Stockholms universitet

SIR epidemic on a configuration model network

Marielle Andersson

Kandidatuppsats i matematisk statistik
Bachelor Thesis in Mathematical Statistics

Kandidatuppsats 2013:1
Matematisk statistik
Juni 2013

www.math.su.se

Matematisk statistik
Matematiska institutionen
Stockholms universitet
106 91 Stockholm

Stockholm University

Mathematical Statistics
Stockholm University
Bachelor Thesis **2013:1**
http://www.math.su.se

SIR epidemic on a configuration model network

Marielle Andersson*

June 2013

Abstract

In this thesis we study Susceptible-Infectious-Removed epidemics on configuration model networks. Networks are used in different kinds of studies, such as the study of the internet, social networks and biological networks, as a simplified model of the real world. We look at a closed population without births, deaths and migration. On that population we look at an SIR epidemic, which divides the population into three different states: susceptible, infectious and removed. Those who are susceptible can be infected if they are in contact with an infectious individual. Those who are infected make contacts at a fixed rate, then they recover and becomes immune or die from the disease. How a disease spreads through the population depends strongly on the connections that occur between infectious and susceptible individuals. By constucting a configuration model network it is possible to investigate when the epidemic may become large and when it will stay small with probability one and how the distribution of the infectious period affects the outbreak. We use generating functions and percolation theory to answer these questions. The early stages of an epidemic outbreak can be approximated by a branching process, we see that this approximation is possible until approximately the $\sqrt{n}$th infection in a population that consist of n individuals. We also show that an epidemic outbreak is possible when the expected number of transmission causing contacts an infectious individual has is above one.

*Postal address: Mathematical Statistics, Stockholm University, SE-106 91, Sweden. E-mail:maan0129@gmail.com . Supervisor: Pieter Trapman.

Abstract

In this thesis we study Susceptible-Infectious-Removed epidemics on configuration model networks. Networks are used in different kinds of studies, such as the study of the internet, social networks and biological networks, as a simplified model of the real world. We look at a closed population without births, deaths and migration. On that population we look at an SIR epidemic, which divides the population into three different states: susceptible, infectious and removed. Those who are susceptible can be infected if they are in contact with an infectious individual. Those who are infected make contacts at a fixed rate, then they recover and becomes immune or die from the disease. How a disease spreads through the population depends strongly on the connections that occur between infectious and susceptible individuals. By constucting a configuration model network it is possible to investigate when the epidemic may become large and when it will stay small with probability one and how the distribution of the infectious period affects the outbreak. We use generating functions and percolation theory to answer these questions. The early stages of an epidemic outbreak can be approximated by a branching process, we see that this approximation is possible until approximately the $\sqrt{n}$th infection in a population that consist of n individuals. We also show that an epidemic outbreak is possible when the expected number of transmission causing contacts an infectious individual has is above one.

Preface and acknowledgements

This is a thesis of 15 ECTS, which will lead to a Bachelor's Degree in Mathematical Statistics at the Department of Mathematics at Stockholm University.

I would like to thank my supervisor Pieter Trapman, Department of Mathematics at Stockholm University, for introducing me to this interesting topic and for all the help and guidance through this thesis.

Contents

1 Introduction 5

2 Networks 5
- 2.1 Random graphs . . . 6
  - 2.1.1 Giant component . . . 7
  - 2.1.2 Small component . . . 8
- 2.2 The configuration model . . . 8
- 2.3 Percolation . . . 10

3 Susceptible-Infectious-Removed epidemics 10
- 3.1 The basic reproduction number . . . 12

4 SIR epidemic on a configuration model network 12
- 4.1 Transmissibility . . . 13
- 4.2 Outbreak of a disease . . . 14
  - 4.2.1 The generating function . . . 14
  - 4.2.2 Branching processes . . . 17
  - 4.2.3 Application of the birthday problem . . . 18
- 4.3 Distribution of the outbreak size . . . 20
- 4.4 Outbreak sizes and epidemic threshold . . . 22
- 4.5 Degree of infected individuals . . . 23

5 Conclusions and Discussion 24

References 25

A The degree distribution of an Erdös-Rényi graph 26

B Random infectious periods result in overestimate with bond percolation 26

1 Introduction

A network is a collection of points connected pairwise by lines. The points in the network are called vertices and the lines are called edges. Several different fields of science use networks as a model for connections, it is common to use them in the study of the internet, neural networks or social networks. A social network is a network of people where the individuals are represented by the vertices and the edges represent some kind of relationships between the individuals.

In this thesis we study Susceptible-Infectious-Removed (SIR) epidemics in a large population. The population is divided into three subgroups: susceptible, infectious and removed. An infected individual can transmit the disease to susceptible individuals before she recovers and become immune or dies from the disease. The population is closed, this means that there are no births, death or migration in the population. The patterns of the spread of the disease are similar to the connection patterns between susceptible and infective individuals. A contact network is a social network where the edges represent possible contacts between individuals. During a day an individual may be in contact with different individuals: family, friends, colleagues and so forth, these people make up this individual's possible connections. There is no guarantee that a connected individual will have a transmission causing contact with the infected individual. The degree of a vertex is the number of edges attached to it, i.e. the number of connections an individual has. We study a configuration model network, that is a stochastic model network with arbitrary degree distribution. We have n individuals in our model and assign independent and identically distributed number of connections to these individuals. We choose one graph with this degree distribution uniformly and consider an SIR epidemic on it. Two vertices that are connected by an edge are called neighbors. Two neighboring vertices make contact at a fixed rate. Those contacting processes are independent for different pairs of vertices. If an infective individual contacts a susceptible individual, the disease will be transmitted and the susceptible individual becomes infectious. The infectious period for an individual is independent of each other individual's infectious periods and is distributed as the random variable τ.

We use a configuration model network in this thesis to examine when an SIR epidemic might become large and when it will stay small with probability one. We also look at after how many infections approximations of the epidemic will become unrealistic and how the distribution of the infection length will affect the outbreak.

2 Networks

In this thesis we consider large networks where the number of vertices is denoted by n and the number of edges is denoted by m. Two vertices may have more than one edge connecting them, these are called parallel edges or multiedges. Edges that connect a vertex to itself are called self-loops [5]. A common terminology for networks is graphs,

or multigraphs in the case when we allow parallel edges.

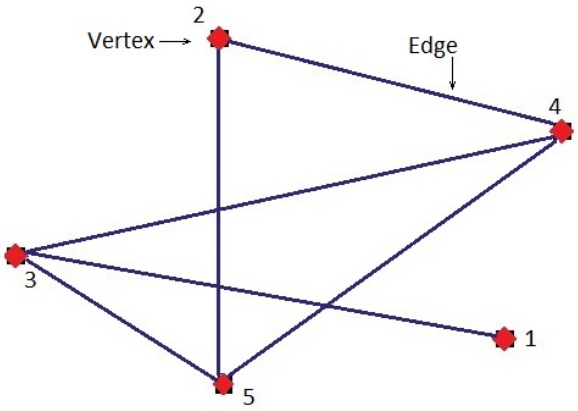

Figure 1: A random network with 5 vertices and 6 edges.

