Individual-based simulation model of the transmission of STIs along a dynamic sexual network.

Merel van Schieveen (3510190)
Utrecht University

Supervised by K.Y. Leung[a], G. Sleijpen[a], M. Kretzschmar[b]

[a] Mathematical Institute, Utrecht University.

[b] Julius Center for Primary Care and Health Sciences,
University Medical Center Utrecht

May 1, 2016

Contents

Introduction

In this master thesis we will simulate the transmission of sexual transmitted infections along dynamic sexual networks. We will focus on the sexual transmitted infection HIV. The Human Immunodeficiency Virus (HIV) is a virus that attacks the immune system, which causes that the body has more difficulty fighting off infections and diseases. Transmission of HIV can take place because of unprotected sex, sharing needles, pregnancy, childbirth, breastfeeding, blood transfusion, transplants etc. To better understeand the spread of HIV and the effect of treatment and control measures, mathematical modeling can be very insightful. The HIV infection, when untreated, is considered to progress in three stages:

- **Acute infection**
  During the acute infection phase, the amount of virus produced in the body is very high compared to the other two stages. Therefore this is also the stage with the highest risk of transmitting HIV.
- **Clinical latency**
  In the clinical latency period the amount of virus in the body is very low, but still active. This stage lasts an average of 10 years.
- **AIDS**
  Acquired Immune Deficiency Syndrome (AIDS), also called advanced HIV infection, is the last stage. An individual's immune system has become too weak to defend its body. This can result in the development of one or multiple diseases and infections, and eventually death. This stage lasts an average of 3 years.

There is not yet a cure for HIV infection, however there is a treatment that can help control the virus. It not only can increase survival, but also reduces the risk of transmission of HIV [1].

HIV has already claimed more than 34 million lives so far [2]. This large impact of HIV has inspired the development of mathematical modeling to better understand the spread of HIV. Since the most common mode of transmission of HIV is through unprotected sex, models are mostly focused on HIV transmission through sexual acts. Sub-Saharan Africa is the most affected region, accounting for almost 70 percent of the global total of new HIV infections [2]. Many reasons for this alarming high prevalence of HIV in Sub-Saharan Africa have been discussed, for example the high migration levels of Sub-Saharan Africa [3].

In the modeling paper [4] it was hypothesized that having concurrent partners, i.e. multiple sexual partners at the same time, could increase the spread of HIV. When a model does not allow concurrent partnerships, it does not take into account the possibility that an initially uninfected partner of a susceptible individual may become infected over the duration of their partnership [4]. This paper was followed by the works of Morris and Kretzschmar, who found using simulation models, that allowing concurrent partnerships increases the speed of the spread of HIV [5],[6],[7].

Simulation models show that concurrent partnerships can increase the speed of the spread of HIV, but some find the empirical evidence limited that concurrency is actually a main driver of the epidemic in Sub-Saharan Africa. Lurie [8] highlights the fact that whether concurrency *can* drive and actually *is* driving the epidemic in Sub-Saharan Africa are two entirely different questions. It is

reasoned that there is no evidence that levels of concurrency are significantly higher in Sub-Saharan Africa than elsewhere and that the observed levels of concurrency cannot explain the sub-Saharan African epidemics.

Many of the modelling studies devoted to the effect of concurrent partnerships on the spread of HIV have made use of simulations. This has motivated the studies of Leung et al. to formulate a mathematical model for the spread of an sexually transmitted infection along a dynamic partnership network and derive analytical results [9], [10]. The network is dynamic because individuals enter and leave the population, but also because partnerships are formed and broken. The model is a deterministic model, in the sense that it concerns expected values for a population of infinite size.

However, to make the system of equations that describe the model solvable, or in other words to close the system of equations, an assumption was made called the 'mean field at distance one' assumption. This assumption has motivated this masterthesis project, in which the deterministic model is compared to a simulation model for which the assumption is not needed. By comparing the results of both model we check the correctness of the assumption. In later work of Leung [11], this deterministic model has been extended to a more general deterministic model with several generalizations involving both the network and the infection. It is desired to also implement the same generalizations in the simulation model.

A property of the simulation model is that the entire network structure is known: we know exactly who has a relationship with whom. This information is not incorporated in the deterministic model. This property of the simulation model gives us the opportunity to study the network structure. Thus in this masterthesis we will focus on: comparing the deterministic model and simulation model and further studying the network structure of the simulation model.

The simulation program is written in the programming language Python. The Python code is desired to be flexible extensible (for possible more generalizations in the future), have the potential to 'switch' the generalizations on and off and should be well documented and easy to read for others.

This masterthesis has the following structure. In section 1 the deterministic model by Leung et al. will be explained. In section 2 an individual-based simulation model, based on the deterministic model without generalizations, will be introduced. In section 3 the simulation model will be extended with generalizations. In section 4 we will explain the structure of the simulation program in Python. Section 5 will be devoted to the results, which will include the comparison of the deterministic model and the simulation model (to check the correctness of the mean field at distance one assumption) and further investigations of the network (such as its connected components). Finally, in section 6 we will perform a senstivity analysis.

1 The deterministic model

Leung et al. have introduced a deterministic model for the spread of an SI sexually transmitted infection (STI) on a dynamic homosexual network. The deterministic model is introduced in two articles. First in [9] a network without infection is introduced, in which concurrent partnerships are allowed, i.e. an individual in the population is allowed to have multiple partners at the same time. Second in [10] an STI is introduced that spreads along the network. Therefore, from now on when discussing the deterministic model, we will distinguish 'the network' and 'the infection'.

