UTRECHT UNIVERSITY

MASTER THESIS MATHEMATICAL SCIENCES

Modelling the Effect of Clustering in Contact Networks on the Effectiveness of Contact Tracing

---

Author:
Isis Marsman
7579837

Supervisor:
dr. M.C.J. Bootsma

Daily Supervisor:
dr. D. Klinkenberg

Second reader
dr. C. Spitoni

Rijksinstituut voor Volksgezondheid
en Milieu
Ministerie van Volksgezondheid,
Welzijn en Sport

November 1, 2023

Abstract

At the request of the Ministry of Health, Welfare and Sport, during the COVID-19 pandemic, The National Institute for Public Health and the Environment (RIVM) conducted a model study into the effectiveness of contact tracing, with the focus on effectiveness of the CoronaMelder. A model was created to simulate the epidemic with and without tracing and the use of the CoronaMelder app. In that model an assumption was made to take the effect of clustering in contact networks into account, not based on earlier publications. In our project we wanted to investigate if this assumption on how to implement the effect of clustering on the effectiveness of contact tracing is justified. Two models were created to simulate clustering in contact networks. The first model is a simplified version of the model used in the CoronaMelder study. In the second model epidemics spread on a clustered contact network to represent the process closer to reality. With these models contact tracing was simulated and compared. We found that the model based on the CoronaMelder study overestimates the effect of clustering in the contact network on the effectiveness of contact tracing.

Acknowledgements

Contents

1   Introduction

The first cases of COVID-19 pandemic, caused by the SARS-CoV-2 virus, were reported in December 2019. In March 2020 the COVID-19 outbreak was declared a global pandemic by the World Health Organisation. In absence of a vaccine, other interventions had to be used to contain the spread of the virus. Besides measures like lockdowns, travel restrictions and social distancing, contact tracing was used to contain the spread of the virus. Contact tracing was a critical strategy to identify and isolate individuals who had been exposed to the virus. Research showed that containing the spread of the SARS-CoV-2 virus by manual contact tracing was infeasible, but the use of a contact tracing app could stop the spread if used by enough people [1].

In April 2020 the Dutch minister of Health, Welfare and Sport announced that the government was considering to introduce a contact tracing app for COVID-19 in The Netherlands. Companies could send in their proposal for the app. Seven of these proposals were considered for the final version. However the ministry decided to develop the app themselves, because of concerns about the use and protection of the personal data of the users. After multiple phases of testing, in October 2020 the CoronaMelder app was launched in The Netherlands. Users of the app got notified when they had been in close or extended contact with an infected user. This allowed more people to act accordingly, i.e., to quarantine and get tested.

At the request of the ministry of Health, Welfare and Sport a model study into the effectiveness of contact tracing, with the focus on effectiveness of the CoronaMelder app, was conducted by The National Institute for Public Health and the Environment (RIVM) [2], [3]. A model was created to simulate the epidemic with and without contact tracing and the use of the CoronaMelder app. The results were used to calculate the number of infections prevented by the app and the reduction in the reproduction number.

Both the spread of infectious diseases and contact tracing takes place over physical contacts within social networks. Research indicates that social networks are fundamentally different from other types of networks [4]. One of the things that distinguishes social networks from other types, is that they are known to be highly clustered. The degree of clustering in a network indicates the probability that contacts of a node are also contacts of each-other. Because of these clusters it is possible that infected people can be traced over a contact which wasn't the source of the infection. Thus when modeling a form of contact tracing, clustering in the contact network should be considered.

A contact network is often represented by a graph, where the nodes represent the individuals and the links the contacts between the individuals. Over this network an epidemic can be simulated. A transmission network describes the spread of the infectious agent. This can also be represented by a graph, where the nodes represent the infected individuals and the edges the infectious contacts, i.e. the contacts over which transmission took place. A transmission network has no clustering if individuals can be infected only once.

When modelling a social network like a contact network clustering should be incorporated. Multiple papers have been written about different ways to incorporate clustering when modelling social networks. The search is for a mathematical model of a clustered network which is a good representation of a social network and which properties can be expressed analytically. Standard random graph models, like the configuration model, can be adjusted to incorporate clustering [5], [6], [7]. Another type of models are the so-called household models [8], [9], where individuals within the same group form a cluster, and have additional contacts outside the group that may be of a different kind. These groups can for example represent households, schools or work.

The CoronaMelder study takes a different approach to incorporate clustering into the model. Instead of creating a contact network with clustering and which is used to simulate an epidemic and contact tracing, clustering in the contact network is modeled by adding contacts to the network of transmission contacts. These extra contacts represent the social contacts between infected nodes which are used for contact tracing. These contacts are added in three steps. In the first step, a contact is added between the nodes at distance 2 with a probability c, where the distance between two nodes represents the minimum number of edges in the network to go from one node to the other. In the next step, between every pair of nodes that are at distance 2 after step 1 a connection is added with a probability c^2. In the third step, again between every pair of nodes that are at distance 2 after step 2 a connection is added with a probability c^3. The question that arises is if

adding the clustering to the transmission network instead of using a contact network is a suitable manner of studying the effect of clustering on contact tracing.

In this thesis we want to investigate if a model which adds contacts to the transmission network approaches the effect of clustering on contact tracing. We do this by introducing two new models: a simplified version of the model used in the CoronaMelder study, and a similar model but in which first a contact network is formed. We want to answer the following questions:

1. What is the effect of clustering in the contact network on the spread of an epidemic?

2. What is the effect of clustering on the effectiveness of contact tracing?

3. What are the differences and similarities of the effectiveness of contact tracing with the two models?

This thesis has the following structure. In Chapter 2 we introduce the two models used in this thesis and compare their structures. In Chapter 3 we give some analytical results describing the properties of the networks that arise form our models. In Chapter 4 we discuss the effect of clustering in a contact network on a epidemic. In Chapter 5 we introduce contact tracing to the networks and discuss the effect of clustering on the effectiveness of contact tracing. In Chapter 6 we compare the effectiveness of contact tracing with both models. We end with the conclusion and discussion in the last two chapters.

2 Models

In this project we compare the effectiveness of contact tracing during an epidemic on a network resulting from 2 different models. Both models contain a form of clustering, i.e., contacts along which tracing takes place without transmission, but they are different. The first model is based on the model used in the study for the contact tracing app where extra contacts were added after simulating the epidemic so that tracing can take place across contacts that did not play a role in the spread of the epidemic. The second model uses an approach closer to the process in reality, by simulating the epidemic on a clustered network.

The networks are realisations of a stochastic process. The nodes in the networks represent the individuals in the population and the edges represent the contacts between these individuals. During this project we looked at multiple network properties. The degree distribution of a network is the probability distribution describing the number of contacts the nodes in the network have. The offspring distribution of a branching process is the distribution describing the number of offspring of a single node. The clustering coefficient is a measure of the degree to which nodes tend to cluster together. All the networks used are static.

We simulate an epidemic with permanent immunity on a generation basis. When describing an epidemic, we are mostly interested in the basic reproduction number, R_0. The basic reproduction number of an epidemic is the expected number infections caused by a typical infected individual in a susceptible population. R_0 is often used in epidemiology for assessing the potential spread of an infectious disease within a population.

When denoting the network properties, we use a subscript to denote to which network a property belongs and a superscript to denote which model was used to create this network. For example, $D_t^{(1)}$ is used to denote the degree distribution of the network used for contact tracing, the tracing network, obtained with model 1.

2.1 Model 1

We first describe the model based on the model used during the study of the contact tracing app. The first step of this model describes the transmission of the pathogen. We assume each individual can be infected only once. Hence, if we start with a single infected individual, the transmission network can be represented by a tree. This means that a branching type process is a natural way to describe the transmission process. If we assume that depletion of susceptible individuals in the population is negligible, each infected individual has the same offspring distribution. The offspring distribution $O_e^{(1)}$ is based on the epidemic it aims to describe, i.e., $\mathbb{E}(O_e^{(1)}) = R_0$.

Next we add extra connections to this transmission network. Between every pair of nodes at distance 2 we add independently a connection with probability $p^{(1)}$, denoted as the clustering probability. These extra connections represent contacts between infected nodes which did not lead to transmission. This resulting network is called the tracing network as contact tracing is simulated over this network. This tracing network has a clustering coefficient, $C_t^{(1)}$, and degree distribution, $D_t^{(1)}$.

So, the input parameters for this model are the offspring distribution of the transmission network, $O_e^{(1)}$, and the clustering probability, $p^{(1)}$. The outcome is the resulting tracing network with its properties such as $C_t^{(1)}$ and $D_t^{(1)}$.

Note that only infected individuals are part of the tracing network. Consequently, this model cannot be used to assess the amount of effort needed to preform the contact tracing, as individuals who did have contact with infectious individuals but did not get infected are not part of the network, but may be traced in reality.

2.2 Model 2

With the second model we want to stay closer to the process as it happens in reality. In this model a clustered contact network is created over which an epidemic is simulated and from this the transmission and tracing network are obtained.

In this model, we start by creating the contact network. For simplicity we use a branching process with offspring distribution $O_b^{(2)}$. Then clustering is added to this branching process in a
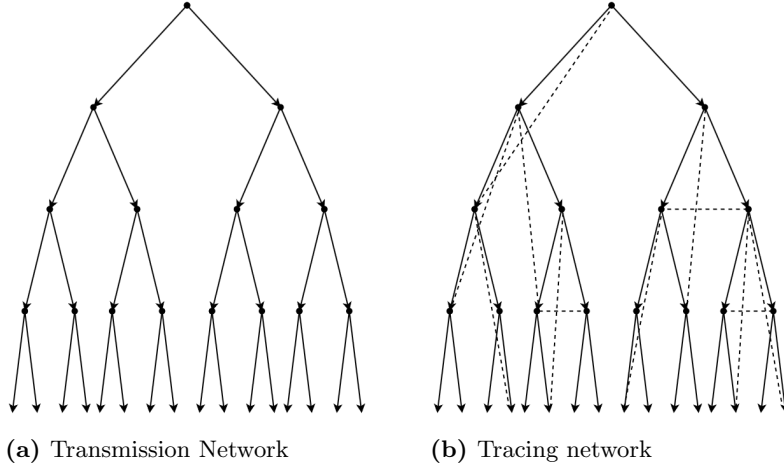
(a) Transmission Network   **(b)** Tracing network

Figure 1: Example of networks obtained using model 1 with an fixed offspring of 2. The arrows represent the transmission contacts and the dashed lines the added connections.

similar way as in model 1. Connections are added between each pair of nodes at distance two. These connections are added independently with probability $p^{(2)}$, the clustering probability. The resulting network is the contact network with a degree distribution, $D_c^{(2)}$, and clustering coefficient, $C_c^{(2)}$.

Over this resulting contact network we simulate the epidemic. For simplicity we simulate the epidemic as follows. For each edge, during a generation transmission occurs with probability q, denoted as the transmission probability, if one of the endpoint of the edge is infectious and the other is susceptible. A node is only infectious during one generation. Every infected node in the contact network independently infects each susceptible neighbour with the transmission probability. We start the epidemic with one infected node as the initial case.

The epidemic over the contact network is described in a transmission network. In the transmission network the connections represent the contacts over which the infections took place. This network has a offspring distribution, $O_e^{(2)}$, from which we can obtain the basic reproduction number of the simulated epidemic, as $R_0^{(2)} = E(O_e^{(2)})$.

After simulating the epidemic we can obtain the tracing network. The tracing network in this model is the sub-network of the contact network consisting of all the infected nodes and connections between them, as susceptible nodes do not have an effect on contact tracing if one neglects that in reality effort is taken to detect all contacts, also the ones which were not infected. The tracing network has degree distribution $D_t^{(2)}$ and clustering coefficient $C_t^{(2)}$.

So, the input parameters for model 2 are the offspring distribution of the branching process, $O_b^{(2)}$, and the clustering and transmission probabilities, $p^{(2)}$ and q. The outcome of this network are the resulting contact, transmission and tracing networks with their properties.

2.3 Structural Differences

The biggest difference between these two models is that in model 1 the transmission tree is directly simulated and thus the transmission path is fixed, while in model 2 the epidemic is simulated over a contact network and thus multiple realisations of the epidemic are possible. The tracing network obtained with model 1 has two different kinds of connections. The connections describing the transmissions and the added connections. Both connections can be used for contact tracing, but transmission is only possible over the transmission contacts.

The connections in the tracing network obtained with model 2 are all equal. Some are used for transmission, but unlike in model 1 the transmission path is not fixed. If any intervention is modeled on these networks, in model 1 the transmission path is stopped, while in model 2 alternative infection paths may become available.

Another difference is that in model 1 the extra connections are always between nodes at distance

two. If we look at the tracing network of model 2 as the transmission tree with added connections, it is possible that the added connections are between nodes at distance larger than two.