The number of edges connected to a vertex is called the degree of the vertex. For example if a certain vertex i has three edges attached, then vertex i has degree three. In Figure 1 we have a network with five vertices and six edges. One vertex has degree one, one vertex has degree two and the other three have degree three. A degree sequence is a set of degrees for all the vertices $\{k_1, k_2, ...\}$ and in this case the degree sequence is $\{1, 2, 3, 3, 3\}$.

A contact network is a social network where the vertices represent individuals, the edges represent connections or possible contacts between two individuals and the degree of a vertex is the number of connections that individual has. Two individuals that are connected are neighbors of each other. When we study epidemics, the connecting edges represent the possible paths a disease can be transmitted along.

A graph can be either directed or undirected. If we have a pair of vertices (v_i, v_j) and the edge between v_i and v_j is identical to the edge between v_j and v_i, for all i and j, we have an undirected graph. In the case of an epidemic process this means that we have a pair of individuals i and j, if the graph is undirected then i can infect j if and only if j can infect i, for all individuals i and j. If this is not true we have a directed graph.

2.1 Random graphs

If a model network is specified by specific set of parameters and the network is random in other respects we have a random graph. A common way to construct a random graph is by keeping the number of vertices, n, and the probability of an edge between vertices, p, fixed. The expected number of connections of a vertex we denote by z and this means that $p = \frac{z}{n-1} \approx \frac{z}{n}$ for large n. We let p_k denote the probability that a vertex has degree k, if we estimate these probabilities from Figure 1 we get: $p_1 = \frac{1}{5},\ p_2 = \frac{1}{5},\ p_3 = \frac{3}{5}$, and

$p_k = 0$ for all other values of k. The probability that a vertex is connected to k of the vertices and not to any of the other $n - k$ vertices is given by

$$p_k = \binom{n}{k} p^k (1-p)^{n-k} \approx \frac{z^k e^{-z}}{k!}. \tag{1}$$

That is, for large n, the number of neighbors has a Poisson distribution, we show the details of the calculations behind equation (1) in Appendix A. This is also called the Erdös-Rényi model. In many cases this model does not give a good approximation of real-world phenomena, particularly because of the degree distribution. The degree distribution in most real-world networks is right skewed, this means that there are many vertices with low degrees and a relatively many with high degrees [4], thus the degree distribution is a power-law distribution of the form:

$$p_k = \Theta(k^{-\alpha})$$

i.e.

$$0 < \liminf_{k\to\infty} \frac{p_k}{k^{-\alpha}} \leq \limsup_{k\to\infty} \frac{p_k}{k^{-\alpha}} < \infty$$

where α is constant. In words this means that the lower bound of $\frac{p_k}{k^{-\alpha}}$ is larger than zero while the upper bound is finite when k goes to infinity.

2.1.1 Giant component

When $p = 0$, there are no edges between any vertices in the network, so the largest component is of size one and is therefore independent of the number of vertices in the network. When $p = 1$ all vertices are connected and form a component of size n. The size of such a component grows with the size of the network. The largest component in the random graph differs for different values of p.

A giant component is a component that grows proportional to n. It is shown in [4] that for the Erdös-Rényi model the fraction of vertices in the giant component is a solution of $S = 1 - e^{-zS}$, where S denotes the size of the giant component as a proportion of the size of the network and z denotes the expected degree of a vertex.

In a graph there can only exist one giant component. This can be shown by first calculating the expected degree of a vertex, z, in a generated random graph. On this generated graph we then add some additional edges between vertices, that are not already connected, with a probability $p' = \frac{z}{(n-1)^{3/2}}$. It is shown on page 409 in [4] that the expected degree is the same after as it were before we added the additional edges when n is large. If there is more than one giant component in the random graph before we add the additional edges, we can choose two of them with sizes $S_1 n$ and $S_2 n$ respectively. There are $S_1 n * S_2 n = S_1 S_2 n^2$ possible pairs of vertices between the two giant components. With probability p' new edges are created. The probability that none of these

new edges is among the $S_1S_2n^2$ pairs of vertices that are connecting the two components is given by $q' = (1-p')^{S_1S_2n^2}$. We have

$$\ln(q') = S_1S_2n^2\ln(1-p') = S_1S_2n^2\ln(1-z(n-1)^{-3/2}) \approx -zS_1S_2\sqrt{n}$$

and thus $q' = e^{-zS_1S_2\sqrt{n}}$. This probability goes to zero as n grows large and the expected degrees stays the same, thus when $z > 1$ there can only be one giant component.

2.1.2 Small component

Small components are the components that do not belong to the giant component. If there exists one giant component, this component usually does not consist of all the vertices. In large networks most of the small components are trees without loops, a tree of s vertices contains $s-1$ edges. When we are looking at a small component of size s in the Erdös-Rényi graph, the probability of another edge, which would result in a loop, is $p = z/(n-1)$. An extra edge could be created in $\binom{s}{2} - (s-1) = 1/2(s-1)(s-2)$ ways. The expected number of extra edges in the component is $\frac{1}{2}(s-1)(s-2) * \frac{z}{n-1}$ which goes to zero as $n \to \infty$, hence the component is a tree without loops.

Not all small components are trees, but most of them are. We assume that the degree distribution has finite variance. The following theorem, given by Durrett on page 71 in [2], says that the number of self-loops and parallel edges in a network is Poisson distributed:

Theorem 1 Let $\mu = \sum_k kp_k$ and $\mu_2 = \sum_k k(k-1)p_k$. As $n \to \infty$, the number of self-loops χ_0 and the number of parallel edges χ_1 are asymptotically independent Poisson$(\mu_2/2\mu)$ and Poisson$((\mu_2/2\mu)^2)$.

In the Erdös-Rényi graph the degree distribution is a Poisson(λ) distribution. To calculate the expected number of self-loops and parallel edges in a Erdös Rényi graph we have $\mu = E[K] = \lambda$ and $\mu_2 = E[K^2] - E[K] = Var(K) + E[K]^2 - E[K] = \lambda + \lambda^2 - \lambda = \lambda^2$, where we have used that $Var(K) = E[K^2] - E[K]^2 = \lambda$. The expected number of self-loops χ_0 are asymptotically Poisson$(\lambda/2)$ and the number of parallel edges χ_1 are asymptotically Possion$(\lambda^2/4)$ as $n \to \infty$.

2.2 The configuration model

In this thesis we are interested in an epidemic spreading process and for this the Erdös-Rényi model is not appropriate. Instead we use a configuration model. The configuration model is a random graph with a general degree distribution. We consider a configuration model where the vertices have independent and identically distributed degrees. The probability that a randomly chosen vertex has degree k is p_k. From this degree distribution we uniformly draw a degree sequence $\{k_i\}$, which is a set of n degrees k_i, $i = 1, ..., n$, where k_i denotes the number of neighbors of vertex i. This sequence we use to generate a network by giving each vertex i a total of k_i "stubs". A stub is an end of an edge, in

total there are $2m = \sum_i k_i$ stubs. We create an edge by choosing two stubs uniformly at random and connecting them to one another. We do the same thing for the remaining $2m - 2$ stubs. For this to work we must have an even number of stubs. To get an even number we can condition on that $E_n = \{k_1 + ... + k_n \ is \ even\}$. We can do this because if the probability $P(E_1) \in (0, 1)$ then $P(E_n) \to 1/2$ when n goes to infinity and this will have a small effect.

