are

$$\mathcal{R}_0 = (\gamma\pi_m + (1-\gamma)\pi)E(K) \quad \text{and} \quad \mathcal{R} = (\gamma\pi_m + (1-\gamma)\pi)E(K_1). \tag{5.4.9}$$

5.5 Discussion

To illustrate the effect of the individual control measures we use networks that describe three different levels of contact's heterogeneity:

1. Poisson(4)

2. Polylogarithmic(2.6762,100)

3. Constant(5)

The constant distribution is the degenerate random variable with all its mass at 5 and represents a regular graph. As shown in Figure 5.1 the p.f. for the polylogarithmic produces the most heterogeneous network and its distribution is close to the power law distribution with parameter 2.7 for small values of k.

In contrast with the power law, the polylogarithmic distribution has all moments for all its parameter values since its tail falls faster as it becomes dominated by the exponential term. See Table A-1 in Appendix A.

The parameters for the three distributions are selected such that the mean excess degrees $E(K_1)$ are equal to 4, resulting in the same replacement number $\mathcal{R} = \pi E(K_1)$ given a fixed value of the transmissibility π.

If the distributions are selected such that the mean degrees $E(K)$ are the same, the basic reproductive number $\mathcal{R}_0$ will be the same but the replacement number $\mathcal{R}$ will vary a lot, especially if we include for consideration highly skewed degree distributions such as the polylogarithmic. See Figure 5.2.

The mean excess degree of a network with Poisson degree distribution is equal to the mean degree, and in the case of the constant network (regular graph), the mean excess degree is the mean degree minus 1.

The means of the degree distributions of the three considered networks are 4, 1.543686 and 5, respectively.

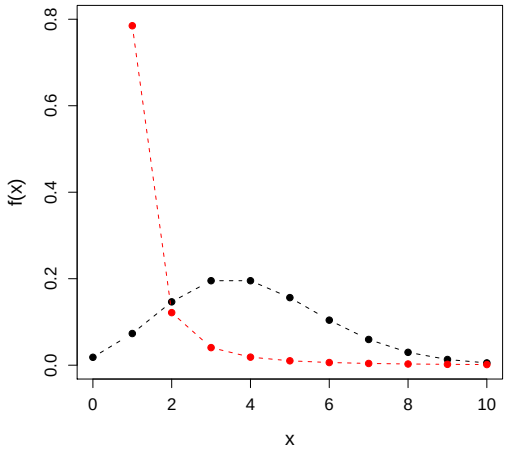


Figure 5.1: *Poisson(4) and polylogarithmic(2.6762,100) distributions.*

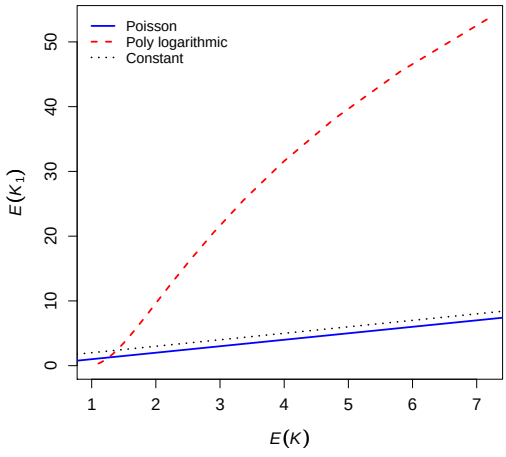


Figure 5.2: *Excess degree mean vs. degree mean.*

As can be observed from (5.1.1) and (5.3.4), the p.g.f. of the occupied excess degree for the mass vaccination and the ring vaccination are the same when parameterized with respect to ν and q. In view of this, the replacement number and the probability for an outbreak becoming an epidemic are equally modified for the value variation of ν and q, reducing the replacement number to $(1-\nu)\mathcal{R}$ or $(1-q)\mathcal{R}$.

Similarly, from (5.4.8) the isolation measure (for the case $m = 0$) reduces the replacement number to $(1-\gamma)\mathcal{R}$; and since the distributions have the same mean excess degree, then the graph for the reciprocal of the threshold transmissibility π for all graphs, in the cases of the mass vaccination, ring vaccination and isolation coincides with the line with y-intercept 4 and slope -4.

Then, for the three considered networks, to increase the transmissibility threshold to 80%, it is necessary to mass vaccinate 50% of the population or vaccinate the exposed individuals with probability 0.5 (ring vaccination) or identify and isolate the infective cases with probability 0.5.

Figure 5.3 presents the threshold for the value $1/\pi$ when the network is treated with the acquaintance vaccination model. In contrast with the isolation and mass and ring vaccination, this threshold for the reciprocal of the transmissibility, parameterized with respect to the percentage of vaccinated population (using 5.2.2 and 5.2.3), varies with the network's degree variance and skewness.

As can be observed in the figure, the plot corresponding to the constant degree distribution agrees with the mass vaccination while for the polylogarithmic degree distribution the figure shows that the transmissibility threshold rises a lot faster. This means that for heterogeneous contact networks a considerably lower number of individuals have to be vaccinated to reach the threshold under which the outbreaks almost surely do not evolve into an epidemic.

This observation agrees with the results obtained by Cohen et al. (2003). The author remarked that the acquaintance vaccination, when implemented in networks with heterogeneous degree, is able to achieve the same levels of containment as mass vaccination with

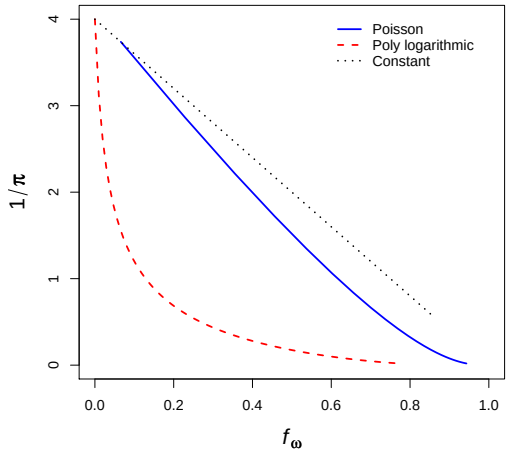


Figure 5.3: *Reciprocal for the transmissibility threshold for the fraction of immunized individual under acquaintance vaccination.*

lesser number of vaccinated nodes.

Although the vaccination control strategies can be evaluated in terms of the number of vaccines that have to be applied, they offer different implementation difficulties. For example, for some agents the definition of contact might be still uncertain, resulting in a low or difficult detection of exposed cases to vaccinate; or the incubation period can be much larger than the latent period, leading to inefficient cases identification and inefficient ring vaccination or isolation.

Having mentioned that the evaluation of the control measures must consider the practical problems related with specific populations and agents, we can still compare the control measures with respect to the expected number of vaccines to be utilized in order to have $\mathcal{R} = 1$.

Figure 5.4 presents the proportion of vaccines for different transmissibility values π to reach the threshold value 1 of the replacement number.

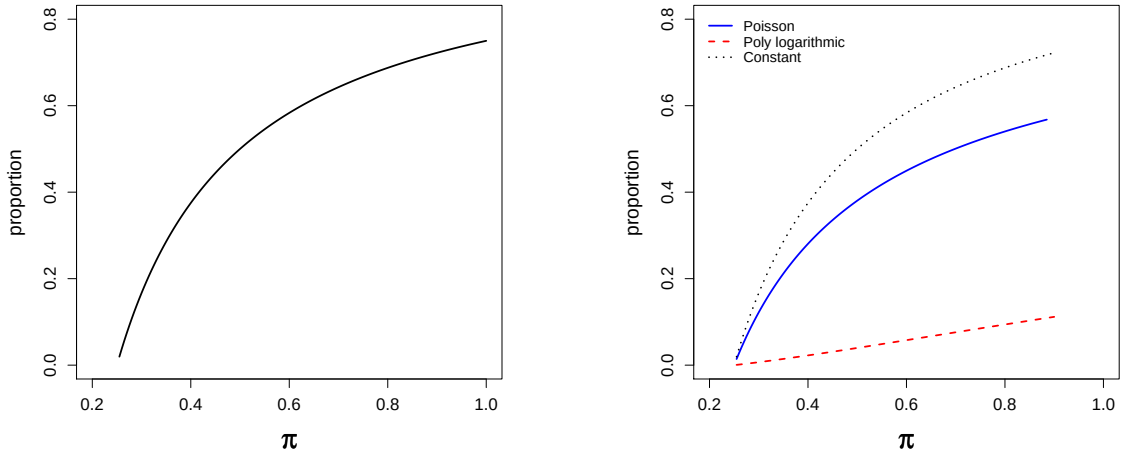


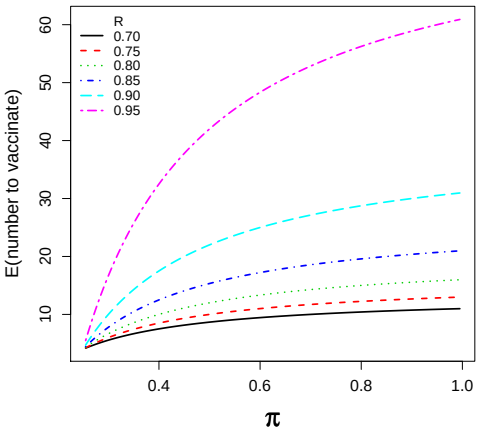
Figure 5.4: *Expected number of individuals to vaccinate to reach* $\mathcal{R} = 1$ *for different transmissibility values* π *and under mass and acquaintance vaccination required .*

The proportion of individuals to be mass vaccinated is the same regardless of the contact network structure when the mean excess degree are the same. However in the case of the acquaintance vaccination, as previously observed, the number of necessary vaccines to be administrated is much less for the networks with polylogarithmic degree distribution than for those with Poisson and the constant distributions.

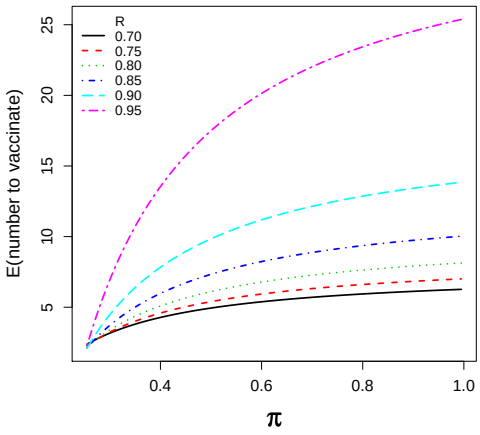
For the ring vaccination ($m = 0$) the expected number of vaccines is $E_{\text{Ring}} = qE(Z)E(K_1)$. If we consider the values of $\mathcal{R}$ below the threshold 1: $\mathcal{R} = 0.70, 0.75, 0.80, 0.85, 0.90$ and 0.95 then the needed number of vaccines (regardless of the population size) is presented for the respective networks in Figure 5.5.

The curves in Figure 5.5 have similar shapes but their scale is different. Thus, for example, to have a reproductive number of 0.95 such that the transmissibility is equal to one, about 60, 25 and 70 individuals must be ring vaccinated in the respective network.

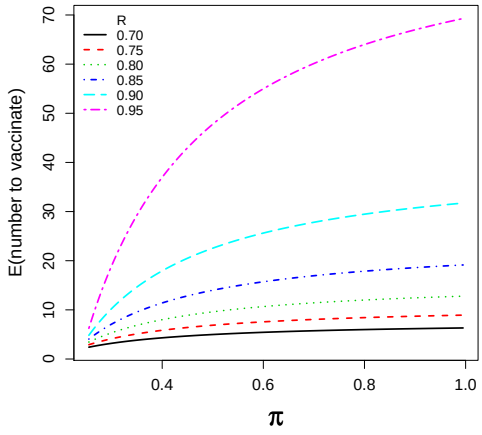
The standard deviation for the number of required vaccines can be obtained from the variance of the final outbreak size given in (3.1.14), the fact that the new occupied degree



Poisson



Polylogarithmic



Constant

Figure 5.5: *Expected number of individuals to vaccinate for different transmissibility values π and under ring vaccination for the networks with Poisson, polylogarithmic and constant degree distribution.*

K_T^* and occupied excess degree K_{T1}^* are equal to $(1-q)K_T$ and $(1-q)K_{T1}$, and

$$\text{Var(total number of vaccines)} = q^2E(K_1)^2\text{Var}(Z) + q^2E(Z)\text{Var}(K_1).$$

The only extra information needed to compute the standard deviations are the third moment of the degree distribution, $E(K^3)$ and $\pi_2 = E_{(R,I)}\left((1-\exp(-RI))^2\right)$.

Owing to the fact that the expected number of isolated individuals, when $m = 0$, is

$$E_{\text{Isolated}} = \gamma E(Z) = \frac{\gamma}{qE(K_1)}E_{\text{Ring}},$$

the threshold curves, when the isolation is implemented, keep the same shape as the curves as the curves in the previous figure, suggesting that isolation, like ring vaccination, is more efficient when the contact network is heterogeneous. See Figure 5.6.

Since the smallest values q_r and γ_r that make $\mathcal{R} = r$ are obtained from the same expression

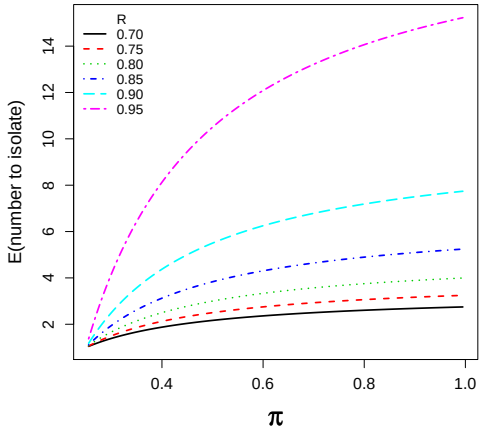
$$1 - \frac{r}{\pi E(K_1)}$$

then $q_r = \gamma_r$.

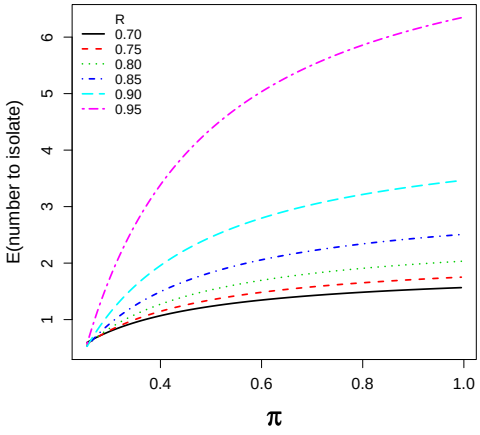
Then, the isolation is more efficient than ring vaccination in any network that has mean occupied excess degree $(E(K_1))$ greater than 1.

In the cases when the incubation period does not end when the infectious period starts, for the ring vaccination and isolation we obtain the values of the probability of transmission and the probability of transmission in m unit of time (π and π_m, respectively), and use (5.3.5), (5.3.6) and (5.4.9) to compute q or γ such that $\mathcal{R} = 0.95$ and the corresponding $E(Z) = 1 + \mathcal{R}_0/(1-\mathcal{R})$ to finally find the expected number of vaccines or individuals to isolate. Figure 5.7 depicts the expected number of vaccines to use when ring vaccination is applied in a network with Poisson(4) degree distribution.

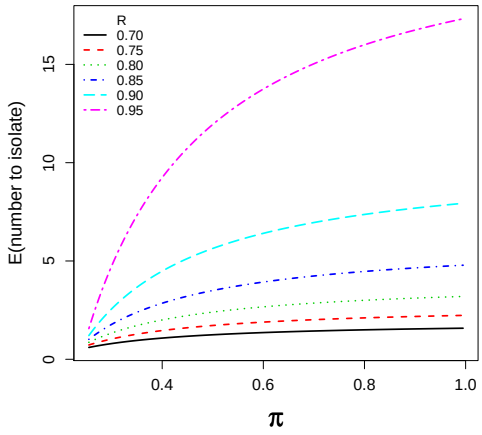
The graphic at the right is the contour representation of the plot at the left and we can observe that for low values of π and π_m no vaccine needs to be applied since the value of the replacement number is already smaller than 0.95. On the other hand for large values of π



Poisson



Polylogarithmic



Constant

Figure 5.6: *Expected number of individuals to isolate for different transmissibility values π and in the networks with Poisson, polylogarithmic and constant degree distribution.*

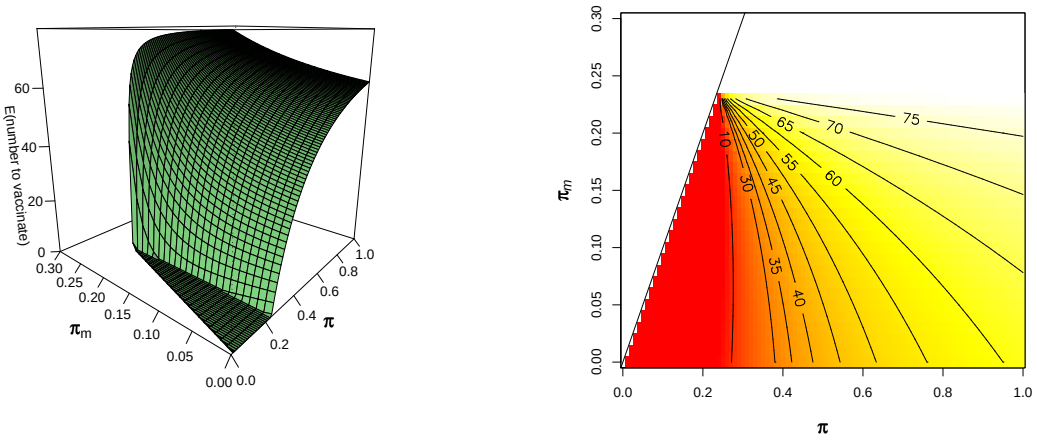


Figure 5.7: *Expected number of individuals to vaccinate for different transmissibilities probabilities π and π_m, and for the network with Poisson degree distribution.*

and π_m the replacement number cannot be reduced to the level r even if all the susceptible individuals exposed to symptomatic infective were vaccinated ($q = 1$).

In the contour plot in Figure 5.7 it is easy to observe that the expected number of vaccines is more sensitive to the variations in π than to the variations in π_m; however, the values of π_m for which we can reach the level $r = 0.95$ are restricted to the interval $[0, 0.95/E(K_1))$.

As in the ring vaccination with $m = 0$, the curves that describe the expected number of vaccines for each of the three different networks have similar shape but in different scale. The contour plots for the networks with polylogarithmic and constant degree distributions are presented in Figure 5.8.

The same plots can be obtained for the bipartite and hierarchical networks using their corresponding mean final size outbreaks. The graphs are similar to those obtained but modified by the extra heterogeneity that their network structure naturally confers.

As we have pointed out, the acquaintance vaccination is more efficient than the mass vaccination, the more heterogeneous the network is. However in both cases the number of required vaccines increases with the population size.

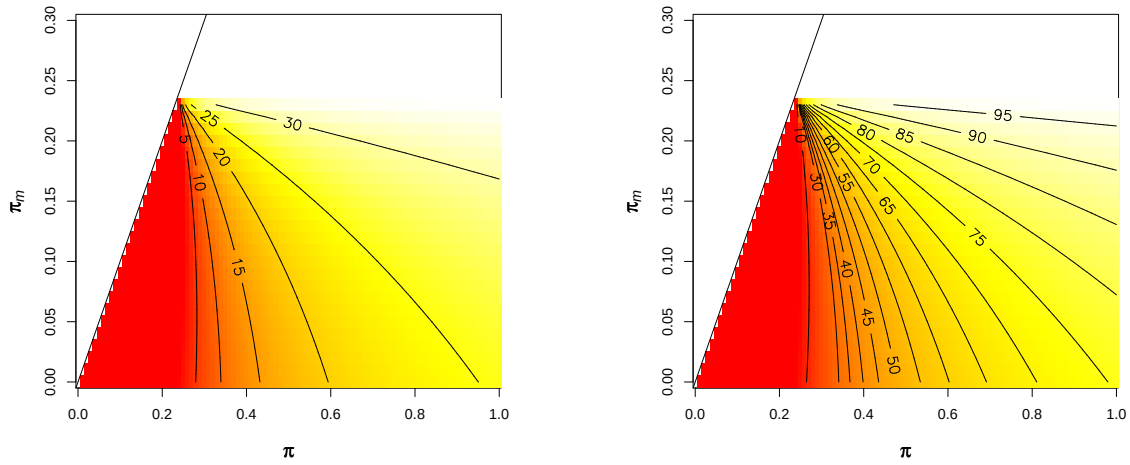


Figure 5.8: *Expected number of individuals to vaccinate for different transmissibilities probabilities π and π_m, and for the networks with polylogarithmic and constant degree distribution.*

In contrast with the mass and acquaintance strategies, the ring vaccination and isolation follow the infection path; thus the mean and variance of the number of required vaccines and number of individuals to isolate, to have $\mathcal{R} = 1$, is functionally independent of the population size, making these two control measures more efficient in large susceptible populations.

When $\pi_m > 0$ we can have that the threshold $\mathcal{R} = 1$ cannot be achieved even if $q = 1$ or $\gamma = 1$. However, ring vaccination or isolation can be combined with mass or acquaintance vaccination to reduce the reproductive number to lie under the threshold value.

If mass vaccination is applied prior to the outbreak to a fraction ν of the population, then the susceptible population becomes $(1 - \nu)N$ and the mean excess degree is reduced to $(1 - \nu)E(K_1)$. Then the ring vaccination with parameters q and $m = 0$, applied over the remaining susceptible population, results in a replacement number $\mathcal{R} = (1 - \nu)(1 - q)\pi E(K_1)$.

Similarly the combination of mass vaccination-isolation, acquaintance-ring vaccination and acquaintance-isolation leads to the reproductive numbers $(1-\nu)(1-\gamma)\pi E(K_1)$, $(1-q)\pi E(K_{A1})$ and $(1-q)\pi E(K_{A1})$, respectively, where $\pi E(K_{A1})$ is the replacement number resulting from the acquaintance vaccination (5.2.2).

When the parameter $m > 0$ in the ring vaccination and isolation then the reproductive numbers for the combinations are:

$$\begin{aligned}
\text{mass-ring vaccination:} \quad \mathcal{R} &= (\pi_m + (1-q)(1-\pi_m)\pi_{m+})(1-\nu)E(K_1), \\
\text{mass vaccination-isolation:} \quad \mathcal{R} &= (\gamma\pi_m + (1-\gamma)\pi)(1-\nu)E(K_1), \\
\text{acquaintance-ring vaccination:} \quad \mathcal{R} &= (\pi_m + (1-q)(1-\pi_m)\pi_{m+})E(K_{A1}), \\
\text{acquaintance-isolation:} \quad \mathcal{R} &= (\gamma\pi_m + (1-\gamma)\pi)E(K_{A1}).
\end{aligned}$$

Chapter 6

Analysis of data

In this chapter we analyze two different infectious agents using information from several outbreaks.

Section 6.1 studies the measles infection from the data given in Bailey (1950) and corresponding to outbreaks in households with two, three and four susceptibles.

In Section 6.1.2 we obtain the point estimates for the parameters that describe the rate of transmission, the latent and infectious period. The estimation is based on the Monte Carlo Maximum-likelihood Estimation as described in McLeish (2005) and the local prediction for the likelihood function.

Based on the measles estimated parameter then we study the effectiveness of control measures in a population with a contact network structure.

In Section 6.2 we study the infections related to the influenza infectious disease. The utilized data corresponds to outbreaks in long term care in the Region of Waterloo. Based on some of the outbreaks we do the statistical inference for the influenza infection and cross-evaluate the obtained parameters with some other outbreaks registered in the data.

6.1 Measles data

In this section we study the measles household data presented in Bailey (1950). In Section 6.1.1 we briefly describe the epidemiology of measles, and in Section 6.1.2 we estimate

the measles parameters using the Monte Carlo-likelihood obtained from the outbreak simulations.

In Section 6.1.3 we explore the Reed-Frost model for the measles data and the estimated transmissibility, and finally in Section 6.1.4 we evaluate the effect of the control measures considered in Chapter 5 and some of their parametric values.

6.1.1 Measles

Measles (also known as rubeola) is a highly contagious and acute infectious disease caused by a virus of the genus Morbillivirus in the family Paramyxoviridae.

This disease is one of the best known and deadliest of all childhood rash and febrile illnesses. In severe cases, complications such as pneumonia, diarrhoea, middle ear infection and encephalitis (a dangerous infection of the brain) may occur (PHAC). Measles mainly affects young children, but can strike older children and adults as well.

Reports of measles go at least as far back as 600 B.C. However, in the 9th-century, the Persian physician Ibn Razi was the first scientist who described in details the symptoms and signs of smallpox and measles based on clinical examination (in his book The Book on Smallpox and Measles), and he was the first who distinguished between these two diseases.

There is no cure for measles, but it can be prevented with vaccination. Licensed vaccines became available in 1963, and in developed countries, most children are immunized against measles at the age of 18 months, generally as part of a three-part MMR vaccine (measles, mumps, and rubella). The vaccination is generally not given earlier than this because children younger than 18 months usually retain anti-measles immunoglobulins (antibodies) transmitted from the mother during pregnancy. The first vaccine is strengthened with a second dose given between the ages of four and five.

In 1952 the World Health Organization observed for measles that "... its almost invariable direct transmission, the relatively fixed duration of infectivity, and the lasting immunity which it generally confers, have made it possible to lay the foundations of a statistical theory of epidemics" (WHO, 1952). These transmission-illness characteristics

made the measles have one of simplest epidemiological behaviors of all the major human viral diseases (Black, 1984).

The measles is a highly contagious airborne pathogen which spreads by airborne droplets. Contagion results from close personal contact with infected persons that expose the susceptible individual to nasal or throat secretions, such as those produced by a coughing patient. The virus is less commonly spread through contact with contaminated articles, such as tissue paper that has been freshly soiled with nose or throat secretions (PHAC).

Measles symptoms generally appear in two stages. The first signs are the onset of fever, red eyes (i.e., conjunctivitis), runny nose, cough and white spots on the inside lining of the mouth. In the second stage, a characteristic red blotchy rash appears beginning on the face and becoming generalized. Generally, the disease is more severe in infants and adults than in children.

One attack of measles usually gives lifetime immunity, but a measles attack does not extend immunity to the similar illness named German measles (also called rubella).

Once an individual has acquired the measles virus, only isolation, bed and rest are recommended since there is no specific treatment against the agent.

According to PHAC, the incubation period for measles is approximately 10 days, varying from 8 to 13 days from exposure to onset of fever, and about 14 days from exposure until the appearance of a rash.

An individual is contagious usually from four days before to four days after the appearance of a rash (PHAC).

6.1.2 Parameter inference

Wilson et al. (1939) developed an extensive investigation into measles epidemics that occurred in Providence, Rhode Island during the years 1929-1934. Bailey (1975, Chap. 14) reported and utilized the final epidemic size in households of four (Table 6.1) and three (Table 6.3) susceptible individuals. All the susceptible individuals were between the ages of 7 months and 10 years, and in each of the houses there was one who was initially infected.

Bailey (1975) also analyzed the collected data by Hope Simpson (in the Cirencester area of England, between 1946-1952) that consist in the times between measles cases for 264 households, each with two susceptible individuals (Table 6.2). In each household both susceptibles were under the age of 15 and one of them was initially infected.

If we simulate the measles evolution for different transmission rates for outbreaks in households with 4 susceptibles we can easily explore which transmission rates are likely under the observed data in Table 6.1.

Total Number of cases	Frequency
1	4
2	3
3	9
4	84
Total	100

Table 6.1: *Total number of cases in households with four susceptibles.*

If we consider that

- the transmission rate is a constant r,
- all the individuals within the household have contact with each member of the family,
- the latent period is Gaussian distributed with parameters μ_L and σ_L, and
- the infectious period is Log normal distributed with parameters μ_I and σ_I,

then for example if $\mu_L = 10, \sigma_L = 1, \mu_I = 1.7, \sigma_I = 0.25$, for different transmission rates r Figure 6.1 presents the box plots for the frequencies for each number (1, 2, 3, 4) of affected individuals in sets of 100 houses (1000 simulations for each set). The red asterisks represent the observed data given in Table 6.1 and the black diamonds are the mean of the simulations for each final epidemic size.

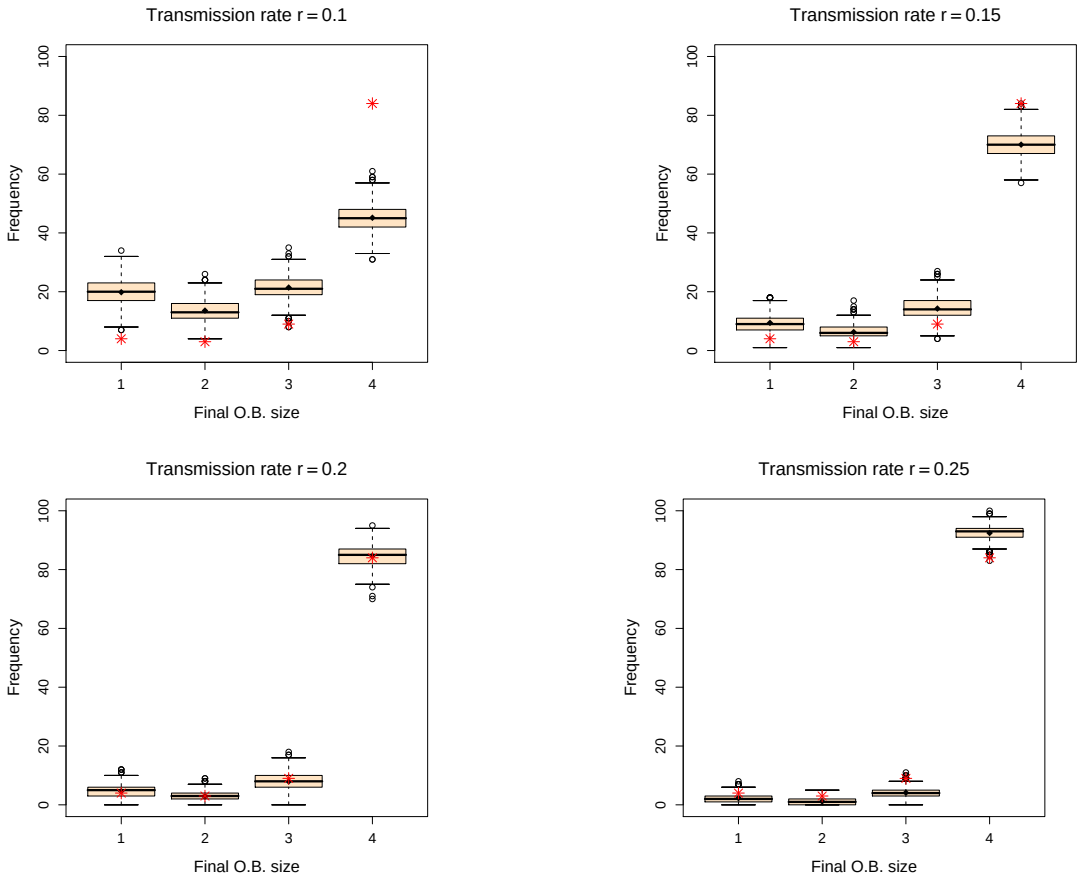


Figure 6.1: *Distribution of final O.B. size in 100 4-susceptible household with infectious period* $\sim LN(1.7, 0.25)$ *and latent period* $\sim N(10, 1)$.