(a) Contact Network   **(b)** Simulated epidemic   **(c)** Tracing network
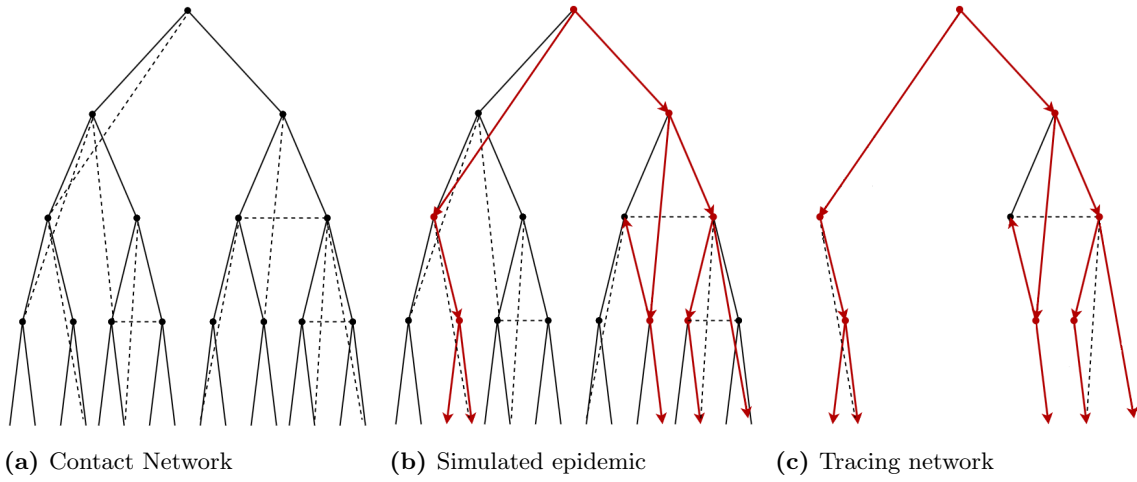
Figure 2: Example of networks obtained using model 2 with with a fixed offspring of 2 used for the branching process. The dashed lines represent the added connections and the red arrows the path of the infection.

3 Network Properties

In both models connections are added between nodes at distance 2 with a clustering probability. We want to know if we can analytical express some properties of a network after these connections are added. The properties discussed in this section are the degree distribution and clustering coefficient.

3.1 Degree Distribution

Let the degree of an individual in the branching process be distributed according to distribution X. In a branching type network every individual has one parent, hence we assume $\mathbb{P}(X = 0) = 0$. Our aim is an expression for the resulting degree distribution Y when connections at distance 2 are added with probability p:

$$\mathbb{P}(Y = k) = \sum_{i=1}^{k} \mathbb{P}(Y = k, X = i).$$

The new degree distribution Y depends on the number of nodes at distance 2. Let L be the sum of all these nodes, i.e., the sum of the number of siblings, grandchildren and grandparents in a branching process.

$$\begin{aligned}
\mathbb{P}(Y = k) &= \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \mathbb{P}(Y = k, X = i, L = l) \\
&= \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \mathbb{P}(Y = k | X = i, L = l) \mathbb{P}(X = i, L = l) \\
&= \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \mathbb{P}(Y = k | X = i, L = l) \mathbb{P}(L = l | X = i) \mathbb{P}(X = i).
\end{aligned}$$

When a node has degree i and has l nodes at distance 2 the probability that its degree after adding the connections is k is the probability that of the l possible connections $k - i$ are added. Each connection is independently added with probability p, so we have that the number of added edges is binomial distributed:

$$\mathbb{P}(Y = k | X = i, L = l) = \binom{l}{k-i} p^{k-i} (1-p)^{l-k+i},$$

where we use the standard conventions that $\binom{i}{j} = 0$ if $j < 0$ or $j > i$.

So the last step is to calculate $\mathbb{P}(L = l | X = i)$. Note that if the degree of a node in a branching process equals i, there are $i - 1$ children and 1 parent. Let S be the number of siblings of a node. The probability that a node has s siblings is the same as the probability that the parent has degree $s + 2$, its own parent, the node itself and the s other children. The degree of the parent is size-biased. So, we have

$$\mathbb{P}(S = s) = \mathbb{P}(\text{parent degree} = s + 2) = \frac{(s+1)\mathbb{P}(X = s+2)}{\sum_{s'}(s'+1)\mathbb{P}(X = s'+2)} = \frac{s+1}{\mathbb{E}(X) - 1} \mathbb{P}(X = s+2).$$

Let G be the number of grandchildren of a node. The probability to have g grandchildren, given that there are $i - 1$ children is

$$\mathbb{P}(G = g | X = i) = \mathbb{P}(X_1 + X_2 + \cdots + X_{i-1} = g + (i-1)).$$

Here $X - 1, \ldots, X_{i-1}$ are independent copies of X. The number of siblings and the number of grandchildren are independent and the number of siblings does not depend on the degree of that node. Therefore we have that:

$$\mathbb{P}(L = l | X = i) = \sum_{g=0}^{l-1} \mathbb{P}(G = g, S = l - g - 1 | X = i) = \sum_{g=0}^{l-1} \mathbb{P}(S = l - g - 1) \mathbb{P}(G = g | X = i).$$

The sum is from 0 to $l-1$, because one of the nodes at distance 2 is the grandparent and every node has exactly one grandparent.

So, we get that the probability that after adding the extra connections with probability p a uniformly at random chosen node has degree k equals:

$$\mathbb{P}(Y = k) = \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \sum_{g=0}^{l-1} \mathbb{P}(Y = k|X = i, L = l)\mathbb{P}(S = l - g - 1)\mathbb{P}(G = g|X = i)\mathbb{P}(X = i).$$

3.1.1 Fixed Offspring

We start by looking at a branching process with a fixed number of offspring. We say that every node has λ offspring. This can be described as a offspring distribution: assume that the number of offspring is distributed according to Z, then

$$\mathbb{P}(Z = k) = \delta_{k,\lambda} = \begin{cases} 1 & \text{if } k = \lambda \\ 0 & \text{if } k \neq \lambda \end{cases},$$

where $\delta_{i,j}$ is the Kronecker delta.

So, we have that $\mathbb{P}(X = k) = \mathbb{P}(Z = k - 1)$ and thus

$$\mathbb{P}(X = k) = \begin{cases} 1 & \text{if } k = \lambda + 1 \\ 0 & \text{otherwise} \end{cases}.$$

When the number of offspring is fixed, the number of nodes at distance 2 is easily found: every node has $\lambda - 1$ siblings, 1 grandparent and λ children, which all have λ children, thus λ^2 grandchildren. So, in total there are $(\lambda + 1)\lambda$ nodes at distance 2.

$$\mathbb{P}(L = l|X = \lambda + 1) = \begin{cases} 1 & \text{if } l = (\lambda + 1)\lambda \\ 0 & \text{otherwise} \end{cases}.$$

This gives us the following expression for the degree distribution after adding the connections at distance 2 with probability p:

$$\begin{aligned} \mathbb{P}(Y = k) &= \mathbb{P}(Y = k|X = \lambda + 1, L = (\lambda + 1)\lambda) \\ &= \binom{(\lambda + 1)\lambda}{k - (\lambda + 1)} p^{k-(\lambda+1)}(1 - p)^{(\lambda+1)\lambda-k+(\lambda+1)}. \end{aligned}$$

3.1.2 Poisson Distribution

Now we assume that the number of offspring is Poisson distributed, i.e. $\mathbb{P}(X = k) = \mathbb{P}(Z = k - 1)$ where Z follows a Poisson distribution with rate λ, thus

$$\mathbb{P}(X = k) = \mathbb{P}(Z = k - 1) = \frac{\lambda^{k-1}e^{-\lambda}}{(k-1)!}.$$

For calculating the chance to have g grandchildren, given degree i we use the fact that the sum of j idd random variables all with rate λ is a Poisson random variable with rate $j\lambda$.

$$\mathbb{P}(G = g|X = i) = \mathbb{P}(Z_1 + Z_2 + \cdots + Z_{i-1} = g) = \mathbb{P}(\text{Pois}((i-1)\lambda) = g) = \frac{((i-1)\lambda)^g e^{-(i-1)\lambda}}{g!}$$

here $Z_1, \ldots, Z_{i-1}$ are idd copies of Z.

Now we look at the number of siblings:

$$\mathbf{P}(S = s) = \frac{s+1}{\mathbb{E}(X) - 1}\mathbb{P}(X = s + 2) = \frac{s+1}{\lambda}\mathbb{P}(Z = s + 1) = \frac{s+1}{\lambda}\frac{\lambda^{s+1}e^{-\lambda}}{(s+1)!} = \frac{\lambda^s e^{-\lambda}}{s!} = \mathbb{P}(Z = s)$$
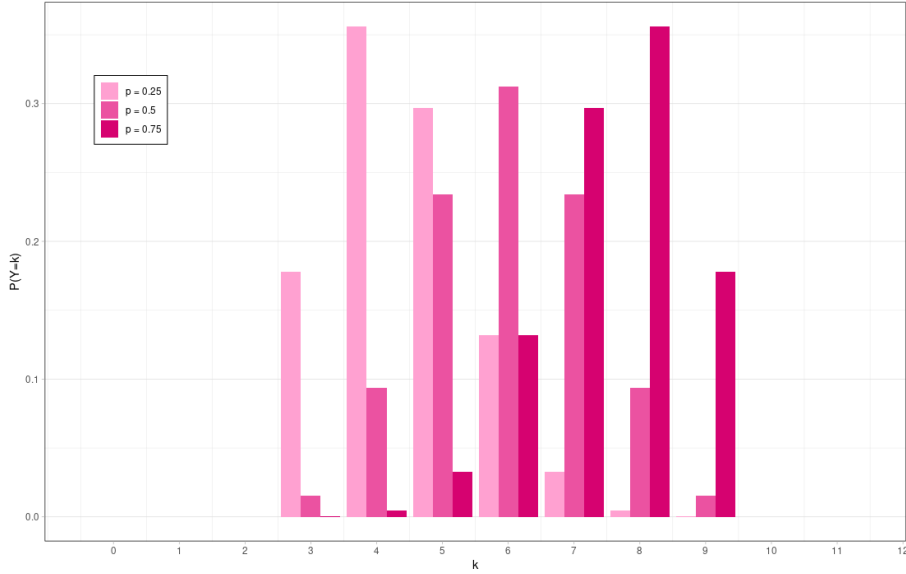
Figure 3: The degree distribution of a network when adding contacts to nodes at distance 2 with probability p to a branching process with a constant offspring of 2.

So, the number of siblings is also Poisson distributed with rate λ

Using that each of the $(i-1)$ children have children distributed according to Z, and the number of siblings is also distributed as Z, and that there is always 1 grandparent, we have that, if there are $i - 1$ children, L will be distributed as $1 + Z_1 + \cdots + Z_i$. Hence we have that:

$$\mathbb{P}(L = l | X = i) = \mathbb{P}(Z_1 + Z_2 + \cdots + Z_i = l - 1) = \mathbb{P}(\mathrm{Pois}(i\lambda) = l - 1) = \frac{(i\lambda)^{l-1}e^{-i\lambda}}{(l-1)!}$$

and thus

$$\mathbb{P}(Y = k) = \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \mathbb{P}(Y = k | X = i, L = l)\mathbb{P}(L = l | X = i)\mathbb{P}(X = i)$$

$$= \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \binom{l}{k-i} p^{k-i}(1-p)^{l-k+i} \frac{(i\lambda)^{l-1}e^{-i\lambda}}{(l-1)!} \frac{\lambda^{i-1}e^{-\lambda}}{(i-1)!}.$$

This can be further simplified, as for the inner summation we have

$$\sum_{l=k-i}^{\infty} \binom{l}{k-i} p^{k-i}(1-p)^{l-k+i} \frac{(i\lambda)^{l-1}e^{-i\lambda}}{(l-1)!} \frac{\lambda^{i-1}e^{-\lambda}}{(i-1)!} = \frac{e^{-\lambda(1+ip)}\lambda^{i-2}(ip\lambda)^{k-i}(k + i(-1 + \lambda(1-p)))}{i!(k-i)!}$$

and hence, we have that

$$\mathbb{P}(Y = k) = \frac{e^{-\lambda}}{\lambda^2} \sum_{i=1}^{k} \frac{(\lambda e^{-\lambda p})^i (ip\lambda)^{k-i}(k - i + i\lambda(1-p))}{i!(k-i)!}.$$

3.1.3 Negative Binomial Distribution

Now assume that the number of offspring is Negative Binomial distributed, i.e. $\mathbb{P}(X = k) = \mathbb{P}(Z = k - 1)$ where Z follows a negative binomial distribution with parameters r and π, where
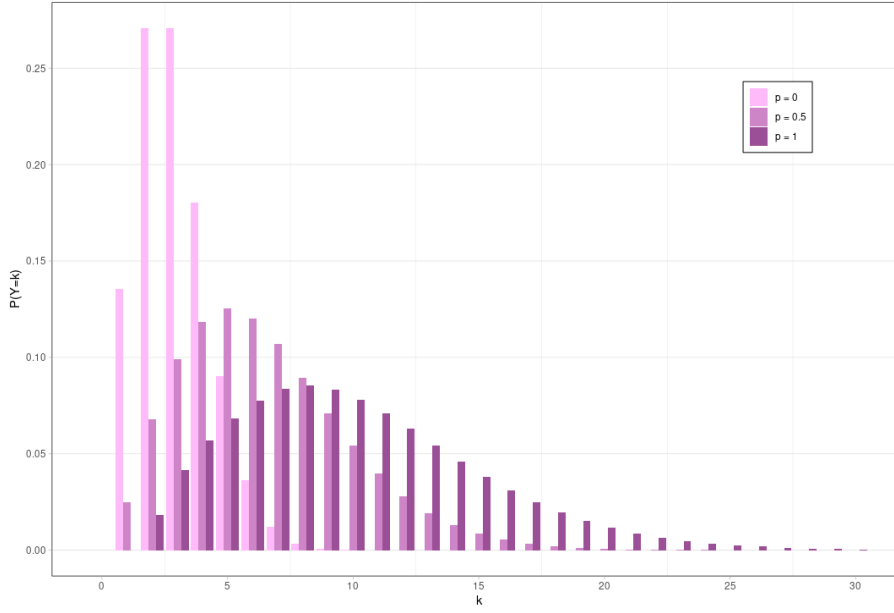
Figure 4: The degree distribution of a network when adding contacts to nodes at distance 2 with probability p to a branching process with offspring distributed according to a Poisson distribution with rate 2.