To illustrate this, we look at Figure 1 again and assume that it is constructed as a configuration model. We have the degree sequence $\{1, 2, 3, 3, 3\}$, each of the five vertices is given k_i number of stubs, as shown on the left hand side of Figure 2. These stubs are paired at random and creating edges, this is illustrated by the dotted lines on the right hand side of the figure.

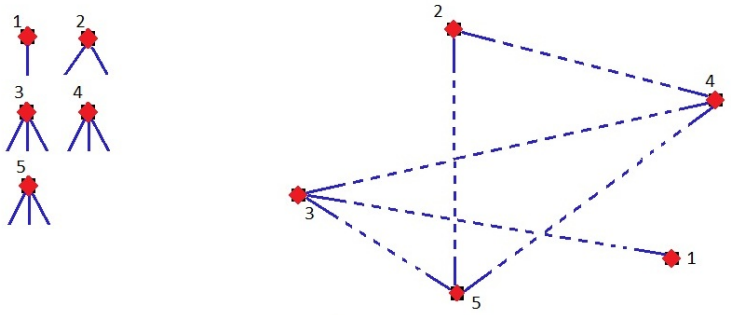

Figure 2: Construction of a configuration model with 5 vertices and 12 stubs. The stubs are on the left hand side, these are paired at random and create the graph on the right hand side.

Since each pair of stubs is equally likely to create an edge, the configuration model can have self-loops and parallel edges. As mentioned in the previous subsection, the number of self-loops and parallel edges is Poisson distributed with a parameter that asyptotically does not depend on n as n grows large. The number of self-loops and parallel edges in the network are therefore sparse if n is large, [2]. What kind of network we get depend on the exact degree distribution, specified by the p_k's. We either get a network with several small clusters or it may contain a giant component and some small components. In a social network where the edges represent connections between individuals, a person A can only be connected to person B if there is at least one connecting edge between them. They have to belong to the same component for it to be a connecting edge between them. A large fraction, S, of the individuals in the network can communicate with one another if there is a giant component in the network. In networks without a giant component there are only small components, so all communication takes place in small groups. The size of a small component has distribution P_s and the expected value $E[s]$ is its typical size.

2.3 Percolation

Percolation is used on networks to see what happens if we remove some vertices and/or edges. There are two kinds of percolation: bond percolation where edges are removed at random and site percolation where vertices are removed at random along with all the edges attached to them.

It is used to study different processes, in the case of transmission of a disease one can use percolation to look at the effects of vaccination or immunization. Diseases spread through the population as a network of connections. If one individual becomes immune against a specific disease, he does not affect the spread since he will not get infected and infect others. Such an individual can be removed from the network. By using site percolation we can randomly "occupy" vertices and calculate how many individuals that need to be vaccinated to avoid an epidemic outbreak. A cluster of occupied edges is called a percolation cluster.

3 Susceptible-Infectious-Removed epidemics

The susceptible-infectious-removed epidemic or SIR epidemic is an epidemic where the population is divided into three different states. An individual is susceptible if he does not have the disease but can be infected when he is in contact with an infectious individual. An individual who has the disease is in the infectious state and can transmit the disease to susceptible individuals. When individuals get infected they stay infected for independent infectious periods, which are distributed as the random variable τ. An individual who has been infected can either recover and become immune to the disease or die and thus make a transition to the removed state. We have a closed population with n individuals, where n is large. We let $S(t)$ denote the number of susceptible individuals at time t, $I(t)$ denote the number of infectious individuals at time t and $R(t)$ denote the removed part of the population at time t, thus $S(t) + I(t) + R(t) = n$.

$$S \Longrightarrow I \Longrightarrow R$$

When the disease is rapidly spreading and the survivors of the disease are immune to further infection of it, this model is appropriate. For example it is appropriate on diseases like influenza or measles.

We first assume that we have a fully mixed model, which means that an individual is equally likely to be in contact with any other individual in the population. An individual in the infectious state has contacts with individuals from any other state at an average rate β per time unit and leaves the infectious state in an average rate γ per time unit. If we let s, i and r denote the fraction of individuals in the state susceptible, infectious and removed respectively, we get the following differential equations, under the assumption that n goes to infinity:

$$\frac{\partial s}{\partial t} = -\beta is, \qquad \frac{\partial i}{\partial t} = \beta is - \gamma i, \qquad \frac{\partial r}{\partial t} = \gamma i. \tag{2}$$

It is necessary that $s+i+r = 1$. In the beginning of the epidemic development, at $t = 0$, the population consist of many susceptible individuals, a few infectious individuals and no yet recovered or deceased individuals. We denote the initial fraction of the population in the different states as s_0, i_0 and r_0 respectively.

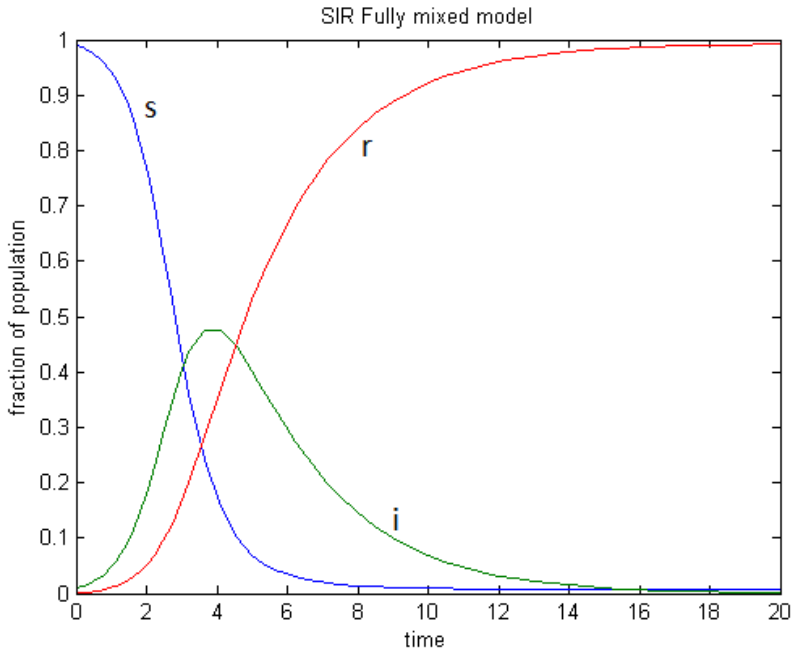

Figure 3: The fully mixed SIR epidemic model, where the lines s, i and r represent the fraction of individuals in the state susceptible, infectious and removed respectively. With inputs $s_0 = 0.99$, $i_0 = 0.01$, $r_0 = 0$, $\beta = 2$ and $\gamma = 0.4$.