We begin by giving a short overview of the deterministic model in section 1.1. After that a more detailed description of the model will be given in two parts. First we will discuss the network in section 1.2, followed by the introduction of an STI along the network, which is discussed in section 1.3. Lastly, in section 1.4, a few generalization of the deterministic model will be discussed [11].

In this section we will use the notations as in the articles of Leung et al. [9], [10], [11].

1.1 Overview

The network

In [9] a model of a network of partnerships for a homosexual population is introduced. This network is dynamic, because individuals enter and leave the population and formation and dissolution of partnerships. Individuals in this network are assumed to have a maximum number n of simultaneous partners, n is called the *partnership capacity*. An individual with partnership capacity n can be thought of as having n binding sites which can be either 'free' or 'occupied'. An important assumption of the model is that the binding sites of an individual behave independently of each other, as far as partnership dynamics are concerned. The model involves two specifications only, one at *individual-level* and the other at *population-level*:

- At the individual-level, since binding sites of an individual behave independently from each other as far as partnership dynamics are concerned, then as long as an individual does not die it is involved in a continuous-time Markov process. In this Markov process the state of the individual is described by its *number of partners* (or 'occupied' binding sites). Therefore there are $n + 1$ possible states, corresponding to $0, \ldots, n$ partners.
- At the population-level, a stationary age distribution is assumed with the probability density function $a \mapsto \mu e^{-\mu a}$, i.e. there is a constant population birth rate and the lifetime of an individual is exponentially distributed with parameter μ.

Changes at these two levels are assumed to be coupled in the sense that the rate at which partnership formation at individual-level occurs, is proportional to the fraction of free binding sites in the total pool of binding sites. Note that this fraction of free binding sites is a population-level quantity.

The infection

An SI infection is introduced along the network [10]. Individuals in the network can be in two possible infection stages: susceptible (S) or infected (I). It is assumed that the infection stage of an individual has no influence whatsoever on partnership formation, partnership dissolution or the

age distribution. Therefore the rate at which partnership formation occurs is still proportional to the fraction of free binding sites in the total pool of binding sites.
Transmission of infection can occur in a partnership between an infectious individual and a susceptible individual. Therefore for a susceptible individual in the network, the probability of becoming infected depends on its number of infectious partners.

When considering just the network (in section 1.1), the state of an individual was described by its number of partners. Introducing the infection along the network implies that the state of an individual is defined by a vector containing its own infection state, its number of susceptible partners and its number of infectious partners. The state of an individual can change due to partner dynamics and transmission of the infection. A susceptible partner of an individual can get infected and the rate at which this occurs depends on the number of infectious partners of this susceptible partner. However, this information about partners of partners is not included in the state of an individual. Therefore, an assumption is made about the partners of partners. This assumption is called the *mean field at distance one* assumption. This assumption is explained in detail in section 1.3.

For a more extensive description of the deterministic model, see section 1.2 and section 1.3.

1.2 The network

In this section we will give a more formal description of the dynamic homosexual network as outlined in section 1.1. The outline in section 1.1 and the more detailed description in this subsection can be read separately form each other. Therefore there will be some repetition.

A sexually active population is considered. Therefore 'birth' and 'death' denotes entering or leaving the sexually active population, respectively. As explained before, the network allows concurrent partnerships. There is a maximum number of simultaneous partners an individual can have, which is called the partnership capacity n. An individual with partnership capacity n can be thought of as having n binding sites which can be either 'free' or 'occupied'. To keep track of concurrent partnerships we say that an individual who has k partners ($0 \leq k \leq n$) is in state k, i.e. an individual in state k has k occupied binding sites and $n - k$ free binding sites. The assumption is made that the binding sites of an individual behave independently from each other, as long as the individual does not die. This means that the binding sites behave independently from one another as far as formation and dissolution of partnerships is concerned.

The individual-level
Let us first look at individual-level. At the individual-level, as long as an individual does not die the changes in its number of binding sites are assumed to be modelled according to a continuous-time Markov process. Therefore the the state of the individual indicates its number of partners. It is assumed that an individual does not die in the period it is under consideration. The state of an individual can change by an occupied binding site that becomes free or a free binding site that becomes occupied, i.e. partnership formation and dissolution:

- Partnership formation:
  Let $F(t)$ define the fraction of free binding sites in the total pool of binding sites at time t. It is assumed that the rate at which a free binding sites becomes occupied at time t is

proportional to $F(t)$ and this rate is called $\rho F(t)$. Note that F is a population-level quantity, which connects the population-level to the individual-level.

- Partnership dissolution:
  An occupied binding site can become free in two ways, either the partner dies with rate μ or the partners separate with rate σ. Note that it was previously assumed that the individual itself does not die in the time period it is under consideration, therefore only the death of the partner is considered.

Therefore an individual with $0 \leq k \leq n$ occupied binding sites and $n-k$ free binding sites at time t can have the state transitions shown in table 1.

transition	**rate**	**description**
$k \to k+1$	$\rho(n-k)F(t)$	gaining a partner
$k \to k-1$	$(\sigma+\mu)k$	losing a partner

Table 1: The rates of the state transitions for the network.

Let $\boldsymbol{A}(F(t))$ be the $(n+1)\times(n+1)$ *transition rate matrix* belonging to the continuous-time Markov process, for which $\boldsymbol{A}_{ij}$ $(i \neq j)$ denotes the rate at which a transition from state j to state i is made. Since $\boldsymbol{A}$ is a transition rate matrix, the diagonal elements are defined as $\boldsymbol{A}_{jj} = \sum_{i \neq j} -\boldsymbol{A}_{ij}$, such that the each column sums up to zero.