Observing the plots in Figure 6.1 we can see that among the parameter values utilized to simulate the outbreaks, the one that best describes the final outbreak size distribution for households of size 4 is $r = 0.2$ when $\mu_I = 1.7$ and $\sigma_I = 0.25$.

Bailey (1950) studied this same data under two chain-binomial models: the Greenwood and the Reed-Frost model. He obtained point estimates for the transmission probability equal to 0.709 and 0.653, respectively. These estimates made the models fit very well and they are close to our probability of transmission

$$\pi = E_{(I)}\left(1 - e^{0.2l}\right) = \int_0^\infty \left(1 - e^{0.2l}\right) dF_I(l) \sim 0.6642.$$

Since the final outbreak size does not contain any information about the time scale of the outbreak, none of the latent parameters can be estimated from this data. In order to estimate μ_L and σ_L we have to use the data corresponding to the days between infections (See Table 6.2).

Time(days)	Frequency
No transmission	45
0	0
1	0
2	0
3	0
4	1
5	2
6	4
7	11
8	5
9	25
10	37

Time (days)	Frequency
11	38
12	26
13	12
14	15
15	6
16	3
17	1
18	3
19	0
20	0
21	1
Total	**235**

Table 6.2: *Times between measles cases in households with two susceptibles.*

Figure 6.2 explores some of the simulated distribution of days between infections for $\mu_I = 1.7$, $\sigma_I = 0.25$ and $r = 0.2$ and different values of μ_L and σ_L. The red line represents

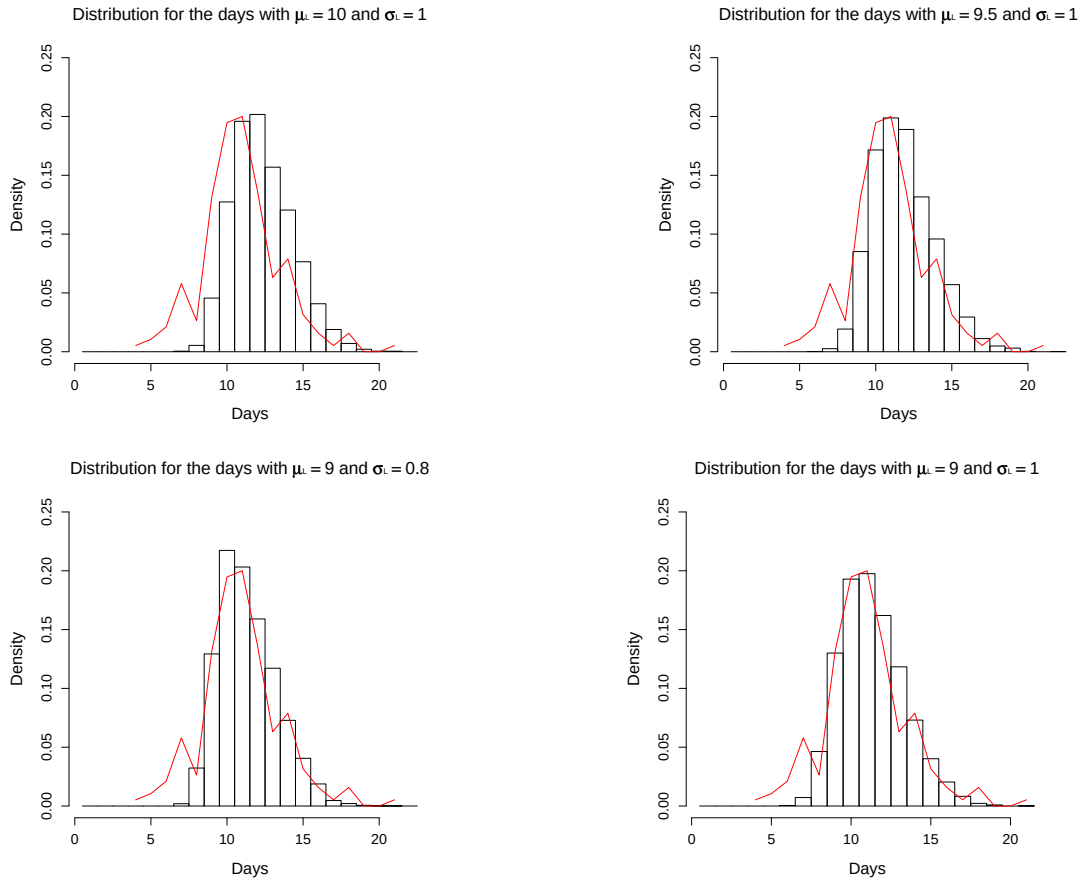


Figure 6.2: *Distribution for the number of days between infections for 2-susceptible households.*

the data values from Table 6.2 and each histograms is calculated from 10,000 simulated outbreaks in households with 2 susceptibles.

For simple observation, the data latent parameters that are close to describing the data are $\mu_L = 9$ and $\sigma_L = 0.8$. For these values once again we simulate the final outbreak size in households of size 4 and 3. See Figure 6.3.

Since the parameter values do not describe the data from the houses with 3 susceptibles (Table 6.3) as well as the data from the hoses with 4 susceptibles, this can suggest the hypothesis the households have different infection rates.

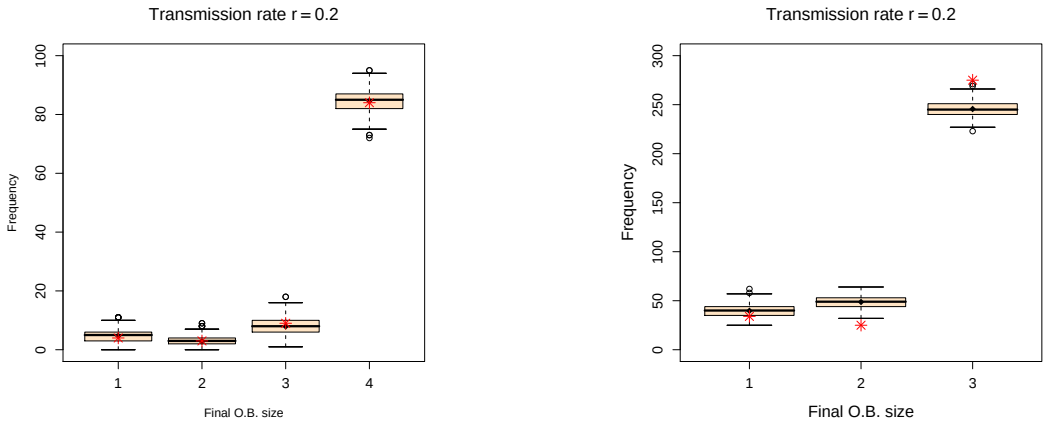


Figure 6.3: *Distribution of final O.B. size for 100 4-susceptible and 334 3-susceptible household for* $r = 0.2$*, infectious period* $\sim LN(1.7, 0.25)$ *and latent period* $\sim N(9, 1)$.

Total Num of cases	Frequency
1	34
2	25
3	275
Total	334

Table 6.3: *Total number of cases in households with three susceptibles.*

Bailey (1950) modeled the rate variation assigning a Beta distribution for the probability of transmission in the chain-binomial models; however this variation can be directly related to the ages distribution of the susceptible individuals in household of different sizes and the immunity that children younger than 18 month usually retain from the mother.

Using the information from the 1930 U. S. Census for Providence, Wilson et al. (1939) studied some of the incidence rates of measles and scarlet fever by family (household) size (number of individuals 21 year or younger). However since they also wanted to consider in their study the measles incidence up to household size they also obtained an approximation to the age distribution of the susceptible individuals in households of different sizes.

Information about the age distribution by family size was not contained in the 1930

U. S. Census. However, Wilson et al. obtained a sample of 8% from a census taken in Boston on January 13 of 1934 (carried out by the Civil Works Administration and Federal Emergency Relief Administration). The authors noticed that the distribution of the family sizes between their sample and the U.S. Census were "remarkably similar". Encouraged by this similarity they obtained the age distribution of children in families by size. The data corresponding to households of between 2 and 5 susceptibles are shown in Table 6.4 and Figure 6.4.

Age (years)	2 susc	3 susc	4-5 susc
0	3.5	2.9	2.3
1	4.3	3.4	2.6
2	5.1	4.4	3.4
3	5.1	5.2	3.5
4	4.3	4.8	4.0
5	4.5	4.4	4.6
6	4.3	4.9	4.8
7	3.9	4.6	5.1
8	3.7	4.9	5.3
9	4.0	4.3	5.5
10	4.8	4.2	5.5
11	4.1	4.6	5.7
12	4.3	5.2	5.9
13	4.3	4.5	5.6
14	4.1	4.9	5.7
15	4.1	4.7	5.1
16	4.9	4.9	5.0
17	4.8	5.1	4.7
18	4.8	4.6	4.1
19	6.0	4.7	4.5
20	5.5	4.6	3.3
21	5.5	4.1	3.6
Total	100	100	100

Table 6.4: *Distribution (in %) by age of children in families of different size (Wilson et al., 1939).*

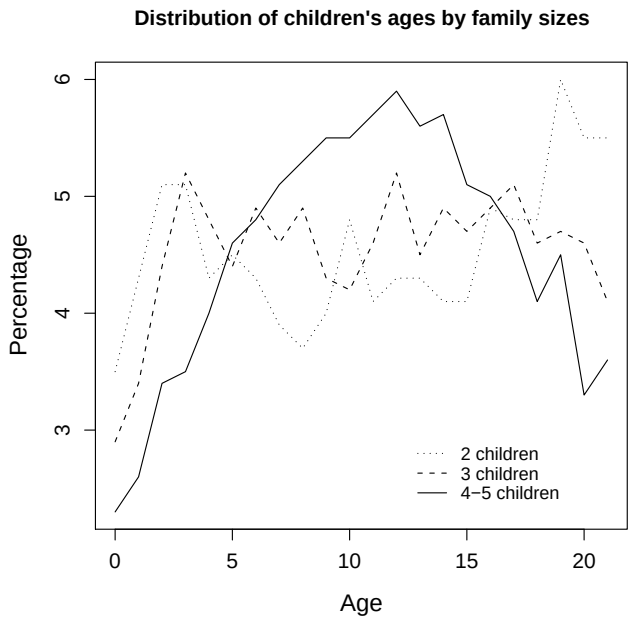


Figure 6.4: *Distribution by age of children in families of different size (Wilson et al., 1939).*

Using the information from Table 6.4 we simulated the distribution for the two groups of ages: "less one year old" and "1 year old and more" for houses with 3, 4 and 2 susceptibles, where all susceptibles are 10 years or younger (in case of the first two household sizes) and 14 years or younger (in the case of households with 2 susceptibles).

In simulating the ages in the households we considered the children to be at least one year apart. For example using the distribution for the ages in families of size 4 to 5 we sampled 4 ages without replacement and we used them if all of them were 10 years or lower. Otherwise, we obtained another set of 4 ages.

In order to estimate the parameters we obtain the Monte Carlo maximum-likelihood estimates. First we simulate the epidemic dynamic in each household 100,000 times for each parameter value. Then we estimate the probability density functions for the final outbreak size in households of size 3 and 4, and the number of days between infections.

These Monte Carlo distributions can then be used to approximate the likelihood based on the data presented in Tables 6.1, 6.2 and 6.3.

The likelihood function of $\boldsymbol{\theta} = (r_1, r_2, \mu_I, \sigma_I, \mu_L, \sigma_L)$ is

$$L(\boldsymbol{\theta}; \{x_{ij}\}_{i\in 1,2,3, j\in 1,\ldots,n_i}\}) \propto \prod_{1}^{334} \Pr\left(x_{1j}; \boldsymbol{\theta}\right) \prod_{1}^{100} \Pr\left(x_{2j}; \boldsymbol{\theta}\right) \prod_{1}^{235} f(x_{3j}; \boldsymbol{\theta})$$

where r_1 and $r = r_1 \times r_2$ are the rates of infection for individuals that are less than 1 year old and 1 year or older, respectively.

The data x_{ij} correspond to the j-th observation from the i-th house size. The index i is 1, 2 or 3 when the data corresponds to the final outbreak size in households of size 3; or in households of size 4; or the days between infections in households of size 2.

Since the final outbreak sizes for households of size 3 and 4 are discrete random variables, the probabilities $\Pr\left(x_{1j}; \boldsymbol{\theta}\right)$ and $\Pr\left(x_{2j}; \boldsymbol{\theta}\right)$ are unbiasedly estimated with the average number of time the values are observed.

The distribution for the days between infections is expressible through

$$f(x; \boldsymbol{\theta}) \;=\; \begin{cases} \Pr\left(\text{not transmission}\right), & \text{if } x = \infty \\ g(x; \boldsymbol{\theta}), & \text{if } x < \infty \end{cases} \tag{6.1.1}$$

where $g(x;\boldsymbol{\theta})$ is the sub-probability density function of a continuous random variable. This function can be estimated using simulated observations and smoothing the resulting empirical density function. In this study we use a Gaussian kernel with optimal smoothing parameter given by Bowman and Azzalini (1997, Section 2.4).

Regardless of the simulation size, the estimated likelihood function $\hat{L}$ is affected by the noise originated from the probability density estimation. To approximate to the real likelihood function, the obtained $\hat{L}$ are regressed on its 6 parameters, and since we are interested in the local values for the regressed function we apply a simple regression procedure.

We use the nonparametric regression implemented in the R package *np* (Hayfield and Racine, 2008) with the the Naradaya-Watson estimate:

$$\mu_h(\boldsymbol{\theta}) := E\left(\hat{L}|\boldsymbol{\theta}\right) = \frac{\sum K_h(\boldsymbol{\theta}-\boldsymbol{\theta}_i)\hat{L}}{\sum K_h(\boldsymbol{\theta}-\boldsymbol{\theta}_i)}$$

where K_h is based on the uniform kernel estimator

$$K(z) = \begin{cases} 1/2 & \text{if } |z| < 1 \\ 0 & \text{otherwise,} \end{cases}$$

as $K_h(z) = K(z/h)$.

The bandwidth value h is obtained with the least-square cross-validation method.

Then the Nelder and Mead method is used on the smoother regressed $\hat{L}$ to obtain the Monte Carlo likelihood estimates. The obtained estimates are shown in Table 6.5 and their fit is illustrated using 100,000 simulations in Figure 6.5.

Parameters	Estimate
r_1	0.1062
r	0.2860
μ_L	9.2138
σ_L	1.3709
μ_I	1.6902
σ_I	0.6243

Table 6.5: *Estimated parameters for measles outbreaks.*

As we can observe, the estimates are in agreement with the observed fact that infants are less susceptible to measles when they are less than one year old; and according to our estimates they are almost half as likely to acquire the illness as those individuals that are 1 year or older.

As we have mentioned before, Bailey estimated the probability of transmission as 0.709 (Greenwood model) and 0.653 (Reed-Frost). Both estimates are between 0.458 and 0.758, that correspond to the probability of transmission in a population of individuals less than one year old and 1 year and older, respectively.

Based on the estimated parameters then we have that 80% of infected individuals have a latent period between 7.5 and 11 days and 80% of the cases have infectious periods that last between 2.4 and 12.1 days.

The latent period originating from our estimation is very close to the one considered by Cliff et al. (8 to 12 days), and the interval for the infectious period includes the typical infectious period contemplated by PHAC (typically 8 days long).

In spite of the fact that the PHAC does not provide any information about any range for the infectious period it is important to notice that in our case the large variance of the infectious period (20.68) implies possible infectious periods that vary a great deal from the mean (6.587) and from the value 8.

The point estimates are computed by locally identifying the region $\mathcal{C}$ where the maximum likelihood estimate lies, obtaining $\hat{L}(\boldsymbol{\theta}_i)$ for several parametric points $\boldsymbol{\theta}_i \in \mathcal{C}$, locally regressing $\hat{L}$, and finding the parametric values that maximize the regressed hypercurve.

The estimation methodology used here can be regarded as maximum smoothed likelihood estimation (Iionides, 2005) and is related to methodology used in computer experiments (Sacks et al., 1989; Welch et al., 1992; Butler, 2001). In extension, the hypercurve that approximates to the likelihood function can be obtained for a wider range of parametric values. Some of the models and methods used for this task are linear regression, kriging and Bayesian spatial prediction (Sacks et al., 1989).

The broader hypercurve that approximates to the likelihood function can be used to

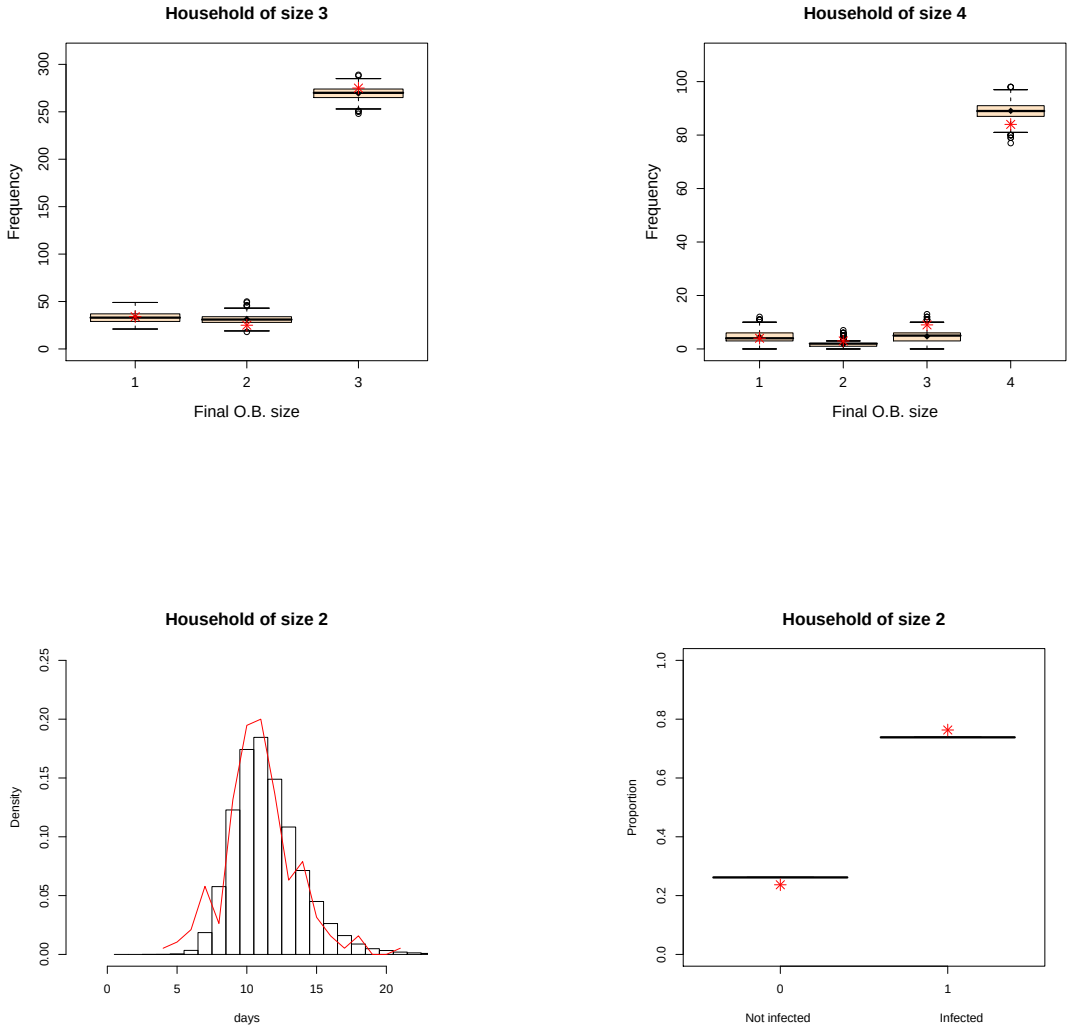


Figure 6.5: *Distributions of final O.B. size for 3 and 4-susceptible households, days between infections and probability of infection for 2-susceptible households.*

obtain an information matrix for assessing uncertainty of the parameter estimates. It can also be used to evaluate possible control strategies, varying the value for the model parameter within likelihood areas at level δ ($\delta \in (0,1)$).

6.1.3 Reed-Frost model

Since the measles has a long latent period compared with its infectious period the Reed-Frost model can be used to approximately describe the number of infected individuals at time t.

As an example of this approximation we consider the estimated transmissibility of measles in a population of 1 year or older ($\tau = 0.758$) and an interconnection described by a network with degree distribution Poisson(2).

In the results presented in Section 4.1 we assume that the population is very large. Then the expressions (4.1.3) and (4.1.5) describe the evolution of the outbreak while the reduction of the susceptible population is almost negligible with respect to the total population.

Figure 6.6 presents 20 simulations of outbreaks that start with a single infected, in a network with degree distribution Poisson(2) and order of 100,000 nodes.

The blue line represents the mean (4.1.3) and the red dotted lines are 2 standard deviations (using (4.1.5)) from the mean.

Since the network model for the Reed-Frost assumes that the number of nodes in the networks is large, it describes the evolution of the outbreak during its first stages, when the number of infected and removed is negligible with respect to the number of susceptibles.

6.1.4 Control measures

As mentioned in Chapter 5, one way of evaluating the control strategies is comparing the threshold values for similar control efforts.

For contact networks that are more complex than those studied here the analysis of control effects can be evaluated using simulation of the outbreaks.

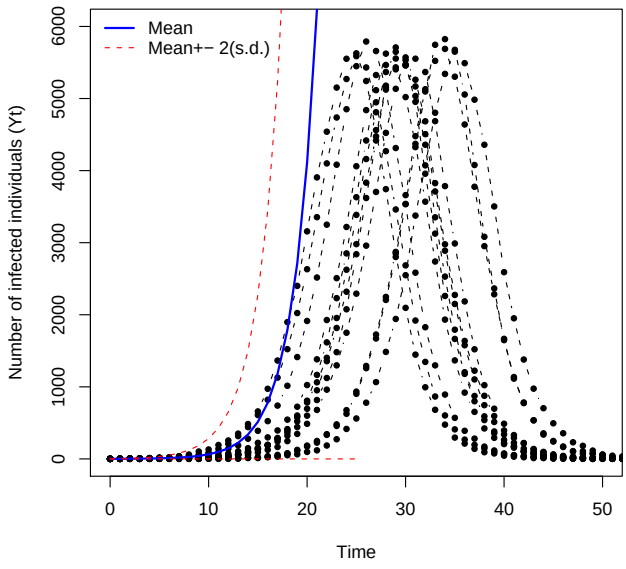


Figure 6.6: *Reed-Frost model in a network with 100,000 nodes and degree distribution Poisson(2).*

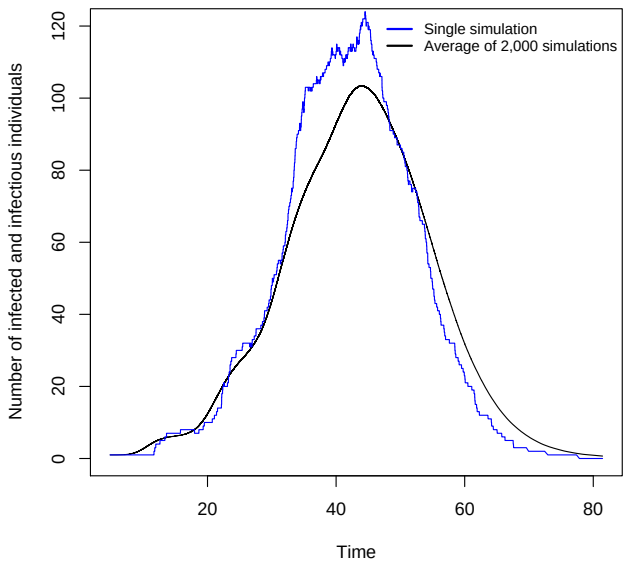


Figure 6.7: *Measles simulation in school.*

For example, we consider a school setting in which the only susceptible population members are 180 children who are divided in 6 classes. We assume that an individual has degree distribution Poisson(6) that connects him/her with individuals within the same class and that he/she also has degree Poisson(0.5) that connects him/her with each other class.

If we simulate the measles outbreak that starts with a single individual, then we can plot the number of infected and infective individuals at time t. This is the blue line of Figure 6.7. After repeating the simulation n times then the average number of infected and infectives at time t can be obtained. This average results in a smoothed line that in Figure 6.7 is represented by the black line and obtained from 2,000 simulations.

As indicated by Yan (2007) there are some public health objectives that are directly related with the curve of the number of infected and/or infectives at time t:

- reducing the initial growth of the curve (delaying the peak),
- reducing the peak burden and
- reducing the final outbreak size.

Then the control measures can be evaluated not only considering the final outbreak size but also considering the first two goals.

If we assume that the measles latent period is such that the first symptom appears after 4 days of being infective, then we can compare the outbreak evolution of the four control methods described in Section 5, comparing the average number of infected and infectives at time t.

Figure 6.8 presents the average for 2,000 simulations of the four control measures with parameters:

1. mass vaccination: $\nu = 0.2$,
2. acquaintance vaccination: $\omega = 0.2$,
3. ring vaccination: $q = 0.8$, $m = 4$,
4. isolation: $\gamma = 0.8$, $m = 4$.

As we can observe from Figure 6.8 the mass and the acquaintance vaccination have a similar effect over the outbreak. They decrement the initial growth of the curve and the total number of cases is lower than with the other two control strategies.

The ring vaccination and the isolation have hardly any effect on the outbreak but this is principally due to the facts that measles's incubation period is shorter than its latent period and the measles has a short infectious period. Thus the infective individuals freely transmit the agent to their neighbors for almost the entirely infectious period, before being detected and the neighbors vaccinated or the infectives isolated.

As pointed out in Section 5 it has been observed that the acquaintance vaccination performs better when reducing the probability of epidemics, over the mass vaccination,

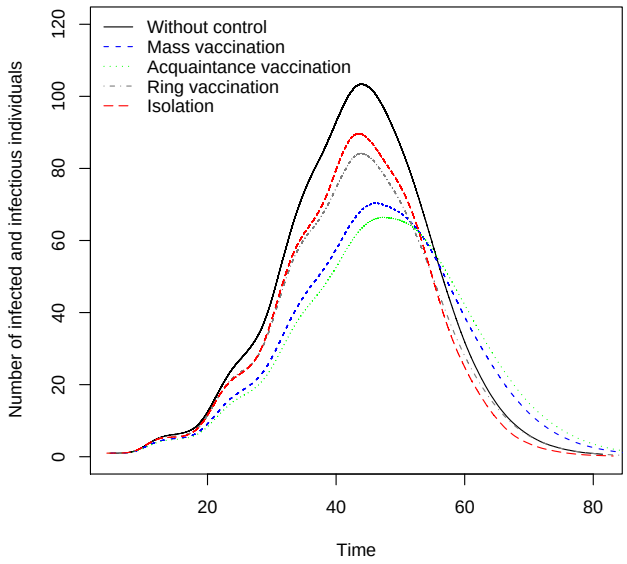


Figure 6.8: *Measles control simulations in school model.*

when the contact network is heterogeneous (Cohen et al., 2003; Takeuchi and Yamamoto, 2006).

The school model that we have described has a higher level of heterogeneity than a simple network with Poisson degree distribution; however this heterogeneity does not result in a important improvement of the acquaintance over the mass vaccination. This similarity prevails when considering different vaccination and selection fractions (ν and ω). Figure 6.9 shows the average final outbreak size for 5 different fractions for vaccination (mass vaccination) and acquaintance selection (acquaintance vaccination).

Figure 6.10 presents the distribution for the proportion of vaccinated individuals f_ω for different levels of selection fractions ω when the acquaintance strategy is implemented in the school model. The distribution is obtained from 2,000 simulation for each level of ω.

Since the values of the vaccination fractions f_ω are very close to the values ω (particularly for $\omega < 0.65$), then the similarities for the average final O.B. size for the mass vaccination and acquaintance vaccination still exist when the average final O.B size is compare with both vaccination fractions ν and f_ω.

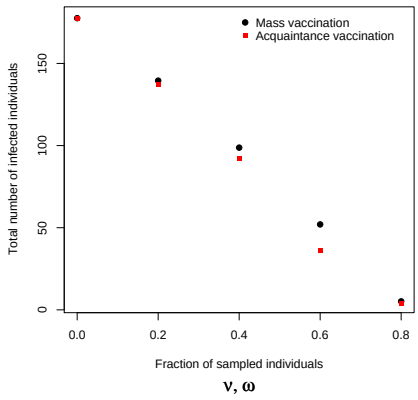


Figure 6.9: *Average final O.B. size for vaccination and selection fractions (mass and acquaintance vaccination strategies).*

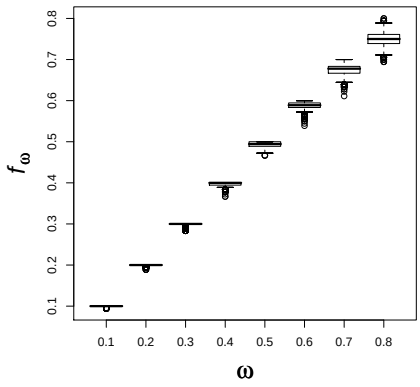


Figure 6.10: *Selection vs. vaccination fraction for acquaintance vaccination in the school model.*

Although the estimated mean final outbreak sizes are very similar for the mass and

acquaintance vaccination for the same parameter values, there can be some other outbreak characteristics that are differently affected by these strategies. For example, Figure 6.11 that examines the final outbreak distributions for the 5 different selection fractions, shows that the distributions are very different for the values $\nu > 0$, and $\omega > 0$.

From Figure 6.11 we can observe that the acquaintance vaccination has lighter tails for most fractions and therefore the acquaintance vaccination is preferable over the mass vaccination with these parameter settings.