In Figure 3 we can see how the SIR model developes over time. The fraction of susceptible decrease over time while the fraction of removed increase over time. The number of infected increase in the beginning and then decrease. As t goes to infinity, the fraction susceptibles does not go to zero. When t grows large, i goes towards zero and there are no infected individuals left who can transmit the disease to the remaining susceptibles. This also implies that the fraction removed does not reach one as $t \to \infty$. If we solve the differential equations (2) above, we get the expressions $s = s_0 e^{-\beta r/\gamma}$ and $\frac{dr}{dt} = \gamma(1 - r - s_0 e^{-\beta r/\gamma})$ [4]. When the population grows large and we initially have $s_0 \approx 1$, the fraction removed individuals is approximately $r = 1 - e^{-\beta r/\gamma}$ which does not reach one. This is the same equation as the one for the size S of the giant component in a Erdös-Rényi graph, we mention in the previous subsection that the size of the giant component is the solution to $S = 1 - e^{-zS}$, when $z = \beta/\gamma$.

3.1 The basic reproduction number

The basic reproduction number, R_0, is the average number of other individuals an infected individual in the early stage of the epidemic development will transmit the disease to. In the early stage there are only few individuals carrying the disease and most individuals are susceptible. When the expected number of people an infected individual will transmit the disease to is larger than one, the epidemic may spread, while when the expectation is less than one the epidemic will die out. Thus $R_0 = 1$ corresponds to the epidemic threshold, the point of transition between the epidemic and non-epidemic regimes.

We denote the time during which an individual is infected by ι. Given the average rate of recovering from the disease, γ, the probability of recovering in any infinitesimal time interval $\partial\iota$ is $\gamma\partial\iota$. The probability that an infectious individual is still infected after a time ι, is given by

$$\lim_{\partial\iota\to 0}(1-\gamma\partial\iota)^{\iota/\partial\iota} = e^{-\gamma\iota}.$$

The probability that an individual has been infected this long and then recovers in the time interval ι and $\iota + d\iota$ is $\gamma e^{-\gamma\iota}d\iota$. This means that the time an individual remains infected is exponentially distributed. The expected number of contacts this individual has during his infectious period is $\beta\iota$. In the early stages of the epidemic development every other individual he has contact with is in the susceptible stage. Hence $\beta\iota$ is also the expected number of people who will get infected by this individual. According to [4] we have

$$R_0 = \beta\gamma\int_0^\infty \iota e^{-\gamma\iota}d\iota = \frac{\beta}{\gamma} \tag{3}$$

and when the basic reproduction number is equal to one, it corresponds to the epidemic threshold. By putting $R_0 = 1$ into equation (3), we can solve it and the solution is $\gamma = \beta$. Thus when the average rate of recovering is larger than the contact rate, $\gamma > \beta$, an epidemic outbreak is not possible.

4 SIR epidemic on a configuration model network

In this section we will look at the SIR epidemic on a configuration model network. The assumption about a fully mixed model is not the real life case. It is not very likely that all individuals meet the same number of other individuals during the infectious time or that it is equally likely to meet everyone in the entire population. We therefore modify the fully mixed model by replacing it with a contact network. The vertices are individuals and all possible transmission causing contacts are represented by the edges between two vertices. The number of possible connections an individual i has is given by the degree of vertex i.

4.1 Transmissibility

The contact network is undirected and has an arbitrary degree distribution. On this contact network we consider an SIR epidemic. The contacting processes are independent for different pairs of vertices. If an infectious vertex contacts a susceptible then the susceptible becomes infectious as well, if an infectious individual contacts an infective or removed individual nothing happens. An infectious individual stays so for independent infectious periods, distributed as the random variable τ.

We want to know the probability of transmission, we therefore look at two randomly chosen individuals that are connected, one of them is infected i and one is susceptible j. The probability of transmission of the disease between these two individuals is denoted by T_{ij}. The average rate of disease causing contact we denote by r_{ij} and the infectious period for individual i by τ_i, where τ_i is random. The probability that j will be infected by i when we have continous time is

$$T_{ij} = \lim_{\partial t \to 0} 1 - (1 - r_{ij}\partial t)^{\tau_i/\partial t} = 1 - e^{-r_{ij}\tau_i}.$$

If we instead use discrete time, we set $\partial t = 1$ and get the expression

$$T_{ij} = 1 - (1 - r_{ij})^{\tau_i},$$

where τ_i is measured in time steps.

Since the values of r_{ij} and τ_i usually will vary between individuals, the probability of transmission also varies. We assume that the r_{ij} are independent and identically distributed (i.i.d.) and distributed with some arbitrary distribution P_r and that τ_i is one i.i.d. taken from some arbitrary distribution P_τ. The $r_{ij}{}'s$ and $\tau_i{}'s$ are independent of each other.

If an infected individual stays infective for some fixed infectious period, then T_{ij} is one i.i.d. random variable and the a priori probability of transmission is the average of T_{ij} over the distributions P_r and P_τ. Thus for the continous case

$$\hat{T} = E[T_{ij}] = 1 - \int_0^\infty dr d\tau P_r P_\tau e^{-r\tau} \tag{4}$$

and for the discrete case

$$\hat{T} = E[T_{ij}] = 1 - \int_0^\infty dr \sum_{\tau=0}^{\infty} P_r P_\tau (1 - r)^\tau. \tag{5}$$

We call $\hat{T}$ "transmissibility", $0 \leq \hat{T} \leq 1$ [6]. When we consider SIR epidemics and individuals are infective for fixed infectious periods, bond percolation can be used on the network. When the epidemic outbreak starts with one infected individual we can watch the spread of the disease across the network by occupying each edge along which the

disease can be transmitted. An individual is transmitted with a probability $\hat{T}$ and the component that consists of the vertices that are connected by occupied edges, including the initial infected individual, corresponds to the size of the outbreak.

If the infection period is random, we can not use bond percolation and the quantities investigated will change [2]. We show this later in the end of section 4.2.1. This is because when we have random infectious periods the transmission probabilities are not independent. For example, the probabilities T_{12} and T_{13} are both functions of the same τ_1 and hence not independent.

In the case when one infected individual either transmits the disease to all of its neighbors or to none, we have a situation that is approximately corresponding to site percolation. If an infected vertex transmits the disease to all of its neighbors, this vertex and all its edges get occupied. But since some of the vertices in the other end of an edge do not spread the disease any further, we can not occupy this edge, which is why the connection between the epidemic outbreak and site percolation is not exact.

4.2 Outbreak of a disease

In this thesis the focus is on large networks. When the population is large we can calculate several quantities of interest exactly. We assume that we initially have one infective individual which is chosen uniformly at random. The outbreak spreads from this initial individual through the population across the network. We can look at the development of the disease with fixed infectious periods as a model equivalent to bond percolation, with the transmission probability $\hat{T}$ as the probability of occupation. This means that if $\hat{T} = 1$ all edges in the network are occupied and thus form a giant component, while if $\hat{T} = 0$ no edges are occupied.

By occupying each edge along which the disease is transmitted, the set of vertices connected by these occupied edges represents the infected individuals. The infected individuals form a percolation cluster along the occupied edges. The size of the outbreak corresponds to the size of the cluster to which the initial vertex is attached.

Our graph is constructed as the configuration model, so the graph is chosen uniformly at random from the set of all graphs with the drawn degree sequence.