Let $p_k(t_b, a)$ denote the probability that an individual, born at time t_b, is in state k at age a. Let $\boldsymbol{p}$ be the vector

$$\boldsymbol{p}(t_b, a) = \begin{pmatrix} p_0(t_b, a) \\ p_1(t_b, a) \\ \vdots \\ p_n(t_b, a) \end{pmatrix}$$

The changes in $\boldsymbol{p}(t_b, a)$ are modelled with a continuous-time Markov process

$$\frac{\partial \boldsymbol{p}}{\partial a}(t_b, a) = \boldsymbol{A}(F(t_b + a))\boldsymbol{p}(t_b, a) \tag{1}$$

Since a newborn individual is assumed to have no partners, when entering the sexually active population, we have the initial condition

$$\boldsymbol{p}(t_b, 0) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{2}$$

The population-level

At the population level it is assumed that there is a constant population birth rate and that the age of an individual is exponentially distributed with parameter μ (i.e. the age distribution has the probability density function $a \mapsto \mu e^{-\mu a}$). Let $P_k(t)$ denote the fraction of the population that is in

state k at time t. Then, when assuming a large population size and using the probability density function $a \mapsto \mu e^{-\mu a}$, $P_k(t)$ equals

$$P_k(t) = \int_0^\infty \mu e^{-\mu a} p_k(t-a, a) da \tag{3}$$

The fraction of free binding sites $F(t)$ was defined as the fraction of free binding sites in the total pool of binding sites at time t. This can be expressed as,

$$F(t) = \sum_{k=0}^{n} \frac{n-k}{n} P_k(t) \tag{4}$$

This population level quantity $F(t)$ causes, as described before, the feedback from population-level to individual-level.

Due to the assumption of the independence of binding sites with respect to partnership dynamics, it can be shown that $F(t) \to \bar{F}$ for $t \to \infty$ (details in [9]). This has motivated us to take $F(t)$ to be constant:

$$F(t) = \bar{F} \tag{5}$$

Which causes a stable network over time: $p_k(t_b, a) = p_k(a)$ and $P_k(t) = P_k$.

1.3 The infection

In this section we will summarize the results from [10], where an SI infection is introduced on the network discussed in the previous subsection. With an SI infection individuals in the population can either be susceptible (S) or infectious (I). A susceptible individual can get infected by an infectious partner. It is assumed that when an individual gets infected, it is immediately infectious and stays infectious for the rest of its life.

By introducing the SI infection, the state of an individual is now described by (x, k_-, k_+), where $x \in \{+, -\}$ indicates whether the individual is susceptible (-) or infectious (+), k_- is its number of susceptible partners and k_+ is its number of infectious partners. Therefore an individual can be in state $l \in \{(x, k_-, k_+) \mid k_-, k_+ \geq 0,\ k = k_- + k_+ \in \{0, \ldots, n\},\ x \in \{+, -\}\}$.

With the more extended state $l = (x, k_-, k_+)$ of an individual, the following transitions should be distinguished concerning partnership dynamics:

- Losing a susceptible partner
- Losing an infectious partner
- Acquiring a new susceptible partner
- Acquiring a new infectious partner

For calculating the rates of these transitions, $F_x(t)$ is introduced as the fraction of free binding sites in the population belonging to individuals in infection stage $x \in \{+, -\}$ in the total pool of free binding sites at time t. Since the total fraction of free binding sites $\bar{F}$ is constant (eq. (5)), it must hold that $\bar{F} = F_+(t) + F_-(t)$. For an individual in state (x, k_-, k_+), assuming the individual does not die in the time period it is under consideration, the state transitions and their corresponding rates are shown in table 2.

transition	**rate**	**description**
$(\pm, k_-, k_+) \to (\pm, k_- - 1, k_+)$	$(\sigma + \mu)k_-$	losing a susceptible partner
$(\pm, k_-, k_+) \to (\pm, k_-, k_+ - 1)$	$(\sigma + \mu)k_+$	losing an infectious partner
$(\pm, k_-, k_+) \to (\pm, k_- + 1, k_+)$	$\rho F_-(t)(n - k_- - k_+)$	gaining a susceptible partner
$(\pm, k_-, k_+) \to (\pm, k_-, k_+ + 1)$	$\rho F_+(t)(n - k_- - k_+)$	gaining an infectious partner

Table 2: The rates of the state transitions involving partnership formation, separation and death of a partner.

Besides the transitions caused by losing or acquiring a partner, the transmission events should also be considered. There are three different transitions describing transmission events:

- The individual itself gets infected
- A susceptible partner gets infected, while the individual itself is susceptible
- A susceptible partner gets infected, while the individual itself is infectious

It is assumed that a susceptible individual, that has a binding site occupied by an infectious partner, gets infected by that partner with a constant rate β. It is assumed that the frequency of sex acts in a partnership does not depend on other concurrent partnerships. Therefore a susceptible individual with k_+ infectious partners gets infected with rate βk_+.

Let us consider an individual u. The rate at which a susceptible partner v, of the individual u, gets infected depends on the number of infectious partners of v. The only known partner of v is u, with u either susceptible or infectious. The rest of the partners of v is unknown. Therefore the assumption is made that the average over all possibilities can be taken, the so called *mean field at distance one* assumption:

Definition 1. Mean field at distance one assumption*:*
$\Lambda_\pm(t)$ *is the expected number of infectious partners of an susceptible individual that has one known partner in infection state* $\pm$ *at time* t*.*

In section 2 we will introduce an individual-based simulation model for which this assumption is not needed, and compare both models in section 5.3 to check the correctness of the assumption.