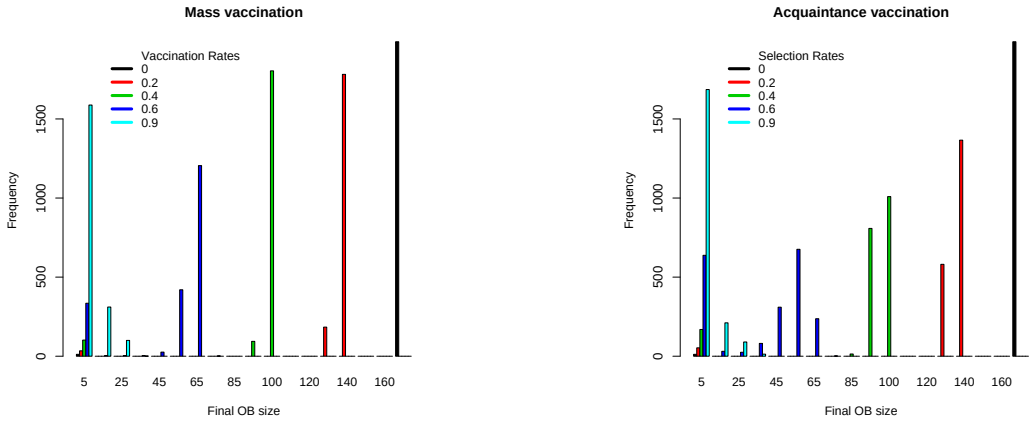


Figure 6.11: *Final O.B. size for mass and acquaintance vaccination.*

The final outbreaks size for the ring vaccination and isolation for the values of $q = 0.2, 0.4, 0.6, 0.8$, $\gamma = 0.2, 0.4, 0.6, 0.8$ and $m = 0, 2, 4$ are simulated 2,000 times and their average is shown in Figure 6.12. Table 6.6 shows the plot values and Figure 6.13 presents the boxplot for the outbreak size in each value of the parameters. From this last figure it can be observed that the variance of the final outbreak size is larger, the smaller the value of m. The ring vaccination and isolation have similar effects over the distribution of the final outbreaks size. However, the isolation is less effective when the delay to identify and isolate the infectives is larger.

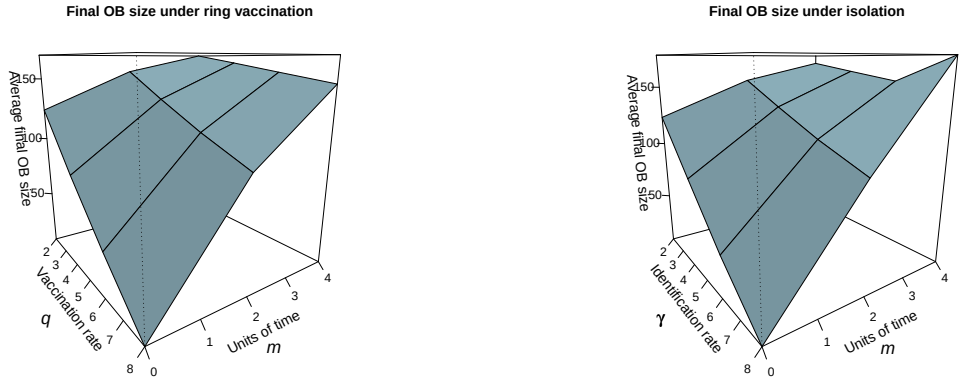


Figure 6.12: *Average final O.B. size for ring vaccination and isolation.*

Probability	Ring vaccination			Isolation		
q, γ	$m = 0$	$m = 2$	$m = 4$	$m = 0$	$m = 2$	$m = 4$
0.2	124.4005	154.251	168.75	123.767	153.1725	168.9145
0.4	81.1125	133.1615	162.137	81.7835	132.7065	161.837
0.6	41.7625	113.4115	155.2175	40.8195	113.224	155.0005
0.8	5.9105	96.2295	148.42	5.702	96.7795	176.3715

Table 6.6: *Average final O.B. size for ring vaccination and isolation.*

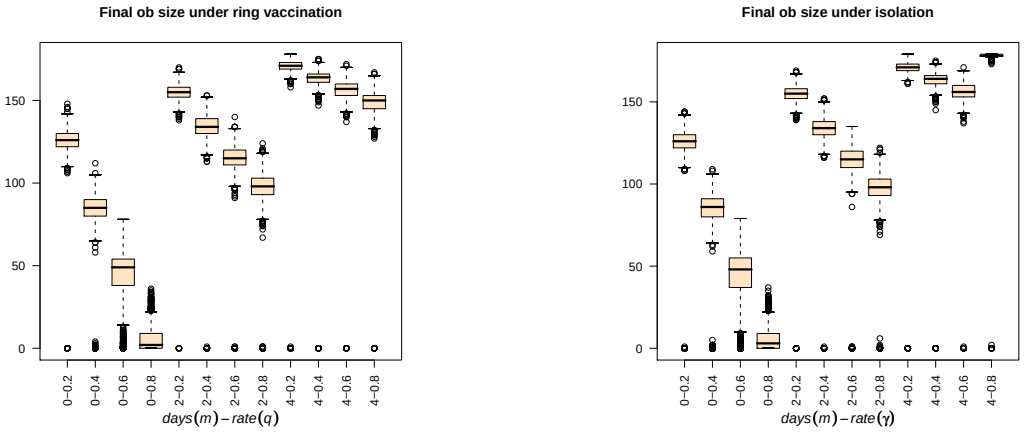


Figure 6.13: *Final O.B. size for ring vaccination and isolation.*

6.2 Analysis of influenza data

This section describes the influenza illness, the influenza data corresponding to outbreaks in long term care in the Region of Waterloo and presents the influenza estimates obtained based on four influenza outbreaks.

Sections 6.2.1 describes the influenza virus, its types and subtypes, their transmission dynamic among humans, the evolutionary stages in an infected individual and the effectiveness of seasonal vaccines.

Section 6.2.2 discusses some of the data characteristics and in Section 6.2.3 we obtain the model point estimates based on the Monte Carlo maximum-likelihood function and the information from four different outbreaks.

In order to do some cross-evaluations of the obtained parameters, utilizing the information from some other outbreaks in the database, in Section 6.2.4 we obtain the clinical diagnosis based on the reported individual symptoms and in Section 6.2.5 we predict the influenza status of the symptomatic individuals.

6.2.1 Influenza

Influenza is the clinical illness caused by viruses of the group of myxoviruses. These viruses can be classified into three types: A, B and C (discovered in 1932, 1940 and 1949, respectively). Type C influenza has been identified widely but infrequently and is not always associated with influenza epidemics (Cliff et al., 1986). This virus appears to be endemic and has not been associated with the regular influenza epidemics. The influenza C virus is morphologically and genetically different from the other two viruses and is generally non-symptomatic, so it is of little medical concern and it will not be considered in this study.

Due to the similarity between viruses A and B, they have been well characterized. Influenza A is the most important virus epidemiologically, and has the widest range of hosts (Stuart-Harris et al., 1985) since it has the ability to infect animals as well as humans

(Guo et al., 1983). Most cases of epidemic or pandemic flu are caused by the influenza A virus, but the B virus, which normally is only found in humans, is responsible for many localized outbreaks (Smith and Palase, 1989).

The influenza virus has a kind of reproduction that makes it able to replicate rapidly and likely produce mutant descendants. Some of the mutations affect the agent's surface that is closely associated with attachment functions of the virus to a host cell, providing the new viruses the ability to bypass any immunity acquired by the individual who was previously exposed to the same virus type with different surface structure. In the case of the A virus, the antigens can present frequent minor changes, named *drifts*, or infrequent major changes, named *shifts*. These modifications can be detected in serological tests and can affect either or both of the surface antigens (Cliff et al., 1986).

The influenza virus type B is more stable than the virus type A since it only presents drifts, so that individuals previously exposed to the B influenza virus will possess some degree of immunity against its next drift. Because of this we might expect the frequency of epidemic caused by the B virus to be related mainly to a buildup though births of a stock of susceptibles in the population who have never had influenza caused by the B virus (Cliff et al., 1986).

Mutations in the surface of the influenza A virus have resulted in a number of different influenza subtypes and strains. Specific varieties of the virus are generally named according to the particular antigenic determinants of the attachment and reproduction: the haemagglutinin (denoted as H and with 13 major types) and the neuraminidase (denoted as N and with 9 major types), respectively.

The A virus first isolated in the 1930's, with its particular H and N surface antigens, was classified by WHO in 1980 as H1N1. On each occasion that a shift has occurred in one of the surfaces antigens to a new subtype, the digit relating to the changing antigen has been incremented by one.

Epidemiology

Influenza viruses pass from person to person as droplets of respiratory secretion exhaled by an infected that are inhaled by susceptible individuals. The big droplets travel only a few feet before falling to the ground and smaller particles (2-4 micrometers) can remain suspended in the air for up to an hour.

Coughing or sneezing will increase droplet emission and some external conditions can enhance the virus survival and infectivity. According to Cliff et al. (1986), the transmission will vary with three factors:

1. Individual factor. Infected individuals may differ in the amount of virus that they shed into the environment.

2. Proximity. The proximity and degree of crowding of the susceptibles in the population will affect the transmission rate.

3. Ambient conditions in the environment outside the host. Clinical observations suggest that both low humidity and temperatures predispose the virus to survive and transfer, but laboratory experiments suggest additional and undefined factors related to the winter season.

The last two factors are often confounded because the winter period may bring together both crowding and low temperature and humidity.

According to Bean et al. (1982) influenza A and B have been shown to survive on hard surfaces such as stainless steel and plastic for 24-48 hours, and on soft surfaces such as cloth, paper and tissue for up to 12 hours.

In contrast to human influenza spread, in many animals species such as ducks, influenza is transmitted by the faecal route rather than by air (Shortridge, 1997).

The transmission of influenza has been long considered as direct, but the outbreaks can be mixed with some other mechanisms as suggested by Hope-Simpson (1979, 1992). He proposes a latent virus hypothesis to explain the long interval between epidemics, the

virtual disappearance of the virus between outbreaks, the relatively rare documentation of secondary cases within households, and the sudden appearance of influenza in isolated locations (Cliff et al., 1986). Hope-Simpson suggests the influenza virus, like the *Herpes simplex* virus, invades the individuals and leaves a lifelong residue of latent virus that is periodically reactivated.

According to Cliff et al. (1986), the infection with influenza virus may or may not lead to clinically manifest disease but may impart immunity in either case. Even with no symptoms the virus may be responsible for some transmission of virus (Fine, 1981). Furthermore, Fine mentions that several authors (Care et al., 1958 and Davis et al., 1970, among some others) have estimated the percentage of asymptomatic infections between 20 and 75%. The proportion probably varies with virus strain, but it is unclear to what extent this occurs.

Symptomatic influenza differs considerably in severity from individual to individual depending on age and existing state of health as well as on the virus strain.

The general evolution of the disease in a host, without considering type or strain variations can be summarized as follows:

1. The incubation period is estimated between 18 and 72 hours. Cliff et al. (1986) assumes a range of one to three days and a time distribution that is typically lognormal.

2. The symptomatic stage (clinical illness) ranges from two to seven days and conventionally is assumed to be normal (Cliff et al., 1986). The total course of an attack from initial receipt to recovery may occur within ten days.

3. The latent period and the infectious period are not synchronous with that of disease. The latent period ends approximately 24 hours before the end of the incubation period and the infectious period is probably of the order of 3 to 6 days (Fine, 1981). See Figure 6.14.

Some of the symptoms the influenza may include are chills, fever (38° to 40° C), fatigue

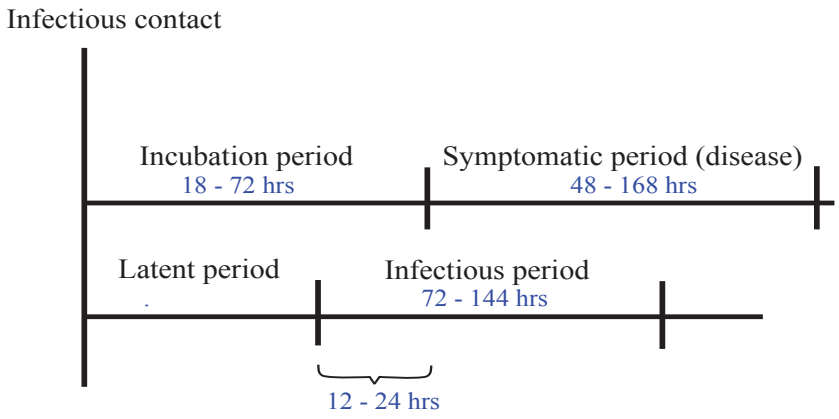


Figure 6.14: *Individual influenza evolution.*

(malaise), muscle aches (myalgia), headache, sore throat, dry cough, sneezing, irritated watering eyes, nasal congestion, redness in the eyes or skin, and nausea.

Due to the fact that several of the influenza symptoms are shared with the common-cold, influenza is difficult to distinguish. However the flu symptoms are more severe than those produced by the common-cold equivalents.

Call et al. (2005) did a meta analysis on signs and symptoms of influenza including 6 studies and 7,105 patients. They computed the positive and negative likelihood ratios (LR+ and LR-, respectively) and the diagnostic odds ratio (DOR). The LR+ is a measure to "rule-in" influenza given that the specific symptom is present. It is the ratio of the probabilities of an individual having and not having the disease given that the symptom is present. Similarly the LR- is a measure to "rule-out" influenza given that the specific symptom is not present.

Call et al. (2005) found that among the symptoms: fever, feverishness, cough, myalgia, malaise, headache, sore throat, sneezing, nasal congestion, chills, vaccine history, fever-

cough, and fever-cough-acute onset "no single clinical symptom consistently had a LR+ high enough to clinically *rule-in* influenza, nor did any single finding have a LR- low enough to clinically *rule-out* influenza. The best symptoms for excluding the diagnosis of influenza from common-cold were fever, cough and nasal congestion".

However, for decreasing the likelihood of influenza, the absence of fever, cough and nasal congestion were the only symptoms with LR- less than 0.5. See Table 6.7.

Symptom \	LR+ (95% CI)	LR- (95% CI)	DOR (95% CI)
Fever	1.8 (1.1-2.9)	0.40 (0.25-0.66)	4.5 (1.8-11.0)
Cough	1.1 (1.1-1.2)	0.42 (0.31-0.57)	2.8 (2.1-3.7)
Nasal congestion	1.1 (1.1-1.2)	0.49 (0.42-0.59)	2.3 (1.9-2.8)

Table 6.7: *Best individual symptoms for excluding influenza, Call et al. (2005).*

Vaccines

In the case of influenza A, the shifts are the result of recombination of the influenza genome. This event can occur when a cell is simultaneously infected by two different strains of type A influenza (Wharton et al., 1989). Since influenza A can infect several animal species, the virus shift likelihood increases. This virus characteristic hinders the vaccine's efficiency against new subtypes. Nevertheless the shift events can be decreased by avoiding individual multi-infection through the vaccination of susceptible animal species against known influenza A strains.

Even though the drifts involve small changes in the virus antigenic structure, the accumulative changes can result in the failure of the immune system to detect the new strain of the virus. However, the yearly influenza vaccine reformulation often enables scientists to take into account any new strains that have emerged so as to increase the probability of the individual to be immune to existing and new strains.

According to the manual published by the Public Health Division (PHD) and Long-Term Care Homes (2004), the protection from the vaccine develops by two weeks after the

shot, and may last up to one year. The vaccine is about 70 to 90% effective in preventing influenza infection in healthy adults. In children, it is about 77 to 91% effective and in elderly people, it is just about 30 to 40% effective. However the vaccine among elderly people can prevent pneumonia and hospitalization in about 50 to 60% or cases, and prevent death in about 80% of cases (PHD, 2004; NACI, 2000).

6.2.2 Data description

The data consist of two databases. One contains the laboratory-confirmed cases of influenza presented in the Region of Waterloo and the other, the respiratory cases in long term care facilities (LTC) in the same region.

The first database (GCCD) was provided by the Public Health Unit of the Region of Waterloo (general confirmed cases) and it contains information for the influenza cases with onsets between January 1995 and December of 2004. For the purpose of this work we consider only the information for the two seasons: Sep 1st, 2003-August 30th, 2004 (153 cases) and Sept 1st, 2004-August 30th, 2005 (3 cases).

The second database (LTCID) was captured from the investigation of respiratory outbreaks in long term care facilities (LTC), compiled by the the Public Health Unit of the Region of Waterloo. It has 2,582 cases registered in the three seasons: 2003-2004 (1,215), 2004-2005 (1,300) and 2005-2006 (67 cases).

A total of 44 cases were repeated in the two databases resulting in a total of 2,694 cases registered in the merged data.

The LTCID has the respiratory cases (influenza confirmed or not) that occurred among the resident and staff in each of the LTC's. Each registered case has variables such as the sex, age, unit, resident (yes, no=staff), day of onset of the symptoms (integer), prior influenza vaccine (yes, no), symptoms, clinical diagnosis and influenza laboratory-confirmed cases (yes, no).

The clinical diagnosis of the respiratory cases is the individual evaluation based on the new respiratory symptoms. The clinical diagnosis can be *upper respiratory tract illness*

(URTI) and *influenza-like illness* (ILI).

In the LTCID there are two influenza periods (2003-2004 and 2004-2005) for which the information includes the not confirmed cases with respiratory symptoms associated to ILI and URTI. That is, the merged information from September 1st, 2003 to August 30th, 2005 contains the confirmed cases of influenza in the Region of Waterloo (mainly LTC's) and the not confirmed cases in LTC's that were associated to ILI or URTI.

The age distribution in the GCC data contains some newborns and the LTCID mostly contains information for elderly people. Figure 6.15 shows the age histogram in the combined database for all the seasons and Table 6.8 shows the number of cases by age groups and seasons.

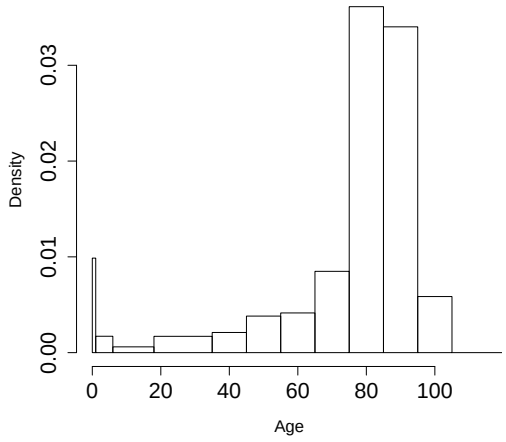


Figure 6.15: *Age distribution in influenza cases.*

Age \ Season	03-04	04-05	05-06	Total
[0,1)	14	1	0	15
[1,6)	13	0	0	13
[6,14)	5	1	0	6
[14,65)	136	64	2	202
[65,75)	77	50	2	129
[75,85)	275	263	11	549
[85,95)	259	247	11	517
[95,100)	38	46	2	86
Missing	507	631	39	1,177
Total	1,324	1,303	67	2,694

Table 6.8: *Age distribution by influenza season.*

Most of the respiratory activity is registered in LTCID but from the GCC database we recover some information of the activity outside LTC's.

In the database there is a total of 46 facilities (LTC and hospital) registered with respiratory activity during the three flu seasons. For each outbreak the information related to the number of residents and staff in each affected facility was captured.

The facilities can be classified into two types according to the isolation measures that are typically taken during an infectious outbreak. This classification is then in close relation with the facilities' size. During the outbreak small facilities tend to isolate the residents from any exterior contact, and in larger facilities the isolation is by units or floors and the staff is divided accordingly into cohorts .

6.2.3 Parameter inference

The clinical diagnosis is the assessment based on a study of the signs and symptoms of a disease. This diagnosis is elaborated by the medical personnel in the facilities considering only the new individual symptoms, since some existing ones can be due to chronic conditions.

For the respiratory cases there exist two possible clinical diagnoses: *influenza-like illness* (ILI) and *upper respiratory infection* (URTI). The first diagnosis is associated with the presence of the influenza virus and the second one is related with the common-cold.

In order to study the evolution of the influenza cases during the outbreaks in the long term care facilities, we propose to analyze the transmission of the cases with clinical diagnosis ILI.

The epidemic model parameters were estimated using the information of four influenza A outbreaks in four different facilities with number of residents and staff presented in Table 6.9.

Facilities	F1	F2	F3	F4
Residents	68	79	40	52
Staff	45	58	21	40

Table 6.9: *Resident and staff population during the outbreaks.*

Since the considered facilities have similar isolation policies and type of rooms (single, semi-private and wards), they are assumed to have similar contact structure.

For the four outbreaks we assume that resident and staff individuals have an average of 3 contacts *within* their respective populations and the staff members have an average of 4 contacts with residents.

The mean inter-population connection between the resident and staff population is adjusted by the population size and presented in Table 6.10.

Facilities	F1	F2	F3	F4
Staff-residents	4	4	4	4
Residents-staff	2.65	2.94	2.1	3.08

Table 6.10: *Mean contact between the resident and staff population.*

If the degree distribution between populations follows the Poisson distribution, the probability of a staff member not having any contact with any resident is greater than zero. However this is not far from reality since some staff personnel included in the reports work in the administration or maintenance and do not have any interaction with residents.

The inference for the model parameters is based on the times between symptom onset for the residents and staff members. For example, suppose the onset times of the resident population are $i+2, i+2, i+5, i+6$ and the onset times of staff members are $i, i+3, i+4$; then the inter-event times configurations for the resident and staff population are {2, 0, 3, 1} and {3, 1}, respectively.

As we have mentioned above, although the outbreaks contain at least one confirmed case with influenza A, not all the symptomatic cases were tested. In the data utilized to estimate the parameter we include the ILI cases that were not tested, or were tested and resulted positives for influenza.

In Table 6.11 we present the inter-events for the four outbreaks. The third and fourth outbreak report initial onsets that are too close to consider that we have actually observed the first outbreak case. We assume that only the initial case is missing and hence that we fail to observe only the first time between onsets.

Given that the almost all individuals in the facilities are vaccinated, but the vaccine

Facilities		Inter event times (days)
F1	Resident	{3, 1, 1}
	Staff	{3, 4}
F2	Resident	{4, 0, 0, 1, 0, 0, 2, 1, 0}
	Staff	{4}
F3	Resident	{Miss, 0, 1, 1, 0}
	Staff	{Miss}
F4	Resident	{Miss, 0, 2, 1, 1, 1, 0, 1, 0, 1, 0}
	Staff	{Miss, 4, 0, 0, 0, 0, 2}

Table 6.11: *Inter-event times for influenza outbreaks.*

has two different effectiveness levels for the residents and staff individuals, we introduce the variable s that describes the staff susceptibility to the agent.

In order to describe the outbreak evolution in the hierarchical network we assume that:

- the transmission rate is a constant r,
- the susceptibility to the agent is 1 for the residents and s for the staff members,
- the degree distributions in the networks are Poisson,
- the latent period is Gaussian distributed with parameters μ_L and σ_L, and
- the infectious period is Gaussian distributed with parameters μ_I and σ_I.

The variable that is used to make inference on the parameters is the time between starts of infectious periods. This variable is reported by program as a continuous variable (in hours) and the difference between consecutive events is rounded into days.

The difference between the starting of the infectious and symptomatic periods has been estimated to be between 12 and 24 hours (See Figure 6.14). If the waiting time for becoming symptomatic after becoming infective is a random variable that has one mode, the times between starts of infectious periods and the times between symptoms onsets are close to being equal.

As in the measles' parameter inference we obtained the model parameter using the Monte Carlo maximum-likelihood estimate. In order to do this for several parametric points we simulated the outbreaks in each facility several times.

Using the population sizes given in Table 6.9 and the model assumptions listed above, we simulate the outbreak 1,000 times for each facility and set of parameter values to obtain the approximation to the likelihood

$$L\left(\theta; \{t_{r_{kj}}\}_{(k,j)}, \{t_{s_{kj}}\}_{(k,j)}\right) = \prod_{k=1}^{4} \left(\prod_{j=1}^{n_k} \Pr(T_{r_{kj}} = t_{r_{kj}}) \prod_{j=1}^{m_k} \Pr(T_{s_{kj}} = t_{s_{kj}}) \right)$$

where $\boldsymbol{\theta} = (r, s, \mu_I, \sigma_I, \mu_L, \sigma_L)$, n_k and m_k are the number of observed inter-event times in facility k, and $T_{r_{kj}}$ and $T_{s_{kj}}$ are the $j-th$ resident and $j-th$ staff observed times between events, respectively.

Each term $\Pr(T_{u_{kj}} = t_{u_{kj}})$, $u = r, s$, $k = 1, 2, \ldots$ is approximated with the frequency estimator, and the resulting $\hat{L}$ is regressed on the parametric values using 195 different parametric points and the R package written by Hayfield and Racine (2008). As in the estimation of the measles parameters we use the Naradaya-Watson estimate.

The fitted regression that describes the likelihood function is then maximized using the Nelder and Mead method. The obtained parameter estimates are shown in Table 6.12.

Parameters	Estimate
r	0.00799
s	0.84085
μ_L	62.62974
σ_L	4.33776
μ_I	127.58110
σ_I	15.73186

Table 6.12: *Estimated parameters for influenza outbreaks.*

We observe that the fitted parameters describe the influenza as an infectious agent with transmissibilities within the resident and staff populations of $\pi_{(B,B)} = 0.6368$ and

$\pi(A, A) = 0.5737$, respectively. This transmissibilities in the within-populations basic reproductive numbers $\mathcal{R}_{0,\mathbf{resident}} = G'_{K_{(B,B)T}}(s) = \pi_{(B,B)}E(K_{(B,B)}) = 1.9104$ and $\mathcal{R}_{0,\mathbf{staff}} = G'_{K_{(A,A)T}}(s) = 1.7211$, and both values are consistent with some studies that estimate $1 < \mathcal{R}_0 < 2$, described next.

The basic reproductive number $\mathcal{R}_0$ has been estimated in diverse populations and population settings (households, schools, day-care centers) and has been shown to be highly variable. For some strains, like the 1918 influenza, $\mathcal{R}_0$ was estimated to have a value between 2 and 3 (Mills et al., 2004), while some other authors have estimated $\mathcal{R}_0$ to be between 1 and 2 (for example Vynnycky et al. 2007).

The Public Health Agency of Canada considers that the influenza virus has a replacement number $\mathcal{R}$ between 1.4 and 1.8. In the cases of networks with Poisson degree distribution $\mathcal{R} = \mathcal{R}_0$ and $\mathcal{R} > \mathcal{R}_0$ in more heterogeneous contact structures.

Based on the fitted parameters in Table 6.12 we have that 95% of the cases have a latent period of between 53.95 and 71.30 hours while that same proportion of cases have an infectious period between 96.12 and 159.04 hours.

Comparing these intervals with the range of values obtained in previous studies (see Figure 6.14), we see that our parameter estimates imply larger values for latent and infectious periods, with less variation between individuals. The high values and smaller variances can be directly related to the fact that most of the observed cases correspond to people in the group of age of sixty or more.

Mean final outbreak size and epidemic probability

The mean for the final outbreak size is given by the (3.3.31) in terms of the functions $N_{\underline{(\mathbf{A},\mathbf{A})}}$ and $N_{\underline{\mathbf{ABAT}}}$. Since all the degrees in the hierarchical network used to model the contacts in the facilities have a Poisson distribution, it is easy to find a expression for them.

Next we use the fact that the mean degree and the mean excess degree distribution are the same when the degrees distributions are Poisson.

$$N_{\underline{(\mathbf{A},\mathbf{A})}} = \frac{\lambda_{(A,A)}\pi_{(A,A)}}{1-\lambda_{(A,A)}\pi_{(A,A)}},$$

now

$$\begin{aligned} G'_{\underline{\mathbf{ABAT}}}(1) &= \lambda_{(A,B)}\pi_{(A,B)}\lambda_{(B,A)}\pi_{(B,A)}\sum_{i=0}^{\infty}\left(\lambda_{(B,B)}\pi_{(B,B)}\right)^i \\ &= \lambda_{(A,B)}\pi_{(A,B)}\lambda_{(B,A)}\pi_{(B,A)}\frac{1}{1-\lambda_{(B,B)}\pi_{(B,B)}} \\ &= G'_{\underline{\mathbf{ABAT1}}}(1), \end{aligned}$$

if $\lambda_{(B,B)}\pi_{(B,B)} < 1$. Hence

$$N_{\underline{\mathbf{ABAT}}} = \frac{\lambda_{(A,B)}\pi_{(A,B)}\lambda_{(B,A)}\pi_{(B,A)}}{1-\lambda_{(B,B)}\pi_{(B,B)}-\lambda_{(A,B)}\pi_{(A,B)}\lambda_{(B,A)}\pi_{(B,A)}}.$$

From the last expression we observe that we can obtain the mean final outbreak size if:

1. $\lambda_{(A,A)}\pi_{(A,A)} < 1$, $\lambda_{(B,B)}\pi_{(B,B)} < 1$, and

$$\lambda_{(A,A)}\pi_{(A,A)}\lambda_{(A,A)}\pi_{(A,A)} < 1-\lambda_{(B,B)}\pi_{(B,B)}, \textit{or}$$

2. $\lambda_{(A,A)}\pi_{(A,A)} < 1$ and $\pi_{(A,B)} = 0$ or $\pi_{(B,A)} = 0$

The parameter estimates for the influenza data in facilities indicate that the influenza has a probability larger than zero, since the mean excess degrees are all greater than one and we estimate

$$\begin{aligned} \pi_{(A,A)} &= \pi_{(B,A)} = 0.5737 \\ \pi_{(A,B)} &= \pi_{(B,B)} = 0.6368. \end{aligned}$$

On the other hand, the basic reproductive numbers for the infections that occur only within the same populations, are

$$\begin{aligned} \mathcal{R}_{0,\mathbf{staff}} &= G'_{K_{(A,A)^T}}(1) = \lambda_{(A,A)}\pi_{(A,A)} = 1.7211 \\ \mathcal{R}_{0,\mathbf{resident}} &= G'_{K_{(B,B)^T}}(1) = \lambda_{(B,B)}\pi_{(B,B)} = 1.9104 \end{aligned}$$

but since $\lambda_{(B,B)}\pi_{(B,B)} > 1$ and $\lambda_{(A,A)}\pi_{(A,A)} > 1$ then the basic reproductive numbers

$$\mathcal{R}_{0,\text{staff}} = G'_{K_{(A,A)T}}(1) + G'_{\underline{\mathbf{ABAT}}}(1)$$

and

$$\mathcal{R}_{0,\text{resident}} = G'_{K_{(B,B)T}}(1) + G'_{\underline{\mathbf{BABT}}}(1)$$

diverge.

If a control measure like isolation and cohort staffing are implemented to reduce the transmissibilities to half their values, then the probability of epidemic is reduced to zero and the expected final outbreak size can be obtained.

6.2.4 Cross analysis based on the clinical diagnosis

In order to evaluate the fitted parameter we use the evolution information of other outbreaks and compare it with the simulated evolution with the estimated parameters. In this subsection we use the onset information of resident and staff cases clinical diagnosed as ILI.