4.2.1 The generating function

To solve the average behaviour of graphs we use generating functions. The generating function is an alternative representation of a probability distribution. It is defined in [3] as

Definition 1 Let K be a nonnegative, integer-valued random variable. The (probability)

generating function of K is

$$g_K(x) = E[x^K] = \sum_{k=0}^{\infty} x^k P(K = k).$$

The generating function has many advantageous properties, we will use the folllowing theorems and corollaries in [3]

Theorem 2 Let $K_1, K_2, ..., K_n$ be independent, nonnegative, integer-valued random variables, and set $S_n = K_1 + K_2 + ... + K_n$. Then

$$g_{S_n}(x) = \prod_{h=1}^{n} g_{K_h}(x).$$

Corollary 2.1 If, in addition, $K_1, K_2, ..., K_n$ are identically distributed, then

$$g_{S_n}(x) = (g_K(x))^n$$

Theorem 3 Let K be a nonnegative, interger-valued random variable, and suppose that $E|K|^h < \infty$ for some $h = 1, 2, ...$ Then

$$E[K(K-1) * ... * (K-h+1)] = g_K^{(h)}(1),$$

where $g_K^{(h)}(1)$ stands for the hth derivative of the generating function $g_K(x)$ at $x = 1$.

Corollary 3.1. Let K be a nonnegative, integer-valued random variable, then
(a) $E[|K|] < \infty \Longrightarrow E[K] = g'_K(1)$, and
(b) $E[K^2] < \infty \Longrightarrow Var(K) = g''_K(1) + g'_K(1) - (g'_K(1))^2$.

The proof of Corollary 3.1 follows from Theorem 3, by inserting $h = 1$ and $h = 2$ we obtain
$h = 1 \Rightarrow g'_K(1) = E[K]$,
$h = 2 \Rightarrow g''_K(1) = E[K(K-1)] = E[K^2 - K] = E[K^2] - E[K]$.
From this we get $Var(K) = E[K^2] - E[K]^2 = g''_K(1) + g'_K(1) - (g'_K(1))^2$.

We denote the generating function for the degree distribution by $G_0(x)$. We let p_k be correctly normalized and denote the probability that a randomly chosen vertex has degree k. This function is defined as

$$G_0(x) = \sum_{k=0}^{\infty} p_k x^k, \tag{6}$$

where we have that $G_0(1) = 1$. Since the degree distribution is normalized and positive definite we have that $G_0(x)$ is absolutely convergent for all $|x| \leq 1$ [5].

If we instead follow a randomly chosen edge to the vertex in one of its ends, the vertex we arrive to has another degree distribution. This is because if the vertex has degree k, it is k times as likely to be chosen as a vertex with degree 1. The degree distribution of a vertex we reach by following a randomly chosen edge is proportional to kp_k, thus the normalized generating function of the degree of a vertex we arrive to by following a randomly chosen edge is

$$\frac{\sum_k kp_k x^k}{\sum_k kp_k} = \frac{\sum_k kp_k x^{k-1} x}{G_0'(1)} = x\frac{G_0'(x)}{G_0'(1)}. \tag{7}$$

In function (7) we include the edge from which we arrived, but we are interested in how many ways we can leave such a vertex. We therefore exclude the vertex through which we arrived, that is the degree minus one. If we divide equation (7) by x we achieve this new generating function

$$G_1(x) = \frac{G_0'(x)}{G_0'(1)} = \frac{1}{z}G_0'(x), \tag{8}$$

where z denotes the expected degree of a vertex, $z = E[K] = \sum_k kp_k = G_0'(1)$.

If we know the generating function we can calculate the probability p_k, since it is given by the kth derivative of G_0

$$p_k = \frac{1}{k!}\frac{\partial^k G_0}{\partial x^k}\bigg|_{x=0}.$$

We need two other generating functions to solve the bond percolation problem, those who correspond to the distribution of occupied edges attached to a vertex as a function of transmission. These generating functions we denote by $G_0(x, \hat{T})$ and $G_1(x; \hat{T})$, where $\hat{T}$ is the transmissibility for an individual when the infectious periods are fixed. The probability that a vertex has exactly m of k occupied edges is given by $\binom{k}{m}\hat{T}^m(1-\hat{T})^{k-m}$, there are $\binom{k}{m}$ different ways of chosing m edges out of k and the probability of m occupied edges and $k-m$ unoccupied edges is given by the other part of the expression. We thus have the generating function:

$$G_0(x;\hat{T}) = \sum_{m=0}^{\infty}\sum_{k=m}^{\infty} p_k \binom{k}{m}\hat{T}^m(1-\hat{T})^{k-m}x^m$$

$$= \sum_{k=0}^{\infty}\sum_{m=0}^{k} p_k \binom{k}{m}(x\hat{T})^m(1-\hat{T})^{k-m} = \sum_{k=0}^{\infty} p_k(1-\hat{T}+x\hat{T})^k = G_0(1+(x-1)\hat{T}). \tag{9}$$

By corresponding calculations we can get the expression for the number of occupied edges leaving a vertex we arrived at by following a randomly chosen edge

$$G_1(x;\hat{T}) = G_1(1+(x-1)\hat{T}). \tag{10}$$

If the infectious period is random the transmission probability is not independent and equations (9) and (10) are false. This can be shown by using Jensen's inequality [7]:

Proposition 1 (Jensen's inequality) If $f(x)$ is a convex function, then

$$E[f(X)] \geq f(E[X])$$

provided that the expectations exists and are finite.

When the contact rate, r, is independent and fixed while the infectious period is random we have transmissbility

$$T = 1 - \int_0^\infty dtP(\tau = dt)e^{-rt}. \tag{11}$$

By looking at the generating function of the number of neighbors that will become infected, where r_k is the probability of having k neighbors, Durrett shows that this is not the same as (9) and (10), [2]. The entire calculation is in Appendix B, but the main result is the following

$$G_0(x;T) = \int_0^\infty dtP(\tau = dt) \sum_{j=0}^\infty x^j \sum_{k=j}^\infty r_k \binom{k}{j} (1 - e^{-rt})^j e^{-r(k-j)t}$$

$$= ... = E[G(1 + (x-1)(1 - e^{-r\tau}))] > G(1 + (x-1)\hat{T}).$$

We have that $(1 - (1-x)(1-e^{-r\tau})) \in [0,1]$ and $G(1-(1-x)(1-e^{-r\tau}))$ is a strictly convex function, since all derivatives are positive, thus if $r_0 + r_1 < 1$ and $E[1 - e^{-r\tau}] = \hat{T}$ this inequality holds. From this we can see that when the infectious periods are random the generating function of the number of neighbors that will become infected is larger than when two individuals are independent. This means that assuming that the transmissibility between individuals is independent, when it is not, will overestimate the number of infected neighbors.

4.2.2 Branching processes

In large populations the early stages of an epidemic can be described as a branching process. In the initial stages of an epidemic most of the contacted individuals are susceptible, which is why we can make the coupling that the number of infectious individuals follows a branching process. In the branching process an individual's birth corresponds to the transition of the disease to susceptible individuals. By analysing a branching process we can determine if a major outbreak is possible. If such an outbreak is possible we can then further analyse the branching process to determine the probability of a major epidemic.