The state transitions describing transmission and their corresponding rates, when assuming the mean field at distance one assumption, are as shown in table 3.

transition	**rate**	**description**
$(-, k_-, k_+) \to (+, k_-, k_+)$	βk_+	individual gets infected
$(-, k_-, k_+) \to (-, k_- - 1, k_+ + 1)$	$\beta \Lambda_-(t) k_-$	a partner gets infected, while the individual is susceptible
$(+, k_-, k_+) \to (+, k_- - 1, k_+ + 1)$	$\beta \Lambda_+(t) k_-$	a partner gets infected, while the individual is infectious

Table 3: The rates of the state transitions involving transmission of infection.

In section 1.2, where the network without the SI infection was described, we denoted by $p_k(t_b, a)$ the probability that an individual, born at time t_b, is in state $k \in \{0, 1, \ldots, n\}$ at age a. Let $p_l(t_b, a)$ denote the probability that an individual, born at time t_b, is in state $l \in \{(x, k_-, k_+) \mid k_-, k_+ \geq 0,\ k = k_- + k_+ \in \{0, \ldots, n\},\ x \in \{+, -\}\}$ at age a. Choosing a way to order the states l, we can again consider the $p_l(t_b, a)$ as coordinates of a vector $\boldsymbol{p}$. Let $\boldsymbol{B}(t) = \boldsymbol{B}(F_+(t), F_-(t), \Lambda_+(t), \Lambda_-(t))$ be the transition rate matrix with the transition rates as described in the tables 2 and 3. The number of states (x, k_-, k_+) is equal to the number of infection stages multiplied with the number of possible combinations of infectious and susceptible partners, under the condition that no more than n partners is allowed. Therefore the number of states l is equal to $2 \cdot \sum_{k=0}^{n} (k+1) = 2 \cdot \frac{1+(n+1)}{2}(n+1) = (n+1)(n+2)$. Therefore the matrix $\boldsymbol{B}$ will be of size $(n+1)(n+2) \times (n+1)(n+2)$. Note that, since $\boldsymbol{B}$ is a transition rate matrix, the columns of $\boldsymbol{B}$ sum up to zero.

Then, as long as an individual does not die in the time period it is under consideration, an individual is involved in the Markov process

$$\frac{\partial \boldsymbol{p}}{\partial a}(t_b, a) = \boldsymbol{B}(F_+(t_b + a), F_-(t_b + a), \Lambda_+(t_b + a), \Lambda_-(t_b + a))\boldsymbol{p}(t_b, a) \tag{6}$$

Note that $F_\pm$ and $\Lambda_\pm$ are yet to be specified. Since a newborn individual is assumed to be susceptible and without any partners, it is in state $(-, 0, 0)$. This newborn state is chosen as the first state for the ordering, therefore the initial condition is

$$\boldsymbol{p}(t_b, 0) = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{7}$$

Let $P_l(t)$ be the fraction of the population that is in state l at time t. Then, in a deterministic description of a large population,

$$P_l(t) = \int_0^\infty \mu e^{-\mu a} p_l(t - a, a) da \tag{8}$$

Use the chosen ordering of states l and let $\boldsymbol{P}$ denote the vector containing the variables P_l.

We have noted before that $F_\pm$ and $\Lambda_\pm$ are not yet specified, only the interpretations were given.

Using these interpretations we can derive the definitions of $F_\pm$ and $\Lambda_\pm$ in terms of population-level fractions.

$F_\pm(t)$ was defined as the fraction of free binding sites in the population belonging to individuals in infection stage $\pm$ at time t. Using this interpretation we can express $F_\pm(t)$ as

$$F_\pm(t) = \sum_{k_+=0}^{n} \sum_{k_-=0}^{n-k_+} \frac{n-k_--k_+}{n} P_{(\pm,k_-,k_+)}(t) \tag{9}$$

$\Lambda_\pm(t)$ was introduced as the expected number of infectious partners of a susceptible partner of an individual in infection state $\pm$. Consider an individual u in infection stage $\pm$, with a susceptible partner v. Let the susceptible partner v be in the state $(-, m_-, m_+)$. Then, using the interpretation, the mean field at distance one rates $\Lambda_\pm(t)$ can be expressed as

$$\Lambda_-(t) = \sum_{m_+=1}^{n-1} m_+ \lambda_-(m_+) \tag{10}$$

$$\Lambda_+(t) = \sum_{m_+=1}^{n} m_+ \lambda_+(m_+) \tag{11}$$

Where $\lambda_\pm(m_+)$ can be interpreted as the conditional probability that susceptible partner v has m_+ infectious partners, given that individual u is infection stage $\pm$. The terms $\lambda_\pm(m_+)$ can be expressed as

$$\lambda_-(m_+) := \frac{\sum_{m_-=1}^{n-m_+} m_- P_{(-,m_-,m_+)}(t)}{\sum_{l_+=0}^{n-1} \sum_{l_-=1}^{n-l_+} l_- P_{(-,l_-,l_+)}(t)} \tag{12}$$

and

$$\lambda_+(m_+) := \frac{\sum_{m_-=0}^{n-m_+} m_+ P_{(-,m_-,m_+)}(t)}{\sum_{l_+=1}^{n} \sum_{l_-=0}^{n-l_+} l_+ P_{(-,l_-,l_+)}(t)} \tag{13}$$

We will not give further explanations of the derivation of these expressions. If the derivation is wanted, we refer to the original paper [10].

In [10] it is also shown, by differentiation of (8) and using (6) and (7), that

$$\frac{d\boldsymbol{P}}{dt} = \mu \mathbf{1}_{(-,0,0)} + B(F_\pm, \Lambda_\pm)\boldsymbol{P} - \mu \boldsymbol{P} \tag{14}$$

where $\mathbf{1}_{(-,0,0)}$ is the indicator function of $(-,0,0)$. Therefore the constant population birth rate (the rate new individuals enter the state $(-,0,0)$) is equal to μ.