In the handwritten reports of the Public Health Unit the clinical diagnosis was obtained together by the health care workers at the LTC facilities and the Public Health Unit personnel. Unfortunately, such a classification is not always present in the reports.

However, since for some cases the most important symptoms were captured, a general classification or diagnosis can be retrieved for the study of ILI transmission.

According to the Public Health Unit of the Region of Waterloo, a general classification or diagnostic evaluation is the following:

- The disease is URTI if we have at least two of the following symptoms: Abnormal temperature (high or low), cough, sore throat, chest congestion, nasal congestion, laryngitis, hoarse voice, runny nose, wheezing, head congestion, sinus pain, and sneezing.

- The disease is ILI if

1. we have at least two of the following symptoms: headache, general aches, or abnormal temperature, and
2. we also have at least two of the following symptoms: general aches, cough, sore throat, chest congestion and nasal congestion.

Notice that a case which is classified as ILI is not necessarily URTI.

In the database there is a total of 650 cases in which no symptoms is recorded ("Miss Symp"), but 299 were classified as ILI and 220 as URTI.

The agreement between the original diagnosis and the diagnosis obtained from these definitions is shown with bold font in Table 6.13.

Original \ Obtained	Not a case	URTI	ILI	Miss Symp	Total
Missing	149	647	122	131	1,049
ILI	18	174	**98**	299	589
URTI	110	**694**	32	220	1,056
Total	277	1,515	252	650	2,694

Table 6.13: *Comparison between clinical diagnosis and diagnostic evaluation.*

Given that in the elderly population with influenza, fever can be absent (PHD, 2004), a second illness definition or diagnosis can require only two of any of the three most important symptoms considered by Call et al. (2005) (fever, cough and nasal congestion). Then we consider the following second diagnostic evaluation based on the respiratory symptoms:

- The disease is URTI if at least two of the following symptoms are present: headache, general aches, cough, sore throat, chest congestion, nasal congestion, tiredness, laryngitis, hoarse voice, runny nose, wheezing, chills, head congestion, sinus, sneezing.
- The disease is ILI if it is URTI and at least two of the following symptoms are present: abnormal temperature, cough, and nasal congestion.

The agreement between the clinical diagnosis and the last diagnostic evaluation is shown in Table 6.14. As we can see, more cases are diagnosed as ILI and fewer as *not respiratory*

cases, but there are a large number of cases originally classified as URTI that are now assigned to ILI.

Original \ Obtained	Not a case	URTI	ILI	Miss Symp	Total
Missing	109	247	454	239	1,049
ILI	19	41	**226**	303	589
URTI	84	**287**	465	220	1,056
Total	212	575	1,145	762	2,694

Table 6.14: *Comparison between clinical diagnosis and second diagnostic evaluation.*

Based on the 334 tested cases for which we have the original diagnosis (See Table 6.15) we have that 48.03% and 51.97 % of the residents and staff diagnosed as ILI were positive and negative for influenza, respectively. It is evident that when clinical diagnosis is ILI, it poorly describes the probability that a case is positive for influenza; nevertheless the probability for negative and positive influenza given that the clinical diagnosis is URTI are 86.67% and 13.33%. Hence the URTI category contains substantial information about the probability of cases being negative for influenza.

	Neg	Pos	Total
ILI	119	110	229
URTI	91	14	105

Table 6.15: *Original clinical diagnosis and test result.*

Now we use the inter-event times of ILI cases and compare them with the simulation outputs that use the Monte Carlo estimates of the influenza parameters (Table 6.12).

First we consider an outbreak that occurred in a facility with 101 residents and 70 staff members. The resident inter-onset times for the cases diagnosed as ILI are $\{\text{Miss}, 2, 1, 1, 2\}$ and only one secondary staff case was registered.

We simulated the outbreak 1,000 times and obtained the mean and variance of the first, second, third, fourth and fifth inter event-times. In Figure 6.16 we show the mean

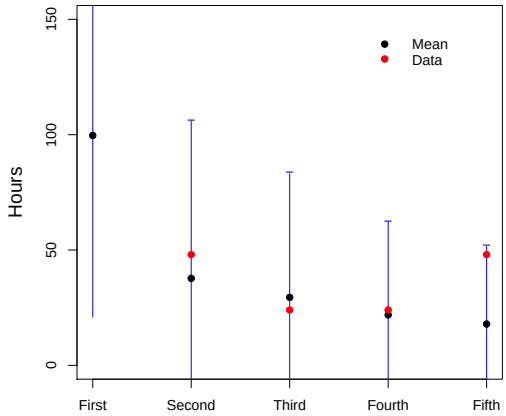


Figure 6.16: *Distribution of residents' inter-event times from simulated outbreaks.*

inter-event times, the points that are two standard deviations around the mean, and the data.

From the figure we have that the second, third and fourth observed times are very close to the estimated mean and the fifth time is further from the estimated mean but less than two standard deviations from it. Hence the observed data do not contradict the estimates obtained for the influenza parameters.

Now we consider another facility with 79 residents and 58 staff members. For this facility three different respiratory outbreaks were reported. They were not reported as flu-outbreaks since the residents tested resulted in negatives.

In Figure 6.17, as in Figure 6.16, we show the mean inter-event times and deviation simulated from 1,000 simulations. Since not all outbreaks evolve into epidemics, the mean and standard deviations of the i-th inter-event is obtained from those outbreaks that at least had $i + 1$ cases.

The connected points in Figure 6.17 are the inter-event times observed in each outbreak for the cases which are classified as ILI according to the second diagnostic evaluation

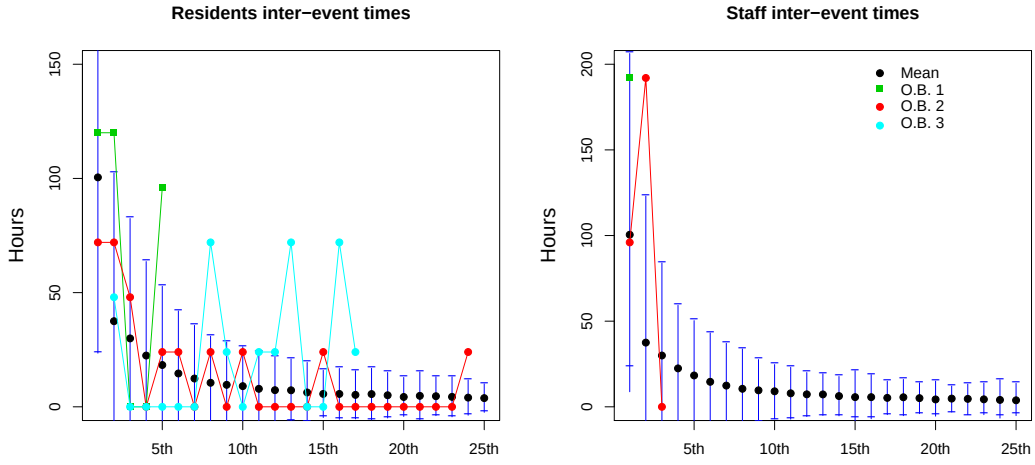


Figure 6.17: *Residents and Staff inter-event times from simulated outbreaks.*

described above.

For the three outbreaks the resident inter-onset times (in days) are {5, 5, 0, 0, 4}, {3, 3, 2, 0, 1, 1, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1} and {Miss, 2, 0, 0, 0, 0, 0, 3, 1, 0, 1, 1, 3, 0, 0, 3, 1}. The staff inter-onset times are {8}, {4, 8, 0} and {ø}.

As can be observed in Figure 6.17 the information from the third outbreak is the most different from the simulated results. The waiting times between the onset of assigned ILI cases (using the second diagnostic) tend to be longer than the simulated waiting times for the resident population.

For the second outbreak the clinical diagnosis for all cases are missing and almost all of them were classified as ILI using the information about their symptoms. In the first and second outbreaks the cases had missing clinical diagnosis or they were URTI (only one case was ILI). Several of the cases in these two outbreaks (but not most of them) were re-classified as ILI.

Since most cases for the second outbreak were re-classified as ILI, the inter-event times describe the waiting times of a respiratory infection.

The longer waiting times, together with the few cases re-classified as ILI in the third outbreak, can indicate that the second diagnostic may classifies as ILI some respiratory cases that actually are not ILI in outbreaks that are not flu-related, or the influenza outbreak occurs simultaneously with another respiratory outbreak and there are some cases that are not classified as ILI when they actually are.

6.2.5 Predicted influenza cases

In order to study the spread of influenza it is crucial to distinguish the influenza and the non-influenza cases in an outbreak. In this section we predict the probability that an individual has influenza based on a logistic regression model with variables corresponding to the clinical diagnosis and the main symptoms presented.

The model utilized to obtain the predicted probabilities of influenza is a logistic regression based on the categorical variables listed in Table 6.16. The variables consist of respiratory symptoms, age group and clinical diagnosis.

It is important to emphasize that for all the variables the missing value "not registered" is considered to be a category value, and that the logistic regression model was selected using the clinically classified data (with a total of 587 registers from January 1st, 2003 to August 30th, 2006) from all sub-models that include the variables listed in Table 6.16.

Two models were obtained with backward elimination and forward selection based on the Akaike information criterion (AIC) and bayesian information criterion (BIC). The fist model has AIC and BIC values equal to 457.43 and 514.306, respectively, but it has divergent estimates for three values in the for final variables.

Since in most of the cases the symptoms were all registered or all missing together, the inclusion of several variables related with the symptoms lead to the collinearity and infinite estimates for some of the model parameters.

The final logistic regression model based on the BIC has AIC=463.28, BIC=507.027 and parameters shown in Table 6.17. We can observe that the standard deviations for the value "not reg" corresponding to the variables "abnormal temperature" and "nasal congestion"

Variables	Values
headache	yes, no, not registered
general aches (myalgia)	
cough	
sore throat	
chest congestion	
nasal congestion	
lethargic (malaise)	
laryngitis	
runny nose	
wheezing	
chills	
head congestion	
sinus	
sneezing	
abnormal temperature	
prior influenza vaccine	
group of age	[0,6), [6,18), [18,65), [65,), not reg
clinical diagnosis	ILI, URTI, not case, not reg

Table 6.16: *Principal respiratory symptoms, age and medical assessment.*

are very large; however the information contained in the other categories for the variables is significant, resulting in a c index equal to 89.72%, with number of concordant pairs of observations equal to 1.540260×10^5, and Somer's Dxy rank correlation 0.7944.

According to Call et al. (2005) the symptoms: fever, cough and nasal congestion are the most important indicators for patients in studies that evaluate the entire age spectrum. However among studies of patients limited to those aged 60 years and older the most relevant indicators were fever, malaise and chill.

As we can observe, in the model results presented in Table 6.17, the category "nasal congestion" decreases the probability for having influenza. This can be due to the fact that the majority of the population included in the data is over 60 years old.

The box plot of the predicted probabilities by the actual influenza status (according to

Coefficient		Estimate	Std. Error
intercept		-2.1878	0.48
abnormal temperature (AT)	yes	0.5408	0.37
	not reg	-16.9189	558.12
nasal congestion (NC)	yes	-0.6261	0.32
	not reg	19.0549	558.12
clinical diagnosis (CD)	not reg	-0.0396	0.32
	not case	-0.9179	1.27
	URTI	-1.6871	0.36
prior vaccine (PV)	no	1.6549	0.49
	yes	1.0335	0.34

Table 6.17: *Estimated parameters for the logistic regression.*

laboratory results) for the complete data is presented in Figure 6.18. As pointed out by the c index, the predicted probabilities tend to be larger for the cases that were influenza positives.

To graphically analyze the fit of the obtained model we present the histograms for the residuals (Figure 6.19). If the model is appropriate to describe the data, it is expected that the deviance residuals are approximately Gaussian distributed.

Based on the model the predicted probabilities for influenza can be obtained for all the cases in the database and by selecting a threshold the cases can be classified as influenza or not influenza. The Table 6.18 shows the misclassification tables for the probability thresholds of 40% and 50%.

Threshold	40%		50%	
Lab result / Predicted result	negative	positive	negative	positive
negative	269	42	276	35
positive	69	207	79	197

Table 6.18: *Misclassification tables.*

The total number of predicted influenza cases (out of 2778) considering the thresholds

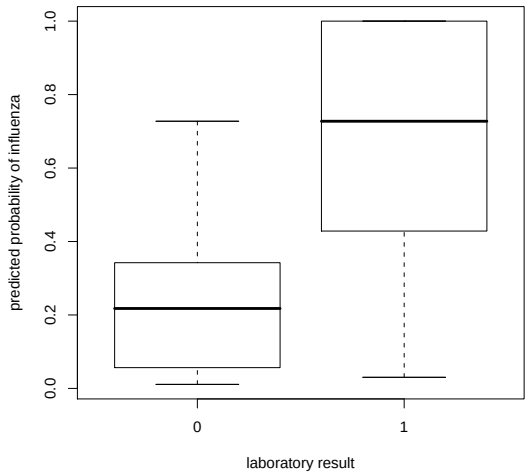


Figure 6.18: *Predicted influenza probabilities. Model with AT, NC,CD and PV.*

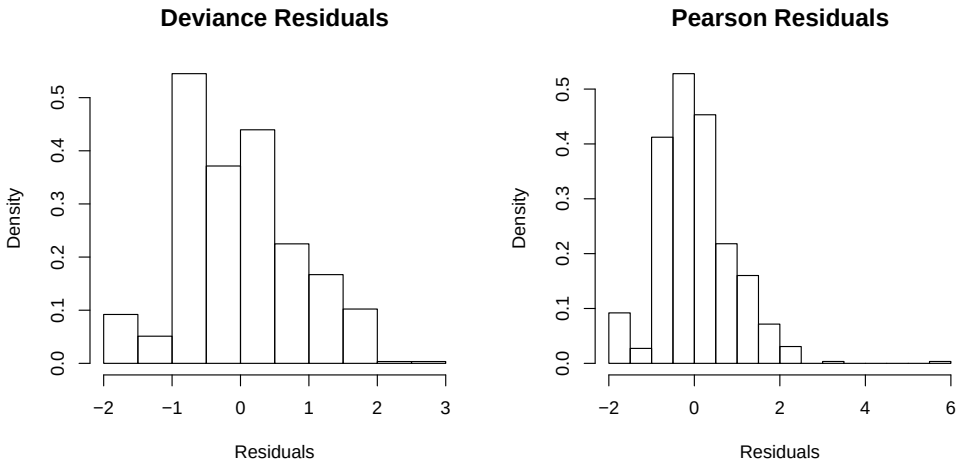


Figure 6.19: *Residuals. Model with AT, NC, CD and PV.*

of 40% and 50% are 596 and 459 respectively.

It is important to notice that most of the respiratory cases registered in the database correspond to individuals that are 60 years and older, and according to Call et al. (2005) the proportion of true positives (*sensitivity*) and the proportion of true negatives (*specificity*) for nasal congestion in patients over 60 years old is just 47% and 50%.

Also the likelihood associated to nasal congestion for individuals that are over 60 years are all close to one (LR+=.95(0.57-1.6), LR-=1.0 (0.67-1.7) and DOR=0.9 (0.3-2.4), Call et al. 2005). Since all these parameters indicate that nasal congestion does not contain any information to "rule-in" or "rule-out" influenza in this population, next we evaluate the model that excludes this variable.

Predicted probabilities based on model without nasal congestion

Since the variable "clinical diagnosis" includes the individual information gathered by the facility and Health Unit about new respiratory symptoms additional to existing ones, we consider it is important to include this variable whenever it is present.

Due to the fact that the clinical diagnosis is not always present it is also important to include some symptoms, but not so many that the collinearity in their category "not reg" makes it impossible to find finite parameter estimates.

As suggested in the last subsection we exclude the variable "nasal congestion". The estimated parameters for this sub model are presented in Table 6.19.

Coefficient		Estimate	Std. Error
intercept		-2.6859	0.3933
abnormal temperature (AT)	yes	0.8674	0.3007
	not reg	2.6296	0.3603
clinical diagnosis (CD)	not reg	1.0636	0.2552
	Not case	-0.8550	1.2781
	URTI	-1.7840	0.3625
prior vaccine (PV)	no	2.9898	0.3622
	yes	1.0227	0.2775

Table 6.19: *Estimated parameters for the logistic regression.*

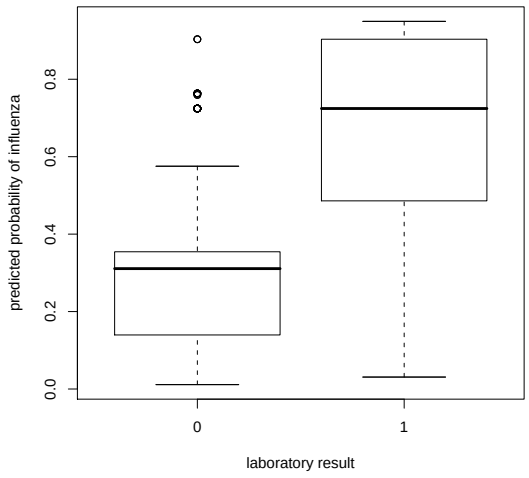


Figure 6.20: *Predicted influenza probabilities. Model with AT, CD and PV.*

It is remarkable that the estimates' standard errors for this model were dramatically reduced and in spite of AIC=600.06 and BIC=635.064, the c index was hardly affected, decreasing its value to 82.1% (Dxy =0.6420 and number of concordant pairs of observations = 1.409410×10^5).

The importance of the remaining variables can be explained from the facts that, As pointed out in the previous section, fever can be absent in the influenza cases for people over 60 years old (sensitivity 34%) but according to Call et al. (2005) fever has a high specificity (91%) among this population.

On the other hand, in the last section we noticed that the clinical diagnosis has a high sensitivity (88.7%) and low specificity (43.33%).

The predicted probability of this sub-model presents a slight improvement (Figure 6.20) but the deviance residuals seem to be less symmetrical (Figure 6.21).

Under this model both misclassification tables, for the training data and the thresholds of 40% and 50% (Table 6.20), show that the sub-model tends to classify more individuals as flu positives, increasing the false positives, and of course decreases the false negatives.

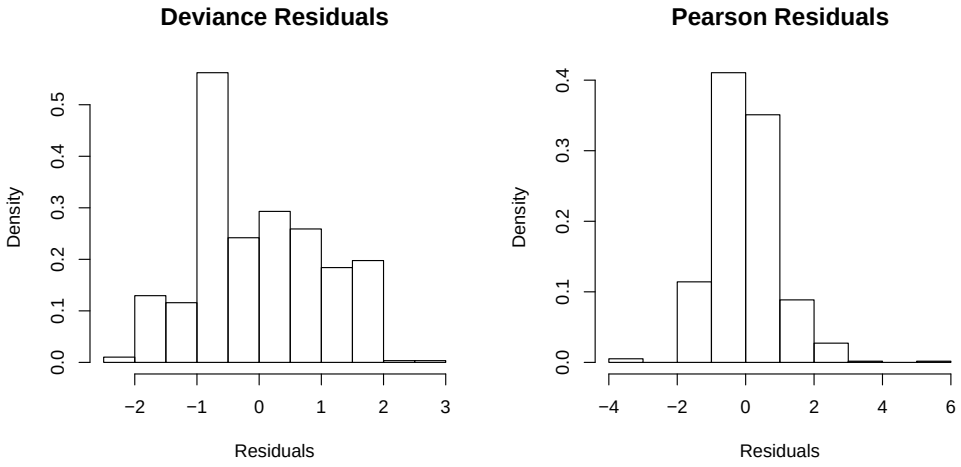


Figure 6.21: *Residuals. Model with AT, CD and PV.*

Threshold	40%		50%	
Lab result / Predicted result	negative	positive	negative	positive
negative	236	75	241	70
positive	63	213	70	206

Table 6.20: *Misclassification tables.*

The total number of predicted influenza cases in the database, considering the thresholds of 40% and 50%, are 734 and 685 respectively.

Considering the same facility as the one use for Figure 6.17, Figure 6.22 shows the inter-onset times for the cases predicted as influenza (with a level of 40%). Since no staff case was predicted to be positive for influenza and the times between the first and second cases are at least of 4 days then we assume that the first case was a resident member and then the 1,000 outbreak simulations are obtained having a resident as patient zero.

The inter-event times in Figure 6.22 are {5, 0} and {4, 1} days.

When the symptoms are not reported, the prediction of flu cases heavily depends on the clinical diagnosis, but when some data of the symptoms is available, the predicted

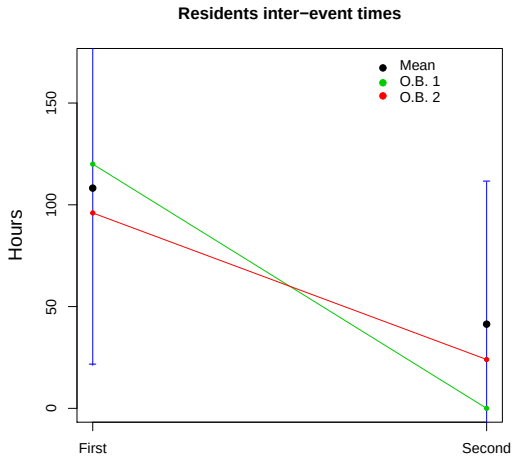


Figure 6.22: *Residents inter-event times from simulated outbreaks.*

influenza cases are reduced with respect to the ILI clinical cases and even more reduced with respect to the ILI cases obtained using the second diagnostic from last section.

However Figure 6.22 does not show any evidence against the fitted model, there exists an important difference between the number of cases obtained based on the prediction of flu cases and the number of ILI cases obtained, either with clinical diagnosis or the diagnostic evaluations presented in the last section.

As mentioned in Section 6.2.3, several studies have reported diverse estimates for the basic reproductive number. The observed variation for the model's point estimates can have two main sources: the difference in the population's connectivity patterns from which the data is obtained, and the sensitivity to the flu cases definition (test positives, cases clinically diagnosed as ILI, etc.).

Chapter 7

Future Research

In this chapter we describe some future analysis that can be done and the research directions that can be taken to improve and expand the epidemic models we have studied.

7.1 Analysis based on the discrete time model from Chapter 4

Given the p.g.f. recursive expression for the number of infected at time t such as (4.1.2) it is feasible to analytically study the evolution for outbreaks with variable transmissibility probabilities τ_t.

The change of the probability of transmission between individuals could be due to the implementation of control measures such as droplet precautions, hand washing, use of antiviral, etc.

From the recursive form of the p.g.f.'s it is also possible to obtain the joint p.g.f. at any two time points and the use of the conditional distributions (or moments) to do the parameter inference for the probabilities of transmission.

7.2 Further use of time-continuous simulation program

Since the simulation program reports the sizes of the susceptible, infected, infective and removed population at every time there is a new event, the program can be used not only to study the final outbreak size or to obtain the estimates for the model's parameters, but also to estimate the transmission incidence over time or other parameters of interest such as the outbreak duration distribution.

Using some few modifications to the simulation program, the effect of some variation of the control measures can be studied. For example, we could simulate conditions where:

- the ring vaccination or isolation strategies are started s units of time after the first case (proposed during a talk by Dr. Yan)
- ring vaccination is implemented not only for the first but also for the second neighbors around infective cases
- not all the infective cases in the ring vaccination strategy are identified
- isolation is not 100% effective.

The implementation of control measures in the simulations do not only allows to evaluate their effect on the outbreaks, but they potentially increase the amount of data that can be use in the inference of the agent's parameters. When the simulated control strategies depend on few parameters (or their parameters are approximated by some other procedures), they can be use to mimic the real control strategies, allowing the inference to use data originated in controlled outbreaks.

The simulation program can run outbreaks with individual values for transmissibility and susceptibility to the agent. In the measles and influenza data we have use different values of susceptibility for two different populations. In the first case the two populations (susceptibles at most one year old, and one year and older) are connected with a simple

graph and in the second case the two populations (residents and staff) have a different connectivity pattern described by a hierarchical network.

The use of different transmissibility and susceptibility constant rates can naturally be generalized to transmissibility and susceptibility random variables with individual probability distributions. The individual distribution can be defined obeying not only individual characteristics but environmental heterogeneities. The environmental effect on the graph can be modeled superimposing a map on the network structure.

It is natural to think that for many infectious agents, the individuals in a local network will be more likely to be under similar environmental situations. However, the map can describe a more global dissimilarity measure among local networks.

One modification that can also be done to the program is allowing the external or environmental variables not only modifying the agent's transmissibility between individuals, but also its latent and infectious distribution in the host.

7.3 Modeling the network structure

As it has been pointed out, the network structure allows us to introduce a non-homogeneous contact structure into the epidemic model. For some agents the contact associated and its the network structure can be naturally defined by the population setting such as in a hospital, school, work units, etc. However, for more complex contact definitions and/or more general settings, the problem of selecting the network structure and its degree distributions is added to the analysis of the outbreaks evolution or the statistical inference problem.

Although the problem of defining a social structure or network of relations has been studied in the area of social network analysis (See Wasserman and Faust (1994); Scott (2000); de Nooy et al. (2005)) more formal statistical techniques can be used to plan and interpret the data collected by sampling some individuals in the population.

It is interesting to point out that in the field of social network analysis the diffusion of ideas in a population has also attracted the attention of several researchers. The "diffusion"

phenomenon is equivalent to a SI epidemic model.

Once the network structure is selected as simple, bipartite or hierarchical (with two or more local networks) and the degree(s) family distribution(s) is chosen, then the parameters of the network could be added to the estimation process that uses the Monte Carlo-likelihood function. Of course, adding more parameters, will very likely lead to the problem of non-unique model estimates.

The suggestion of simultaneously doing the estimation of the network and transmission process is more viable if the network is a simple graph with distribution defined by a single parameter and the data contains substantial information of the infectious dynamic over time.

Including a sensitivity analysis of the outbreaks based on the most likely network structures may be desirable. The results can indicate if further data (related to the network structure or outbreaks) have to be collected.

7.4 Incorporating the demographic dynamics

In this work we consider that the disease evolves very rapidly and the demographical changes are not important. However, this is not true for several infectious diseases such as tuberculosis and HIV.

The connectivity changes that we consider occurring in the graph are due to the change of the nodes to the "removed" status, and they do not include the migration of individuals by births, territorial migration, deaths for reasons other that being infected, etc.

The epidemic models for closed populations are the most simple, and those that assume the mass action can include the immigration and emigration dynamics, without having to modify the standard analysis. In the compartmental deterministic model the inclusion of migration is usually translated into more ordinary differential equations and parameters that are analyzed using the Euler algorithm or other numerical methods.

As studied by Barabási and Albert (1999), a network with power law distribution (with

parameter 2.9 ± 0.1) can be regarded as a growing graph with preferential attachment. However, the nodes' status of the growing network do not change over time.

As the epidemic spreads, it preferentially hits the highly connected nodes, then the incoming individuals would be included in a network that has a modified degree distribution. The effect of changing the network is magnified when the illness spreads as fast or faster than the arrival of new nodes.

As described in Appendix A the simulation program incorporates the immigration of susceptible, infected and infective individuals. The algorithm assigns a degree (with the same degree distribution a that of the host local network) and uniformly connects it with the local susceptible, infected and infective nodes.

However the simulation program's algorithm does not obey any observed or studied social behavior and it can be looked upon as an exercise to include nodes immigration. It is one of the first attempts to incorporate the immigration into populations with a network structure and during an outbreak.

A different evolution in the network that does not involve the migration of individuals, and that has attracted the attention of researchers in social networks is the connectivity change of close populations. Wasserman and Faust (1994, section 2.4.3) describe several studies that involve the recollection of longitudinal social network data.

7.5 Modeling missing and misclassified data

The problem of missing data is very common in the epidemic data, and as in the case of the influenza this problem also combines with the misclassification of cases that were clinically diagnosed and not tested, or tested but yielded false positives or false negatives.

For some infectious diseases missing data and misclassification can lead to severely biased estimates since they can be correlated with some population characteristics such as age and sex.

It is also important to distinguish the different sources of missing data. Some of them

can have particular importance such as those that arise because of individuals that are infected or infectives and do not become symptomatic (silent cases).

As in Sections 6.2.4 and 6.2.5 several case definitions can be explored. As observed, the inference can not only be affected by the number of cases that disagree in each classification, but mostly by their distribution in time within each outbreak.

7.6 Further applications of distribution theory

In this work we obtained the point estimates for measles and influenza. However this information should be complemented with the estimates properties or asymptotic properties such as bias, consistency or efficiency.

Appendices

Appendix A: Description of simulation program

In this chapter we describe the main algorithms utilized to simulate outbreaks in populations with a predetermined contact structure defined by their degree distribution, like those in Chapter 3 and Section 4.1.

The simulation program developed is written in the R language (R Development Core Team, 2007) and it is fully presented in the Appendix B.

An earlier version of the code in Appendix B was available as the R package "InfNet" V.0.1, and the updated version V.1.0 will be uploaded to the R site in the next weeks after testing it in the Unix and Mac platforms.

Section A.1 describes the main subroutine or function to generate simple networks. The function local.network generates a simple undirected graph of order n with a given degree (fixed) or with one of several degree distributions.

In Section A.2 the principal characteristics of the function called epidemic.sim are described. The algorithm implemented in epidemic.sim runs the evolution in time of the infectious process given a contact network, and allows us to simulate epidemic processes in which the infectious agent can have latent and infectious periods that follow a distribution other than exponential.

Section A.3 presents a brief description of the complete simulation algorithm.

A.1 Network construction

Constructing a simple graph from a given degree sequence is a classical problem in graph theory and computer science. However, generating a graph that meets the given degree, uniformly at random from all the possible graphs with that degree sequence, is still an open problem (Aiello et al., 2001; Gkantsidis et al., 2003; Mihail and Vishnoi, 2002).

A.1.1 Local network

The function local.network builds simple random graphs with order n and degrees that can follow any of the distributions shown in Table A-1.

Tail	Name		Probability Density Function	
Light				
	Poisson	$=$	$\lambda^x \frac{\exp(-\lambda)}{x!}$,	$x = 0, 1, 2, \ldots$
	Zero truncated Poisson	$=$	$\lambda^x \frac{\exp(-\lambda)}{x!(1-\exp(-\lambda))}$,	$x = 1, 2, \ldots$
Exponential				
	Geometric	$=$	$\left(1 - \frac{1}{p+1}\right)^x \frac{1}{p+1}$,	$x = 0, 1, 2, \ldots$
	Zero truncated Geometric	$=$	$\left(1 - \frac{1}{p+1}\right)^{x-1} \frac{1}{p+1}$,	$x = 1, 2, \ldots$
	Negative Binomial	$=$	$\binom{x+r-1}{x} p^r q^x$,	$x = 0, 1, 2, \ldots$
Power Law				
	Polylogarithmic (Gutenberg-Richter Law)	$\propto$	$x^{-\alpha} \exp(-x/\beta)$,	$x = 1, 2, \ldots$
	Logarithmic	$=$	$-\frac{\exp(-x/p)}{x \ln(1-\exp(-1/p))}$,	$x = 1, 2, \ldots$
	Power Law (Zeta distribution)	$\propto$	$x^{-\alpha}$,	$x = 1, 2, \ldots$

Table A-1: *Degree probability distributions.*

The routine local.networks obtains a sample of size n from the specified distribution to be the network's degree sequence and constructs the graph trying to meet this sequence.