$r \in \mathbb{R}$ is the number of successes and $\pi \in [0,1]$ the probability of success. In this section we only consider the situation where $r \in \mathbb{N}$. We get that the probability that a node has degree k is

$$\mathbb{P}(X = k) = \mathbb{P}(Z = k - 1) = \binom{k+r-2}{k-1}(1-\pi)^{k-1}\pi^r.$$

Use the fact that if $Y_1 \sim \text{NB}(r_1, \pi)$ and $Y_2 \sim \text{NB}(r_2, \pi)$, then $Y_1 + Y_2 \sim \text{NB}(r_1 + r_2, \pi)$ to get that the chance to have g grandchildren is

$$\mathbb{P}(G = g | X = i) = \mathbb{P}(Z_1 + Z_2 + \cdots + Z_{i-1} = g) = \mathbb{P}(\text{NB}((i-1)r, \pi) = g)$$
$$= \binom{g + (i-1)r - 1}{g}(1-\pi)^g \pi^{(i-1)r}.$$

here $Z_1, \ldots, Z_{i-1}$ are idd copies of Z.

For the number of siblings we get:

$$\mathbf{P}(S = s) = \frac{s+1}{\mathbb{E}(X) - 1}\mathbb{P}(X = s + 2) = \frac{s+1}{r(1-\pi)/\pi}\mathbb{P}(Z = s + 1) = \frac{(s+1)\pi}{r(1-\pi)}\binom{s+r}{s+1}(1-\pi)^{s+1}\pi^r$$
$$= \frac{(s+1)}{r}\pi^{r+1}(1-\pi)^s\binom{s+r}{s+1} = \binom{s+r}{s}\pi^{r+1}(1-\pi)^s = \binom{s+(r+1)-1}{s}\pi^{r+1}(1-\pi)^s$$
$$= \mathbb{P}(\text{NB}(r+1, \pi) = s).$$

So, we see that the number of siblings is negative binomial distributed with size $r+1$ and probability π. And thus it follows that

$$\mathbb{P}(L = l | X = i) = \mathbb{P}(\text{NB}((ir+1, \pi) = l-1) = \binom{l-1+(ir+1)-1}{l-1}(1-\pi)^{l-1}\pi^{ir+1}$$
$$= \binom{l+ir-1}{l-1}(1-\pi)^{l-1}\pi^{ir+1}.$$

9

When we combine all these probabilities we get that

$$\mathbb{P}(Y = k) = \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \mathbb{P}(Y = k | X = i, L = l) \mathbb{P}(L = l | X = i) \mathbb{P}(X = i)$$

$$= \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \binom{l}{k-i} p^{k-i}(1-p)^{l-k+i} \binom{l+ir-1}{l-1}(1-\pi)^{l-1}\pi^{ir+1} \binom{i+r-2}{i-1}(1-\pi)^{i-1}\pi^{r}$$

$$= \sum_{i=1}^{k} \sum_{l=k-i}^{\infty} \binom{l}{k-i}\binom{l+ir-1}{l-1}\binom{i+r-2}{i-1} p^{k-i}(1-p)^{l-k+i}(1-\pi)^{l+i-2}\pi^{(i+1)r+1}.$$
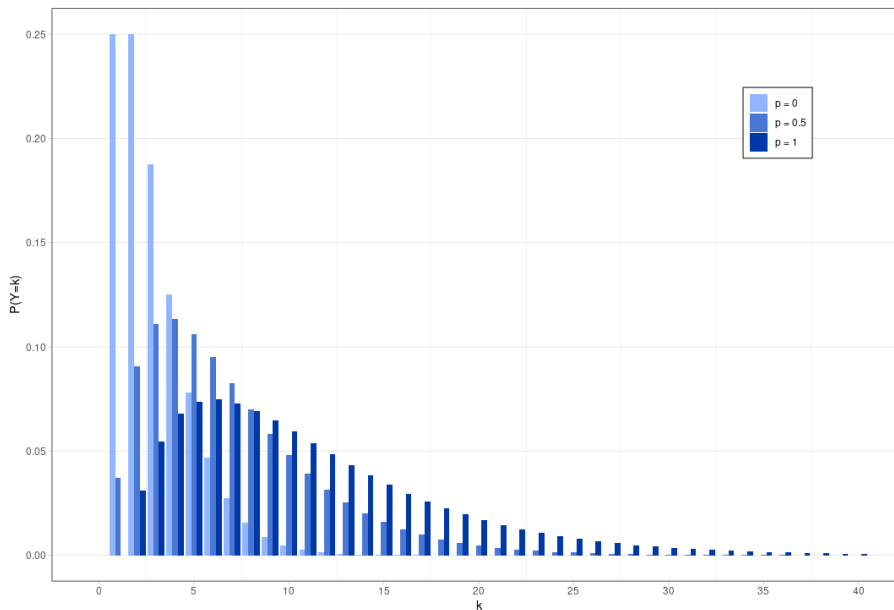

Figure 5: The degree distribution of a network when adding contacts to nodes at distance 2 with probability p to a branching process with offspring distributed according to a negative binomial distribution with size 2 and probability 0.5.

3.2 Expected Degree

It is also interesting to look at how the expected degree changes after adding the extra connections with the clustering probability p.

Lemma 3.1. *Consider a branching process with offspring distribution Z, and thus degree distribution $X = Z + 1$. If connections are independently added with probability p between all pairs of nodes at distance 2, the expected degree $\mathbb{E}(X)$ in the resulting network equals:*

$$\mathbb{E}(Y) = \mathbb{E}(X) + p\left(\mathbb{E}(Z)^2 + \mathbb{E}(Z) + \frac{Var(Z)}{\mathbb{E}(Z)}\right).$$

Proof. The new degree is the old degree and p times the number of nodes at distance 2, hence for the expected degree we get

$$\mathbb{E}(Y) = \mathbb{E}(X) + p\mathbb{E}(L)$$
$$= \mathbb{E}(X) + p\left(\mathbb{E}(G) + \mathbb{E}(N) + \mathbb{E}(S)\right)$$
$$= \mathbb{E}(X) + p\left(\mathbb{E}(Z)^2 + 1 + \mathbb{E}(S)\right),$$

where G is the number of grandchildren, N the number of grandparents and S the number of siblings.

Hence, we rewrite $\mathbb{E}(S)$:

$$\mathbb{E}(S) = \sum_{s=0}^{\infty} s\mathbb{P}(S = s)$$

$$= \sum_{s=0}^{\infty} s\frac{s+1}{\mathbb{E}(X)-1}\mathbb{P}(X = s+2)$$

$$= \sum_{s=0}^{\infty} s\frac{s+1}{\mathbb{E}(Z)}\mathbb{P}(Z = s+1)$$

$$= \frac{1}{\mathbb{E}(Z)}\sum_{k=1}^{\infty} k(k-1)\mathbb{P}(Z = k)$$

$$= \frac{1}{\mathbb{E}(Z)}\left(\sum_{k=2}^{\infty} k^2\mathbb{P}(Z = k) - \sum_{k=1}^{\infty} k\mathbb{P}(Z = k)\right)$$

$$= \frac{1}{\mathbb{E}(Z)}\left(\sum_{k=0}^{\infty} k^2\mathbb{P}(Z = k) - \mathbb{P}(Z = 0) - \sum_{k=0}^{\infty} k\mathbb{P}(Z = k) + \mathbb{P}(Z = 0)\right)$$

$$= \frac{1}{\mathbb{E}(Z)}\left(\mathbb{E}(Z^2) - \mathbb{E}(Z)\right)$$

$$= \frac{\mathbb{E}(Z^2)}{\mathbb{E}(Z)} - 1.$$

Substituting this in to the term for $\mathbb{E}(Y)$ gives the desired result:

$$\mathbb{E}(Y) = \mathbb{E}(X) + p\left(\mathbb{E}(Z)^2 + 1 + \frac{\mathbb{E}(Z^2)}{\mathbb{E}(Z)} - 1\right)$$

$$= \mathbb{E}(X) + p\left(\mathbb{E}(Z)^2 + \frac{\mathbb{E}(Z^2)}{\mathbb{E}(Z)}\right)$$

$$= \mathbb{E}(X) + p\left(\mathbb{E}(Z)^2 + \mathbb{E}(Z) + \frac{\text{Var}(Z)}{\mathbb{E}(Z)}\right).$$

$\square$

Figure 6 shows that for the three offspring distributions the obtained expressions for the new degree distribution agree with the found expression for the expected degree.

3.3 Clustering Coefficient

The addition of the connections at distance 2 to a branching type network adds clustering to the network. A way to measure this is by calculating the clustering coefficient of the network.

The clustering coefficient of a network can be defined as

$$C = \frac{3N_\triangle}{N_3},$$

where $N_\triangle$ is the number of triangles and N_3 is the number of connected triplets. A triplet is a set of three nodes that are connected by either two or three connections. If the triplet is connected by three connections it is also a triangle.

While the clustering coefficient is an important property of the network, it doesn't describe the network. It is not possible to replicate the network only from the clustering coefficient. Networks with very different structures can have the same clustering coefficient, so while it is nice to know this property of a network, we should remember its shortcomings.

Our aim is to find the clustering coefficient in a branching type network where individuals have a degree X (offspring X - 1) and where we added additional connections at distance 2 with probability p. The clustering coefficient is the expected number of triangles a random node in the
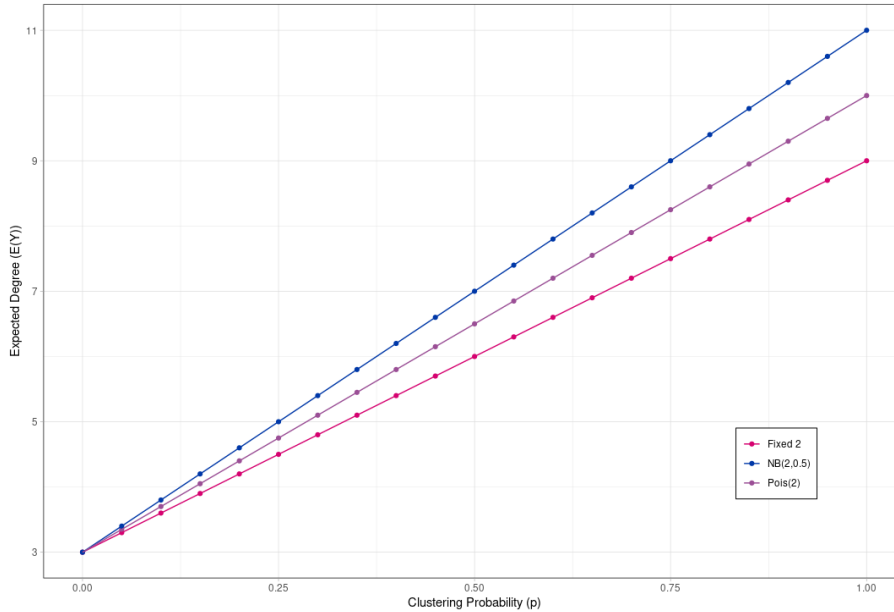
Figure 6: The expected degree of a network when adding contacts to nodes at distance 2 with probability p to a branching process with offspring distributed with distributions Pois(2) and NB(2,0.5), and a fixed offspring of 2 calculated using the expressions for $\mathbb{P}(Y = k)$.

network is part of divided by the expected number of triplets this node is the central node of. This node has degree distributed according to X, i.e., $X-1$ children, it has S siblings, G grandchildren, 1 parent and 1 grandparent. The number of connections after we add connections at distance 2 with probability p is given by Y. Hence an individual is the central node of a triplet between $\binom{Y}{2}$ pairs. We want to find out how many of these pairs are also directly connected to each other, such that there is a triangle in the network. Let an individual has s siblings (s), $x - 1$ children (c), g grandchildren (g), 1 parent (p) and 1 grandparent (n). These are all individuals at distance at most 2 in the original network.