1.4 Generalizations of the baseline deterministic model

In [9] it was already mentioned that the network model can be extended by some meaningful generalizations. Later, in [11], some generalizations were implemented in the model, which will be discussed briefly in this section. Until now a homosexual population was considered with a fixed partnership capacity n for the entire population and only one type of partnership. The following generalizations of the network were introduced in [11]:

- Partnership capacity as a random variable
- Heterosexual population
- Steady and casual partnerships

And one generalization of the infection was considered:

- Multiple stages of infection

We will give a brief description of each of these generalizations in the sections 1.4.1 - 1.4.4.

1.4.1 Partnership capacity as a random variable

Originally it was assumed that the partnership capacity n was a fixed number, the same value for all the individuals in the population. This is extended by allowing individuals in the population to have different partnership capacities. This added heterogeneity between individuals hopefully gives a better approximation of the reality.

Let N be the random variable that describes the partnership capacity. Let $\mathbb{P}(N = n)$ be the probability that a random individual has partnership capacity n, $n = 1, 2, \ldots$.

1.4.2 Heterosexual population

One of the modifications of the model is to consider a heterosexual population instead of a homosexual population, i.e. male and female individuals are distinguished. Considering a heterosexual population gives us the opportunity to investigate the effect heterogeneity in sexual behavior between men and women has on the spread of HIV.

The assumption is made that in a heterosexual population only heterosexual partnerships can be formed. Let n_m and n_f be the partnership capacities of males and females respectively. And let $1 : x$ denote the ratio male : female in the population.

1.4.3 Steady and casual partnerships

Originally no distinction was made between different types of partnerships, only one type was considered. But in reality partnerships can vary greatly in their duration. Also, the probability of transmission within a partnership per time unit can differ between partnerships (caused for example by the difference in the number of sex acts per time unit). Therefore the model is extended by considering two types of partnerships: steady and casual partnerships.

Let n_s and n_c be the number steady and casual binding sites respectively. It is assumed that binding sites can only bind with a binding site of the same type. Just like in the original model, the assumption is still made that all binding sites (both casual and steady) of an individual behave independently of one another where partnership formation and separation is concerned.

1.4.4 Multiple stages of infection

Lastly, the model is extended by not considering only one infection stage (I), but a finite number of infection stages $I_1, I_2, \ldots, I_m$. This could be very useful for an infection such as HIV, because HIV progresses in different phases such as a very infectious acute infection phase and a longer but less infectious clinical latency phase. Also a treatment phase could be considered, a phase for HIV in which an individual is less infectious, and therefore investigating the influences of treatment to spread of HIV in the population. The assumption is made that an individual can only go through the infection stages in sequential order, so an individual can only go from infection stage I_1 to I_2, from I_2 to I_3, and so on. Each infection stage $i \in \{I_1, I_2, \ldots, I_m\}$ has its own infection rate β_i. Also each infection stage, except for the last one, has its own transition rate α_i. The transition rate α_i denotes the rate at which an individual in infection stage i enters the next infection stage $i+1$.

In [11] three stages were considered for the deterministic model: I_1, I_2, T, an acute phase, a chronic phase and a treatment phase. We will consider the same infection stages in the simulation model.

In [11] not only are these generalizations considered separately, but also in combination.

2 The individual-based model

In this section we will introduce the stochastic individual-based simulation model based on the deterministic model discussed in section 1. We will focus on a homosexual population with fixed partnership capacity, only one type of partnership and one infection stage I, i.e. the model without generalizations. Later, in section 3, we will introduce generalizations on the individual-based model. To make this masterthesis easier to read it is divided into two separate sections.

First we will motivate in section 2.1 why we consider an individual-based model. Then we give a general description of individual-based modeling and event-based modeling in sections 2.2 and 2.3. Lastly we will introduce the individual-based model in section 2.4.

2.1 Motivation

We want to introduce an individual-based simulation model based on the deterministic model. An individual-based model keeps track of each individual in the network through time. Considering an individual-based simulation model has certain advantages compared to the deterministic model.

- For one, by keeping track of each individual in the network, we know for each individual in the network its partners, and its partners' partners etc. In the deterministic model this was not incorporated in the model and instead it was assumed that the average could be taken over all possibilities (the *mean field at distance one* assumption, see definition 1). No longer needing to take the mean field at distance one assumption in the individual-based model, is one of its greatest advantages. We will compare both models and check the correctness of the mean field at distance one assumption in section 5.3.
- Another advantage is that there are no longer restrictions on the population size. In the deterministic model the assumption of a large population was needed, while in the individual-based model no such assumption is needed and we can vary the population size.
- Thirdly, the individual-based model is easier to extend than the deterministic model. In section 1.4 a few possible generalizations for the model were given. But the mathematical tractability will become more and more complicated when adding generalizations, a problem that does not arise in an individual-based model.

However the individual-based model also has certain disadvantages compared to the deterministic model. For example the deterministic model gives one solution for a specific parameter set, while the individual-based model gives different results with every run. This is not a disadvantage in itself, since the simulation model gives a distribution for the outcome variables instead of just a number, which makes it more informative. But to obtain statistically relevant statements the individual-based model will need to be run many times and the results need to be averaged before being able to make comparisons. This makes the simulation model more costly. It also causes that the effect of small changes in input parameters can be investigated very easily in the deterministic model, while in the individual-based model this is much harder because of the stochasticity.

2.2 Individual-based modeling

Before introducing the individual-based model, let us give a short general description of *individual-based modeling*. We follow [12] to explain individual-based modeling.