This function also includes three deterministic degree sequences. The first one, called "full", assigns a degree of $m-1$ to all the nodes in a graph with m nodes. The second one, invoked as "none", assigns degree 0 to each node in the network, and the third option, called "fixed" allows the user to directly state the degree distribution to be use.

In order to construct the network with the designated degree sequence, the function local.network considers the following four algorithms:

1. **Havel-Hakimi algorithm (HH)** The Havel-Hakimi Theorem (Havel, 1955; Hakimi, 1962) is a existence-constructive result for graphs given a degree sequence. It involves a recursive algorithm that decreasingly orders the degrees that have not been used yet to build the graph (*residual degrees*) $(\mathrm{d}_{(1)} \geq \mathrm{d}_{(2)} \geq \cdots \geq \mathrm{d}_{(n)})$ in each step and selects the two nodes with highest residual degree to set an edge between them if this does not already exist. After this step the residual degrees are updated, or another pair of nodes is selected.

   The theorem states that if the algorithm fails to obtain a graph with the specified degree set, then no graph exists with the original degree sequence.

   The Havel-Hakimi result is an important existence-constructive theorem, but the topologies of the obtained graphs using this algorithm have very *dense cores*, failing in creating uniformly random graphs among those which meet the degree specifications.

   If the algorithm is changed to connect the two nodes, one with the highest and the other with the lowest residual degree, then the topologies of the obtained graphs tend to have very *sparse cores* (Gkantsidis et al., 2003).

2. **Modified Havel-Hakimi algorithm (MHH)** In order to achieve more uniformly random graphs, the condition of meeting a given degree sequence must be relaxed to obtain graphs whose nodes have degrees close to those specified.

This modification to the Havel-Hakimi algorithm consists in assigning an edge between the node with highest residual degree and a randomly selected node with positive residual degree.

3. **Weighted selection algorithm (MR)** This algorithm was suggested by Molloy and Reed in 1995. In each step the edge (i, j) is selected with probability proportional to the product of the residual degrees $\mathrm{d}_{(i)}$ and $\mathrm{d}_{(j)}$.

4. **Random selection algorithm** In each step this algorithm connects any two nodes with positive residual degree that are not yet connected. As we will see, this algorithm is the one that allocates the degree sequence worst in the resulting graph, but is given to be compared with the previous three algorithms.

A.1.2 Evaluation of the algorithms

With the aim to evaluate the MHH and the MR algorithms, the HH and the random algorithms were implemented. It is desirable to use an algorithm that meets the degree distribution as closely as possible (as in HH) while at the same time approximately generates uniformly random graphs.

In order to compare the performance of the algorithms, 1,000 graphs were obtained with each of the four algorithms and

1. two different sizes: 50 and 500 nodes;

2. two different distributions with the same expected value: Poisson(2) and Power law(2.46).

For each variable combination, three plots were obtained. The first evaluates the allocation of the degree distribution, and the second and third plots display measures of the randomness of the obtained networks.

The first plot corresponds to the histogram for the proportion of not allocated degrees (PNAD). That is, the total of degrees that were not used in the final created graph over the total of degrees that should have been used in each random graph.

The second and third plots are the frequencies for the degree distribution of one randomly chosen node that is connected to the node with largest (DDCL) and smallest (DDCS) (but positive) degree, respectively.

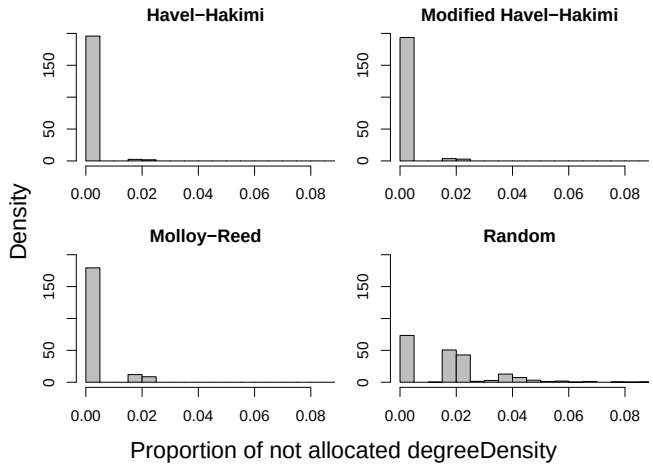
These plots are respectively labeled as **(a)**, **(b)** and **(c)** in Figures A-1, A-2, A-3 and A-4; and **(b)** and **(c)**'s graphs show the theoretical degree distribution of a node that is selected by randomly choosing one of its edges (black dotted line).

In the case of the graphs with power law degree distribution, the DDCL and DDCS are presented as log-log plots.

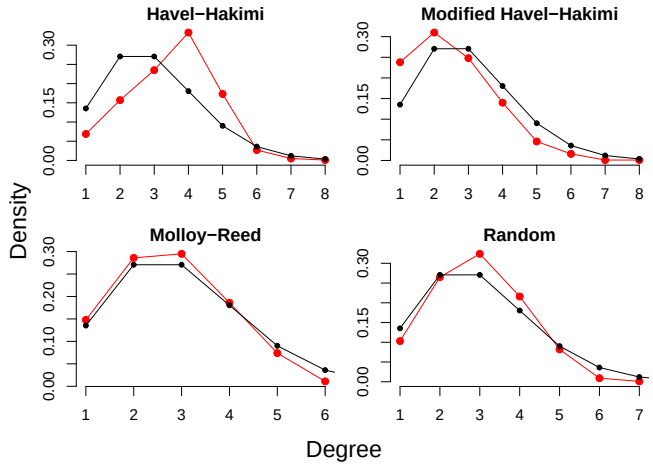
It can be observed that in any case the HH and the MHH algorithm are the most efficient in allocating the degree (histograms **(a)** in Figures A-1, A-2, A-3 and A-4) whilst the random algorithm is the least efficient by far.

As observed by previous works (Gkantsidis et al., 2003) the HH algorithm creates clusters of nodes with higher degree. This characteristic is shown by plots labeled as **(b)** and **(c)**. Comparing with the degree distribution of a node selected by choosing one of its edges (black line) the observed degree in graphs created with the HH algorithm gives evidence that if we select a node connected to a node with high degree, this will have higher probability of having a large degree compared to the degree of a node that is selected by randomly choosing one of its edges.

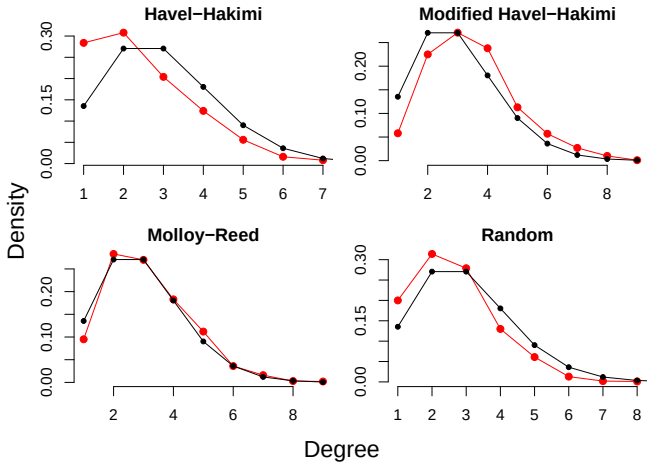
After obtaining the network under any algorithm, some authors suggest to apply a rewiring algorithm to randomize the network while conserving the degree of the nodes. This algorithm changes the contact pattern with probability p, in each step randomly choosing a couple of edges that have not any common endpoints (i_1, i_2) and (i_3, i_4). If the edges (i_1, i_3) and (i_2, i_4) do not already exist in the graph, then the first couple of edges is removed and the second couple of edges included with probability p. The new network has the same degree sequence as the original and after many steps it converges to a graph



(a) Distribution of proportion of not allocated degrees

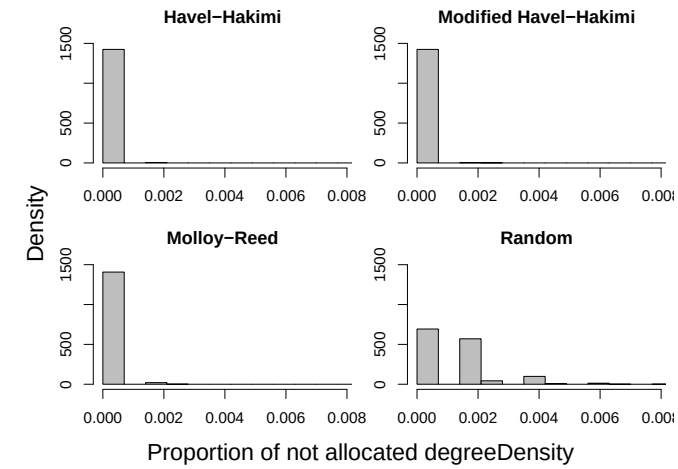


(b) Degree distribution of a node randomly chosen from the neighbors to the degree with highest degree

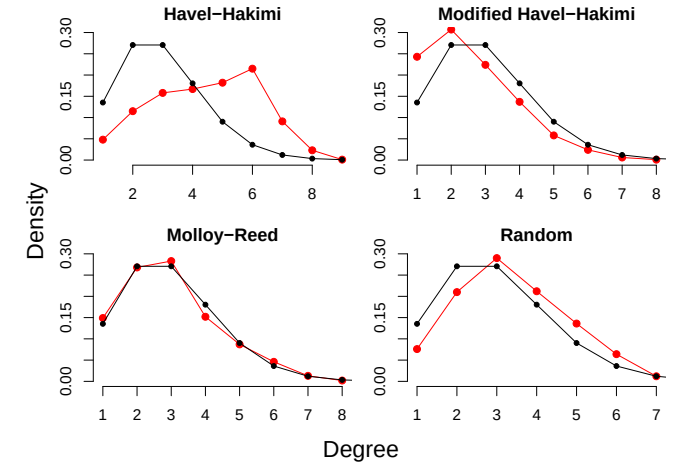


(c) Degree distribution of a node randomly chosen from the neighbors to the degree with smallest positive degree

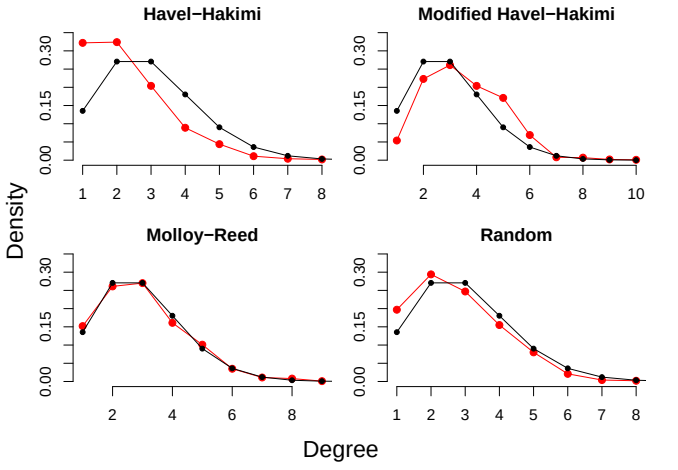
Figure A-1: *PNAD, DDCL and DDCS for 50-nodes-graphs with degree distribution Poisson(2). Theoretical values for DDCL and DDCS are presented in black.*



(a) Distribution of proportion of not allocated degrees

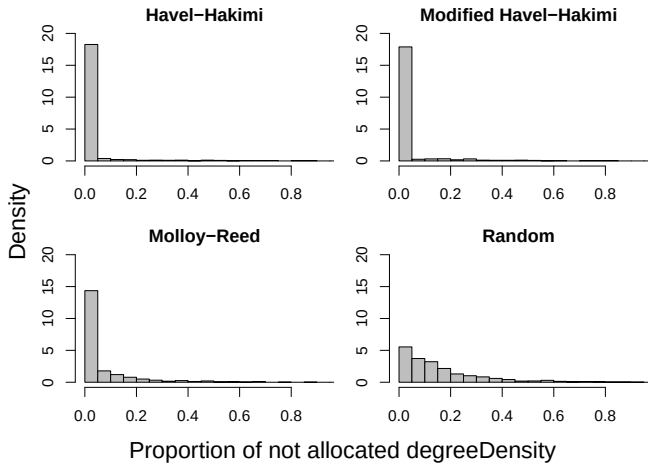


(b) Degree distribution of a node randomly chosen from the neighbors to the degree with highest degree

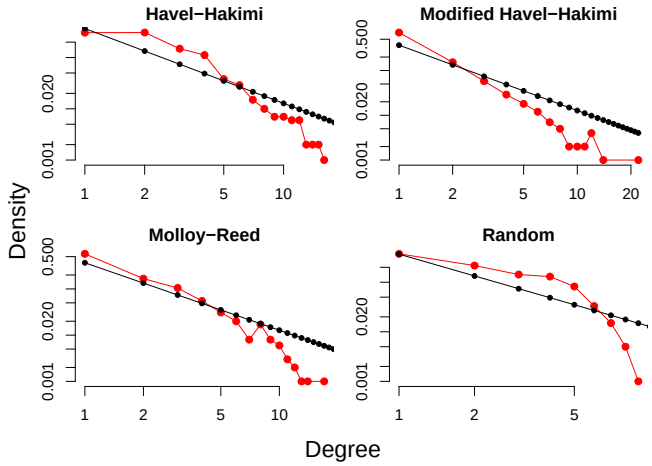


(c) Degree distribution of a node randomly chosen from the neighbors to the degree with smallest positive degree

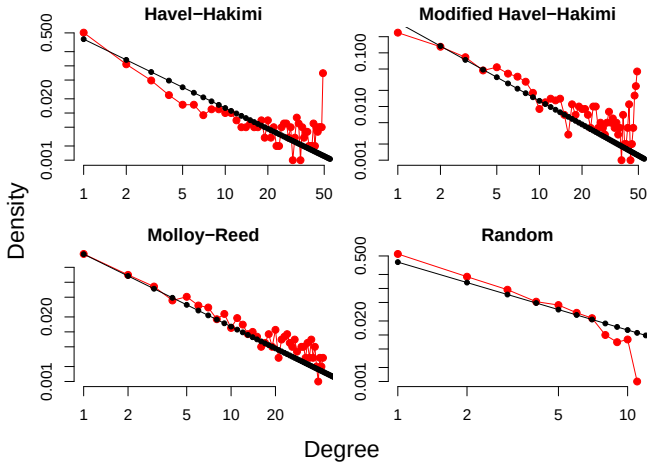
Figure A-2: *PNAD, DDCL and DDCS for 500-nodes-graphs with degree distribution Poisson(2). Theoretical values for DDCL and DDCS are presented in black.*



(a) Distribution of proportion of not allocated degrees

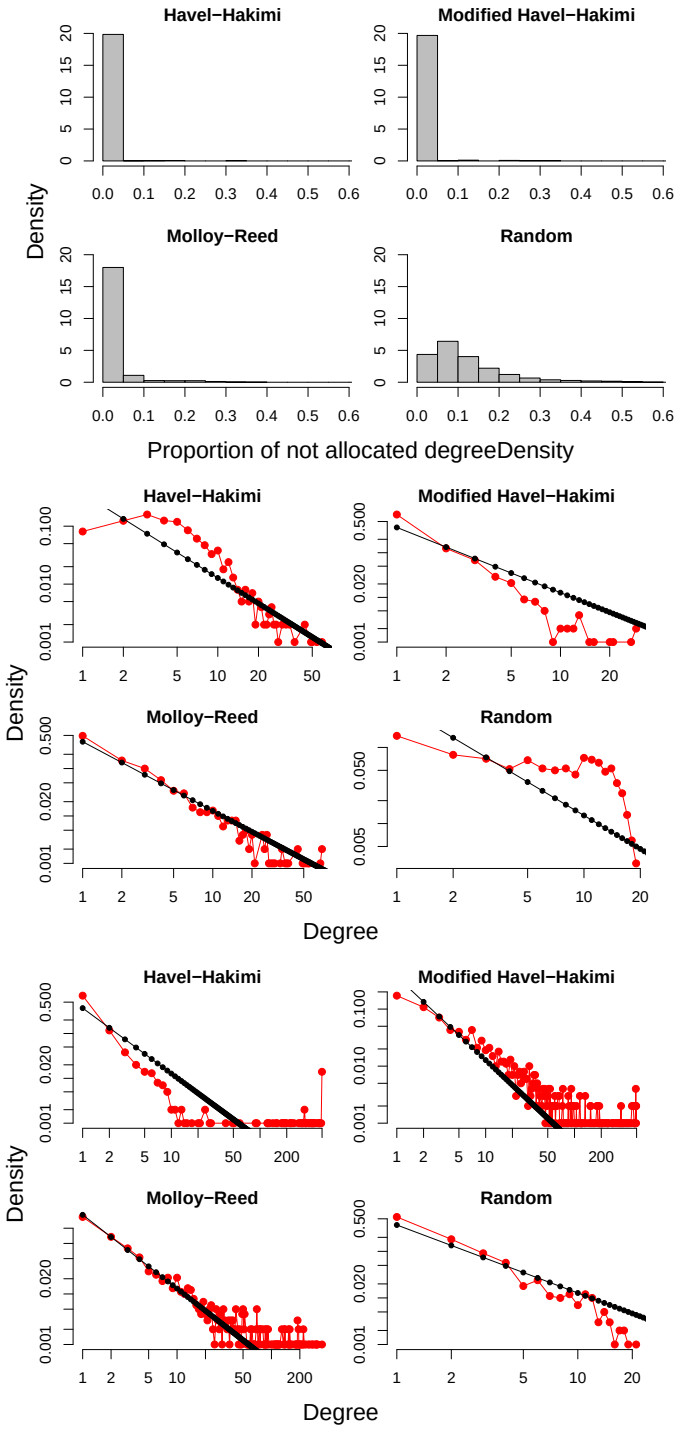


(b) Degree distribution of a node randomly chosen from the neighbors to the degree with highest degree



(c) Degree distribution of a node randomly chosen from the neighbors to the degree with smallest positive degree

Figure A-3: *PNAD, DDCL and DDCS for 50-nodes-graphs with degree distribution power law(2.46). Theoretical values for DDCL and DDCS are presented in black.*



(a) Distribution of proportion of not allocated degrees

(b) Degree distribution of a node randomly chosen from the neighbors to the degree with highest degree

(c) Degree distribution of a node randomly chosen from the neighbors to the degree with smallest positive degree

Figure A-4: *PNAD, DDCL and DDCS for 500-nodes-graphs with degree distribution power law(2.46). Theoretical values for DDCL and DDCS are presented in black.*

that is random among those with the initial degree sequence (Kannan et al., 1999; Maslov and Sneppen, 2002; Gkantsidis et al., 2003).

Since the HH algorithm is the one that better allocates any degree sequence, it could be recommended to obtain the initial graph with this procedure and then rewire it as described. However the less random the initial graph is, the more steps (and resources) would be needed in the second process.

The random algorithm has an important lack of efficiency in allocating the degree sequences as observed in histograms **(a)**. This can be corrected by a similar algorithm to the rewiring process described above. In each step an edge (i_1, i_2) and a couple of nodes with positive residual degrees i_3, i_4 are randomly chosen. If the edges (i_1, i_3) and (i_2, i_4) do not exist in the graph, then the edge (i_1, i_2) is removed and the formers are added to the network.

Although the random graph obtained using the random algorithm can be improved by the described process, this is not only time demanding but for more complex networks, as bipartite and hierarchical, it becomes more complicated to implement.

It is important to note that in spite of the fact that the random algorithm performs the worst among above algorithms described, some authors (Newman et al., 2001) have directly used it (without any rewiring) to generate and study the properties of random graphs.

The MHH tends to allocate the degree better than the MR algorithm, especially for heavy tailed degree distributions, but the MR algorithm creates graphs with DDCL and DDCS that almost perfectly agrees with the theoretical distribution $kP_k/E(K)$. The only exception is the DDCL for the 50-nodes-graphs that have power law degree distribution.

Hence the MR looks to create fewer clusters than the MHH, however the MR is more time-consuming to run for large networks. Since the MHH generates acceptable random graphs and it is more machine-efficient than the MR algorithm, we use it in preference to the MR to create networks with a large number of nodes.

No rewiring procedure is implemented on the obtained networks.

A.1.3 Global network

Based on the local network generation function, a two level (or hierarchical) network can be built. In this random graph an individual i is associated to n local networks.

If w is the local network node i belongs to, then i has n degrees $\mathrm{d}_w(i)$, $\mathrm{d}_{(w,j_1)}(i)$, $\mathrm{d}_{(w,j_2)}(i), \ldots, \mathrm{d}_{(w,j_{n-1})}(i)$, where $w \neq j_k$, for $k \in \{1, \ldots, n-1\}$.

The degree $\mathrm{d}_w(i)$ is sampled from the degree distribution of the local network w, and the degrees $\mathrm{d}_{(w,j_k)}(i)$ are sampled from the distributions that connect the local networks w and j_k.

The degree $\mathrm{d}_w(i)$ is used to build the n sub networks with the function local.network and the degrees $\{\mathrm{d}_{(w,j_k)}(i)\}_{i \in w} \cup \{\mathrm{d}_{(j_k,w)}(l)\}_{l \in j_k}$ are incorporated in a function similar to local.network that connects the two sub networks w and j_k.

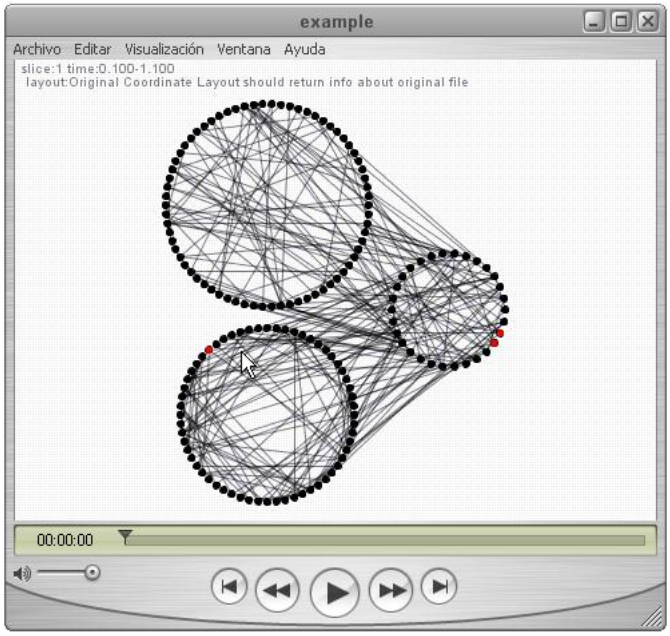
Part (a) of Figure A-5 shows a network built with 3 local networks of 30, 60 and 70 nodes (clockwise). For this example local networks 2 and 3 are not connected and the rest of the utilized degree distributions are given in part (b) of the same figure.

The parameter of the degree distribution among the local networks pairwise fulfills the condition (3.2.22). When this is not the case, only the degree distribution that generates a smaller total of edges can be considered as simulated; the other degree distribution is upper truncated to build the graph.

Using this algorithm we can construct the networks with the structures studied in Chapter 3 and 4.2:

- simple random graphs, setting $n = 1$,
- bipartite networks, selecting n=2, 2 degree distribution between the sub networks and "none" for the degrees within them, and
- hierarchical networks with 2 populations, setting $n = 2$.

(a)



.

(b)

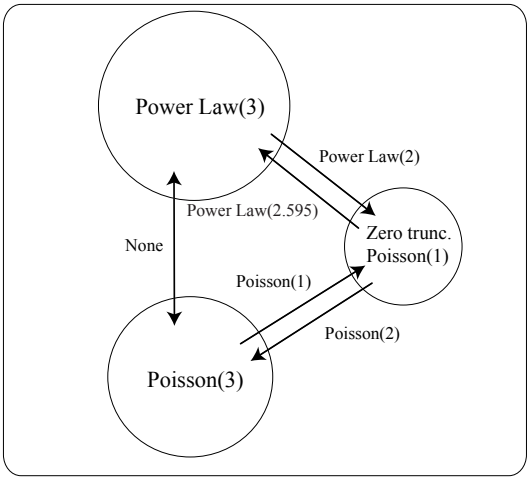


Figure A-5: *Example of a hierarchical network. Visualized with SoNIA (Bender-deMoll and McFarland, 2003).*

A.2 Epidemic evolution

The simulation program runs an SIR or SEIR epidemic process in continuous time, considering independent transmission processes among individuals, latent and infectious periods, and immigration rates associated to each local network. These are further described next.

Transmission rate The time to transmission between any two pair of vertices that are connected (one infective and the other one susceptible) is exponentially distributed with mean equal to the reciprocal of the transmission rate. The rate can vary from each pair of individuals and it is computed as the product of two individual-specific rates that enter the program as vectors. The first vector contains the individual transmissibility rates and the second contains the individual susceptibility rates.

Latent and infectious periods The latent (SEIR model) and infectious periods are independent and with a distribution which may be exponential, normal or log normal.

Immigration rates The times between immigration of susceptibles and infectives into each local network can also be simulated as random variables exponentially distributed. The immigration rates for susceptible and infectives can be different, and they can also vary according to the local network the individuals are arriving to. The degree for a new individual is drawn from the degree distribution of its local network and the nodes at the end of each edge are randomly selected from the susceptibles and infectives in the local network at the moment of the arrival.

The program does not include the connection of the new individuals with nodes in other local networks other than the one they arrive to.

Since the distribution of the event times that are not the latent and infectious periods, are exponentially distributed, the simulated dynamic is Markovian if the latent (if SEIR model) and infectious periods also follow the exponential distribution.

A.3 Algorithm general description

1. [**global.network**] This function builds the graph and the user must specify:

   (a) number of local networks (n).

   (b) size of susceptible population in each network ($p_1, \ldots, p_n$).

   (c) distributions and parameters for the connections within the local networks.

   (d) distribution and parameter for the connections among local networks. These distributions are entered following the order for the pairs of local networks: $(1,2), (1,3), \ldots, (m, m-1), (m, m+1), \ldots, (n, n-1)$. If only one distribution is specified, this is copied for each pair of local networks.

   (e) algorithm to build the network (MHH or MR).

   Example: *network3<-global.network(n=3, p=c(30,70,140), distrib=c("pois", "poly.log", "poly.log"), param=list(3,c(1.7,2),c(1.7,3)), distrib.among="pois", param.among=2, method="MR")*

2. [**epidemic.sim**] Runs the simulation program on the resulting network. The required parameters are:

   (a) contact network.

   (b) SEIR=T (false if epidemic model is SIR).

   (c) number of initial infected individuals in each local network (if SEIR).

   (d) number of initial infectives in each local network.

   (e) the length of the observation time.

   (f) the rate for the transmissibility of the infectious agent. Can be different for each node in the graph.

   (g) the rate for the susceptibility to the agent. Can be different for each node in the graph.

(h) the distribution and parameters for the length of the latent period.

(i) the distribution and parameters for the length of the infectious period.

(j) the immigration rate of susceptibles (default value is NULL).

(k) the immigration rate of infectives (default value is NULL).

The main steps of that the routine follows are:

(a) Random (and without replacement) selection of the initial infected and infective individuals from each local network. If $n > 1$ and the number of initial infected or infectives is entered as a scalar, the same number is assigned to each local network.

(b) Simulation of the latent periods and assignation to each infected node.

(c) Simulation of the infectious periods and assignation to each infective individual.

(d) While the observation time is not exceeded:

   i. The time of infection is obtained for every pair of connected nodes in which one is infective and the other is susceptible. The time of infection is exponentially distributed with rate of transmission equal to the product of the transmissibility rate (of the infective individual) and the susceptibility rate (of the susceptible individual).

   ii. The time for immigration is simulated for each of the n local networks as realizations of exponential distributions with the indicated immigration rates.

   iii. If the smallest time of infection is less than all the immigration times and the infected and infectious periods left (*latent period + time of infection - present time*, and *infectious period + starting time of infectious period - present time*, respectively), then infection takes place for the pair of nodes corresponding to the minimum time of infection.

The status of the infected node is changed to "infected" ("infective", if SIR) and its length of the latent period (infectious period) is obtained from the indicated distribution.

iv. If the latent period left (the infectious period left) is less than the minimum time of infection, immigration events and infectious period left (latent period left), the node's status is changed to "infectious" ("removed") and a infectious period is obtained from the specified distribution and assigned to the node.

v. If the smallest of the times for the immigration of infected (infective) is smaller than any other event time then a new node is introduced to the associated network and the degree for the new node is obtained from the degree distribution indicated for the host network. The neighbors nodes are sampled without replacement and a latent period (infectious period) is assigned to the new node.

vi. The information of the network is recorded in a couple of variables (*.hist) and updated to run the cycle that started in i. The loop is repeated until the sum of the inter-event times is greater or equal than the observation time. If the immigration is not included in the simulation then the loop continues until the observation time or until new events (infections, initiation of infectious periods or removals) can take place.

Example: *hierarchical3<-epidemic.sim(network=network3, seir=T, ini.infected= c(1,1,0), ini.infective=0, obs.time=10000, BETA1=0.008, BETA2=1, distrib.lat= "norm", LAMBDA= list(c(62.63,4.33)), distrib.inf="norm",GAMMA=list(c(127,15)))*

3. [**sonia.format**] If the user wants to visualize the resulting simulation with the software SoNIA (Bender-deMoll and McFarland, 2003), then the routine sonia.format is called, indicating the name for the output file.

Example: *temp<-sonia.format(hierarchical3,F, filename.son="C:\\sonia.son")*

The function **epidemic.control.sim** is a variation of **epidemic.sim** that does not include the immigration nor the option to visualize with SoNIA. However, it is capable to simulate the outbreak evolution including the four control measures studied in Chapter 5.

The function **epidemic.reedfrost.sim** is also a variation of **epidemic.sim** that simulates the outbreak in discrete time.