To describe a branching process we look at a population that generates offspring of the same kind. At time zero there exists an initial number of individuals, denoted by X_0. In the end of its lifetime an individual has generated j offspring with probability $P_j,\ 0 \leq j$. The number of offspring an individual generates is independent of other individuals offspring. The offspring from X_0 result in a new generation denoted X_1. The

size of the nth generation is denoted by X_n. Thus $\{X_n,\ n = 0, 1, ...\}$ is a Markov chain, i.e. the next generation conditioned on past generations only depends on the current generation. As long as $P_0 > 0$ the population will either die out or converge to infinity. The expected number of offspring an individual generates is given by $\mu = \sum_{j=0}^{\infty} jP_j$ and the variance is $\sigma^2 = \sum_{j=0}^{\infty}(j-\mu)^2 P_j$ [8].

A starting vertex has k neighbors with probability p_k. The number of neighbors corresponds to the first generation of a branching process. The second generation has a different distribution, since a vertex with a high degree is more likely to be chosen than one with a low degree. The distribution of the number of offspring of a first generation vertex has the probabilities for $k \geq 1$

$$q_{k-1} = \frac{kp_k}{\sum_k kp_k} \tag{12}$$

since we use one edge when connecting to the vertex it is $k-1$ on the left hand side. The mean of q is equal to $v = \sum_k k(k-1)p_k/\mu$ where $\mu = \sum_k kp_k$, which is finite when we assume that p has finite second moment. We can use equations (6) and (8) to calculate the transition point of the existence of a giant component. The growth of a cluster is a two-phase branching process where the first generation has distribution p, the initial individual is a randomly chosen vertex and has degree k with probability p_k, and later generations has distribution q, which is the random variable determined by the q_k's. We denote X_n as the number of vertices in generation n, where $n \geq 1$, $E[X_n] = \mu v^{n-1}$ so when $v < 1$

$$E\left(\sum_{n=0}^{\infty} X_n\right) = 1 + \sum_{n=1}^{\infty} \mu v^{n-1} = 1 + \frac{\mu}{1-v}.$$

When the mean of the number of offspring is less than one, the probability that the population will die out is one, which means that the probability of not having an giant component is one. When $v > 1$ there exists a giant component. We denote the extinction probability of the homogenous branching process with offspring distribution q by ρ. The probability that this two phase branching process will die out is the probability that all k independent first generation families will die out and thus $\sum_{k=0}^{\infty} p_k \rho^k = G_0(\rho)$, where ρ is the smallest fixed point of G_1 in $[0, 1]$. The giant component does not cover the entire graph, the fraction of vertices in the giant component is asymptotically $1 - G_0(\rho)$. We show this in section 4.4.

4.2.3 Application of the birthday problem

The birthday problem is the following: among N random individuals, what is the probability that at least two of them share the same birthday? We assume that there is 365 possible birthdays and all of them is equally probable, so $n = 365$ [1]. We denote E as the event that none of the N persons share birthday, the probability of event E is

$$P(E) = 1 * \left(1 - \frac{1}{365}\right) * \left(1 - \frac{2}{365}\right) * ... * \left(1 - \frac{N-1}{365}\right)$$

$$= \frac{365 * 364 * ... * (365 - N + 1)}{365^N} = \frac{365!}{365^N(365 - N)!}.$$

The probability of at least two people sharing a birthday is then $1 - P(E)$. We generalize the birthday problem, we have n different possible outcomes and N people, then the probability of at least two people sharing an outcome is

$$1 - P(E) = 1 - \frac{n!}{n^N(n - N)!}.$$

To calculate after how many infections the branching process approximation of the epidemic is unreliable we will use a version of the birthday problem. Whenever a birth occur in the branching process, this corresponds to a contact in the epidemic process. If the individual contacted at the ith contact is susceptible, he becomes infected. On the other hand if he is not susceptible he already is or has already been infected so she and all her offspring are ignored in the epidemic process because we have a loop. Thinking about the epidemic outbreak as a branching process, we need to determine the number of contacts made before a previously infected individual is in a transmission causing contact again.

In the configuration model we have n vertices and an even number of stubs that will be paired uniformly at random and create m edges. That is we have $2m$ stubs in our network. We have a degree sequence and let k_i denote the degree of the vertex of ith contact or ith birth in terms of the branching process. In the branching process approximation we first choose a vertex uniformly at random among all the vertices. This vertex has k_1 number of stubs, each stub will be paired with one of the other stubs in the network. We use the generalized birthday problem, where we have $2m$ possible stubs that can be drawn. We draw stubs with replacement and when a drawn stub is attached to a vertex to which a previously drawn stub already has been attached we can no longer use the branching process approximation. The probability of drawing a stub attached to a high degree vertex is larger than the probability of drawing a stub attached to a low degree vertex, it is in fact proportional to the degree of the vertex.

We know that $\sum_{i=1}^{n} k_i = 2m$ and that $E[K] = \sum_k kp_k = \frac{2m}{n}$, where p_k denote the degree distribution. This means that the fraction of the vertices with degree k is p_k. All stubs are equally likely to be chosen, so the stub we draw has probability $\frac{k}{2m}$ of belonging to any particular vertex with degree k. The total number of vertices with degree k is approximately np_k, by the law of large numbers. The probability of chosing any stub attached to a vertex with degree k is $\frac{k}{2m} * np_k = \frac{kp_k}{E[K]}$. From this we can calculate the expected degree of a vertex we reach by chosing a stub, namely

$$E[K'] = \sum_k k\frac{kp_k}{E[K]} = \frac{E[K^2]}{E[K]}.$$

We have N infected persons and will calulate when the number of infected individuals is large enough so the probability of infecting the same individual more than once is close

to one. We let k_i denote the degree of the vertex to which the ith drawn stub is attached and by the law of large numbers the average value of k_i goes towards the expected value. The probability that at least one of the N drawn individuals are drawn twice is

$$1 - P(no\ match) = 1 - 1 * \left(1 - \frac{k_1}{2m}\right) * \left(1 - \frac{(k_1 + k_2)}{2m}\right) * ... * \left(1 - \frac{(k_1 + ... + k_{n-1})}{2m}\right)$$

$$\approx 1 - \prod_{i=0}^{N-1} \left(1 - \frac{iE[K']}{2m}\right)$$

The first order Taylor series expansion of the exponential function when x is small is $e^x \approx 1 + x$. We put $x = -\frac{iE[K']}{2m}$ where i is replaced with non-negative integers for each term in the formula of $1 - P(no\ match)$ until $i = N - 1$, we get

$$P(match) \approx 1 - \exp\left(-\frac{1}{2m}(\sum_{i=0}^{N-1} iE[K'])\right) = 1 - \exp\left(-\frac{1}{2m}E[K'](N(N-1)/2))\right)$$

$$\approx 1 - \exp\left(-\frac{1}{2m}\left(\frac{N^2}{2} * \frac{E[K^2]}{E[K]}\right)\right) = 1 - \exp\left(-\frac{1}{nE[K]}\left(\frac{N^2}{2} * \frac{E[K^2]}{E[K]}\right)\right),$$

so the probability that all the drawn stubs are attached to different vertices is approximately $\exp\left(-\frac{1}{2m}\left(\frac{N^2}{2} * \frac{E[K^2]}{E[K]}\right)\right)$. The probability of drawing the same vertex more than once is above ε, where ε is small, when

$$1 - P(match) = 1 - \varepsilon \approx \exp\left(-\frac{1}{nE[K]}\left(\frac{N^2}{2} * \frac{E[K^2]}{E[K]}\right)\right)$$

$$-2\ln(1-\varepsilon) * n * \frac{E[K]^2}{E[K^2]} \approx N^2 \Longrightarrow N \approx C * \sqrt{n},$$

where $C = \sqrt{-2\ln(1-\varepsilon)E[K]^2/E[K^2]}$. Thus after $C\sqrt{n}$ infections approximations become unreliable, during the course of a major epidemic.