Individual-based modeling looks at how the properties and behavior of individuals determine the properties of the population they compose. Each individual is a uniquely identifiable individual and will have its own characteristics. For the individual's characteristics think for example of size, location or history (or in our case for example the infection state of the individual). Sometimes the interactions of an individual can be limited to a subgroup defined by for example geographical space, or in our case a network. The behavior of an individual will depend on and change with their own current state, the current states of the other individuals and maybe even their environment.

2.3 Event-based modeling

For the individual-based model we will use so called *event-based modeling*. In this section we will give a short general description of event-based modeling.

Let m be the number of types events that can take place in a population occuring with rates $\lambda_1, \lambda_2, \ldots, \lambda_m$ respectively. Every time an event has taken place the rates of all events can change, i.e. $\lambda_1, \lambda_2, \ldots, \lambda_m$ are dependent on time. We assume that events occur according to a Poisson process. We can combine these m independent Poisson processes into one Poisson process with rate $\lambda = \sum_{i=1}^{m} \lambda_i$ [13]. Therefore the total rate with which any event will hapen is λ.

The exponential distribution describes the time between events that occur according to a Poisson process. Therefore the cumulative distribution function of the exponential distribution, $F(t, \lambda) = 1 - e^{-\lambda t}$, describes the waiting time to the next event. To calculate the time till the next event occurs, let r_1 denote a random number that is uniformly distributed between 0 and 1. Let r_1 be equal to the probability $F(t, \lambda)$ and find the corresponding elapsed time t:

$$r_1 = 1 - e^{-\lambda t}$$
$$t = \frac{-\log(-r_1 + 1)}{\lambda}$$

Note that $-r_1 + 1$ is also a random number uniformly distributed between 0 and 1.

When the time till the next event has been calculated, an event has to be chosen. We follow the Roulette Strategy: Let r_2, a random number uniformly distributed between 0 and 1, decide which event will take place. When $r_2 < \frac{\lambda_1}{\lambda}$ event 1 will take place, when $\frac{\lambda_1}{\lambda} \leq r_2 < \frac{\lambda_1 + \lambda_2}{\lambda}$ event 2 will take place etc. When the chosen event has taken place the rates $\lambda_1, \ldots, \lambda_m$ need to be recalculated to determine the new total rate λ and to repeat this process.

2.4 Structure of the individual-based model

As explained before, an individual-based model keeps track of each individual in the population and its characteristics. For our individual-based model we have the following characteristics of individuals

- infection state (S or I) at time t

- identities of an individual's partners at time t

Each characteristic can change with time: the individual can get infected or it can gain or lose a partner. Besides that the characteristics of an individual can change, individuals can also enter or leave the population. Just like in the deterministic model the following events can take place in the individual-based model:

- Birth of an individual
- Death of an individual
- Formation of a partnership
- Dissolution of a partnership
- Transmission of infection

To chose which events will occur and the time that elapses before the event occurs will be calculated with event-based modeling, previously explained in section 2.3. To give an idea of the structure of the individual-based model, using event-based modeling, see table 4.

(0) $t = 0$

(1) Calculate $\lambda_1, \lambda_2, \ldots, \lambda_m$
$\lambda = \sum_{i=1}^{m} \lambda_i$

(2) $\Delta t = \frac{-\log r_1}{\lambda}$
where r_1 is a random number between 0 and 1

(3) **if** $0 \leq r_2 < \frac{\lambda_1}{\lambda}$ **then**
Event 1
if $\frac{\lambda_1}{\lambda} \leq r_2 < \frac{\sum_{i=1}^{2} \lambda_i}{\lambda}$ **then**
Event 2
$\vdots$
if $\frac{\sum_{i=1}^{m-1} \lambda_i}{\lambda} \leq r_2 \leq 1$ **then**
Event m
where r_2 is a random number between 0 and 1

$t = t + \Delta t$
Repeat (1), (2) and (3) until t reaches the end time

Table 4: Time steps of simulation for m events.

2.4.1 The rates

Before we calculate the rates $\lambda_1, \ldots \lambda_5$ at which the events occur, let us introduce a few new notations that we will use in the individual-based model in table 5.

notation	description
N	the total population size
$[S, S]$	the number of partnerships between two susceptible individuals
$[S, I]$	the number of partnerships between a susceptible and an infectious individual
$[I, I]$	the number of partnerships between two infectious individuals
X	the total number of free binding sites
Y	the total number of partnerships

Table 5: A few notations and their descriptions used for the simulation model

Note that we work with numbers in the individual-based model, in stead of the fractions of the deterministic model. Note that the total number of partnerships Y is equal to

$$Y = [S, S] + [S, I] + [I, I].$$

Note that the total number of free binding sites Nn is equal to

$$Nn = X + 2Y.$$

The adding of $2Y$ is caused by the fact that every partnership consists of two binding sites. The fraction of free binding sites F, used in the deterministic model, can be calculated in the individual-based model by

$$F = \frac{X}{Nn}.$$

We will now discuss the rates of all types of events in the deterministic model and how to translate them to the individual-based model. We want to make sure that the meaning of the parameters are equivalent in both models, such that we can compare the two models easily.

Death

As mentioned before, in the deterministic model it was assumed that the lifetime of an individual is exponentially distributed with parameter μ. The same is assumed for the individual-based model. So every individual dies with rate μ per unit of time. Let there be N individuals in the population, the total death rate is then μN. In other words, the rate that the event 'Death' occurs is μN. Keep in mind that the population size N changes with time.

Birth

The deterministic model had a constant population birth rate, we assume the same for the individual-based simulation model. Let B denote the constant population birth rate for the individual-based model. Let N_0 be the initial population size. We want the population size over time to remain constant at N_0, apart from stochastic fluctuations, therefore we take $B = \mu N_0$. This causes the population size N to oscillate around the initial population size N_0 over time. Thus the rate that the event 'Birth' occurs is μN_0, constant over time.