Appendix B: Simulation program code

```
1
2  library(gsl)
3  #- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
4  li<-function(delta,b){   #b can be a vector
5  lim<-10000
6  res<-rep(0,length(b))
7  if(length(delta)!=1) stop("delta debe ser escalar")
8  else res[b==1]<-zeta(delta)
9  #else if(length(delta)==length(b))res[b==1]<-zeta(delta[b==1])
10 a<-which(b!=1)
11 if(length(a)) for (j in a) res[j]<- sum(b[j]^(1:lim)/(1:lim)^delta)
12 res}
13 #- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
14 #Suggested function to improve the sample function in R
15 resample <- function(x, size, rep=FALSE,prob=NULL){
16 if(length(x) <= 1) {
17     if(!missing(size) && size == 0) x[FALSE] else x
18 }
19 else sample(x, size, replace=rep,prob=prob)}
20 #- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
21 #I created NDIM,to use with either matrices or vectors
22 #takes a vector as column matrix if as.row=F and row matrix if as.row=T
23 NDIM<-function(a,as.row=FALSE){
24 if((is.vector(a) | is.list(a)) & as.row) n.dim<-c(1,length(a))
25 else n.dim<-c(NROW(a),NCOL(a))
26 n.dim}
27 #- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
28 #function that returns a vector of zeros and ones. The ones are in positions given
29  by q
30 #when q==numeric(0)  r is a vector of zeros
31 vector.one<-function(n,q){
32 r<-rep(0,n)
33 r[q]<-1
34 r}
35 #- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
36 #"matrix" can be a matrix or a vector representation of a matrix with
37 rows=rows.mat.
38 #If cols=TRUE This function returns: a) The (row,column) position of the
39 elements in "matrix" with value "esc", or
40 # b)transform the element "case"-th in the vector "matrix" into (row, column)
41 for "matrix" as a matrix
42 #if cols=FALSE only returns the rows
```

```
#When "matrix" is a vector, it can be considered column (as.row=F) or row
(as.row=T)
#cases is the position in the matrix of the elements equal to esc (by columns)
elements.matrix<-function(matrix,esc,cases=0,rows.mat=0,as.row=FALSE,cols=FALSE){
if(cases==0) cases<-which(matrix==esc)
if(rows.mat==0) rows.mat<-NDIM(matrix,as.row)[1]
rows<-(cases-1)%% rows.mat +1
result<-rows
if(cols){
     columns<-(cases-1)%/% rows.mat +1
     result<-list(rows=rows,cols=columns)
}
result}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function that deals with special cases for the function rexp (exponential
distributed random numbers)
newrexp<-function(n,p){
if(is.null(p)| n==0) a<-Inf
else{
    if(n!=length(p)) stop("dimensions do not match in newrexp")
    a<-rep(0,n)
    a[p>0]<-rexp(length(p[p>0]),p[p>0])
    a[p==0]<-Inf
}
a}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function that add new edges to an immigrant individual (id)
#distib.ln is the distribution of the local network
#param.ln is the parameter of the distribution
#act.indexesID.ln are the id active nodes in the local network (it is a vector or
numeric(0))
new.edges<-function(id, distrib.ln, param.ln, act.indexesID.ln){       #param.ln
is a list
new.degree<-local.network(1,distrib.ln,unlist(param.ln))$degree
active.p<-length(act.indexesID.ln)
true.degree<-min(new.degree,(active.p-1))
index.conn<-resample(act.indexesID.ln[act.indexesID.ln!=id],true.degree)  #!=
 to exclude itself  -samp without repl
if (true.degree>0) {
    new.conn<-cbind(id,index.conn)
    new.conn<-order.edges(new.conn,T)
}
else new.conn<-NULL     #important to be kept NULL for the use con concatenate
new.conn }              #the output is vector or matrix
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#this function is to order the network matrix (or vector). Output is always a
matrix
#order the network.edges so the rows are increasing
#ord.col=T if we want the program to order columns as well, so in the first
column appears the smaller ID's
order.edges<-function(edges,ord.col=TRUE){
if(is.matrix(edges)){
    if(ord.col){
        keep<-edges[,1]<edges[,2]
        edges[keep==0,]<-cbind(edges[keep==0,2],edges[keep==0,1])
    }
    edges<-edges[order(edges[,1],edges[,2]),]
}
else edges<-t(as.matrix(sort(edges)))
```

```
dimnames(edges)<-NULL
edges}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#I define this function us use it with "sapply"
compare.vectors<-function(y, x){
a<-sum(y==x)
a}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function to paste vec.mat (a matrix or vector) as new rows (or row) to a matrix
 edges.hist
#constant is a vector of constants (usually: time,arc.width,color)
#edges.hist is data.frame
concatenate<-function(edges.hist, vec.mat, constant=NULL){
if(sum(NDIM(vec.mat)>0)==2){                #vec.mat is not empty (is.null is not
good)
    if (is.vector(vec.mat)) vec.mat<-t(as.matrix(vec.mat))   #vector (scalars are
vectors) is now a matrix with one row
    vec.mat<-
cbind(vec.mat,matrix(rep(constant,nrow(vec.mat)),nrow(vec.mat),length(consta
nt),byrow=TRUE))
    vec.mat<-as.data.frame(vec.mat)
    names(vec.mat)<-names(edges.hist)           #to be able to use rbind next
    edges.hist<-rbind(edges.hist, vec.mat)
}
edges.hist}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function that creates a column of numbers telling to which group each element
(row) belongs to
#vec.mat is a matrix and groups is a vector (numeric or character) with length
=dim(vec.mat)[1] used to form the groups
grouping<-function(vec.mat, groups){
cbind(vec.mat,match(groups,groups))}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#I define this function to use within "sapply" in the function "coordinates"
angles<-function(X){
(0:(X-1))*(2*pi/X)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#I define this function us use it with "sapply" #Number of elements in column
col that are equal to the value x.
number.elements<-function(mat,col, x){
length(mat[mat[,col]==x,col])}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function so simulate random numbers from:
#Zero-truncated Poisson distribution P(N=k)=(lambda^k*exp^{-lambda
})/(k!*(1-e^{-lambda})), for k=1,2,3,..
rztpois <- function(n,lambda){
cdf <- (cumsum(dpois(0:50,lambda))-exp(-lambda))/(1-exp(-lambda))
cut(runif(n),unique(c(0,cdf)),labels=FALSE,right=FALSE)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function to generate random numbers from the Logarithmic distribution (poly-
logarithmic when alpha=1) where param is a positive real number.
rlog<-function(n,param){
if(length(param)!=1 | param<0) stop("the Logarithmic parameter must be a
positive real number")
else{
    pdf.s<-(1:500)^(-1)*exp(-(1:500)/param)/(-log(1-exp(-1/param)))
    sim<-cut(runif(n),unique(c(0,cumsum(pdf.s))),labels=FALSE,right=FALSE)
}
sim}
```

```
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function to generate random numbers from the power law k^{-param} where
param a real number >1
rpower.law<-function(n,param){
if(length(param)!=1 | param<=1) stop("the power law parameter must be a real
number grater than one")
else{
    pdf.s<-(1:1000)^(-param)*(zeta(param))^(-1)
    sim<-cut(runif(n),unique(c(0,cumsum(pdf.s))),labels=FALSE,right=FALSE)
}
sim}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function to generate the distribution for the power law k^{-param}, where
param is real number
dpower.law<-function(x,param){
if(length(param)!=1) stop("the power law parameter dimension is wrong (it
must be of length 1)or the parameter value is incorrect")
else pdf.s<-x^(-param)*(zeta(param))^(-1)
pdf.s}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function to generate random numbers with the distribution Gutenberg-Richter
law ck^{-delta}e^{-k/lambda}
#c=(li_lambda(e^{-1/lambda}))
#param=c(delta,lambda)
rpoly.log<-function(n,param){
if(length(param)!=2) stop("the Gutenber-Richter law parameter is wrong")
else sim<-cut(runif(n),unique(c(0,cumsum(dpoly.log(1:1000,param)))),labels=FALSE,rig
ht=FALSE)
sim}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Function to generate the distribution for the Gutenberg-Richter law ck^{-
delta}e^{-k/lambda}
#c=(li_lambda(e^{-1/lambda}))
#param=c(delta,lambda)
dpoly.log<-function(x,param){
if(length(param)!=2) stop("the Gutenber-Richter law parameter is wrong")
else pdf.s<-x^(-param[1])*exp(-x/param[2])/li(param[1],exp(-1/param[2]))
pdf.s}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#rutine to generate the degrees based on the specified distribution
#n is the number of indivudual to connect
#distrib is the name of the distribution
#param is the unlist distribution parameter
sdegree<-function(n, distrib,param=0){
degree<-3
while(sum(degree)%%2==1){   #first step to realizable graph (Newman,
Stogatz and Watts, 2001)
    if(distrib=="fixed"){
        if(length(param)!=n | sum(param)%%2==1) stop("the length of the degrees
must be equal to the number of nodes and the sum of degree must to be even")
    degree<-param
    }
    else if(distrib=="pois") degree<-rpois(n,param)
    else if(distrib=="ztpois") degree<- rztpois(n,param)
    else if(distrib=="geom") degree<- rgeom(n,1/(param+1))           #sometimes
called exponential graph. param is then the expected value
    else if(distrib=="ztgeom") degree<- rgeom(n,1/(param+1))+1
    else if(distrib=="nbinom" && length(param)==2) degree<-
rnbinom(size=param[1], prob=param[2])
```

```
    else if(distrib=="poly.log") degree<-rpoly.log(n,param)
    else if(distrib=="logarithmic") degree<- rlog(n,param)
    else if(distrib=="power.law") degree<- rpower.law(n,param)
    else if(distrib=="full") degree<-rep(n-1,n)
    else if(distrib=="none") degree<-rep(0,n)
    else stop("incorrect distribution specification")
}
degree}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#rutine to generate the degrees based on the specified distribution
# n is the number of indivudual to connect
#distrib is the name of the distribution
#param is the unlist distribution parameter
ccheck<-function(nodes,edges,degree.left){
    other<-0
    flag<-1
    nodes.degree.left<-nodes[degree.left>0]
    position<- lapply(nodes.degree.left,elements.matrix,matrix=edges) #position
 (row) in edges for each node
    m<-NDIM(nodes.degree.left)[1]
    indexes<-cbind(unlist(mapply(rep,1:(m-1),(m-1):1)),
unlist(mapply(seq,2:m,m)),-1)
    for(i in 1:(m*(m-1)/2)) indexes[i,3]<-
sum(unlist(sapply(position[[indexes[i,1]]],compare.vectors,position[[indexes[i,2
]]])))
    paired.inf<-indexes[indexes[,3]==0,1:2] #index of the nodes.degree.left that
are not already connected
    paired.inf<-matrix(nodes.degree.left[paired.inf],NDIM(paired.inf,as.row=TRUE)[1],2)
    if(NDIM(paired.inf,TRUE)[1]>1){         #at leat one pair can be connected
        new.edge<-sort(paired.inf[resample(1:dim(paired.inf)[1],1),])
        other<-1
    }
    else {
    new.edge<-NULL
    flag<-0                                 #no more nodes can be connected
    }
    list(new.edge=new.edge,other=other,flag=flag)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
edge.check<-function(edges,new.edge){
new.edge<-sort(new.edge)
a<-which(edges==new.edge[1])
b<-elements.matrix(edges,new.edge[2])
if(length(a)==0|length(b)==0)comp<-0
else comp<-sapply(a,compare.vectors,b)
comp}

#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -local.network-#
#network creates the network based in the edge distribution and the number of
nodes
#n number of individuals (susceptible and infective)
#distrib is the degree distribution
#param is the distribution parameter (if the function is "fixed" it is a vector of
degrees)
#distrib can be "fixed" or "pois" or "ztpois" or "geom" or "nbinom","ztgeom" or
"poly.log" or "logarithmic"
#or "power.law" or "full" (fully connected) or "none" (no element connected)
#one.connection is TRUE when only one connection is allowed between two nodes.
#method specifies the algorithm to build the network: Havel-Hakim ("HH") or
Modified Havel-Hakim ("MHH"), Molloy-Reed ("MR") or Random
```

```
("Random")
local.network<-function(n,distrib, param=NULL, one.connection=TRUE,
method="MHH"){       #param is not a list
degree<-sdegree(n,distrib,param)
flag<-1
nodes<-1:n
edges<-NULL
other<-0
degree.left<-degree
count<-counthh<-0
while(length(degree.left[degree.left>0])>1 & flag==1){     #at lest two ind to
connect with new edge
    if (other==0){
        if(method=="HH"){                          #---Havel-Hakimi algorithm---###
            count<-6
            if(counthh==0){
                newnodes<-cbind(nodes[degree.left>0],degree.left[degree.left>0])
                newnodes<-newnodes[order(runif(n=dim(newnodes)[1],0,1)),]#so
next ord doen't depend on index name
                newnodes<-newnodes[order(newnodes[,2],decreasing=T),]    #order
by degree.left(cols:node & degree.left)
                m<-NDIM(newnodes)[1]
                new.edge <- sort(c(newnodes[1,1],newnodes[2,1]))         #2 nodes
with highest degree.left
                count<-0
             }
            else if(counthh>0 && m>2){
                if(counthh==1) indexes2<-cumsum(c(0,seq(2,m-1)))   #1-3 2-3 1-4 2-
4 3-4 1-5 2-5 3-5 4-5..
                if(counthh<=max(indexes2)){
                    b<-cut(counthh,indexes2,label=F)
                    new.edge<-sort(c(newnodes[counthh-indexes2[b],1],newnodes[b+2,1]))
                    count<-0
                }
            }
        }
        else if(method=="MHH"){        #---Modified Havel-Hakimi algorithm-###
            new.edge<-nodes[which(degree.left==max(degree.left))]
            if(length(new.edge)>1) new.edge<-resample(new.edge,1)    #the node
(or one of them) with max degree.left
            new.edge1<-resample(nodes[degree.left>0 & nodes!=new.edge],1)
            new.edge<-sort(c(new.edge,new.edge1))
        }
       else if(method=="MR"){                      #---Molloy and Reed---###
            newnodes<-cbind(nodes[degree.left>0],degree.left[degree.left>0])
            m<-dim(newnodes)[1]
            indexes<-cbind(unlist(mapply(rep,1:(m-1),(m-1):1)),
unlist(mapply(seq,2:m,m)))
            products<-newnodes[indexes[,1],2]*newnodes[indexes[,2],2]
            new.edge<-resample(1:(m*(m-1)/2),1,prob=products)   #randomly selec
a pair of nodes with positive degree.left and weighted according to their degree
product
            new.edge<-sort(c(newnodes[indexes[new.edge,1]],newnodes[indexes[new.edge,2]]))
        }
        else if(method=="Random"){                 #---Random selection of any to
vertices---###
            new.edge<-sort(resample(nodes[degree.left>0],2))    #randomly select a
pair of nodes with positive degree.left
        }
```

```
        if(!is.null(edges) && NROW(t(edges))>1 && one.connection==T){
#avoid to repeat new edges when whe have more to compare and one.connection==TRUE
            a<- which(edges==new.edge[1])              #since new.edges sorted, a is
always in the first column
            b<- elements.matrix(edges, new.edge[2])
            if(length(a)==0 | length(b)==0) comp<-0
            else comp<- sapply(a,compare.vectors,b)
        }
        else comp<-0
    }
    else{
        other<-0
        comp<-0
    }

    if(sum(comp)==0){               #new.edges are not repeated (or it doesn't matter
if they are repeated)
        count<-0
        counthh<-0
        edges<-rbind(edges, new.edge)
        degree.left[new.edge]<-degree.left[new.edge]-1
    }
    else{
        if(method!="HH") count<-count+1              #Number of times we select
two conncected nodes
        else if(method=="HH") counthh<-counthh+1
        if(count>=6){               #check if the all possible connections are already
made but after count>=6
            evl<-ccheck(nodes,edges,degree.left)
            new.edge<-evl$new.edge
            other<-evl$other
            flag<-evl$flag
            counthh<-0
        }
    }
}        #end while
if(!is.null(edges)) edges<-order.edges(edges)
list(edges=edges,degree=degree,degree.left=degree.left)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
# Connects two populations (local networks) with the specified degrees using
the "MHH" algorithm
# p is the number of individuals in each local network
# degree is in the format of list
connect.two.ln<-function(p,degree){
flag<-1
nodes<-rbind(cbind(1:p[1],0),cbind((p[1]+1):sum(p),1)) #second colum point
the local network they belong to
edges<-NULL
degree.left<-unlist(degree)
count<-0
while((length(nodes[degree.left>0 & nodes[,2]==0,1])>=1) &&
(length(nodes[degree.left>0 & nodes[,2]==1,1])>=1) && flag==1){
        new.edge<-nodes[which(degree.left==max(degree.left)),1]
        if(length(new.edge)>1) new.edge<-resample(new.edge,1)    #the node (or
one of them) with max degree.left
        if(nodes[new.edge,2]==0) new.edge1<-resample(nodes[degree.left>0 &
nodes[,2]==1,1],1)  #randomly select a node with degree.left>0
        else new.edge1<-resample(nodes[degree.left>0 & nodes[,2]==0,1],1)
        new.edge<-sort(c(new.edge,new.edge1))
```

```
397         if(!is.null(edges) && NROW(t(edges))>1){    #avoid to repeat new edges
398 when whe have more to compare
399             a<- which(edges==new.edge[1])           #since new.edges sorted, a is
400 always in the first column
401             b<- elements.matrix(edges, new.edge[2])
402             if(length(a)==0 | length(b)==0) comp<-0
403             else comp<- sapply(a,compare.vectors,b)
404         }
405         else comp<-0
406 
407     if(sum(comp)==0){           #new.edges are not repeated (or it doesn't matter if
408 they are repeated)
409         count<-0
410         edges<-rbind(edges, new.edge)
411         degree.left[new.edge]<-degree.left[new.edge]-1
412     }
413     else{
414         count<-count+1              #Number of times we select two connected nodes
415         if(count>=10) flag<-0
416     }
417 }       #end while
418 if(!is.null(edges)) edges<-order.edges(edges)
419 list(edges=edges,degree=degree,degree.left=degree.left)}
420 
421 #- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - global.network-#
422 # Constructs the global networks as in Newman
423 # n: number of local networks
424 # p: number of individuals in each local network
425 # distrib: different distribution for each of the local networks
426 # param: parameter distributions for each of the local networks (format of list)
427 # distrib.among: distribution for the connection among local networks (it is not
428 in format of list)
429 # param.among: parameter distribution for the connection among local networks
430 (list)
431 
432 global.network<-function(n,p,distrib,param=0,distrib.among="none",param.among=0,
433 method="MHH"){
434 if(n==1){                           #when we have only one local network
435     result<-local.network(p,distrib,param=unlist(param),method=method)
436     global.edges<-result$edges      #at most one connection among two indiv.
437 within a local net.
438     localnet.degree=result$degree
439     localnet.degree.left=result$degree.left
440     among.degree<-NULL
441     among.degree.left<-NULL
442 }
443 else{
444     if (length(p)==1) p<-rep(p,n)
445     if (length(distrib)==1) distrib<-rep(distrib,n)         #Copy the same
446 distribution for all
447     if (length(param)==1)param<-rep(param,n)
448     if (length(distrib)!=n | length(param)!=n ) {
449         stop("Any, the length of the vectors distrib or the parameters are
450 incorrect")
451     }
452     csum.p<-cumsum(c(0,p))                  #the cumulative sum of p
453     global.edges<-c(0,0)                    #A matrix that will contain the effective
454 connections
455     localnet.degree<-localnet.degree.left<-as.list(matrix(0,1,n))
```

```
    #---The contact within local networks---
    for (i in 1:n){
        result<-local.network(n=p[i],distrib=distrib[i], param=unlist(param[i]),
one.connection=TRUE,method=method)
        local.net<-result$edges
        global.edges<-rbind(global.edges, local.net+csum.p[i])
        localnet.degree[[i]]<-result$degree
        localnet.degree.left[[i]]<-result$degree.left
    }
    #---The contacts between local networks---distrib.among 1-2,1-3,...,1-n,2-1,2-
3,2-4,...,2-n,...,n-1,...,n-(n-1)
        cutn<-n*(n-1)
        edges<-among.degree.left<-as.list(matrix(0,cutn/2,1))
        among.degree<-as.list(matrix(0,cutn,1))
    if(sum(distrib.among!="none")){ #at least a pair of local networks to connect
        if (length(distrib.among)==1) distrib.among<-rep(distrib.among,cutn)
        if (length(param.among)==1) param.among<-rep(param.among, cutn)
        if (length(distrib.among)!=cutn | length(param.among)!=cutn )
stop("Incorrect length of the vectors distrib or its parameters (between local networks)")
        pairsln.all<-cbind(rep(1:n,each=n),rep(1:n,n))
        pairsln.all<-pairsln.all[pairsln.all[,1]!=pairsln.all[,2],] #dimension n(n-1),2:
1-2, 1-3,...1-n,2-1,2-3,...2-n,3-1,3-2,3-4,...3-n
        pairsln<-cbind(unlist(mapply(rep,1:(n-1),(n-
1):1)),unlist(mapply(seq,2:n,n))) #matrix with the pairs of local networks (n(n-
1))/2,2
        for(i in 1:cutn){
            if(distrib.among[i]!="full") among.degree[[i]]<-
sdegree(p[pairsln.all[i,1]],distrib.among[i],unlist(param.among[i]))
            else among.degree[[i]]<-rep(p[pairsln.all[i,2]],p[pairsln.all[i,1]])
#cbind(rep(1:p[pairs[j,1]],each=p[pairs[j,2]]),rep(1:p[pairs[j,2]],p[pairs[j,1]]))
        }
        indexes<-t(matrix(rep(0:n,n)+rep((0:(n-1))*n,each=(n+1)),n,n)) #indexes
de pairsln.all
        diag(indexes)<-0
        for(j in 1:NDIM(pairsln)[1]){
            resj<-connect.two.ln(p=c(p[pairsln[j,1]],p[pairsln[j,2]]),degree=list(among.degree[[in
dexes[pairsln[j,1],pairsln[j,2]]]],among.degree[[indexes[pairsln[j,2],pairsln[j,1]]
]]))
            if(!is.null(resj$edges)){
                edges[[j]]<-resj$edges
                if(is.vector(edges[[j]]))edges[[j]]<-t(as.matrix(edgnoes[[j]]))
#############
                edges[[j]]<-cbind(edges[[j]][,1]+csum.p[pairsln[j,1]],edges[[j]][,2]-
p[pairsln[j,1]]+csum.p[pairsln[j,2]])
                global.edges<-rbind(global.edges,edges[[j]])
            }
            among.degree.left[[j]]<-resj$degree.left
        }
    }       #if(distrib.among!="none")
    global.edges<-global.edges[-1,]
    dimnames(global.edges)<-NULL
    global.edges<-order.edges(global.edges,ord.col=TRUE)
}
list(global.edges=global.edges,localnet.degree=localnet.degree,
localnet.degree.left=localnet.degree.left, among.degree=among.degree,
among.degree.left=among.degree.left, n=n, p=p, distrib=distrib, param=param)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
# This function returns the time to have a infective event and the effective
connections for transmission.
```

```
# Those connections are the one that are between and infective and a susceptible
stime.infection<-function(network.edges, indexes, infectives,
BETA1,BETA2,infected,seir){
infection<-Inf
eff.connections<-NULL
    if(NDIM(network.edges)[1]>0 & length(infectives)>0){      #at least one connection
        no.erase<-rowSums(matrix(sapply(network.edges,compare.vectors,infectives),NDIM(net
work.edges)))
        if(seir && length(infected)>0)erase<-
rowSums(matrix(sapply(network.edges,compare.vectors,infected),NDIM(netwo
rk.edges)))
        else erase<-0
        no.erase<-(1-erase)*no.erase
        eff.connections<-network.edges[no.erase==1,]
        if(NDIM(eff.connections)[1]>0){
            if(is.vector(eff.connections)) eff.connections<-
t(as.matrix(eff.connections))
            con.infectives<-
matrix(sapply(eff.connections,is.element,infectives),NROW(eff.connections),2)
            con.infectives<-con.infectives[,1]*1+con.infectives[,2]*2
            con.susceptibles<-sapply(con.infectives,switch,2,1)
            con.infectives<-
eff.connections[cbind(1:NROW(eff.connections),con.infectives)] #the index of
infectives
            eff.Beta1<-BETA1[con.infectives]  #who transmits
            con.susceptibles<-
eff.connections[cbind(1:NROW(eff.connections),con.susceptibles)]#the index
of susceptibles
            eff.Beta2<-BETA2[con.susceptibles] #who adquires
            infection<-newrexp(NDIM(eff.connections,TRUE)[1],eff.Beta1*eff.Beta2)
        }
    }
list(infection=infection,eff.connections=eff.connections)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#to use with sapply
rp.norm<-function(param){
rnorm(1,param[1],param[2])}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#to use with sapply
rp.lnorm<-function(param){
rlnorm(1,param[1],param[2])}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
# This function returns the latent period for the new infective(s). Only positive
numbers allowed
latent.period<-function(distrib.lat, indexes, new.infected, LAMBDA){
lat.time<--1
while(sum(lat.time<0)>0){
if(distrib.lat=="exp") lat.time<-
newrexp(length(new.infected),unlist(LAMBDA[indexes[new.infected,1]]))
else if(distrib.lat=="norm") lat.time<-
sapply(LAMBDA[indexes[new.infected,1]],rp.norm)
else if(distrib.lat=="lnorm") lat.time<-
sapply(LAMBDA[indexes[new.infected,1]],rp.lnorm)
else lat.time<-0
}
lat.time}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
# This function returns the infectious period for the new infective(s). Only
positive number allowed
```

```
infective.period<-function(distrib.inf, indexes, new.infectives, GAMMA){
inf.time<--1
while(sum(inf.time<0)>0){
if(distrib.inf=="exp") inf.time<-
newrexp(length(new.infectives),unlist(GAMMA[indexes[new.infectives,1]]))
else if(distrib.inf=="norm") inf.time<-
sapply(GAMMA[indexes[new.infectives,1]],rp.norm)
else if(distrib.inf=="lnorm") inf.time<-
sapply(GAMMA[indexes[new.infectives,1]],rp.lnorm)
else inf.time<-0
}
inf.time}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
# This function removes the edges of the specified node
# sub.id  subset of id's
remove.edges<-function(network.edges,sub.id){
edges.removed<-NULL
if(sum(dim(network.edges)>0)==2){                       #network.edges is a non
empty matrix
    to.remove<-rowSums(matrix(is.element(network.edges,sub.id),dim(network.edges)[1],2))
    edges.removed<-network.edges[to.remove>0,]
    network.edges <-network.edges[to.remove==0,]
}
list(network.edges=network.edges,edges.removed=edges.removed)}

#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -  epidemic.sim -#
# network: network structure (output of global.network) tambien debe darme
informacion de sus parametros para inmigracion!!!!!!
# ini.infected: number or inicially infected in each newtork
# seir: it is False if model is sir
# ini.infective: number of initially infectives in each network
# obs.time: observation period
# BETA1: parameter of transmission to susceptible
# BETA2: parameter of transmission from an infective
# distrib.lat: distribution of the latent period. The default is poisson resulting in
an overal Markov process
# LAMBDA: parameter for the latent period
# distrib.inf: distribution of the infectious period. The default is poisson resulting
in an overal Markov process
#       other options are Nomal "norm" and lognormal "lnorm".
# GAMMA: parameter of distrib.inf. When distrib.inf is Poisson, it is the
reciprocal of the removal rate of infectives
#       gamma is a scalar or vector of length n (Poisson), or a vector of length 2
(normal, lognormal  list(c(1,2)) ) or a matrix of nX2
# imm.s rate of immigration of susceptibles
# imm.e                         exposed (infected not infective)
# imm.i                         infectives
# emm: Emmigration rate that is equal to the sum of the three immigration rates

epidemic.sim<-
function(network=NULL,seir=F,ini.infected=0,ini.infective=0,obs.time=0,BET
A1,BETA2=1,distrib.lat="norm",LAMBDA=NULL,distrib.inf="exp",GAMMA
,imm.s=NULL,imm.e=NULL,
imm.i=NULL,vac.par=NULL,eff.exp=0,eff.inf=0){
if(is.null(network)) stop("A network structure is needed")
if(distrib.inf!="exp" & distrib.inf!="norm" & distrib.inf!="lnorm")
stop("Infectious period is incorrect")
if (obs.time<=0) stop("You must select the period to observe")
#else cat("Period to observe: form 0 to ",obs.time,"\n")
```

```
n<-network$n
if(sum(network$p)>1){
    if(length(BETA1)==1) BETA1<-rep(BETA1,sum(network$p))
    if(length(BETA2)==1) BETA2<-rep(BETA2,sum(network$p))
    if(length(LAMBDA)==1) LAMBDA<-rep(LAMBDA,sum(network$p))
}
if(n>1){
    if(length(ini.infected)==1) ini.infected<-rep(ini.infected,n)
    if(length(ini.infective)==1) ini.infective<-rep(ini.infective,n)
    if(length(GAMMA)==2 & distrib.inf!="exp") GAMMA<-list(GAMMA)
    if(length(GAMMA)==1) GAMMA<-rep(GAMMA,n)
    if(length(imm.s)==1) imm.s<-rep(imm.s,n)
    if(length(imm.e)==1) imm.e<-rep(imm.e,n)
    if(length(imm.i)==1) imm.i<-rep(imm.i,n)
    if(length(vac.par)==1) vac.par<-rep(vac.par,n)
}
if(length(ini.infective)!=n |length(BETA1)!= sum(network$p)|length(BETA2)!=
sum(network$p)) stop("parameter dimension is wrong")
if(seir  && is.null(LAMBDA)) stop("Set the initial number of infected
individuals")
#******************Initial and constant values**********************
p.hist<-p<-network$p
susp.hist<-susp<-p-ini.infective-ini.infected
infectedp.hist<-infectedp<-ini.infected
infectivep.hist<-infectivep<-ini.infective
remp.hist<-remp<-vacp.hist<-vacp<-isolp.hist<-isolp<-rep(0,n)
popp<-rbind(p,susp,infectedp,infectivep,remp,vacp,isolp)
time.hist<-event.time<-0

#INDEXES*****************
#local network, status(0-removed 1-susceptible 2-infected(not infective) 3-
infective 4-vaccinated and removed),ID,infected time, latent period,infectious
period, number of infected*****
indexes<-as.data.frame(cbind(LocalNetwork=rep(1:n,p),Status=rep(1,
sum(p)),NodeId=1:sum(p),InfectedTime=0,LatentPeriod=0,InfectivePeriod=0,L
abel=0, Control=0, TimeVac=0))
infected<-0
if(seir){
    for (j in 1:n) if (ini.infected[j]>0) infected<-
c(infected,resample(indexes[indexes[,1]==j, \3], ini.infected[j]))
    infected<-sort(infected)[-1]
}
infectives<-0
for (j in 1:n){
    if(ini.infective[j]>0) infectives<-
c(infectives,resample(indexes[setdiff(which(indexes[,1]==j),infected),3],ini.infe
ctive[j])) #very importat to include "which"
}
infectives<- sort(infectives)[-1]
if (length(infected)>0)indexes[infected,2]<-2
if (length(infectives)>0)indexes[infectives,2]<-3
if(seir && length(infected)>0) indexes[infected,5]<-latent.period(distrib.lat,
indexes, new.infected=infected, LAMBDA)
if(length(infectives)>0) indexes[infectives,6]<-infective.period(distrib.inf,
indexes, new.infectives=infectives, GAMMA)
#NODES*****************
nodes.hist<-as.data.frame(cbind(NodeId=1:sum(p),Label=0,StartTime=0,Status=indexes[,2]
))
#EDGES*****************
```

```
network.edges<-network$global.edges
if (!is.null(network.edges)){
    if(is.vector(network.edges)) network.edges<-t(as.matrix(network.edges))
    edges.hist<-
as.data.frame(cbind(FromId=network.edges[,1],ToId=network.edges[,2],StartTi
me=0,EStatus=1))
}
else edges.hist<-NULL
#**********************
event.time<-0
more.events<-T       #flag that turns false when no change of network and no
more events are possible
while(event.time<=obs.time & more.events){
        susceptibles<-indexes[indexes[,2]==1,3]         #the Id of active susceptibles
        infected<-indexes[indexes[,2]==2,3]             #the Id of active infected
        infectives<-indexes[indexes[,2]==3,3]           #the Id of active infectives
        if(is.vector(network.edges)) network.edges<-t(as.matrix(network.edges))
        #--------Time Events--------
        infection.net<-stime.infection(network.edges,indexes,infectives,BETA1,BETA2,infected,seir)
        infection.time<-infection.net$infection
        if(length(infected)>0) latent<-indexes[infected,5]+indexes[infected,4]-
event.time
        else latent<-Inf
        if(length(infectives)>0) removal<-
indexes[infectives,6]+indexes[infectives,5]+indexes[infectives,4]-event.time
        else removal<-Inf
        #cat("immsus")
        immsus<-newrexp(n,imm.s)
        immexp<-newrexp(n,imm.e)
        imminf<-newrexp(n,imm.i)
        #cat("vaccination")
        vac.time<-newrexp(n,vac.par)
        #infected, infective, remov, imm.susc, imm.exposed, imm.infective,
vaccination.time:
        time.events<-c(infection.time, latent, removal,
immsus,immexp,imminf,vac.time)
        time.af<-suppressWarnings(min(time.events[is.finite(time.events)]))
        event.time<-time.af+sum(time.hist)      #Includes transmission, start
infective, removals and migrations
        #--------------------------
        if (!is.finite(time.af) || (sum(popp[1,])==0 & is.null(imm.s) &
is.null(imm.e) & is.null(imm.i))) {
            #print("no new event is possible")
            more.events<-F
        }
        else{
            eff.connections<-infection.net$eff.connections
            time.hist<-cbind(time.hist, time.af)        #the historical times between
consecutive events
            which.event<-which(time.events==time.af)
            criteria<-unique(cumsum(c(0, length(infection.time), length(latent),
length(removal), length(immsus),length(immexp),length(imminf),length(vac.time))))
            event<-cut(which.event,criteria,label=F)
            if (event==1){
            #cat("*** transmission \n")
                who.trans<-sapply(eff.connections[which.event,],compare.vectors,
infectives)
                who.inf<-eff.connections[which.event,who.trans>0]           #the
infective indiv who transmits the infection
```

```
                indexes[who.inf,7]<-indexes[who.inf,7]+1
                who.is.inf<-eff.connections[which.event,who.trans==0]        #the
susceptible that becomes infected
                if(seir){  #2-status , 4-infected time, 5-latent period
                    indexes[who.is.inf,c(2,4,5)]<-c(2,event.time,latent.period(distrib.lat,
indexes, new.infected=who.is.inf,
LAMBDA))
                    change<-c(0,-1,1,0,0,0,0)
                }
                else{
                    indexes[who.is.inf,c(2,4,6)]<-
c(3,event.time,infective.period(distrib.inf, indexes, new.infectives=who.is.inf,
GAMMA))
                    change<-c(0,-1,0,1,0,0,0)
                }
                popp[,indexes[who.is.inf,1]]<-popp[,indexes[who.is.inf,1]]+change
                nodes.hist<-rbind(nodes.hist, c(who.inf, indexes[who.inf,7],event.time,3),
c(who.is.inf,indexes[who.is.inf,7],event.time,indexes[who.is.inf,2]))
                edges.hist<-
concatenate(edges.hist,c(min(who.inf,who.is.inf),max(who.inf,who.is.inf)),c(ev
ent.time,2))
            }
            else if (event==2){
            #cat("*** start infectious period \n")
                now.infective<-infected[which.event-criteria[2]]
                indexes[now.infective,c(2,6)]<-c(3,infective.period(distrib.inf,
indexes, new.infectives=now.infective, GAMMA))
                popp[,indexes[now.infective,1]]<-
popp[,indexes[now.infective,1]]+c(0,0,-1,1,0,0,0)
                nodes.hist<-
rbind(nodes.hist,c(now.infective,indexes[now.infective,7],event.time,3))
            }
            else if (event==3){
            #cat("*** removal \n")
                removed<-infectives[which.event-criteria[3]]
                indexes[removed,2]<-0
                popp[,indexes[removed,1]]<-popp[,indexes[removed,1]]+c(-1,0,0,-
1,1,0,0)
                nodes.hist<-rbind(nodes.hist, c(removed,
indexes[removed,7],event.time, 0))