In Figure 4 we can see that when $n = 1000000$ and $E[K]^2/E[K^2] = 1/2$ the probability of the same vertex being drawn at least twice is close to one when $N \approx 2000$.

4.3 Distribution of the outbreak size

The outbreak size of the epidemic corresponds to the size of the cluster of vertices connected by occupied edges in the percolation model. The distribution of the size, s, of the outbreak of the disease on our network is a function of transmission. When the contact rate is constant the transmissibility is given by equation (11) and is a function of the infectious period τ which is either random or fixed, in the latter the transmissibility $T = \hat{T}$. Since the size distribution is a function of transmissibility, we denote the size

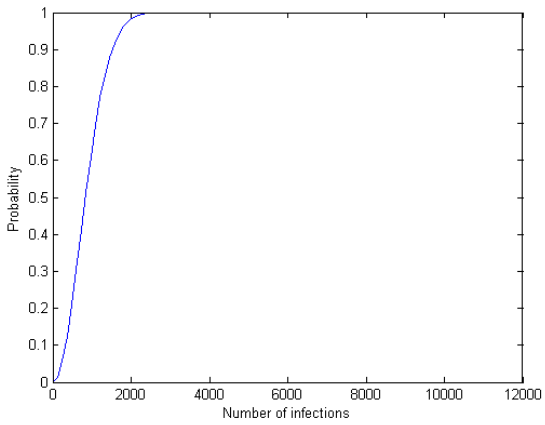

Figure 4: The probability of chosing stub attached to the same vertex more than once when $n = 1000000$ and $E[K^2]/E[K]^2 = 1/2$.

distribution of the outbreak by $P_s(T(\tau)) = P_s(T)$. For this distribution $H_0(x;T)$ is the generating function,

$$H_0(x;T) = \sum_{s=0}^{\infty} P_s(T)x^s.$$

We also need another generating function for the size of the cluster of connected vertices we reach by following a randomly chosen edge, we denote it by $H_1(x;T)$. There are two possible clusters we can reach by following a random edge:

- a single vertex without any occupied edges,
- a single vertex with m occupied edges, where $m \geq 1$, which all lead to another cluster with size distribution generated by H_1.

The probability of an edge between two finite clusters that already are attached to the same vertex goes as $\frac{1}{n}$ with the size n of the graph. The probability of these clusters having loops goes to zero as n grows large and we have a tree structure. If we denote the probability that the initial vertex has k edges coming out of it other than the edge we came along by q_k, then [5]

$$H_1(x;T) = xq_0 + xq_1H_1(x;T) + xq_2[H_1(x;T)]^2 + ... = xG_1(H_1(x;T);T). \tag{13}$$

When we randomly choose an initial vertex, the size of the cluster to which it is attached is distributed as

$$H_0(x;T) = xG_0(H_1(x;T);T). \tag{14}$$

When we have H_0 we can get $P_s(T)$

$$P_s(T) = \frac{1}{s!}\frac{\partial^s H_0}{\partial x^s}\bigg|_{x=0}$$

which can be solved by the Cauchy formula [6].

4.4 Outbreak sizes and epidemic threshold

With the generating functions we can calculate the mean outbreak size, which is the derivative of $H_0(x;T)$ at $x = 1$ and can be expressed as

$$E[s] = H_0'(1;T) = 1 + G_0'(1;T)H_1'(1;T)$$

$$H_1'(1;T) = 1 + G_1'(1;T)H_1'(1;T) = \frac{1}{1 - G_1'(1;T)}$$

$$\implies \quad E[s] = 1 + \frac{G_0'(1;T)}{1 - G_1'(1;T)} = 1 + \frac{TG_0'(1)}{1 - TG_1'(1)}. \tag{15}$$

From equation (15) we can see that when $TG_1'(1) = G_1'(1;T) = 1$ the equation diverges and the disease outbreak might become an epidemic. This point is called the epidemic threshold, T_c, and is equal to

$$T_c = \frac{1}{G_1'(1)} = \frac{G_0'(1)}{G_0''(1)} = \frac{\sum_k kp_k}{\sum_k k(k-1)p_k}.$$

Above T_c, i.e. for $T > T_c$, we can have an epidemic. An epidemic corresponds to a giant component which is widely spread and can therefore contain loops, thus equation (13) is no longer valid. The part of the graph that is covered by this giant component we denote by $S(T)$. We redefine H_0 as the generating function for outbreaks that are not epidemics. These clusters do not cover the entire graph. Above the epidemic threshold

$$H_0(1;T) = \sum_s P_s = 1 - S(T),$$

$$S(T) = 1 - G_0(u;T), \tag{16}$$

where u is the probability that a vertex we reach by following a randomly chosen edge remains uninfected during an epidemic. We have used the equation (14) so $u \equiv H_1(1;T)$ which is the solution to

$$u = G_1(u;T). \tag{17}$$

The probability of an epidemic is not one even above the epidemic transition, the probability that the disease is an epidemic is $S(T)$.

When the infectious periods are random, we can show that by assuming that the transmissibility is independent between individuals we will have the same epidemic threshold as when we calculate it correctly. We look at the branching process approximation when the infectious time τ is constant, to make it easier we assume it is equal to one and that we initially have one infected individual. The transmission probability is $T = (1 - e^{-r})$ and we denote the probability that an infected individual transmit the disease to j of its k neighbors by $\hat{p}_j$, which is equal to

$$\hat{p}_j = \sum_{k=j}^{\infty} p_k \binom{k}{j} (1 - e^{-r})^j e^{-(k-j)r}.$$

The mean of $\hat{p}$ is $\hat{\mu} = \mu(1 - e^{-r})$, where μ is the mean of p. The probability of having k neighbors in the next generations is $q_k = (k+1)p_{k+1}/\mu$ for $k \geq 0$. The probability that j of the k neigbors will become infected in these generations is

$$\hat{q}_j = \sum_{k=j}^{\infty} q_k \binom{k}{j} (1 - e^{-r})^j e^{-(k-j)r}.$$

The mean of q is v, thus the mean of $\hat{q}$ is $\hat{v} = v(1 - e^{-r})$. When $\hat{v} > 1$ we may have an epidemic outbreak. Since $\hat{v} = vT$ the epidemic threshold is the same.