Forming a partnership

In the deterministic model it was assumed that the probability per unit of time that a free binding site becomes occupied is $\rho\bar{F}$. Where $\bar{F}$ has been defined as the fraction of free binding sites in the total pool of binding sites, which was assumed to be constant (see eq. (5)). In terms of the individual-based model this rate is equal to $\rho F = \rho\frac{X}{Nn}$. The fraction of free binding sites in the individual-based model is, unlike in the deterministic model, not a constant value. However it will become constant up to stochastic fluctuations when the network has been stabilized. Therefore we perform the so called *spin-up* to make sure the network reaches that stochastic equilibrium, which will be described in section 4.1.

When a binding site becomes occupied, another binding site automatically becomes occupied as well, since a partnership is formed between two binding sites. Therefore to calculate the total rate of forming a partnership, we do not multiply $\rho\frac{X}{Nn}$ with the total number of free binding sites X, but only with half of the total number of free binding sites, to prevent counting double. This leads to the total rate $\frac{1}{2}\rho\frac{X^2}{Nn}$ of forming a partnership.

Separation

In the deterministic model the rate per unit of time that a partnership dissolves because of separation was equal to σ. We again assume the same for the individual-based model. To calculate the total rate of separation we must multiply σ with the total number of partnerships. Thus the total rate of separation becomes σY.

Transmission

In the deterministic model three different state transitions were considered that described transmission: the individual itself gets infected, a susceptible partner gets infected while the individual itself is susceptible or a susceptible partner gets infected while the individual itself is infectious (see section 1.3). To calculate the rate for the last two transitions, the mean field at distance one assumption was made. Since the individual-based model keeps track of each individual, we need not consider the last two transitions and therefore the mean field at distance one assumption is not needed. As explained before this is one of the greatest advantages of using the individual-based model.

The rate that a susceptible individual gets infected by an infectious partner was denoted by β. Since the infection can only be transmitted in a partnership between a susceptible and an infectious individual, the total rate in the entire population of transmission is $\beta \cdot [S, I]$.

To summarize, we have the following events and their rates for the baseline model (homosexual population without generalizations):

Event	rate
Birth	$\lambda_1 = \mu N_0$
Death	$\lambda_2 = \mu N$
Separation	$\lambda_3 = \sigma Y$
Forming a partnership	$\lambda_4 = \frac{1}{2}\rho\frac{X^2}{Nn}$
Transmission	$\lambda_5 = \beta \cdot [S, I]$

2.4.2 The events

As previously explained, one of the events will be chosen. In this section we will explain which state transition will occur if a specific event is chosen. But before we explain each event let us remember that each individual is characterized by: its infection state and the identities of its partners. Note that an event can change the characteristics of multiple individuals, for example the dissolution of a partnership changes the states of both individuals involved, because both individuals lose a partner.

Birth

In case of the event 'Birth' a new individual enters the population. This new individual will be susceptible, it will have n free binding sites and no partners. Note that the partnership capacity n is not a characteristic of the individual, since it does not vary between individuals, but is a fixed constant for the entire population.

Death

In case of the event 'Death' a random individual leaves the population. Not only will the individual itself leave the population, but also all the partnerships in which the individual is involved will dissolve.

Formation of a partnership

In case of the event 'Formation of a partnership' two random free binding sites are chosen from the pool of all free binding sites. However it is not allowed that these two binding sites both belong to the same individual or to two individuals which already have a partnership with each other. If the two chosen binding sites do not meet these criteria, no partnership is formed and a new event is chosen (see the program structure in appendix C.4 for more detail).

Separation

In case of the event 'Separation' a random partnership is chosen from the pool of all partnerships. This partnership dissolves, i.e. the individuals involved in the partnership both lose the other individual as their partner.

Transmission

In case of 'Transmission' a random partnership between a susceptible and an infectious individual is chosen. Transmission occurs and the susceptible individual becomes infectious.

3 Generalizations of the baseline model

In section 2.4 only a homosexual population was considered with partnership capacity n for the entire population, no distinction was made between different types of partnerships and only one infection stage I was considered. Now the generalizations for the deterministic model discussed in section 1.4 will be introduced in the simulation model. In addition, the model will be extended by a disease-induced death rate.

First we will consider each generalization separately in sections 3.1 - 3.5 to see what changes they bring to the individual-based model. So keep in mind that the changes discussed in one generalization will not yet affect the others. After discussing them separately we will combine all generalizations in section 3.6 and discuss the changes they will bring together.

3.1 Partnership capacity as a random variable

Till now the partnership capacity n was considered to be the same for all individuals in the population. We now extend the model to incorperate varying partnership capacities by considering the partnership capacity as a random variable Z. At birth every individual receives a partnership capacity $n = 1, 2, \ldots$ distributed according to Z. The partnership capacity varies between individuals, but the capacity of an individual stays the same during its life.

Because of this varying partner capacity n between individuals, we need to save this information for each individual. Therefore we consider the partnership capacity n as another characteristic for the individuals in the population. Note that by changing the partnership capacity n between individuals we can no longer say that the total number of binding sites is equal to nN. We however can still express the total number of binding sites as $X + 2Y$ (see table 5).

3.2 Heterosexual population

Another generalization is to distinguish male and female individuals in the population. The gender of each individual will be one of its characteristics. It is assumed that only heterosexual partnerships can be formed, i.e. only the binding sites of a male and female individual can bind to form a partnership.