We obtain the expected number of triangles by calculating the number of triangles of a random node given its number of siblings, children and grandchildren, $\triangle(s, c, g)$, and then multiplying it by the probability of the node having that number of siblings, children and grandchildren, $\mathbf{P}(s, c, g)$. The expected number of triplets can be calculated in the same manner, instead of the number of triangles, we calculate the number of triplets a random node is the central node of, given its number of siblings, children and grandchildren.

$$C = \frac{\sum_{s,c,g} \mathbf{P}(s, c, g) \triangle(s, c, g)}{\sum_{s,c,g} \mathbf{P}(s, c, g) \mathbf{T}(s, c, g)}.$$

We have already seen that an individual is the central node of a triplet between $\binom{Y}{2}$ pair, hence

$$\sum_{s,c,g} \mathbf{P}(s, c, g) \mathbf{T}(s, c, g) = \sum_k \binom{k}{2} \mathbf{P}(Y = k).$$

To calculate $\triangle(s, c, g)$ we consider all possible combinations of triplets a node can make with nodes at distance at most 2 in the original network. For every triplet, we look at the number of those triplets exists, the probability the triplet exists and the chance that the triplet is a triangle. The results are shown in Table 1.

For example a triple consisting of a random node its parent and one of its children, a pic triple, exists with probability 1, as connections are at distance 1. There are c of these triples. The probability that this triple is a triangle is the chance that a child is connected to its grandparent, which is p.

12

There are two kinds of triples where the probability of forming a triangle is a bit more complicated: gig and cig triples. For the gig triples, there exist $g(g-1)/2$ triples and the triple exist with probability p^2. The chance that two nodes of of type g are connected, and thus form a triangle, is more complicated. If the two grandchildren have the same parent, this probability is p. If they have different parents, this probability is 0. The chance that the second g has the same parent as the other g is denoted by $\mathbb{P}(c_{g_1} = c_{g_2}|c, g)$, here c_{g_1} denotes the child which is the parent of the first

For the cig triples, there are cg triples. These triples exist with probability p. The chance that c and g are connected is again more complicated. If g is a child of c, this probability is 1. If g is a child of a sibling of c, the probability is 0. Hence we need the chance that g is a child of c, which is $1/c$.

triple	# triples	triple probability	triangle probability
pic	c	1	p
pig	g	p	0
pis	s	p	1
pin	1	p	1
nic	c	p	0
nig	g	p^2	0
nis	s	p^2	p
sis	$s(s-1)/2$	p^2	p
sig	sg	p^2	0
sic	sc	p	0
cic	$c(c-1)/2$	1	p
gig	$g(g-1)/2$	p^2	$p\mathbb{P}(c_{g_1} = c_{g_2}|c, g)$
cig	cg	p	$1/c$

Table 1: Table showing the number of triplets, the probability the triplet exist and the probability that it forms a triangle for the different kinds of triplet.

Now the only thing left to do to obtain the number of triangles a random node is part of, is to obtain an expression for $\mathbb{P}(c_{g_1} = c_{g_2}|c, g)$.

$$\mathbb{P}(c_{g_1} = c_{g_2}|C = c, G = g) = \sum_{\gamma \in \Gamma_{c,g}} \mathbb{P}(\gamma|G = g, C = c)\mathbb{P}(c_{g_1} = c_{g_2}|C = c, G = g, \gamma), \qquad (1)$$

here $\Gamma_{c,g}$ is the set of all possible combinations of dividing g grandchildren over c children.

The second probability in the expression 1 is the probability that two grandchildren have the same child as parent, when the number of grandchildren per child are given. This probability is given by

$$\mathbb{P}(c_{g_1} = c_{g_2}|C = c, G = \{g_1, g_2, \ldots, g_c\}, \textstyle\sum_i g_i = g) = \sum_{i=1}^{c} \frac{g_i}{g} \frac{g_i - 1}{g - 1} = \frac{1}{g(g-1)} \sum_{i=1}^{c} g_i(g_i - 1),$$

here g_i is the number of children of child i. First there is a chance $\frac{g_i}{g}$ that the first grandchild is from child i and then a chance $\frac{g_i - 1}{g - 1}$ that the second grandchild is also from child i. As we only want the two grandchildren to have the same parent and not that the parent is a specific child, we sum this chance over all children.

Now we rewrite the first probability of expression 1:

$$\mathbb{P}(\gamma|G = g, C = c) = \frac{\mathbb{P}(\gamma)}{\mathbb{P}(G = g, C = c)} = \frac{\prod_{i=1}^{c} f(g_i)}{\mathbb{P}(G = g|C = c)\mathbb{P}(C = c)}.$$

here $f(g_i)$ is the density function. The probabilities in the denominator are obtained in Section 3.1.

$$\mathbb{P}(c_{g_1} = c_{g_2}|C = c, G = g) = \frac{1}{g(g-1)} \frac{1}{\mathbb{P}(G = g|C = c)\mathbb{P}(C = c)} \sum_{\gamma \in \Gamma_{c,g}} \prod_{i=1}^{c} f(g_i) \sum_{j=1}^{c} g_j(g_j - 1),$$

$$(2)$$

We look at the resulting clustering coefficient for a fixed number of offspring and offspring distributed according to the Poisson and negative binomial distributions.

3.3.1 Fixed Offspring

Again, we start with the situation of a branching process where every node has to same number of offspring λ, i.e.,

$$\mathbb{P}(X = k) = \begin{cases} 1 & \text{if } k = \lambda + 1 \\ 0 & \text{otherwise} \end{cases}.$$

For this offspring distribution we know that every node has λ children, λ^2 grandchildren and $\lambda - 1$ siblings. This makes it easy to calculate $\mathbb{P}(c_{g_1} = c_{g_2}|c, g)$, as $\Gamma_{c,g}$ has only one element, which is $g_1 = \lambda, g_2 = \lambda, \ldots, g_\lambda = \lambda$. Hence

$$\mathbb{P}(c_{g_1} = c_{g_2}|c, g) = \frac{1}{g(g-1)} \sum_{j=1}^{c} g_j(g_j - 1) = \frac{cg_1(g_1 - 1)}{g(g-1)}.$$

Thus, the number of triangles a random node is part of is

$$\sum_{s,c,g} \mathbf{P}(s,c,g)\triangle(s,c,g) = \lambda p + (\lambda - 1)p + p + (\lambda - 1)p^3 + (\lambda - 1)(\lambda - 2)p^3/2 + \lambda(\lambda - 1)p/2$$

$$+ \lambda^2(\lambda^2 - 1)p^3 \frac{\lambda^2(\lambda - 1)}{2\lambda^2(\lambda^2 - 1)} + \lambda^3 p/\lambda$$

$$= \frac{1}{2}\lambda p(3\lambda + \lambda^2 p^2 - p^2 + 3).$$

Using the degree distribution obtain in Section 3.1.1 we get

$$\sum_k \binom{k}{2} \mathbf{P}(Y = k) = \frac{1}{2}\lambda(1 + \lambda)(1 + 2(1 + \lambda)p + (-1 + \lambda + \lambda^2)p^2).$$

Combining our results we get that the clustering coefficient after adding the extra connections becomes

$$C = \frac{\frac{1}{2}\lambda p(3\lambda + \lambda^2 p^2 - p^2 + 3)}{\frac{1}{2}\lambda(1 + \lambda)(1 + 2(1 + \lambda)p + (-1 + \lambda + \lambda^2)p^2)} = \frac{p((\lambda - 1)p^2 + 3)}{(\lambda^2 + \lambda - 1)p^2 + 2(\lambda + 1)p + 1}.$$

3.3.2 Poisson Distribution

For a branching process for which the offspring which is Poisson distribution with rate λ we have shown in Section 3.1.2 that the number of children is Poisson distributed with rate λ and the number of grandchildren given i children is Poisson distributed with rate $i\lambda$ and the number of siblings is again Poisson distributed with rate λ. Because of this the probability that a second grandchild has the same parent as the other grandchild given the number of grandchildren and children, $\mathbb{P}(c_{g_1} = c_{g_2}|c, g)$ equals to $1/c$.

We obtain that the number of triangles a random node is part of is $\frac{1}{2}\lambda p(3\lambda + (\lambda^2 + \lambda + 2)p^2 + 4) + p$.

$$\sum_{s,c,g} \mathbf{P}(s,c,g)\triangle(s,c,g) = \lambda p + \lambda p + p + \lambda p^3 + \lambda(\lambda - 1)p^3/2 + \lambda(\lambda - 1)p/2$$

$$+ \lambda^2(\lambda^2 - 1)p^3/2\lambda + \lambda^3 p/\lambda$$

$$= \frac{1}{2}\lambda p(3\lambda + (\lambda^2 + \lambda + 2)p^2 + 4) + p.$$

The number of triplets with the node as central node is $\sum_k \binom{k}{2} \mathbf{P}(Y = k)$, which equals for the offspring distribution obtained in Section 3.1.2 to $\frac{1}{2}(\lambda(\lambda + 2) + \lambda(\lambda + 2)(\lambda^2 + \lambda + 1)p^2 + 2(\lambda(\lambda + 1)(\lambda + 2) + 1)p)$. Hence the clustering coefficient C equals to

$$C = \frac{p(\lambda(3\lambda + (\lambda^2 + \lambda + 2)p^2 + 4) + 2)}{(\lambda(\lambda + 2) + \lambda(\lambda + 2)(\lambda^2 + \lambda + 1)p^2 + 2(\lambda(\lambda + 1)(\lambda + 2) + 1)p)}.$$
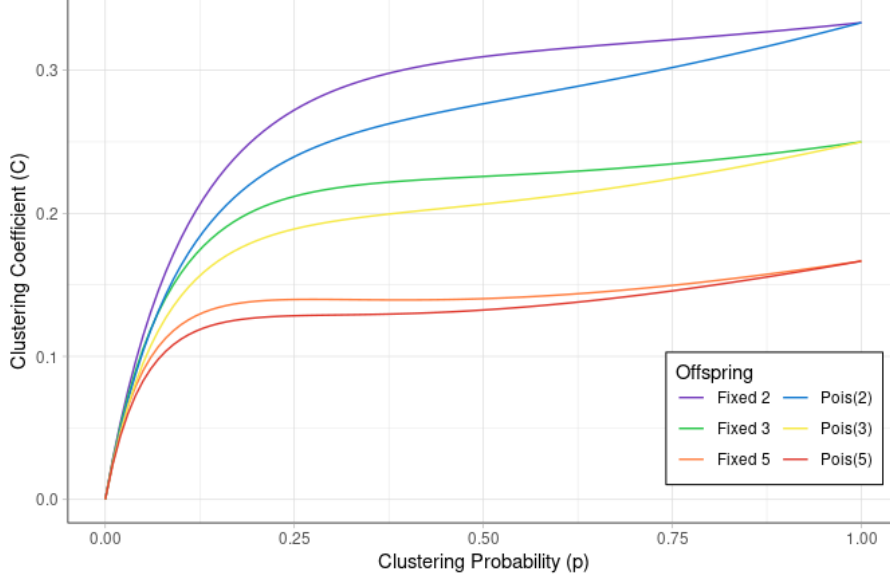

Figure 7: The clustering coefficient of a network when adding contacts to nodes at distance 2 with probability p to a branching process with a fixed offspring and a offspring distributed according to a Poisson distribution.

3.3.3 Negative Binomial Distribution

To be able to express the clustering coefficient for a network where connections are added to a branching process which offspring follows a negative binomial distribution, we have to calculate $\mathbb{P}(c_{g_1} = c_{g_2} | C = c, G = g)$. Equation 2 states that we need $\Gamma_{c,g}$, which is the set of all possible combinations of dividing g grandchildren over c children. The number of offspring of a single node in a branching process which offspring is distributed according to the negative binomial distribution has no upper bound, thus it is not possible obtain $\Gamma_{c,g}$. Because of this, we weren't able to express the clustering coefficient when the number of offspring is negative binomial distributed.

4 Epidemic

In model 2 an epidemic is simulated on the contact network. In this chapter we look into the effect of the clustering added to the contact network on the epidemic. We do this by determining the basic reproduction number, R_0, of the epidemic simulated for different values of the transmission and clustering probabilities.