                a<-remove.edges(network.edges,removed)
                network.edges<-a$network.edges
                edges.hist<-concatenate(edges.hist,a$edges.removed,c(event.time,0))
            }
            else if (event==4 || event==5 ||event==6){
                chlocal.net<- which.event-criteria[event]
                new.id<-NROW(indexes)+1
                nstatus<-event-3
                indexes<-rbind(indexes, c(chlocal.net,nstatus,new.id,0,0,0,0,0,0))
                if(event==4){
                    #cat("*** susceptible arrival \n")
                    change<-c(1,1,0,0,0,0,0)
                }
                if(event==5){
                    #cat("*** exposed arrival \n")
                    change<-c(1,0,1,0,0,0,0)
                    indexes[new.id,4:5]<-c(event.time,latent.period(distrib.lat, indexes,
new.infected=new.id, LAMBDA))
                }
```

```
                if(event==6){
                    #cat("*** infective arrival \n")
                    change<-c(1,0,0,1,0,0,0)
                    indexes[new.id,c(4,6)]<-c(event.time,infective.period(distrib.inf,
indexes, new.infectives=new.id, GAMMA))
                }
                popp[,chlocal.net]<-popp[,chlocal.net]+change
                nodes.hist<- rbind(nodes.hist, c(new.id,0,event.time,nstatus))
                new.network.edges<-new.edges(new.id,
network$distrib[chlocal.net],network$param[chlocal.net],
indexes[indexes[,1]==chlocal.net & indexes[,2]==1,3])
                edges.hist<-concatenate(edges.hist,new.network.edges,c(event.time,1))
                network.edges<-rbind(network.edges,new.network.edges)
            }
            else if (event==7) {
            #cat("*** vaccination \n")
                chlocal.net<- which.event-criteria[7]
                if(popp[1,chlocal.net]>0){
                    suslocal.net<-indexes[indexes[,2]==1 &
indexes[,1]==chlocal.net,3]
                    expnovac<-indexes[indexes[,2]==2 & indexes[,1]==chlocal.net &
indexes[,8]==0,3]
                    infnovac<-indexes[indexes[,2]==3 & indexes[,1]==chlocal.net &
indexes[,8]==0,3]
                    b<-c(length(suslocal.net),length(expnovac),length(infnovac))
                    popn<-c(1,eff.exp,eff.inf)*b
                    popn<-cumsum(popn)/sum(popn)
                    selected<-cut(runif(1),unique(c(0,popn)),label=F)
                    if(length(c(0,popn))>length(unique(c(0,popn)))) selected<-
which(b>0)[selected]
                    if(selected==1){
                        vac<-resample(suslocal.net,1)
                        change<-c(-1,-1,0,0,0,1,0)
                        nodes.hist<-rbind(nodes.hist,c(vac,indexes[vac,7],event.time,4))
                        indexes[vac,2]<-indexes[vac,8]<-4
                        a<-remove.edges(network.edges,vac)
                        network.edges<-a$network.edges
                        edges.hist<-concatenate(edges.hist,a$edges.removed,c(event.time,0))
                    }
                    else{
                        if(selected==2) vac<-resample(1,expnovac)
                        else if(selected==3) vac<-resample(1,infnovac)
                        indexes[vac,8]<-4
                        change<-c(0,0,0,0,0,1,0)
                    }
                    indexes[vac,9]<-event.time
                    popp[,chlocal.net]<-popp[,chlocal.net]+change
                }
            }
            else{
                print("Bug: something wrong with variable -event-")
                cat("event= ", event, "\n")
            }
            p.hist<-rbind(p.hist,popp[1,])
            susp.hist<-rbind(susp.hist,popp[2,])
            infectedp.hist<-rbind(infectedp.hist,popp[3,])
            infectivep.hist<-rbind(infectivep.hist,popp[4,])
            remp.hist<-rbind(remp.hist,popp[5,])
            vacp.hist<-rbind(vacp.hist,popp[6,])
```

```
            isolp.hist<-rbind(isolp.hist,popp[7,])
        }               #end case when time.af is finite
}                       #end while
dimnames(network.edges)<-NULL               #to eliminate headers
time.hist <- as.vector(time.hist)
dimnames(p.hist) <- dimnames(susp.hist) <-dimnames(infectedp.hist)<-
dimnames(infectivep.hist) <- dimnames(remp.hist) <- NULL
list(indexes=indexes,time.hist=time.hist,nodes.hist=nodes.hist,edges.hist=edges.
hist,p.hist=p.hist,susp.hist=susp.hist,infectedp.hist=infectedp.hist,infectivep.hist
=infectivep.hist,remp.hist=remp.hist, n=n)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#Internal
#Function that is complementary to sformat.output to prepare the output of R as
input of SONIA
#Assigns coordinates to the nodes
#n number of networks
#indexes is the last indexes result in data.frame format
coordinates<-function(n,indexes,random=FALSE){
if(n>1){
    indexes<- indexes[order(indexes[,1]),]              #ordering the nodes wrt local
network
    p.all<-sapply(X=1:n, FUN=number.elements, mat=indexes, col=1)
    rad<-matrix(c(1,1,rep(0,n-1)),n,n+1)
    rad[,n]<-rad[,n]+rad[,n+1]
    rad<-rad[,1:n]
    rad<-t(p.all)%*%rad
    t<-(0:(n-1))*(2*pi/n)
    rad2<-2*sin(t[2]/2)*((1/rad)*c(p.all[2:n],p.all[1]))
    rad<-pmin(2*sin(t[2]/2)*((1/rad)*p.all),c(rad2[n],rad2[1 :(n-1)]))
    Xcent<-rep(cos(t), p.all)
    Ycent<-rep(sin(t), p.all)
    if(random) X<-runif(sum(p.all),0.8*rep(rad,p.all),rep(rad,p.all))
    else X<-rep(rad*.90,p.all)
    Y<- sapply(X=p.all,FUN=angles)
    w<-NULL
    for(i in Y) w<-c(w,i)
    Y<-w
    r<-X
    X<-r*cos(Y)+Xcent
    Y<-r*sin(Y)+Ycent
    coord<-as.data.frame(cbind(NodeId=indexes$NodeId,X,Y))
}
else if(n==1){
    t<-(0:(NROW(indexes)-1))*(2*pi/NROW(indexes))
    if(random) r<-runif(NROW(indexes),.7,1)
    else r<-1
    X<-r*cos(t)
    Y<-r*sin(t)
    coord<-as.data.frame(cbind(NodeId=indexes$NodeId,X,Y))
}
coord}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
# Internal
# this function prepares the simultaion output as a SoNIA input
# a is output of epidemic.sim
# random is T or F
sformat.output<-function(a,random=FALSE){
last.time<-sum(a$time.hist)
#**************prepare nodes
```

```
nodes<-a$nodes.hist
nodes<-nodes[order(nodes[,1],nodes[,3]),]
s<-sapply(1:max(nodes[,1]),compare.vectors,nodes[,1])
EndTime <-rep(last.time, nrow(nodes))
cut.time<-nodes[-1,3]
EndTime [-cumsum(s)]<-cut.time[-cumsum(s[-length(s)])]
nodes<-merge(nodes, coordinates(a$n,a$indexes,random),by="NodeId")
nodes<-cbind(nodes[,1:2],nodes$X,nodes$Y,nodes[,3],EndTime,5,nodes[,4])
nodes[nodes[,8]==0,7]<-1
nodes[nodes[,8]==0,8]<-"LightGray"
nodes[nodes[,8]==1,8]<-"black"
nodes[nodes[,8]==2,8]<-"blue"
nodes[nodes[,8]==3,8]<-"red"
names(nodes)<- c("NodeId", "Label", "X","Y","StartTime","EndTime",
"NodeSize", "ColorName")
#**************prepare edges
edges<-a$edges.hist[order(a$edges.hist[,1],a$edges.hist[,2], a$edges.hist[,3]),]
  #order: "FromId" - "ToId" -  "StartTime"
edges<-grouping(edges, paste(edges[,1],edges[,2]))
s<-sapply(1:max(edges[,5]),compare.vectors, edges[,5])
EndTime <-rep(last.time, nrow(edges))
cut.time<- edges [-1,3]
EndTime [-cumsum(s)]<-cut.time[-cumsum(s[-length(s)])]
edges<-cbind(edges[,1:3],EndTime,5,edges[,4])
edges<-edges[edges[,6]!=0,]
edges[edges[,6]==1,6]<-"black"
edges[edges[,6]==2,6]<-"red"
names(edges)<-c("FromId","ToId", "StartTime", "EndTime","ArcWidth",
"ColorName")
#**************preparing the SoNIA's file fomat
edges.for<-rbind(names(edges),edges)
edges.for <- cbind(edges.for,"","")
names(edges.for)<-names(nodes)
res<-rbind(nodes,edges.for)
res}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#
#External
#This function prepares the output from epidemic.sim as an sonia file and save
it
sonia.format<-function(result,random=FALSE,filename.son="sonia.file"){
    for.sonia<-sformat.output(result,random)
    write.table(for.sonia, file=filename.son, row.names=FALSE,quote=FALSE,
sep="\t")
cat(filename.son, "saved")
for.sonia}

#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - epidemic.control.sim  -#

# this program implements the cotrol measures. Parameters are as in
epidemic.sim
# mvac fraction of susceptible nodes that are vaccinated before the epidemic
[0,1]
# aquint fraction of susceptible nodes that are selected to select one of their
neighbors [0,1]
# ring.p The probability that the nodes around an infective are vaccinated
# ring.m The number of days after the nodes are vaccinated around a node who
was infective for m units of time.
# isol.p The probability that a node who become infective is isolated
# isol.m The number of days after becoming infective a node is isolated
```

```
987
988 epidemic.control.sim<-
989 function(network=NULL,seir=F,ini.infected=NULL,ini.infective=0,obs.time=1
990 e6,BETA1,BETA2=1,distrib.lat="norm",LAMBDA=NULL,distrib.inf="exp",G
991 AMMA,vac.par=NULL,eff.exp=0,eff.inf=0,mvac=NULL,
992 aquint=NULL,ring.p=NULL, ring.m=NULL,isol.p=NULL, isol.m=NULL){
993 if(!is.null(mvac) & !is.null(aquint)) stop("Only one pre-outbreak vaccination
994 process is allowed")
995 if(is.null(network)) stop("A network structure is needed")
996 if(distrib.inf!="exp" & distrib.inf!="norm" & distrib.inf!="lnorm")
997 stop("Infectious period is incorrect")
998 #if (obs.time<=0) stop("You must select the period to observe") Elimino porque
999 si no hay inmigracion el proceso va a parar.
1000 n<-network$n
1001 if(sum(network$p)>1){
1002     if(length(BETA1)==1) BETA1<-rep(BETA1,sum(network$p))
1003     if(length(BETA2)==1) BETA2<-rep(BETA2,sum(network$p))
1004     if(length(LAMBDA)==1) LAMBDA<-rep(LAMBDA,sum(network$p))
1005 }
1006 if(n>1){
1007     if(length(ini.infected)==1) ini.infected<-rep(ini.infected,n)
1008     if(length(ini.infective)==1) ini.infective<-rep(ini.infective,n)
1009     if(length(GAMMA)==2 & distrib.inf!="exp") GAMMA<-list(GAMMA)
1010     if(length(GAMMA)==1) GAMMA<-rep(GAMMA,n)
1011     if(length(vac.par)==1) vac.par<-rep(vac.par,n)
1012 }
1013 if(length(ini.infective)!=n
1014 |length(BETA1)!=sum(network$p)|length(BETA2)!=sum(network$p))
1015 stop("parameter dimension is wrong")
1016 if(seir  && is.null(LAMBDA)) stop("Set the initial number of infected
1017 individuals")
1018 #EDGES*****************
1019 network.edges<-network$global.edges
1020 #if (!is.null(network.edges)){
1021 #    if(is.vector(network.edges)) network.edges<-t(as.matrix(network.edges))
1022 #    edges.hist<-
1023 as.data.frame(cbind(FromId=network.edges[,1],ToId=network.edges[,2],StartTi
1024 me=0,EStatus=1))
1025 #}
1026 #else edges.hist<-NULL
1027 #INDEXES*****************
1028 #local network, status(0-removed 1-susceptible 2-infected(not infective) 3-
1029 infective 4-vaccinated and removed) 5 6-isolated,ID,infected time, latent period,
1030 infectious period, number of infected, control, timeVac, ring.vac, TimeIsol*****
1031 indexes<-as.data.frame(cbind(LocalNetwork=rep(1:n,network$p),Status=rep(1,
1032 sum(network$p)),NodeId=1:sum(network$p),InfectedTime=0,LatentPeriod=0,I
1033 nfectivePeriod=0,Label=0, Control=0, TimeVac=0,ring.vac=0, TimeIsol=0))
1034 if(!is.null(mvac)){
1035 indexes[resample(1:sum(network$p),ceiling(mvac*sum(network$p))),2]<-4
1036 #select the mass vaccinated nodes
1037 network.edges<-
1038 remove.edges(network.edges,indexes[indexes[,2]==4,3])$network.edges
1039 }
1040 if(!is.null(aquint)){
1041     selected<-sort(resample(indexes[indexes[,2]==1,3],ceiling(aquint*sum(network$p)),rep=
1042 T)) #resample <- function(x, size, rep=FALSE,prob=NULL)
1043     #buscar vecinos
1044     sel.edges<-remove.edges(network.edges,selected)$edges.removed
1045     tovac<-0
```

```
    for(i in selected){
        if(sum(sel.edges==i)>0){
            tovacedge<-sel.edges[rowSums(sel.edges==i)==1,]
            if(is.vector(tovacedge))tovacedge<-t(as.matrix(tovacedge))
            tovacedge[tovacedge[,1]!=i,]<-tovacedge[tovacedge[,1]!=i,2:1]
            if(length(tovacedge[!is.element(tovacedge[,2],tovac),2])>0){                #
                node.tovac<-resample(tovacedge[!is.element(tovacedge[,2],tovac),2],1)   #
                tovac<-c(tovac,node.tovac)                                              #
                if(!is.element(node.tovac,selected)||i>node.tovac) sel.edges<-
remove.edges(sel.edges,node.tovac)$network.edges
            }                                                                           #
        }
    }
    indexes[tovac[-1],2]<-4
    network.edges<-remove.edges(network.edges,indexes[indexes[,2]==4,3])$network.edges
}
#selection of infected and infectives:

infected<-0
if(seir && sum(ini.infected)>0){
    for (j in 1:n){ if (ini.infected[j]>0) infected<-
c(infected,resample(indexes[indexes[,1]==j & indexes[,2]==1,3],ini.infected[j]))}
    if(length(infected)>1){
        indexes[infected[-1],2]<-2
        if(!is.null(ring.p) & !is.null(ring.m)) indexes[infected[-1],10]<-ring.m
        if(!is.null(isol.p) & !is.null(isol.m)) indexes[infected[-1],11]<-isol.m
        indexes[sort(infected)[-1],5]<-latent.period(distrib.lat, indexes,
new.infected=sort(infected)[-1], LAMBDA)
    }
}

infectives<-0
if (sum(ini.infective)>0){
    for (j in 1:n){
        if(ini.infective[j]>0) infectives<-c(infectives,resample(indexes[indexes[,1]==j
&indexes[,2]==1,3],ini.infective[j]))
    }
    if(length(infectives)>1){
        indexes[infectives[-1],2]<-3
        if(!is.null(ring.p) & !is.null(ring.m)) indexes[infectives[-1],10]<-ring.m
        if(!is.null(isol.p) & !is.null(isol.m)) indexes[infectives[-1],11]<-isol.m
        indexes[sort(infectives)[-1],6]<-infective.period(distrib.inf, indexes,
new.infectives=sort(infectives)[-1], GAMMA)
    }
}
#******************Initial and constant values**********************
p.hist<-p<-as.vector(table(c(indexes[is.element(indexes[,2],c(1,2,3)),1],1:n))-1)
#susc, infected and infective
susp.hist<-susp<-p-ini.infective-ini.infected
if(seir) infectedp.hist<-infectedp<-ini.infected
else infectedp.hist<-infectedp<-0
infectivep.hist<-infectivep<-ini.infective
vacp.hist<-vacp<-as.vector(table(c(indexes[indexes[,2]==4,1],1:n))-1)
remp.hist<-remp<-isolp.hist<-isolp<-rep(0,n)
popp<-rbind(p,susp,infectedp,infectivep,remp,vacp,isolp)
time.hist<-event.time<-0
#NODES*****************
#nodes.hist<-as.data.frame(cbind(NodeId=1:sum(p),Label=0,StartTime=0,Status=indexes[,2]
))
```

```
#**********************
event.time<-0
more.events<-T       #flag that turns false when no change of network and no
more events are possible
while(event.time<=obs.time & more.events){
        susceptibles<-indexes[indexes[,2]==1,3]         #the Id of active susceptibles
        if(seir) infected<-indexes[indexes[,2]==2,3]    #the Id of active infected
        infectives<-indexes[indexes[,2]==3,3]           #the Id of active infectives
        removed<-indexes[indexes[,2]==0,3]
        if(!is.null(ring.p) & !is.null(ring.m)){
            infectivestoring<-intersect(infectives,which(indexes[,8]!=5))
            removedtoring<-indexes[indexes[,2]==0 & indexes[,8]!=5,3]
        }
        if(!is.null(isol.p) & !is.null(isol.m)){
            infectivestoisol<-intersect(infectives,which(indexes[,8]!=6)) #infective
who failed to be detected before
            removedtoisol<-indexes[indexes[,2]==0 & indexes[,8]!=6,3]
        }
        if(is.vector(network.edges)) network.edges<-t(as.matrix(network.edges))
        #--------Time Events--------
        infection.net<-stime.infection(network.edges,indexes,infectives,BETA1,BETA2,infected,seir)
        infection.time<-infection.net$infection
        if(seir && length(infected)>0) latent<-indexes[infected,5]+indexes[infected,4]-
event.time
        else latent<-Inf
        if(length(infectives)>0) removal<-
indexes[infectives,6]+indexes[infectives,5]+indexes[infectives,4]-event.time
        else removal<-Inf
        vac.time<-newrexp(n,vac.par)   #este es el evento de vacunar por tasa de
vacunados
        ring.vac.inf<-ring.vac.rem<-isol.inf<-isol.rem<-Inf
        if(!is.null(ring.p) & !is.null(ring.m)){
            if(length(infectivestoring)>0){
                ring.vac.inf<-
indexes[infectivestoring,10]+indexes[infectivestoring,5]+indexes[infectivestori
ng,4]-event.time
                if(ring.vac.inf<0 && indexes[infectivestoring,10]==0 &&
abs(ring.vac.inf)<1e-10)  ring.vac.inf<-0
            }
            if(length(removedtoring)>0) ring.vac.rem<-
indexes[removedtoring,10]+indexes[removedtoring,5]+indexes[removedtoring,
4]-event.time
        }
       if(!is.null(isol.p) & !is.null(isol.m)){
            if(length(infectivestoisol)>0){
                isol.inf<-
indexes[infectivestoisol,11]+indexes[infectivestoisol,5]+indexes[infectivestoiso
l,4]-event.time
                if(isol.inf<0 && indexes[infectivestoisol,10]==0 && abs(isol.inf)<1e-10)
isol.inf<-0
            }
            if(length(removedtoisol)>0) isol.rem<-
indexes[removedtoisol,11]+indexes[removedtoisol,5]+indexes[removedtoisol,4]
-event.time
        }
        #infected, infective, remov, vaccination.time:
        time.events<-c(infection.time, latent, removal, vac.time, ring.vac.inf,
ring.vac.rem,isol.inf,isol.rem)
        if(sum(time.events<0)>0) browser()
```

```
        time.af<-suppressWarnings(min(time.events[is.finite(time.events)]))
        event.time<-time.af+sum(time.hist)      #Includes transmission, start
infective, removals
        #--------------------------
        if (!is.finite(time.af) || sum(popp[1,])==0) {
            #print("no new event is possible")
            more.events<-F
        }
        else{
            time.hist<-cbind(time.hist, time.af)        #the historical times between
consecutive events
            which.event<-which(time.events==time.af)
            if(length(which.event)>0) which.event<-
which.event[order(runif(length(which.event)))][1]
            criteria<-unique(cumsum(c(0, length(infection.time), length(latent),
length(removal), length(vac.time),length(ring.vac.inf),length(ring.vac.rem),
 length(isol.inf),length
(isol.rem))))
            #browser()
            event<-cut(which.event,criteria,label=F)
            if (event==1){
            #cat("*** transmission \n")
                eff.connections<-infection.net$eff.connections
                who.trans<-sapply(eff.connections[which.event,],compare.vectors,
infectives)
                who.inf<-eff.connections[which.event,who.trans>0]            #the
infective indiv who transmits the infection
                indexes[who.inf,7]<-indexes[who.inf,7]+1
                who.is.inf<-eff.connections[which.event,who.trans==0]        #the
susceptible that becomes infected
                if(seir){  #2-status , 4-infected time, 5-latent period
                    indexes[who.is.inf,c(2,4,5)]<-c(2,event.time,latent.period(distrib.lat,
indexes, new.infected=who.is.inf,
LAMBDA))
                    change<-c(0,-1,1,0,0,0,0)
                }
                else{
                    indexes[who.is.inf,c(2,4,6)]<-c(3,event.time,infective.period(distrib.inf,
indexes, new.infectives=who.is.inf,
GAMMA))
                    change<-c(0,-1,0,1,0,0,0)
                }
                popp[,indexes[who.is.inf,1]]<-popp[,indexes[who.is.inf,1]]+change
                if(!is.null(ring.p) & !is.null(ring.m)) indexes[who.is.inf,10]<-ring.m
                if(!is.null(isol.p) & !is.null(isol.m)) indexes[who.is.inf,11]<-isol.m
                #nodes.hist<-rbind(nodes.hist, c(who.inf, indexes[who.inf,7],event.time,3),
c(who.is.inf,indexes[who.is.inf,7],event.time,indexes[who.is.inf,2]))
                #edges.hist<-
concatenate(edges.hist,c(min(who.inf,who.is.inf),max(who.inf,who.is.inf)),c(ev
ent.time,2))
                #cat("***",who.inf, "transmit to", who.is.inf,"at", event.time,"\n")
            }
            else if (event==2){
            #cat("*** start infectious period \n")
                now.infective<-infected[which.event-criteria[2]]
                indexes[now.infective,c(2,6)]<-c(3,infective.period(distrib.inf,
indexes, new.infectives=now.infective, GAMMA))
                popp[,indexes[now.infective,1]]<- popp[,indexes[now.infective,1]]+c(0,0,-
1,1,0,0,0)
```

```
                #nodes.hist<-
rbind(nodes.hist,c(now.infective,indexes[now.infective,7],event.time,3))
                #cat("***",now.infective," start infectious period at",event.time,"\n")
            }
            else if (event==3){
            #cat("*** removal \n")
                removed<-infectives[which.event-criteria[3]]
                indexes[removed,2]<-0
                popp[,indexes[removed,1]]<-popp[,indexes[removed,1]]+c(-1,0,0,-
1,1,0,0)
                #nodes.hist<-rbind(nodes.hist, c(removed, indexes[removed,7],event.time, 0))
                a<-remove.edges(network.edges,removed)
                network.edges<-a$network.edges
                #edges.hist<-concatenate(edges.hist,a$edges.removed,c(event.time,0))
                #cat("***",removed,"is removed at",event.time,"\n")
            }
            else if (event==4) {
            #cat("*** vaccination \n")
                chlocal.net<- which.event-criteria[7]
                if(popp[1,chlocal.net]>0){
                    suslocal.net<-indexes[indexes[,2]==1 &
indexes[,1]==chlocal.net,3]
                    expnovac<-indexes[indexes[,2]==2 & indexes[,1]==chlocal.net &
indexes[,8]==0,3]
                    infnovac<-indexes[indexes[,2]==3 & indexes[,1]==chlocal.net &
indexes[,8]==0,3]
                    b<-c(length(suslocal.net),length(expnovac),length(infnovac))
                    popn<-c(1,eff.exp,eff.inf)*b
                    popn<-cumsum(popn)/sum(popn)
                    selected<-cut(runif(1),unique(c(0,popn)),label=F)
                    if(length(c(0,popn))>length(unique(c(0,popn)))) selected<-
which(b>0)[selected]
                    if(selected==1){
                        vac<-resample(suslocal.net,1)
                        change<-c(-1,-1,0,0,0,1,0)
                        #nodes.hist<-rbind(nodes.hist,c(vac,indexes[vac,7],event.time,4))
                        indexes[vac,2]<-indexes[vac,8]<-4
                        a<-remove.edges(network.edges,vac)
                        network.edges<-a$network.edges
                        #edges.hist<-concatenate(edges.hist,a$edges.removed,c(event.time,0))
                    }
                    else{
                        if(selected==2) vac<-resample(1,expnovac)
                        else if(selected==3) vac<-resample(1,infnovac)
                        indexes[vac,8]<-4
                        change<-c(0,0,0,0,0,1,0)
                    }
                    indexes[vac,9]<-event.time
                    popp[,chlocal.net]<-popp[,chlocal.net]+change
                }
            }
            else if (event==5 | event==6){
                if(event==5){
                    #cat("*** ring. vaccination around infective\n")
                    vac.neigh<-infectivestoring[which.event-criteria[5]]      #the node
to vaccinate neigh around
                }
                else{
                    #cat("*** ring. vaccination around removed\n")
```

```
                    vac.neigh<-removedtoring[which.event-criteria[6]]
                }
                indexes[vac.neigh,8]<-5 #flag to indicate that its neighbours have
been selected for vaccination (with prob p)
                sel.edges<-remove.edges(network.edges,vac.neigh)$edges.removed
#the edges to which vac.neigh is connected
                if(length(sel.edges)>0){
                    if(is.vector(sel.edges))sel.edges<-t(as.matrix(sel.edges))
                    sel.edges<-
remove.edges(sel.edges,setdiff(indexes[indexes[,2]!=1,3],vac.neigh))$network.e
dges #keep the nodes that still are susceptible
                    if(is.vector(sel.edges)) sel.edges<-t(as.matrix(sel.edges))
                    if(dim(sel.edges)[1]>0){
                        tovaccinate<-runif(dim(sel.edges)[1])<=ring.p
                        #now remove the selected neighbours
                        vaccinate<-
setdiff(unique(as.vector(sel.edges[tovaccinate,])),vac.neigh)#id of the nodes to
be vaccinated
                        if(length(vaccinate)>0){
                            network.edges<-remove.edges(network.edges,vaccinate)$network.edges
                            indexes[vaccinate,8]<-4
                            #vaccinate and remove from the sistem the one that were
susceptibles:
                            rem.vac<-intersect(vaccinate, which(indexes[,2]==1))
                            indexes[rem.vac,2]<-4
                            indexes[vaccinate,9]<-event.time
                            change<-as.vector(table(c(indexes[rem.vac,1],1:n))-1)
                            popp[1,]<-popp[1,]-change
                            popp[2,]<-popp[2,]-change
                            popp[6,]<-popp[6,]+change
                        }
                    #if(event==5)cat("-> nodes vaccinated",vaccinate,"neighbours of
infective", vac.neigh,"at",event.time,"\n")
                    #if(event==6)cat("-> nodes vaccinated",vaccinate,"neighbours of
removed", vac.neigh,"at",event.time,"\n")
                    }
                }
            }
            else if (event==7 | event==8){
                if(event==7){
                #cat("*** isolation of infective\n")
                    isolated<-infectivestoisol[which.event-criteria[7]]
                }
                if(event==8){
                #cat("*** isolation of removed\n")
                    isolated<-removedtoisol[which.event-criteria[8]]
                }
                indexes[isolated,8]<-6  #flag to indicate node has been considered for
isolation
                if(event==7 && runif(1)<=isol.p){
                    indexes[isolated,2]<-6    #isolated and remove from sistem
                    network.edges<-remove.edges(network.edges,isolated)$network.edges
                    popp[,indexes[isolated,1]]<-popp[,indexes[isolated,1]]+c(-1,0,0,-
1,0,0,1)
                }
                #cat("-> isolated node",isolated,"at",event.time,"\n")
            }
            else{
                print("Bug: something wrong with variable -event-")
```

```
                #cat("event= ", event, "\n")
            }
            p.hist<-rbind(p.hist,popp[1,])
            susp.hist<-rbind(susp.hist,popp[2,])
            infectedp.hist<-rbind(infectedp.hist,popp[3,])
            infectivep.hist<-rbind(infectivep.hist,popp[4,])
            remp.hist<-rbind(remp.hist,popp[5,])
            vacp.hist<-rbind(vacp.hist,popp[6,])
            isolp.hist<-rbind(isolp.hist,popp[7,])
        }               #end case when time.af is finite
}                       #end while
dimnames(network.edges)<-NULL               #to eliminate headers
time.hist <- as.vector(time.hist)
dimnames(p.hist) <- dimnames(susp.hist) <-dimnames(infectedp.hist)<-
dimnames(infectivep.hist) <- dimnames(remp.hist) <- NULL
list(indexes=indexes,time.hist=time.hist,p.hist=p.hist,susp.hist=susp.hist,infecte
dp.hist=infectedp.hist,infectivep.hist=infectivep.hist,remp.hist=remp.hist,vacp.h
ist=vacp.hist,isolp.hist=isolp.hist, n=n)}


#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - epidemic.reedfost.sim -#

# network: network structure (output of global.network) tambien debe darme
informacion de sus parametros para inmigracion!!!!!!
# ini.infected: number or inicially infected in each newtork
# seir: it is False if model is sir. When seir=F the infected are automatically
infective.
# ini.infected: if seir ini.infected will not be considered
# ini.infective: number of initially infectives in each network.
# obs.time: observation period
# BETA1: parameter of transmission to susceptible
# BETA2: parameter of transmission from an infective
# mvac fraction of susceptible nodes that are vaccinated before the epidemic
[0,1]
# aquint fraction of susceptible nodes that are selected to select one of their
neighbors [0,1]

epidemic.reedfrost.sim<- function(network=NULL,ini.infective=0,obs.time=1e6,transm.tau=0,vac.par=N
ULL,eff.exp=0,eff.inf=0,mvac=NULL, aquint=NULL){
if(!is.null(mvac) & !is.null(aquint)) stop("Only one pre-outbreak vaccination
process is allowed")
if(is.null(network)) stop("A network structure is needed")
n<-network$n
#if(sum(network$p)>1){
#    if(length(transm.tau)==1) transm.tau<-rep(transm.tau,sum(network$p))
#}
if(n>1){
    if(length(ini.infective)==1) ini.infective<-rep(ini.infective,n)
    if(length(vac.par)==1) vac.par<-rep(vac.par,n)
}
if(length(ini.infective)!=n ) stop("parameter dimension is wrong")
#EDGES*****************
network.edges<-network$global.edges
#if (!is.null(network.edges)){
#    if(is.vector(network.edges)) network.edges<-t(as.matrix(network.edges))
#    edges.hist<-as.data.frame(cbind(FromId=network.edges[,1],ToId=network.edges[,2],StartTi
me=0,EStatus=1))
#}
#else edges.hist<-NULL
```

```
#INDEXES*****************
#local network, status(0-removed 1-susceptible 2-infected(not infective) 3-
infective 4-vaccinated and removed) 5 6-isolated,ID,infected time, latent period,
infectious period, number of infected, control, timeVac, ring.vac, TimeIsol*****
indexes<-as.data.frame(cbind(LocalNetwork=rep(1:n,network$p),Status=rep(1,sum(netw
ork$p)),NodeId=1:sum(network$p),InfectedTime=0,LatentPeriod=0,InfectivePe
riod=0,Label=0, Control=0, TimeVac=0,ring.vac=0, TimeIsol=0))
if(!is.null(mvac)){
indexes[resample(1:sum(network$p),ceiling(mvac*sum(network$p))),2]<-4
#select the mass vaccinated nodes
network.edges<-remove.edges(network.edges,indexes[indexes[,2]==4,3])$network.edges
}
if(!is.null(aquint)){
    selected<-sort(resample(indexes[indexes[,2]==1,3],ceiling(aquint*sum(network$p)),rep=
T)) #resample <- function(x, size, rep=FALSE,prob=NULL)
    #buscar vecinos
    sel.edges<-remove.edges(network.edges,selected)$edges.removed
    tovac<-0
    for(i in selected){
        if(sum(sel.edges==i)>0){
            tovacedge<-sel.edges[rowSums(sel.edges==i)==1,]
            if(is.vector(tovacedge))tovacedge<-t(as.matrix(tovacedge))
            tovacedge[tovacedge[,1]!=i,]<-tovacedge[tovacedge[,1]!=i,2:1]
            node.tovac<-resample(tovacedge[!is.element(tovacedge[,2],tovac),2],1)
            tovac<-c(tovac,node.tovac)
        if(!is.element(node.tovac,selected)||i>node.tovac) sel.edges<-
remove.edges(sel.edges,node.tovac)$network.edges #elimino al que se vacuna,
si no es un selected que todavia este en el juego
        }
    }
    indexes[tovac[-1],2]<-4
    network.edges<-remove.edges(network.edges,indexes[indexes[,2]==4,3])$network.edges
}
#selection of infected and infectives:
infectives<-0
if (sum(ini.infective)>0){
    for (j in 1:n){
        if(ini.infective[j]>0) infectives<-
c(infectives,resample(indexes[indexes[,1]==j &indexes[,2]==1,3],
ini.infective[j]))
    }
    if(length(infectives)>1){
        indexes[infectives[-1],2]<-3
        if(!is.null(ring.p) & !is.null(ring.m)) indexes[infectives[-1],10]<-ring.m
        if(!is.null(isol.p) & !is.null(isol.m)) indexes[infectives[-1],11]<-isol.m
    }
}
#******************Initial and constant values**********************
p.hist<-p<-as.vector(table(c(indexes[is.element(indexes[,2],c(1,2,3)),1],1:n))-1)
#susc, infected and infective
susp.hist<-susp<-p-ini.infective
infectivep.hist<-infectivep<-ini.infective
vacp.hist<-vacp<-as.vector(table(c(indexes[indexes[,2]==4,1],1:n))-1)
remp.hist<-remp<-isolp.hist<-isolp<-rep(0,n)
popp<-rbind(p,susp,infectivep,remp,vacp,isolp)
time.hist<-event.time<-0
#NODES*****************
#nodes.hist<-as.data.frame(cbind(NodeId=1:sum(p),Label=0,StartTime=0,Status=indexes[,2]
))
```

```
#**********************
event.time<-0
more.events<-T       #flag that turns false when no change of network and no
more events are possible
while(event.time<=obs.time & more.events){
        susceptibles<-indexes[indexes[,2]==1,3]         #the Id of active susceptibles
        infectives<-indexes[indexes[,2]==3,3]           #the Id of active infectives
        removed<-indexes[indexes[,2]==0,3]
       # browser()
        if(!is.null(ring.p) & !is.null(ring.m)){
            infectivestoring<-intersect(infectives,which(indexes[,8]!=5))
            removedtoring<-indexes[indexes[,2]==0 & indexes[,8]!=5,3]
        }
        if(!is.null(isol.p) & !is.null(isol.m)){
            infectivestoisol<-intersect(infectives,which(indexes[,8]!=6)) #infective
who failed to be detected before
            removedtoisol<-indexes[indexes[,2]==0 & indexes[,8]!=6,3]
        }
        if(is.vector(network.edges)) network.edges<-t(as.matrix(network.edges))
        #--------Time Events--------
        time.af<-1
        event.time<-time.af+sum(time.hist)
        #---------------------------
        if (!is.finite(time.af) || sum(popp[1,])==0) {
            #print("no new event is possible")
            more.events<-F
        }
        else{
            time.hist<-cbind(time.hist, time.af)        #the historical times between
consecutive events
                #cat("*** transmission \n")
                eff.connections<-remove.edges(network.edges,infectives)$edges.removed
                if(!is.null(eff.connections) && length(eff.connections)>0){
                    if(is.vector(eff.connections)) eff.connections<-t(matrix(eff.connections))
                    #browser()
                    transmit<-
rowSums(matrix(is.element(eff.connections,infectives),dim(eff.connections)))
#considerar solo conexiones entre infective and susceptible
                    transmit<-transmit*(runif(length(transmit))<=transm.tau)
                    if(sum(transmit==1)>0){
                        new.infectives<-
unique(as.vector(eff.connections[which(transmit==1),])[!is.element(eff.connecti
ons[which(transmit==1),],infectives)])
                        indexes[new.infectives,2]<-3  #new infectives
                    }
                    if(sum(transmit==1)==0) more.events<-F
                }
                else more.events<-F
                #removal
                network.edges<-remove.edges(network.edges,infectives)$network.edges
                indexes[infectives,2]<-0
                popp[1,]<-as.vector(table(c(indexes[is.element(indexes[,2],c(1,3)),1],1:n))-1)
#p
                popp[2,]<-as.vector(table(c(indexes[indexes[,2]==1,1],1:n))-1)
#susceptibles
                popp[3,]<-as.vector(table(c(indexes[indexes[,2]==3,1],1:n))-1)
#infectives
                popp[4,]<-as.vector(table(c(indexes[indexes[,2]==0,1],1:n))-1)
#removed
```

```
            p.hist<-rbind(p.hist,popp[1,])
            susp.hist<-rbind(susp.hist,popp[2,])
            infectivep.hist<-rbind(infectivep.hist,popp[3,])
            remp.hist<-rbind(remp.hist,popp[4,])
            vacp.hist<-rbind(vacp.hist,popp[5,])
            isolp.hist<-rbind(isolp.hist,popp[6,])
        }               #end case when time.af is finite
}                       #end while
dimnames(network.edges)<-NULL               #to eliminate headers
time.hist <- as.vector(time.hist)
dimnames(p.hist) <- dimnames(susp.hist) <-dimnames(infectivep.hist) <-
dimnames(remp.hist) <- NULL
list(indexes=indexes,time.hist=time.hist,p.hist=p.hist,susp.hist=susp.hist,infectivep.hist
=infectivep.hist,remp.hist=remp.hist,vacp.hist=vacp.hist,isolp.hist=isolp.hist, n=n)}
#- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -#