4.5 Degree of infected individuals

In this section we calculate the expected degree of the vertices outside and inside the giant component. A vertex does not get infected by any of its edges if the edge is unoccupied, the probability of this is $1-T$, or if it is occupied but the vertex to which it is connected is uninfected, the probability of this is Tu, where u denote the probability of a vertex we reach by following a randomly chosen edge do not belong to the giant component. We let v denote the probability that the vertex does not get infected by any of its edges then $v = 1 - T + Tu = 1 + (u-1)T$, so v^k is the probability of an uninfected vertex of degree k. The probability of having k neighbors given that the vertex is uninfected is

$$\frac{p_k v^k}{\sum_k p_k v^k} = \frac{p_k v^k}{G_0(v)}.$$

The distribution of having degree k while beeing uninfected is generated by the function $\frac{G_0(vx)}{G_0(v)}$. We want to calculate the expected degree of vertices outside the giant component, we get this by differentiating and setting $x = 1$

$$z_{out} = \frac{vG_0'(v)}{G_0(v)} = \frac{vG_1(v)}{G_0(v)} z = \frac{[1 + (u-1)T]G_1(1 + (u-1)T)}{G_0(1 + (u-1)T)} z$$

$$= \frac{[1 + (u-1)T]G_1(u;T)}{G_0(u;T)} z = \frac{[1 - T + Tu]u}{1 - S} z,$$

in the last step we use equation (16) and (17), as before S denotes the fraction of the population affected by the epidemic. The degree distribution of an infected vertex is given by $\frac{\sum_k p_k(1-v)^k x^k}{\sum_k p_k(1-v)^k} = \frac{G_0(x) - G_0(vx)}{1 - G_0(v)}$, therefore in the giant component the expected degree for vertices is

$$z_{in} = \frac{G_0'(1) - vG_0'(v)}{1 - G_0(v)} = \frac{1 - vG_1(v)}{1 - G_0(v)} z = \frac{1 - u[1 - T + Tu]}{S} z.$$

We have that $1 - S = G_0(u;T) \leq u$ which implies $z_{out} \leq z \; and \; z \leq z_{in}$. This means that the mean degree of infected individuals is greater than or equal to the mean degree of uninfected individuals when there is a large outbreak.

5 Conclusions and Discussion

In this thesis we have studied networks, in particular the configuration model network, and how we can use them to investigate the outbreak of an SIR epidemic. We have been using a model which allow self-loops and parallel edges and where the neighboring vertices make contacs at a fixed rate. The transmission of the disease depend on the infectious periods, if the infectious periods are fixed for all individuals the transmission is independent and identically distributed while if the infectious periods are random this is not the case. If the infectious periods are random the extinction probability is smaller when we assume that the transmission is independent than when we do not. By using generating functions and percolation theory we saw that if the transmission probability is above the epidemic threshold we have a giant component in our network and thus an epidemic outbreak is possible. The epidemic threshold is at the point where transmissibility is equal to $\sum_k kp_k / \sum_k k(k-1)p_k$, where k denote the degree of a vertex and p_k specifies the degree distribution. If we have a transmission probability below this threshold there are only small components in our network and the outbreak will not become an epidemic. We have also seen that up to approximately the $\sqrt{n}$th infected individual, we can look at the epidemic outbreak as a branching process.

References

[1] Alm, S.E. and Britton, T. (2008): *Stokastik: Sannolikhetsteori och statistikteori med tillämpningar*, Liber, 2 and 27-29

[2] Durrett, R. (2006): *Random Graph Dynamics*, Cambridge University Press, 70-89

[3] Gut, A. (2009): *An Intermediate Course in Probability*, Springer, 59-60

[4] Newman, M.E.J. (2012): *Networks: An introduction*, Oxford University Press

[5] Newman, M.E.J., Strogatz, S.H. and Watts, D.J. (2001): *Random graphs with arbitrary degree distributions and their applications*, Physical Review, **64**

[6] Newman, M.E.J. (2002): *Spread of epidemic disease on networks*, Physical review, **66**

[7] Ross, S. (1988): *A first course in probability, third edition*, Macmillian Publishing Company, 356

[8] Ross, S.M. (2010): *Introduction to Probability Models 10th Edition*, Academic Press, 245-249

Appendix

A The degree distribution of an Erdös-Rényi graph

In large networks, when the number of vertices n grows large, and we have an Erdös-Rényi graph the degree distribution is a Poisson distribution. The probability of an edge between two vertices is denoted by p, while the expected degree of a vertex we denote by z. Starting from a vertex, the probability that it has degree k is the probability that it is attached to k of the other vertices and not attached to the other $n-1-k$ vertices. There are $\binom{n-1}{k}$ ways to choose these k vertices. The probability of a vertex having degree k is

$$p_k = \binom{n-1}{k} p^k (1-p)^{n-1-k}. \tag{18}$$

Since $p = \frac{z}{n-1}$ will become arbitrary small when n grows large, by expanding the logarithm as a Taylor series we have

$$\ln[(1-p)^{n-1-k}] = (n-1-k)\ln\left(1-\frac{z}{n-1}\right) \approx -(n-1-k)\frac{z}{n-1} \longrightarrow -z,\ n\to\infty$$

$$(1-p)^{n-1-k} \to e^{-z}$$

when $n \to \infty$. We also have

$$\binom{n-1}{k} = \frac{(n-1)!}{(n-1-k)!k!} \approx \frac{(n-1)^k}{k!}$$

and by inserting these results to equation (18) we get

$$p_k \to \frac{(n-1)^k}{k!} p^k e^{-z} = \frac{(n-1)^k}{k!}\left(\frac{z}{n-1}\right)^k e^{-z} = e^{-z}\frac{z^k}{k!}$$

when $n \to \infty$.

B Random infectious periods result in overestimate with bond percolation

When the infectious periods is random it can be shown that the generating functions we calculate by analysing bond percolation methods, the equations (9) and (10) are overestimates. The number of neighbors that will become infected have generating function $G_0(x;T)$ and in the following calculations is r_k the probability distribution of having k neighbors:

$$G_0(x;T) = \int_0^\infty dt P(\tau = dt) \sum_{j=0}^\infty x^j \sum_{k=j}^\infty r_k \binom{k}{j} (1-e^{-rt})^j e^{-r(k-j)t}$$

$$= \int_0^\infty dtP(\tau = dt) \sum_{k=0}^{\infty} r_k \sum_{j=0}^{k} \binom{k}{j} (x(1-e^{-rt}))^j e^{-r(k-j)t}$$

$$= \int_0^\infty dtP(\tau = dt) \sum_{k=0}^{\infty} r_k (e^{-rt} + x(1-e^{-rt}))^k = E\left[\sum_{k=0}^{\infty} r_k (e^{-r\tau} + x(1-e^{-r\tau}))^k\right]$$

$$= E[G(1+(x-1)(1-e^{-r\tau}))] > G(1+(x-1)\hat{T}) = \sum_{k=0}^{\infty} r_k (1+(x-1)\hat{T})^k.$$

We can use Jensen's inequality if $r_0 + r_1 < 1$, because $G(1-(1-x)(1-e^{-r\tau}))$ is strictly convex in $1-e^{-r\tau}$ and $E[1-e^{-r\tau}] = \hat{T}$.