As stated in Chapter 2, we simulate a generation based epidemic. Every infected node independently infects each susceptible neighbour with the transmission probability q. Each node is infectious during one generation. The epidemic is started with one initial case.

4.1 Analytical Results

Ideally we want an analytical expression for the basic reproduction number of the epidemic simulated in model 2, $R_0^{(2)}$. Because of the structure of the contact network, we did not succeed in obtaining this expression. The added contacts make it possible for the infection to climb back up the tree like structure of the contact network after multiple infection generations.

However we are able to give an expression for $R_0^{(2)}$ for some values of q and p. For the trivial case $q = 0$, it is obvious that $R_0^{(2)} = 0$.

In the case of $p^{(2)} = 0$, the contact network has a branching type structure. This makes it easy to calculate $R_0^{(2)}$. An infected node will infect its susceptible contacts with probability q. In a branching type network, the number of susceptible contacts is the number of offspring a node has. This is because one of its contacts is the contact over which the node is infected. Because of the structure none of its other contacts will be infected by another node. Thus we get that if $p^{(2)} = 0$,

$$R_0^{(2)} = q\mathbb{E}(O_b^{(2)})$$

It is also possible to get an analytical expression for $R_0^{(2)}$ if $q = 1$. In this case every individual will become infected. We want to know the expected number of individuals a single individual infects. Because of the added contacts, it is possible that a susceptible node has multiple possible sources of infection. The possible infectors are always the parent and grandparent in the contact tree, as sibling contacts do not play a role when $q = 1$.

Suppose that the grandparent is always assigned to be the infector of their grandchild if the connection between them exists. Then, a node will infect one of its grandchildren with probability $p^{(2)}$, i.e., if that connection is added, and a child with prob $1 - p^{(2)}$, i.e., if the child does not have a connection with their grandparent. A node is expected to have $\mathbb{E}(O_b^{(2)})^2$ grandchildren and $\mathbb{E}(O_b^{(2)})$ children. Thus, we get that if $q = 1$ for all values of $p^{(2)}$

$$R_0^{(2)} = \mathbb{E}(O_b^{(2)})^2 p^{(2)} + \mathbb{E}(O_b^{(2)})(1 - p^{(2)}).$$

4.2 Simulation Results

In this section we will look at the simulation results. From the simulations the reproduction number can be estimated by dividing the number of infected nodes of each generation i with the number of infected nodes in generation $i - 1$. When multiple simulations are used, the reproduction number is estimated by

$$R_0^{(2)} = \frac{\sum_{j=1}^M \mathrm{Inf}_i^{(j)}}{\sum_{j=1}^M \mathrm{Inf}_{i-1}^{(j)}},$$

where M is the number of iterations and $\mathrm{Inf}_i^{(j)}$ the number of infected in generation i in iteration j.

To know from which generation the estimated $R_0^{(2)}$ is stable, we looked at the behaviour of $R_0^{(2)}$ per generation. This is shown in Figure 8. This figure shows that after infection generation 5 the estimates are stable. Based on these findings, we decided to take the result of infection generation 7 and use 10000 iterations of the simulation to estimate the reproduction number for all the results in this project.
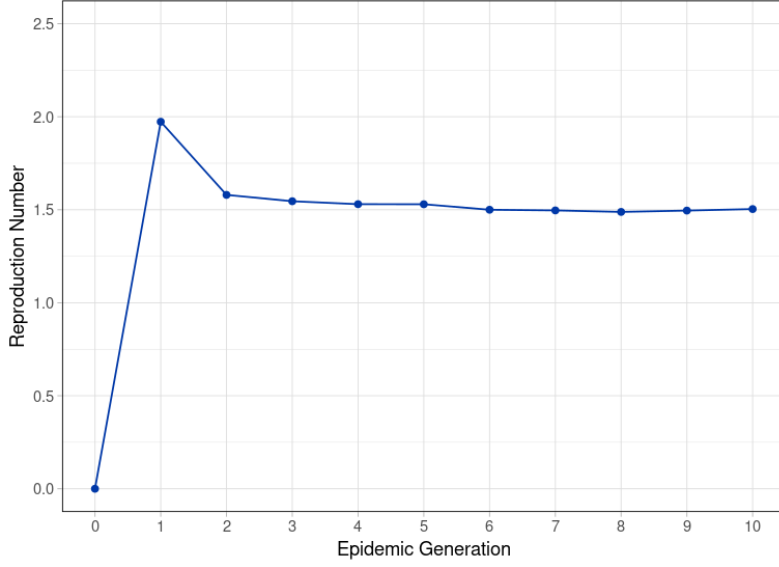
Figure 8: $R_0^{(2)}$ estimated for every epidemic generation using model 2 with $p^{(2)} = 0.5$, $q = 0.326$ and $O_b^{(2)} = \text{Pois}(2)$ using 1000 iterations of the simulation.

We simulated epidemics on networks created with three offspring distributions $O_b^{(2)}$: fixed 2, Pois(2) and NB(2, 0.5), all with an expected value of 2 and a variance of 0, 2 and 4, respectively. Figures 9, 10 and 11 show the reproduction numbers for different values of the transmission and clustering probability.

First it is good to take note of that the simulation results depicted in Figures 9, 10 and 11 agree with our analytical findings for $p = 0$ or $q = 1$, which are independent of the offspring distribution $O_b^{(2)}$. Note that when $q = 1$, $R_0^{(2)} = 2 + 2p^{(2)}$ for all three offspring distributions. This agrees with our analytical finding, as

$$R_0^{(2)} = \mathbb{E}(O_b^{(2)})^2 p^{(2)} + \mathbb{E}(O_b^{(2)})(1 - p^{(2)}) = 2^2 p^{(2)} + 2(1 - p^{(2)}) = 2 + 2p^{(2)}.$$

These analytical results create boundaries for $R_0^{(2)}$ with other values of $p^{(2)}$ and q. As a consequence, the differences between results for the different offspring distributions are not that large. Looking at Figures 9b, 10b and 11b, we see that the curve of the reproduction number initially grows faster when increasing the transmission probability if the variance of the offspring is larger.
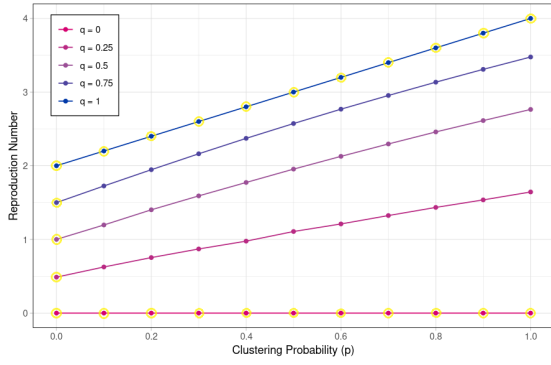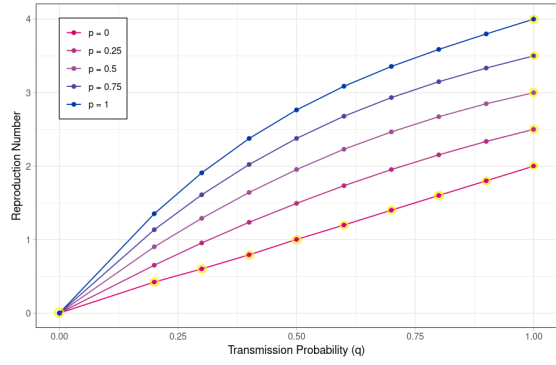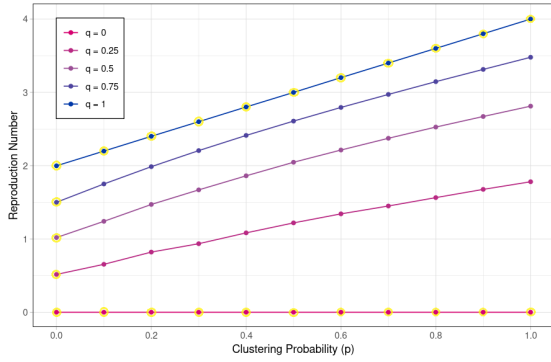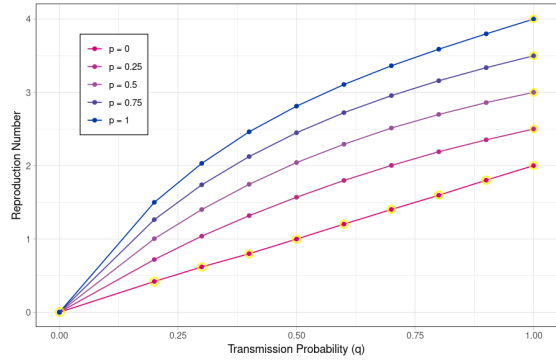
Figure 9: The basic reproduction number of the simulated epidemics over a contact network resulting from model 2 with clustering probability p, transmission probability q and a branching process with a fixed number of offspring of 2. The highlighted nodes agree with our analytical findings.
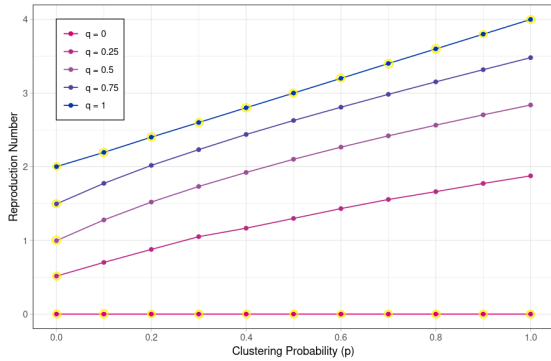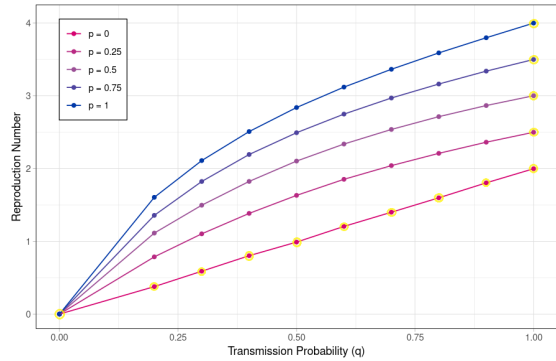

Figure 10: The basic reproduction number of the simulated epidemics over a contact network resulting from model 2 with clustering probability p, transmission probability q and a branching process with offspring distributed according to the Poisson distribution with rate 2. The highlighted nodes agree with our analytical findings.

Figure 11: The basic reproduction number of the simulated epidemics over a contact network resulting from model 2 with clustering probability p, transmission probability q and a branching process with offspring distributed according to a negative binomial distribution with size 2 and success probability 0.5. The highlighted nodes agree with our analytical findings.

5 Contact Tracing

5.1 Description

Contact tracing is a strategy used to contain the spread of a infectious agent. When an infected individual is detected, they report their contacts to find potentially infected individuals. Those contacts get warned and take the needed measures to prevent further spread. In previous publications it has been shown that the effectiveness of contact tracing depends on testing and tracing delays [10],[11], [12]. These publications state that the effectiveness improves if these time delays are short.

Contact tracing can be modeled in various ways. Taking account of these time delays, and other aspects such as adherence to isolation, requires a detailed model and additional assumptions. We chose to implement a very simplified form of contact tracing. The main reason for this that within this project we wanted to focus on the differences between the networks resulting from the different models, instead of aspects such as time delays and adherence which are already known to be important.

For the implementation of contact tracing we introduce two new probabilities: a detection probability, d, and a tracing probability, t. The detection probability is the chance that an infected node will be detected, for example if a infected individual will show symptoms after infection. The tracing probability is the probability that a contact of a detected node will be successfully traced.

We implemented contact tracing in our models as follows. An infected node will be detected with the detection probability, d. The contacts of this node will be traced. Tracing is successful with the tracing probability, t, and all tracing successes are independent. The successfully traced contacts will be warned and won't infect any new contacts. So, detection is not fast enough to prevent a detected node from infecting its contacts, but tracing prevents secondary cases from the traced contacts.

We assumed that a successfully traced node will have zero new contacts, i.e., the individual will be in perfect quarantine. A natural extension would be to add imperfect quarantine by introducing another probability which will determine the chance that a contact of a traced node will be prevented by the quarantine of a traced individual.

5.2 Results

In this section we look at the simulation results of contact tracing on both models separately. We do this, because the networks resulting from both models are not similar and thus their results can not be compared.

5.2.1 Model 1

The first thing to note is that in model 1 sibling contacts have no impact on tracing as they are in the same infection generation. In our implementation of tracing, when tracing a node in the same infection generation, the traced node will be found after it already has infected its contacts.