```

References

Abbey, H. (1951). Examination of Reed-Frost epidemic equations. *Human Biology*, 24:201–233. 10

Aiello, W., Chung, F. R. K., and Lu, L. (2001). A random graph model for power law graphs. *Experimental Mathematics*, 10:53–66. 142

Albert, R., Jeong, H., and Barabási, A.-L. (1999). Diameter of the world-wide web. *Nature*, 400:130–131. 24

Amaral, L., Scala, A., Barthélémy, M., and Stanley, H. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America (PNAS)*, 97(21):11149–11152. 23

Ancel-Meyers, L. (2007). Contact network epidemiology: Bond percolation applied to infectious disease prediction and control. *Bulletin of the American Mathematical Society*, 44:63–86. 17

Ancel-Meyers, L., Newman, M., Martin, M., and Schrang, S. (2003). Applying network theory epidemics: Control measures for outbreaks of *Mycoplama pneumoniae*. Technical report, Santa Fe Institute. xv, 3, 33, 44, 45

Anderson, R. and May, R. M. (1991). *Infectious Diseases of Humans: Dynamics and Control*. Oxford Science Publications. 6, 9, 17, 26, 67

Bailey, N. T. J. (1950). A simple stochastic epidemic. *Biometrika*, 37:193–202. 11, 83, 88, 89, 95

Bailey, N. T. J. (1975). *The Mathematical Theory of Infectious Diseases*. Charles Griffin & Company LTD, second edition. 9, 11, 26, 85

Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286:509–512. 24, 25, 135

Bartlett, M. (1946). Stochastic processes. Course Notes at the Univ. of North Carolina. 10

Bartlett, M. S. (1949). Some evolutionary stochatic processes. *Journal of the Royal Statistical Society. Series B*, 11:211–229. 10, 18

Bartlett, M. S. (1953). Stochastic process of the statistics of change. *Applied Statistics.*, 2:44–64. 11

Bartlett, M. S. (1955). *An Introduction to Stochastic Processes*. Cambridge University Press. 18

Bartlett, M. S. (1956). Deterministic and stochastic model for recurrent epidemics. In Neyman, J., editor, *Third Berkeley Symposium on Mathematical Statistics and Probability*, volume IV, pages 81 – 109, CA. University of California Press. 11

Bartlett, M. S. (1957). Measles periodicity and community size. *Journal of the Royal Statistical Society. Series A (General)*, 120(1):48–70. 11

Bean, B., Moore, B., Peterson, L., Gerding, D., and Balfour, H. (1982). Survival of influenza viruses on envirnmental surfaces. *Journal of Infectious Diseases*, 146:47–51. 107

Bender-deMoll, S. and McFarland, D. (2003). SoNIA: Social network image animator. 152, 156

Bernoulli, D. (1760). Essai d'une nouvelle analyse de la mortalité causée par la petite vérole et des advantages de l'inoculation pour la prévenir. *Mém. Math. Phys. Acad. Roy. Sci.*, pages 1–45. Paris. 7

Black, F. (1984). Chapter 17: Measles. In: *Viral Infections of Humans: Epidemiology and Control.* Ed. Evans A. S. Plenum Medical. 85

Bollobás, B. (1985). *Random Graphs.* Academic Press, New York. 21

Bowman, A. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: the Kernel Approach With S-Plus Illustrations.* Oxford University Press. 94

Brownlee, J. (1906). Statistical studies in immunity: The theory of an epidemic. *Proceedings Royal Society Edinburgh*, 26:484–521. 7

Butler, N. A. (2001). Optimal and orthogonal latin hypercube designs for computer experiments. *Biometrika*, 88(3):847–857. 95

Call, S. A., Vollenweider, M. A., Hornung, C. A., Simel, D. L., and McKinney, W. P. (2005). Does this patient have influenza? *Journal of the American Medical Association*, 293:987–997. xiii, 109, 110, 120, 125, 128, 129

Callaway, D., Newman, M., Strogatz, S., and Watts, D. (2000). Network robustness and fragility: Percolation on random graphs. *Physical Review Letters*, 85(25):5468–5471. 67

Care, D., Dunn, F., Robinson, R., Jensen, K., and Martin, J. (1958). Community-wide epidemic of asian strain influenza: clinical and subclinical illnesses among school children. *J. Amer. Med. Assoc.*, 167:1459–63. 108

Cliff, A., Haggett, P., and Ord, J. (1986). *Spatial Aspects of Influenza Epidemics.* Pion Limited. 95, 105, 106, 107, 108

Cohen, R., Erez, K., ben Avraham, D., and Havlin, S. (2000). Resilience of the internet to random breakdowns. *Physical Review Letters*, 85(21):4626–8. 67

Cohen, R., Havlin, S., and ben Avraham (2003). Efficient immunization strategies for computer networks and populations. *Physical Review Letters*, 91(24). 247901. 67, 68, 69, 74, 102

Davis, L., Caldwell, G., Lynch, R., Bailey, R., and Cin, T. (1970). Hong Kong influenza: the epidemiologic features of high school family study analyzed and compared with a similar study during the 1957 Asian influenza epidemic. *American Journal of Epidemiology*, 92:240–7. 108

de Nooy, W., Mrvar, A., and Batagelj, V. (2005). *Exploratory social network analysis with pajek*. Cambridge University Press. 134

Diekmann, O. and Heesterbeek, J. A. P. (2000). *Mathematical Epidemilogy of Infectious Diseases: Model Building, Analysis and Interpretation*. Wiley Series in Mathematical and Computational Biology. 26

Erdős, P. and Rényi, A. (1959). On random graphs. *Publicationes Mathematicae*, 6:290–297. 20, 21

Erdős, P. and Rényi, A. (1960). On the evolution of random graphs. *Publications of the Mathematical Institute of the Hungarian Academy of Science*, 5:17–61. 20, 21

Erdős, P. and Rényi, A. (1961). On the strength of connectedness of random graph. *Acta Mathematica Scientia Hungary*, 12:216–267. 20

Evans, G. H. (1875). Some arithmetical considerations of the progress of epidemis. *Transactions epidemiology society 1873-75*, 551. 7

Faloutsos, M., Faloutsos, P., and Faloutsos, C. (1999). On power-law relationship of the internet topology. *Computer Comunication Review*, 29:251–263. 24

Farr, W. (1840). Progress of epidemics. *Second Report of the Registrar General of England and Wales*, pages 91–98. 7

Ferrer i Cancho, R. and Solé, R. V. (2001). The small world of human languages. *Proc. Roy. Soc. London B*, 268:2261–2265. 24

Fine, P. (1981). Influenza Models: Prospects for Development and use. In: *Applications of mathematical models to the epidemiology of influenza: a critique*. Ed. Shelby, P. Sandoz Inst. for Health and Socio-economic studies. 108

Gkantsidis, C., Mihail, M., and Zegura, E. (2003). The Markov chain simulation method for generating connected power law random graphs. In *5th Workshop on Algorithm Engineering and Experiments (ALENEX), SIAM*, pages 16–25. 142, 143, 145, 150

Greenwood, M. (1931). On the statistical measure of infectiousness. *Journal of Hygiene (Cambridge)*, 31:336–351. 10

Guo, Y., Jin, F., Wang, M., and Zhu, J. (1983). Isolation of influenza C virus from pigs and experimental infection of pigs with influenza C virus. *Journal of General Virology*, 64:177. 106

Hakimi, S. (1962). On the realizability of a set of integers as degrees of the vertices of a graph. *SIAM Journal on Applied Mathematics (SIAP)*, 10. 143

Hamer, W. H. (1906). Epidemic disease in England. *Lancet*, I:733–739. 8, 9

Havel, V. (1955). A remark on the existence of finite graphs. *Caposis Pest. Matematika*, 80. 143

Hayfield, T. and Racine, J. S. (2008). *np: Nonparametric kernel smoothing methods for mixed datatypes*. R package version 0.14-2. 94, 116

Hope-Simpson, R. E. (1979). Epidemic mechanisms of type A influenza. *Journal of Hygiene (now Epidemiology and Infection)*, 83:11–26. 107

Hope-Simpson, R. E. (1992). *The Transmission of Epidemic Influenza*. Plenum Press. 107

Iionides, E. L. (2005). Maximum smoothed likelihood estimation. *Statistica Sinica*, 15:1003–1014. 95

Isham, V. (2004). Stochastic models for epidemics. Research report, University College London. 10

Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., and Barabási, A. (2000). The large-scale organization of metabolic networks. *Nature*, 407:651–654. 24

Kannan, R., Tetali, P., and Vempala, S. (1999). Simple Markov chain algorithms for generating bipartite graphs and tournaments. *Random Structures and Algorithms*, 14:293–308. 150

Kendall, D. G. (1956). Deterministic and stochastic epidemics in closed populations. In: *Proc. of the 3rd Berkeley Symp. Math. Stat. and Prob.* Ed. Neyman, J. University of California Press. 19

Kermack, W. O. and McKendrick, A. G. (1927, 1932, 1933). Contributions to the mathematical theory of epidemics. *Proceedings of the Royal Society of London A*, 115, 138, 141:700–721, 55–83, 94–112. 9, 14

Kermack, W. O. and McKendrick, A. G. (1937, 1939). Contributions to the mathematical theory of epidemics. *Journal of Hygiene (Cambridge)*, 37, 39:172–187, 271–288. 9

Liljeros, F., Edling, C. R., Amaral, L. A. N., Stanley, H. E., and Åberg, Y. (2001). The web of human sexual contacts. *Nature*, 411:907–908. 2, 24

Maslov, S. and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science*, 3(296):910–913. 150

McKendrick, A. G. (1926). Applications of mathematics to medical problems. *Proceedings of the Edinburgh Mathematical Society*, 14:98–130. 9, 18

McLeish, D. (2005). *Monte Carlo Simulation and Finance*. Wiley Finance. 83

Mihail, M. and Vishnoi, N. (2002). On generating graphs with prescribed vertex degrees for complex network modeling. Position Paper, Approx. and Randomized Algorithms for Communication Networks (ARACNE). 142

Milgram, S. (1967). The small world phenomenon. *Psychology Today*, 2:60–67. 23

Mills, C. E., Robins, J. M., and Lipsitch, M. (2004). Transmissibility of 1918 pandemic influenza. *Nature*, 432:904–906. 117

Molloy, M. and Reed, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures and Algorithms*, 6(2-3):161–180. 144

NACI (2000). Statement on influenza vaccination for the 1999-2000 season. Technical report, National Advisory Committe on Immunization; Canada Comunicable Disease Report. (ACS-2). 111

Newman, M., Barabási, A.-L., and Watts, D. J. (2006). *The Structure and Dynamics of Networks*. Princeton. 23

Newman, M. E. J. (2002). Spread of epidemic disease on networks. *Physical Review, E*, 66:016128. v, 3, 4, 26, 27, 28, 31, 33, 35, 40, 41, 45

Newman, M. E. J. (2003). Ego-centered networks in the ripple effect. *Social Networks*, 25:83–95. 22

Newman, M. E. J., Strogatz, S. H., and Watts, D. J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64:026118. 21, 22, 150

Pastor-Satorras, R. and Vespignani, A. (2001). Immunization of complex networks. *Physical Review.*, 65:036104. 33, 67

Pastor-Satorras, R. and Vespignani, A. (2003). Chapter 5: Epidemics and immunization in scale-free networks. In: *Handbook of Graphs and Networks: From the Genome to the Internet.* Ed. Bornholdt, S. and Georg Schuster, Heinz. Wiley-VCH. 33

PHAC. Public Health Agency of Canada, http://www.phac-aspc.gc.ca/tmp-pmv/info/. 10, 84, 85, 95

PHD (2004). *A guide to the control of respiratory infection outbreaks in long-term care homes*. Public Health Division and Long-Term Care Homes Branch. Ministry of Health and LTC, Ontario. 110, 111, 120

R Development Core Team (2007). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. 4, 5, 141

Ross, R. (1911). *The Prevention of Malaria*. Murray, London, second edition. 8

Ross, R. (1916). An application of the theory of probabilities to the study of *a priori* pathometry. *Proceedings Royal Society of London A.*, 92:204–230. Part I. 7

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4:409–435. 95

Scott, J. (2000). *Social network analysis*. SAGE Publications. 134

Serfling, R. E. (1952). Historical review of epidemic theory. *Human Biology*, 24(3):145–166. 9, 10

Shortridge, K. (1997). The influenza conundrum. *Journal of Medical Microbiology*, 46:813–815. 107

Smith, F. I. and Palase, P. (1989). Chapter 7: Variation in influenza virus genes. In: *The Influenza Viruses*. Ed. Krug, R. M. Plenum Press. 106

Solomonoff, R. and Rapoport, A. (1951). Connectivity of random nets. *Bulletin of Mathematical Biophysics*, 13:107–117. 21

Soper, H. E. (1929). Interpretation of periodicity in disease-prevalence. *Journal of the Royal Statistical Society*, 92:34–73. 9

Stuart-Harris, S. C. H., Schild, G. C., and Oxford, J. S. (1985). *Influenza: The Viruses and the Disease*. Edward Arnold, second edition. 105

Takeuchi, F. and Yamamoto, K. (2006). Effectiveness of realistic vaccination strategies for contact networks of various degree distributions. *Journal of Theoretical Biology*, 243:39–47. 69, 102

Vynnycky, E., Trindall, A., and Mangtani, P. (2007). Damping of epidemic waves. *International Journal of Epidemiology*, 4:881–889. 117

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press. 134, 136

Wasserman, S., Scott, J., and Carrington, P. J. (2005). Chapter 1: Introduction. In: *Models and methods in social network analysis*. Ed. Carrington, P. J., Scott, J. and Wasserman, S. Cambridge University Press. 22

Watts, D. J. (1999). *Small Worlds*. Princeton University Press. 21

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of 'small-world' netwoks. *Nature*, 393:440–442. 21, 23

Welch, W. J., Buck, R. J., Sacks, J., Wynn, H. P., Mitchell, T. J., and Morris, M. D. (1992). Screening, predicting, and computer experiments. *Technometrics*, 34(1):15–25. 95

Wharton, S. A., Weis, W., Skehel, J. J., and Wiley, D. C. (1989). Chapter 3: Structure, function, and antigenicity of the hemagglutinin of influenza virus. In: *The Influenza Viruses.* Ed. Krug, R. M. Plenum Press. 110

Whittle, P. (1955). The outcome of a stochastic epidemic - a note on Bailey's paper. *Biometrika*, 42:116–122. 11

WHO (1952). Epidemiological and vital statistics report. Technical Report 332, World Health Organization. 84

Wilson, E. B., Bennett, C., Allen, M., and Worcester, J. (1939). Measles and scarlet fever in providence, R. I. 1929-34 with respect to age and size of family. *Proceedings of the American Philosophical Society*, 80(3):357–476. 85, 90, 91, 92

Wilson, E. B. and Worcester, J. (1945). Damping of epidemic waves. *The Proceedings of the National Academy of Sciences*, 31:294–298. 9

Yan, P. (2007). Chapter 10: Distribution theory, stochastic processes and infectious disease modelling. In: *Lecture Notes in Mathematical Epidemiology* Ed. Brauer, F; Van den Driessche, P. and Wu, J. Springer-Verlag. 13, 99

Zanette, D. and Kuperman, M. (2002). Effects of immunization on small-world epidemics. *Physica A*, 309:445–452. 67

Glossary

Antibodies Body proteins produced in response to exposure to an antigen foreign substance; antibodies neutralize antigens and render them harmless.

Antigen Any molecule that is recognized by the immune system and that triggers an immune response, such as relase of antibodies.

Antiviral Drugs that inhibit either the life cycle of replication of viruses, resulting in deceasing the severity and duration of a viral infection.

Agent In epidemiology, the cause of a disease; in infectious disease often the agent is a microbe such as a virus or bacterium.

Basic reproductive rate (number) Denoted by $\mathcal{R}_0$, it is the expected number of new infected individuals produced by a single infected case who is in contact only with susceptible individuals.

Direct transmission The spread of infection through individual-to-individual contact.

Epidemic The occurrence in a community or region of cases of an illness (or an outbreak) clearly in excess of expectancy. The term epidemic refers not only to infectious diseases.

Endemic Description of a disease or infectious agent that is habitually present in a community, geographic area, or population group. Often an endemic disease maintains a low but continuous incidence.

Final outbreak size Is defined as the number of initially susceptible individuals that ultimately become infected.

Isolate In microbiology, to obtain a pure strain from a source such as a clinical specimen that may have been part of a mixed primary culture.

Herd immunity The resistance to a specific infectious disease of an entire community due to the immunity developed in a large proportion of individuals in that community. When an individual becomes immune to the disease, prevents the internal replication of the infectious agent so it can be transmitted to any other susceptible.

Host An individual who permits lodgment of an infectious disease agent under natural conditions.

Incidence rate The frequency or new cases or occurrence of some event per individual during a period of time.

Incubation period The period of time in which an individual is infected but asymptomatic. During this period the individual may be infective.

Indirect transmission The agent passes through one or more species of intermediate host in order to complete its life cycle.

Infectivity The capacity of an agent to enter and multiply in a susceptible host and thus produce infection.

Infectious contact A contact between a susceptible and an infective individual in which the transmission of the infection takes place.

Infectious period The period in which an infected individual is referred as infective and is able to pass on the infection to susceptibles

Latent period The period of time in which an individual is infected but not yet infective.

Morbidity The occurrence of an illness or illnesses in a population.

Mortality The occurrence of a death in a population.

Optimal immunization programme The optimal immunization schedule as that which minimizes the overall fraction which must be immunized to eradicate infection.

Period prevalence The number of cases of illness during a time period divided by the average size of the population.

Point prevalence The number of cases of illness in a group of population at a point in time divided by the total number of persons in that group of population.

Prevalence The number of existing cases of a disease or health condition in a population at some designated time.

Proband The individual in a family who brings a disease to the attention of the investigator. In a case family, the proband is likely to be the person who is affected with the disease of interest.

Replacement number Denoted by $\mathcal{R}$, it is the expected number of new infected individuals produced during an outbreak by a typical infected case.

Resistance The ability of an agent to survive adverse environmental conditions.

Surveillance The systematic collection, analysis, interpretation, dissemination and consolidation of data pertaining to the occurrence of a specific disease.

Vaccine A specific substance that elicits an immune response to prevent infection by a foreign agent.

Vector In epidemiology, the intermediate specie or species than can pass the agent to the population of interest when indirect transmission of the agent is possible.

Virus One of a group of submicroscopic infectious agents.

Index