If there is no clustering in the network, i.e. $p^{(1)} = 0$, we can find an expression for the reproduction number under tracing. In this case the fraction d of infected nodes will be detected and the fraction t of their contacts will be traced. Thus further spread is halted in a fraction td of the secondary cases. So, we get that if $p^{(1)} = 0$ for all values of t and d,

$$R^{(1)}_{\text{tracing}} = R^{(1)}_0 (1 - td).$$

Suppose $t = 1$, then all contacts of a detected node will be traced. The added contacts in this situation will not add to preventing more infections, because these added contacts are with nodes of whose infection will be prevented, as their infector will be traced before infecting this node (see Figure 12). Thus the fraction d of infections will be prevented. So, if $t = 1$ we have for all values of $p^{(1)}$,

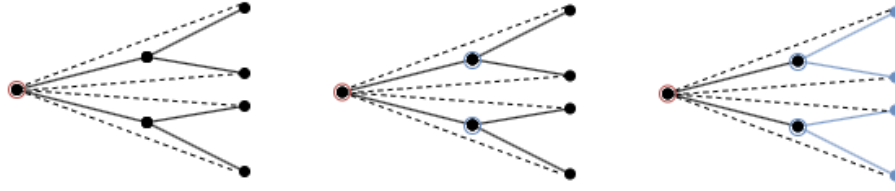$$R^{(1)}_{\text{tracing}} = R^{(1)}_0 (1 - d).$$

19

Figure 12: Contact tracing on a tracing network created with model 1 when $t = 1$. Transmission contacts depicted with solid lines and possible added contacts depicted with dashed lines. The node circled in red is infected and detected. This node infects its contacts, circled in blue. If the red case is detected, these contacts are traced. The contacts of the traced nodes are prevented, depicted in blue. Tracing over added contacts (dashed lines) will not prevent any infections, because the nodes traced using these contacts are not infected. Thus clustering has no effect on the effectiveness of contact tracing.

The simulation results for contact tracing with model 1 are depicted in Figures 13 and 14. In Figure 13 the reproduction number is shown when contact tracing is simulated for different values of t and d with model with the offspring distribution fixed 2, Pois(2) and NB(2, 0.5) all with clustering probability 0.5. The dashed lines show the reproduction number when there is no clustering.
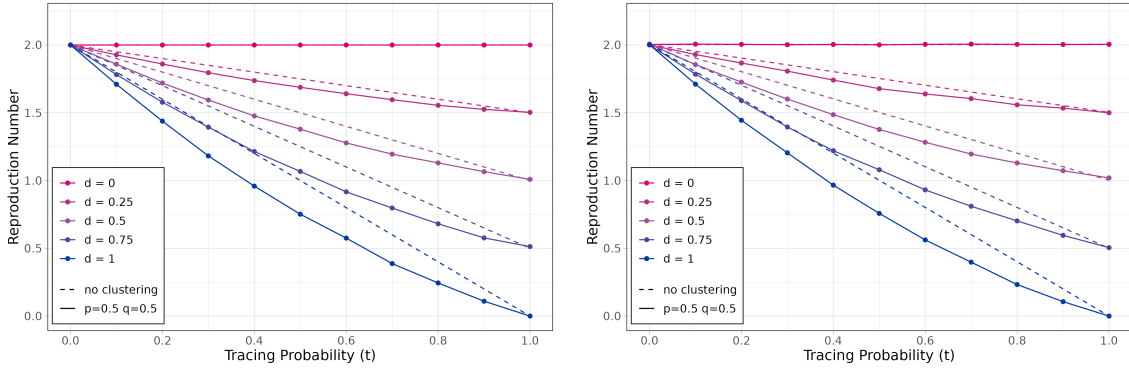
(a) Offspring distributed with a fixed number of 2

(b) Offspring distributed according to Pois(2)
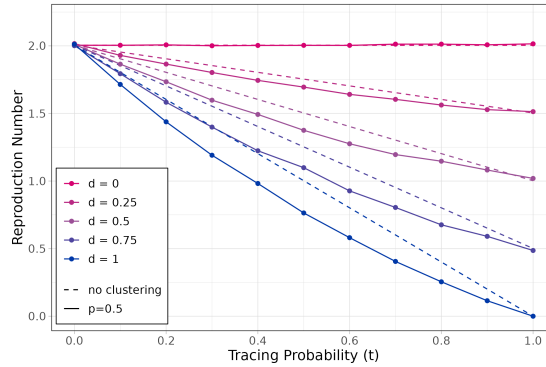
(c) Offspring distributed according to NB(2, 0.5)

Figure 13: The reproduction number with contact tracing for different tracing and detection probabilities using model 1 with different offspring distributions. The solid lines depict the situation with $p^{(1)} = 0.5$ and the dashed lines the situation without clustering.

First we notice that the results depicted in Figures 13 and 14 agree with our analytical findings for $p^{(1)} = 0$ or $t = 1$. We also see that the figures look really similar for the different offspring distributions. Based on our simulations we can not rule out that there are no differences between

the results for different offspring distributions.

We know from the analytical results that there is no effect of the clustering on contact tracing when $t = 1$ and we also know that per definition there is no tracing effect at all when $t = 0$. These results show that the clustering has an effect when t is intermediate.

In Figure 14 the reproduction number under tracing is depicted for different values of $p^{(1)}$ and t. The detection probability is 0.5 in all cases and the results for the three different offspring distributions are shown. Again, based on our simulations we can not rule out that there are no differences between the results for different offspring distributions.

(a) Offspring distributed with a fixed number of 2   **(b)** Offspring distributed according to Pois(2)
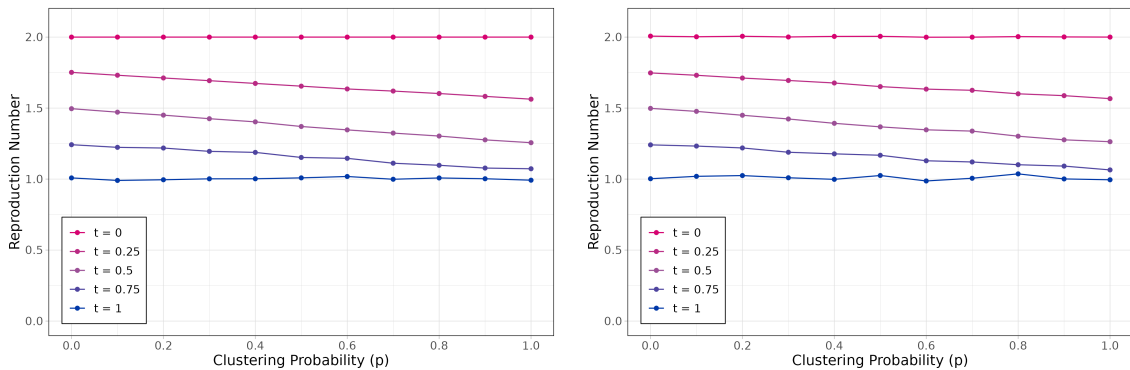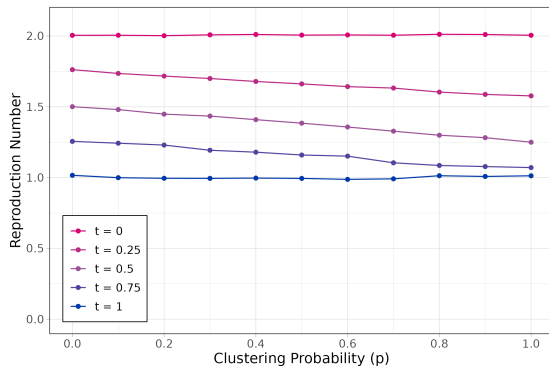
(c) Offspring distributed according to NB(2, 0.5)

Figure 14: The reproduction number with contact tracing for different tracing and clustering probabilities using model 1 with different offspring distributions. The detection probability is fixed at 0.5.

5.2.2   Model 2

In model 2 we also have that if $p^{(2)} = 0$ then $R^{(2)}_{\text{tracing}} = R^{(2)}_0 (1 - td)$. This is because if there is no clustering in the contact network, the tracing and transmission network are equal and have a branching type structure. So, the network has the same structure as with model 1, and thus the same arguments apply when $p^{(2)} = 0$.

The simulation results for contact tracing with model 2 are shown in Figures 15 and 16. In Figure 15 contact tracing is simulated for different values of t and d with model with the offspring distribution fixed 2, Pois(2) and NB(2, 0.5) all with $p^{(2)} = 0.5$ and $q = 0.5$. Also the case in absence of clustering but with the same reproduction number is depicted with the dashed lines.

In Figure 16 the reproduction number under tracing is depicted for different values of clustering and tracing probability. The detection and transmission probability are both 0.5 in all cases. The simulations are done for the three different offspring distributions.

Where we see in Figure 14, which describes the reproduction number using model 1, that for $t = 0$ and $t = 1$ the reproduction number remains constant for increasing $p^{(1)}$ and the other
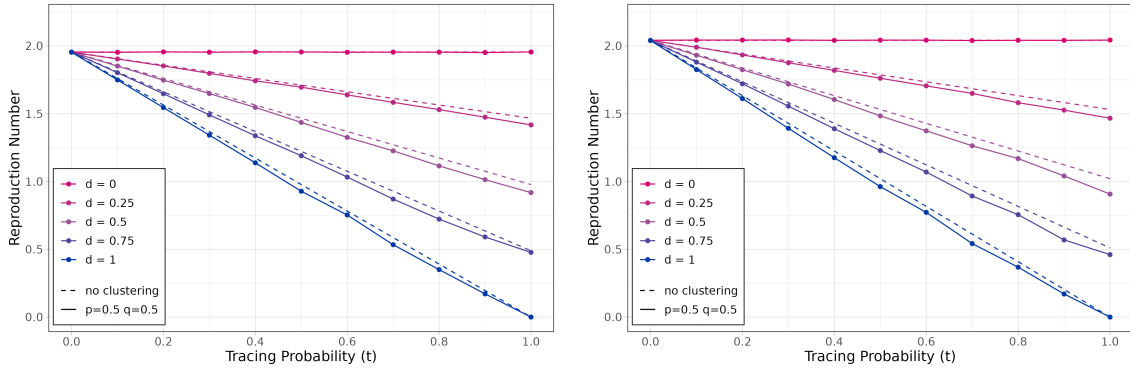
(a) Offspring distributed with a fixed number of 2

(b) Offspring distributed according to Pois(2)
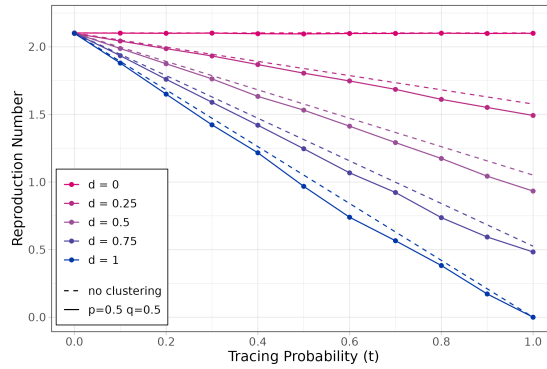
(c) Offspring distributed according to NB$(2, 0.5)$

Figure 15: The reproduction number with contact tracing for different tracing and detection probabilities using model 2 using different offspring distributions creating the branching process. The solid lines depict the situation with $p^{(2)} = 0.5$ and $q = 0.5$. The dashed lines the situation for a network with the same reproduction number without clustering.

lines decrease within these boundaries, we see in 16 that the reproduction number increases for all values of t. This difference can be explained by the fact that when clustering probability in model 2 increases, the number of contacts in the contact network also increases and thus the number of transmissions increases, while if the clustering probability increases in model 1, only the number of contacts used for tracing will increase.

Unlike the results using model 1, we do see some differences between the effect on the reproduction number using different offspring distributions. This can be due to the effect of the variance of the offspring distributions on the reproduction number without tracing (see Chapter 4).

(a) Offspring distributed with a fixed number of 2

(b) Offspring distributed according to Pois(2)

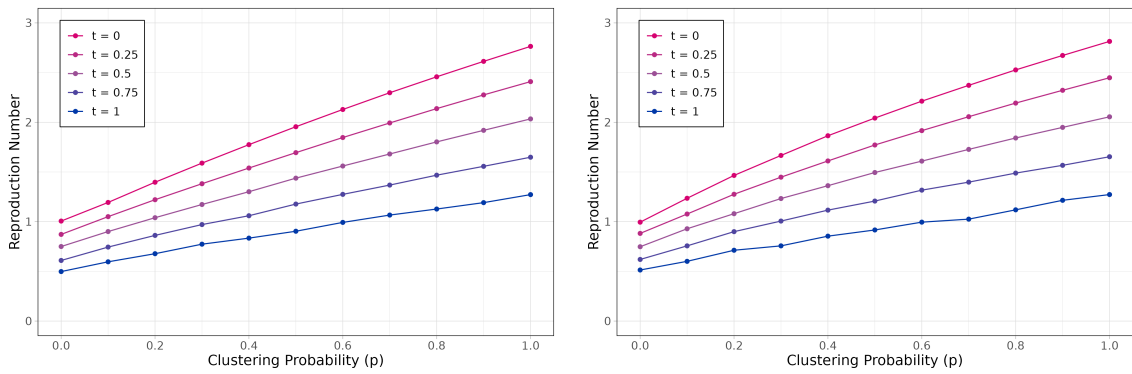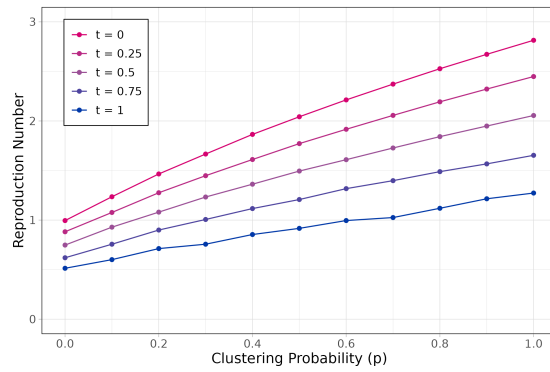**(c)** Offspring distributed according to NB(2, 0.5)

Figure 16: The reproduction number with contact tracing for different tracing and clustering probabilities using model 2 with different offspring distributions. The detection and transmission probabilities both fixed at 0.5.

6 Comparing the Models

In this chapter we we want to answer the main question: Does a model that adds contacts to the transmission network approach the effect of clustering in the underlying contact network on contact tracing as seen in a model in which the infection spreads on a clustered contact network?

6.1 Choosing parameters

The main objective was to compare the effectiveness of tracing in models 1 and 2 that were introduced and analysed in the previous sections. To this end, we need to simulate epidemics in both models that are identical in some aspects and differ almost only in how the networks are created. We wanted the resulting network of model 1 to have a similar tracing network as model 2, as to be able to compare the effectiveness of contact tracing in both models. Thus we chose the input parameters of model 1 based on the resulting networks of model 2.

6.1.1 Parameters Model 1

Model 1 has two input parameters: the offspring distribution of the transmission network $O_e^{(1)}$ and the clustering probability $p^{(1)}$. We try to choose the parameters such that some properties of the transmission and tracing networks of model 1 are identical to those of model 2.

First we look at the offspring of the transmission network $O_e^{(1)}$. A logical way to assign an offspring distribution to model 1 is using the empirical offspring distribution of the transmission network resulting from model 2, $O_e^{(2)}$. This would result in the most similar network. A downside of this method is that we do not have an analytical expression for this distribution and thus we will not have any analytical results to compare with.

We assign a distribution to $O_e^{(1)}$ by estimating the mean and variance of the empirical distribution and choosing a known probability distribution with that mean and variance. With this method we obtain a offspring distribution for model 1, which results in a similar network and we can use all the known relations and properties of this distribution. After looking at the simulation results of model 2, we found that a good approximation of the empirical offspring distribution of the transmission network resulting from model 2 depends on the chosen offspring distribution used for making the contact network. When a Poisson or negative binomial distribution is used for obtaining the contact network in model 2, a negative binomial distribution can be chosen for $O_e^{(1)}$ with the mean and variance equal to those of $O_e^{(2)}$. If a fixed offspring is used, a binomial distribution is a better choice.

We fit the negative binomial distribution by expressing the success probability and number of successes in terms of the mean and variance of the distribution of the transmission network in model 2. The mean μ and variance σ^2 of a random variable following a negative binomial distribution can be expressed in terms of the parameters: $\mu = \frac{r(1-\pi)}{\pi}$ and $\sigma^2 = \frac{r(1-\pi)}{\pi^2}$, where π is the success probability and r the number of successes. Hence, the parameters can be expressed in terms of the mean and variance:

$$\pi = \frac{\mu}{\sigma^2}, \qquad r = \frac{\mu^2}{\sigma^2 - \mu}.$$

We use these expressions to fit the negative binomial distribution such that the distribution has the same mean and variance as the empirical distribution.

We fit the binomial distribution in a similar way. The mean μ and variance σ^2 of a random variable following a binomial distribution can be expressed in terms of the parameters: $\mu = n\pi$ and $\sigma^2 = n\pi(1 - \pi)$, where π is the success probability and n the number of trials. Hence, the parameters can be expressed in terms of the mean and variance:

$$n = \frac{\mu^2}{\mu - \sigma^2}, \qquad \pi = 1 - \frac{\sigma^2}{\mu}$$

However, using the mean and variance from our empirical distribution, we obtain $n \notin \mathbb{N}$. We prioritized that the mean of the distribution is equal to the empirical distribution, thus we assign

the parameters with

$$n = \left\lceil \frac{\mu^2}{\mu - \sigma^2} \right\rceil, \qquad \pi = \frac{n}{\mu}.$$

The other input parameter in model 1 is the clustering probability $p^{(1)}$. We calculate this value using Lemma 3.1. We find that

$$p = \frac{\mathbb{E}(Y) - \mathbb{E}(X)}{\mathbb{E}(Z)^2 + \mathbb{E}(Z) + \frac{\text{Var}(Z)}{\mathbb{E}(Z)}},$$

where $\mathbb{E}(Y)$ is the expected degree of the tracing network, $\mathbb{E}(X)$ the expected degree of the transmission network and $\mathbb{E}(Z)$ and $\text{Var}(Z)$ the expected degree and variance of the offspring of the transmission network. The offspring distribution of the transmission network of model 1 is chosen such that its mean and variance equals those of the transmission network of model 2, thus $\mathbb{E}(O_e^{(1)}) = \mathbb{E}(O_e^{(2)}) = R_0^{(2)}$, $\text{Var}(O_e^{(1)}) = \text{Var}(O_e^{(2)})$, and $\mathbb{E}(D_e^{(1)}) = R_0^{(2)} + 1$. We want the degree of the tracing network resulting from model 1 to be similar to the degree of tracing network resulting from model 2.We obtain this by setting the expected degree of the tracing network in model 1 to be equal to the expected degree of the tracing network in model 2, so we have $\mathbb{E}(D_t^{(1)}) = \mathbb{E}(D_t^{(2)})$. This gives us the following value for the clustering probability used for model 1,

$$p^{(1)} = \frac{\mathbb{E}(D_t^{(2)}) - (R_0^{(2)} + 1)}{(R_0^{(2)} + 1)R_0^{(2)} + \frac{\text{Var}(O_e^{(2)})}{R_0^{(2)}}},$$

where $\mathbb{E}(D_t^{(2)})$, $\text{Var}(O_e^{(2)})$ and $R_0^{(2)}$ are obtained from the simulations of model 2.

6.1.2 Parameters Model 2

Model 2 has three input parameters: the offspring distribution of the generating network $O_g^{(2)}$, the clustering probability $p^{(2)}$ and the transmission probability q. We want to assign values to these parameters such that the effects of different aspects of the networks are shown in the results of the two models.

We want to examine the effect of the variance in the offspring distribution. In Section 3.2 we saw that the variance has an effect on the expected degree of the contact network and its clustering coefficient, so it is interesting to see how it influences the effectiveness of contact tracing. Hence, we choose three different offspring distributions all with the same mean, but different variances. Namely, a fixed offspring of 2, which variance equals 0, an offspring which is Poisson distributed with rate 2, which variance equals 2, and an offspring which is distributed according to the negative binomial distribution with size 2 and probability 0.5, off which the variance equals 4.

Another effect we want to examine is the effect clustering has on the effectiveness of contact tracing. Hence, we want networks with different degrees of clustering. This way we see what happens if there is no clustering, moderate clustering and high clustering. When there is no clustering the contact network of model 2 is a branching process and thus the transmission network and tracing network will be the same. The offspring of the transmission network in model 1 will be based on the offspring of this transmission network and no clustering will be added. Thus the resulting tracing networks of model 1 and 2 will be very similar. So, we don't have to compare the networks for the case $p^{(2)} = 0$. Thus, we want to choose two values of $p^{(2)}$ for which there is some clustering and a lot of clustering.

To be able to compare the effect these parameters have on the effectiveness of the contact tracing, we want that the reproduction number of the simulated epidemics is constant over all simulations. We aimed for a reproduction number of 1.5, as an epidemic with a reproduction number under 1 is already dying out, hence contact tracing is not necessary. To obtain this reproduction number we used simulations to find the transmission probability matching each combination of offspring distribution and clustering probability for the clustering probability and offspring distribution.

So, the simulations will be done for the three different $O_b^{(2)}$ and for each offspring distribution two different $(p^{(2)}, q)$-pairs are used, where $p^{(2)} = 0.5, 1$ and q such that $R_0^{(2)}$ equals 1.5. Thus, in total the results of 6 different simulations will be compared.

6.2 Simulation Results

$O_b^{(2)}$	$p^{(2)}$	q	$R_0^{(2)}$	$\mathrm{Var}(R_0^{(2)})$	$C_c^{(2)}$	$\mathbb{E}(D_t^{(2)})$	$C_t^{(2)}$	$R_{\mathrm{tracing}}^{(2)}$	$p^{(1)}$	$\mathbb{E}(D_t^{(1)})$	$C_t^{(1)}$	$R_{\mathrm{tracing}}^{(1)}$
Fixed 2	0.5	0.358	1.5019	1.2417	0.3098	4.2186	0.3795	1.1003	0.3745	4.2217	0.3160	1.0642
	1	0.224	1.5009	1.2431	0.3333	4.8121	0.4238	1.0935	0.5044	4.8185	0.3368	1.0364
Pois(2)	0.5	0.326	1.5071	2.1040	0.2776	4.6876	0.3883	1.0937	0.4214	4.7014	0.3097	1.0497
	1	0.201	1.5061	2.1135	0.3354	5.7781	0.4876	1.0862	0.6319	5.7956	0.3435	1.0081
NB(2, 0.5)	0.5	0.301	1.5003	2.7576	0.2676	5.1920	0.4016	1.0893	0.4816	5.2097	0.3151	1.0490
	1	0.182	1.5000	2.7158	0.3535	6.8771	0.5470	1.0643	0.7872	6.9012	0.3731	0.9886

Table 2: Table showing the results for the six different simulations of the matched models. Contact tracing simulated using $t = 0.5$ and $d = 0.5$. Results are based on 10000 generated networks on which we ran an epidemic with contact tracing.

In Table 2 the results are summarised. We see that we did succeed in getting similar expected degrees of the resulting tracing networks. To obtain this, $p^{(1)}$ is assigned a smaller value than $p^{(2)}$, but $p^{(1)}$ does increase when an offspring distribution with a higher variance is used.

It is interesting to note that with a similar expected degree, $C_t^{(1)}$ is significant lower then $C_t^{(2)}$. This could be explained by the fact that the nodes in the tracing network from model 2 are nodes that got infected during the simulated epidemic, so these nodes are not equal to the nodes in the tracing network from model 1.

For the reproduction numbers with tracing in the table $t = 0.5$ and $d = 0.5$ are used. To obtain more information about the difference between the effectiveness of tracing, we simulated tracing over the obtained networks for more values of t. In Figure 17 the reproduction number with contact tracing for the 6 different situations is depicted for different values of t while the detection probability is fixed at 0.5.

These results show that contact tracing is more effective using model 1 for almost all tracing probabilities. This difference gets bigger when the clustering probabilities are higher.

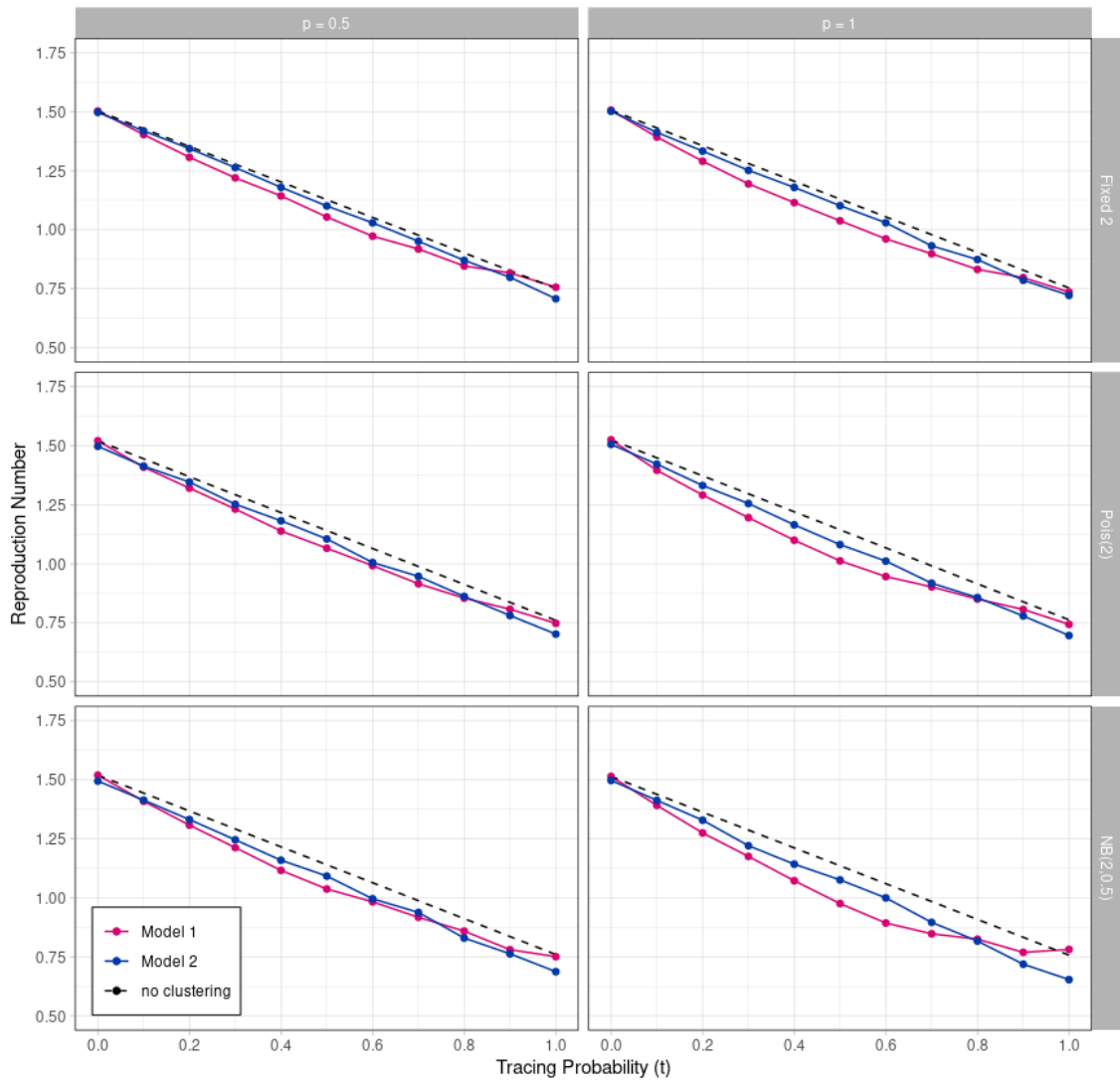**Figure 17:** The reproduction number with contact tracing using the 6 different simulations of the matched models for different tracing probabilities. Detection probability is fixed at 0.5. The dotted line depicts the situation without clustering.

7 Discussion

In this project we introduced two models with the aim to compare the effectiveness of contact tracing using these models. Model 1 is based on the model used in the CoronaMelder study and adds extra connections to the transmission network to include the effect of clustering on contact tracing. Model 2 creates a clustered contact network by adding connections to a branching process. Over this contact network an epidemic is simulated.

We succeeded in expressing some network properties of a branching process with added contacts at distance 2 with probability p. We obtained an expression for the degree distribution and the expected degree. We were also able to express the clustering coefficient if the offspring is a fixed number or distributed according to a Poisson distribution.

Model 2 simulates an epidemic over its clustered contact network. We were able to obtain some analytical and simulation results on the effect of the added clustering on the reproduction number of this epidemic. When the clustering probability increases, the reproduction number also increases, as more contacts are added to the network, but the increase in the reproduction number is not linear in the clustering probability. We also found that for a offspring distribution with a higher variance but equal mean, the reproduction number is greater or equal then when a offspring distribution wit a lower variance is used for equal values of the transmission and clustering probability.

We introduced and implemented a simplified form of contact tracing that captured only the essential factors, namely detection and tracing of contacts. We saw that the added clustering does have an effect on the effectiveness of contact tracing using both models. Using model 1, we saw that the effect was largest for an intermediate tracing probability and based on the results we could not rule out that there are no differences between the results for different offspring distributions. Using model 2, we saw a more conservative effect, which could be explained by the possibility of alternative infection paths. There were also some differences between the results for different offspring distributions, which could be due to the effect the variance of the offspring distribution has on the reproduction number without tracing.

To be able to compare the effectiveness of contact tracing using both models, we matched the resulting transmission and tracing networks. We assigned the input parameters for model 1 such that the offspring distribution of the transmission network has a similar mean and variance and the tracing network has a similar expected degree as the respective networks of model 2.

The results showed that the effectiveness of contact tracing was higher using model 1 for almost all values of the tracing probability. This difference increases when more clustering is introduced into the networks. This suggest that model 1 overestimates the effect of clustering in the contact network on the effectiveness of contact tracing. This could be due to that in model 2 clustering causes alternative infection paths to become available when an intervention is introduced.

In the CoronaMelder study it was found that the effectiveness of contact tracing was strongly dependent on the degree of clustering: the inclusion of clustering resulted for a additional decrease in the reproduction number of approximately 5% in the baseline analysis [3]. Our results show that the effect of clustering on the effectiveness of contact tracing using model 1 is on average 0.0960 for the six simulated situations using $t = 0.5$ and $d = 0.5$, and 0.0338 using model 2. Thus using model 1 instead of model 2 leads to an overestimation of the effect of clustering, in these simulations it is more than 100% This suggests that results obtained in the study into the effectiveness of the CoronaMelder contact tracing app are too optimistic.

The CoronaMelder study did take into account that clustering in the contact network makes it possible to trace infected nodes from a contact which is not the source of infection, but it did not include the possibility of alternative infection paths. Our results suggest that the effect of these other infection paths is significant on the effectiveness of contact tracing. However, during this project we made multiple simplifying assumptions and it would be interesting to know how these assumptions affected our results.

With model 2 a clustered contact network is created by adding extra connections at distance 2 to a branching process. Using a branching process as underlying structure suggest that one contact of a node, the contact with its parent in the tree, is different from its other contacts. Instead of using a branching process, a configuration model could be used. Another extension to this project could be to create a clustered contact network according to a model from publications

about incorporating clustering in social networks.

We did not succeed in expressing the reproduction number of an epidemic over the clustered contact network. This is because the infection generations do not correlate to the generations of the branching process, due to the added connections. It could be possible that choosing a different model to create the contact network, makes it possible to express the reproduction number. An analytical expression of the reproduction number contributes to the understanding of the dynamics of the spread of a virus and might even lead to obtaining an expression for the reproduction number under contact tracing.

Another simplifying assumption we made was that we modelled the epidemic and contact tracing using discrete time steps. Because of this simplifying assumption, it was easier to obtain the analytical results and create the simulations, this gave us the opportunity to concentrate on the effect of the clustering, but this simplifying assumption is not realistic. It would be interesting to see if the effect of clustering changes when continuous time is introduced into the models.

Besides the discrete time steps, we made other simplifying assumptions when implementing contact tracing. We assumed that contact tracing is fast enough to prevent secondary cases of the traced nodes and that the quarantine of the traced nodes is 100% effective. We also did not take into account the amount of effort needed to perform contact tracing. It is difficult to assess how these assumptions did influence our conclusions, thus additional research is needed to determine how the results would change if these assumptions were not made.

We suggested that the difference between model 1 and model 2 could be due to the possibility of alternative infection paths when contact tracing is introduced when using model 2. A possible solution could be to adjust model 1 such that other infection paths will be possible over the added contacts in the tracing network when simulating contact tracing. This adds new complexities to model 1. The alternative transmission paths have to take place and thus the added contacts with grandchildren cannot be transmission contacts. It also remains that model 1 and 2 are different.

In conclusion, this project has shown that the model based on the CoronaMelder study does not approach the effect of clustering in the contact network on the effectiveness of contact tracing. The model based on the study overestimates effectiveness of contact tracing. This suggests that results obtained on the effectiveness of the contact tracing app in the CoronaMelder study are too optimistic. However, further research is needed to determine how the assumptions we made affected our results.

References

[1] L. Ferretti et al. "Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing". In: *Science* 368.6491 (2020), eabb6936. DOI: 10.1126/science.abb6936. eprint: https://www.science.org/doi/pdf/10.1126/science.abb6936. URL: https://www.science.org/doi/abs/10.1126/science.abb6936.

[2] D. Klinkenberg, K.Y. Leung, and J. Wallinga. *CoronaMelder – modelstudie naar effectiviteit. Digitaal contactonderzoek in de bestrijding van COVID-19*. 2021. DOI: 10.21945/RIVM-2021-0092. URL: http://hdl.handle.net/10029/624996.

[3] K.Y. Leung et al. *Effectiveness of a COVID-19 contact tracing app in a simulation model with indirect and informal contact tracing*. June 2023. DOI: 10.1101/2023.06.15.23291010.

[4] M. E. J. Newman and J. Park. "Why social networks are different from other types of networks". In: *Phys. Rev. E* 68 (3 Sept. 2003), p. 036122. DOI: 10.1103/PhysRevE.68.036122. URL: https://link.aps.org/doi/10.1103/PhysRevE.68.036122.

[5] M. E. J. Newman. "Random Graphs with Clustering". In: *Physical Review Letters* 103.5 (July 2009). DOI: 10.1103/physrevlett.103.058701. URL: https://doi.org/10.1103%2Fphysrevlett.103.058701.

[6] C. Fransson and P. Trapman. "SIR epidemics and vaccination on random graphs with clustering". In: *Journal of Mathematical Biology* 78 (June 2019). DOI: 10.1007/s00285-019-01347-2.

[7] A. Hackett, Se. Melnik, and J. P. Gleeson. "Cascades on a class of clustered random networks". In: *Phys. Rev. E* 83 (5 May 2011), p. 056107. DOI: 10.1103/PhysRevE.83.056107. URL: https://link.aps.org/doi/10.1103/PhysRevE.83.056107.

[8] M. E. J. Newman. "Properties of highly clustered networks". In: *Phys. Rev. E* 68 (2 Aug. 2003), p. 026121. DOI: 10.1103/PhysRevE.68.026121. URL: https://link.aps.org/doi/10.1103/PhysRevE.68.026121.

[9] C. Stegehuis, R.W. van der Hofstad, and J.S.H. van Leeuwaarden. "Epidemic spreading on complex networks with community structures". English. In: *Scientific Reports* 6 (2016). ISSN: 2045-2322. DOI: 10.1038/srep29748.

[10] Billy J. Q. et al. "Quarantine and testing strategies in contact tracing for SARS-CoV-2: a modelling study". In: *The Lancet Public Health* 6.3 (2021), e175–e183. ISSN: 2468-2667. DOI: https://doi.org/10.1016/S2468-2667(20)30308-X. URL: https://www.sciencedirect.com/science/article/pii/S246826672030308X.

[11] D. Klinkenberg, C. Fraser, and H. Heesterbeek. "The Effectiveness of Contact Tracing in Emerging Epidemics". In: *PLOS ONE* 1.1 (Dec. 2006), pp. 1–7. DOI: 10.1371/journal.pone.0000012. URL: https://doi.org/10.1371/journal.pone.0000012.

[12] M. E. Kretzschmar et al. "Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study". In: *The Lancet Public Health* 5.8 (2020), e452–e459. ISSN: 2468-2667. DOI: https://doi.org/10.1016/S2468-2667(20)30157-2. URL: https://www.sciencedirect.com/science/article/pii/S2468266720301572.

A Definitions and Notations

Table 3: Definitions

Basic reproduction number	expected number of infections caused by one infected node
Clustering coefficient	measure of the degree to which nodes in the network tend to cluster together
Contact network	network describing the contacts in the population
Degree distribution	probability distribution of the number of edges from a node
Offspring distribution	probability distribution of the number of offspring of a node in a branching process
Transmission network	network describing the transmission process
Tracing network	network used to simulate the contact tracing

Table 4: Notations

Parameters model 1:	
$C_t^{(1)}$	clustering coefficient of tracing network of model 1
$D_t^{(1)}$	degree distribution of tracing network of model 1
$p^{(1)}$	clustering probability in model 1
$R_0^{(1)}$	basic reproduction number of the epidemic used model 1
$O_e^{(1)}$	offspring distribution of transmission network of model 1
Parameters model 2:	
$C_c^{(2)}$	clustering coefficient of contact network of model 2
$C_t^{(2)}$	clustering coefficient of tracing network of model 2
$D_c^{(2)}$	degree distribution of contact network of model 2
$D_t^{(2)}$	degree distribution of tracing network of model 2
$p^{(2)}$	clustering probability in model 2
q	transmission probability used in model 2
$R_0^{(2)}$	basic reproduction number of transmission network of model 2
$O_b^{(2)}$	offspring distribution of the branching process used for creating contact network in model 2
$O_e^{(2)}$	offspring distribution of transmission network in model 2
Parameters contact tracing:	
d	detection probability
t	tracing probability
R_{tracing}	Reproduction number of the epidemic when contact tracing is used

B Code

All simulations used in this thesis was done in **R** and can be found using the following link: https://github.com/IsisMarsman/MasterThesis-IsisMarsman-01112023.