Birth
In the deterministic model it is assumed that the ratio male : females in the population is $1 : x$. We assume the same for the individual-based model, but this will only be an approximate ratio since randomness is involved. That is to say when an individual enters the population it is a male individual with probability $\frac{1}{1+x}$ and a female individual with probability $\frac{x}{1+x}$.

Forming a partnership
In the deterministic model it is assumed that male and female individuals have partnership capacities n_m and n_f. The rates at which a free binding site of a certain gender becomes occupied are $\rho_m F^f$ and $\rho_f F^m$, where F^g with $g \in \{m, f\}$ is the fraction of free binding sites of gender g in the pool of all binding sites of gender g. The parameters ρ_m and ρ_f have to fulfill the following condition:

$$\rho_m n_m = \rho_f n_f x$$

The influence of sex ratio x, n_m and n_f on the rate at which free binding sites become occupied are incorporated in the parameters ρ_m and ρ_f, this is explained in [11]. For the individual-based model we only need one partnership formation rate, instead of the two rates $\rho_m F^f$ and $\rho_f F^m$ defined in the deterministic model. Let us introduce $\tilde{\rho}$ for the individual-based model, and let the relation of $\tilde{\rho}$ to the deterministic values ρ_m and ρ_f be defined as

$$\tilde{\rho} := \rho_m \frac{1}{n_f N_f} = \rho_f x \frac{1}{n_m N_f} = \rho_f \frac{1}{n_m N_m}$$

where N_g denotes the number of individuals of gender g. Remember that the rate at which a free binding sites of gender g becomes occupied is equal to $\rho_m F^f$ and $\rho_f F^m$ in the deterministic model. With the use of the previously defined $\tilde{\rho}$ for the individual based model, the total rate of forming a partnership in the individual based model is equal to

$$\tilde{\rho} X_m X_f$$

Where X_g denotes the total number of free binding sites of gender g in the population.

Transmission
We not only assume that male and female individuals may have different partnership capacities n_m and n_f, but also different transmission rates β_m and β_f. This implies that we have to distinguish two types of transmission events: an infectious male infects a susceptible female (with rate $\beta_m \cdot [I, S]$) or an infectious female infects a susceptible male (with rate $\beta_f \cdot [S, I]$). Here $[i_m, i_f]$ defines the number of partnerships where the male is in infection stage i_m and the female is in infection stage i_f.

To summarize we have the following events and rates with a heterosexual population

Event	**rate**
Birth male	$\lambda_1 = \frac{1}{1+x}\mu N_0$
Birth female	$\lambda_{\tilde{1}} = \frac{x}{1+x}\mu N_0$
Death	$\lambda_2 = \mu N$
Separation	$\lambda_3 = \sigma Y$
Forming a partnership	$\lambda_4 = \tilde{\rho} X_m X_f$
Transmission from male to female	$\lambda_5 = \beta_m \cdot [I, S]$
Transmission from female to male	$\lambda_{\tilde{5}} = \beta_f \cdot [S, I]$

3.3 Steady and casual partnerships

The deterministic model can also be extended to consider multiple types of partnerships. We will do the same for the individual-based model. Two types of partnerships are considered: casual and steady partnerships. Let an individual have n_s binding sites for steady partnerships and n_c binding

sites for casual partnerships. Only partnerships can be formed between two steady or two casual binding sites.

Separation
It is assumed that the two types of partnerships have different separation rates σ_s and σ_c. So we should separate the events of separation of a steady partnership (with rate $\sigma_s Y_s$) and the separation of a casual partnership (with rate $\sigma_c Y_c$). Here Y_h denotes the total number of partnerships of type $h \in \{c, s\}$.

Forming a partnership
It is assumed that a binding site can only form a partnership with a binding site of the same type (s or c). Therefore we should take into account while forming a new partnership, that we choose two free binding sites of the same type. Another assumption is that the two types of partnerships form with two different rates. In the deterministic model a steady free binding site becomes occupied with rate $\rho_s F_s$ and a casual free binding site becomes occupied with rate $\rho_c F_c$. Where F_s is the fraction of free steady binding sites in the pool of all steady binding sites, same for F_c for the casual binding sites.

In the individual-based model we can calculate F_h for the type of partnership $h \in \{c, s\}$ by

$$F_h = \frac{X_h}{X_h + 2Y_h}$$

Therefore, when following the same reasoning as before in section 2.4.1, the total rate for the formation of a partnership of type h is

$$\frac{1}{2}\rho_h \frac{(X_h)^2}{X_h + 2Y_h}$$

Transmission
Let β_s denote the transmission rate in a steady partnership and let β_c denote the transmission rate in a casual partnership. Then the total rate of transmission in steady partnerships is $\beta_s \cdot [S, I]_s$ and the total rate of transmission in casual partnerships is $\beta_c \cdot [S, I]_c$. Where $[i_1, i_2]_h$ is the number of partnerships of type $h \in \{c, s\}$ with the individuals in stages $i_1, i_2 \in \{S, I\}$.

This gives us the following events with the total rates:

Event	rate
Birth	$\lambda_1 = \mu N_0$
Death	$\lambda_2 = \mu N$
Separation steady	$\lambda_3 = \sigma_s Y_s$
Separation casual	$\lambda_{\tilde{3}} = \sigma_c Y_c$
Forming a steady partnership	$\lambda_4 = \frac{1}{2}\rho_s \frac{(X_s)^2}{X_s+2Y_s}$
Forming a casual partnership	$\lambda_{\tilde{4}} = \frac{1}{2}\rho_c \frac{(X_c)^2}{X_c+2Y_c}$
Transmission in steady partnership	$\lambda_5 = \beta_s [S, I]_s$
Transmission in casual partnership	$\lambda_{\tilde{5}} = \beta_c [S, I]_c$

3.4 Multiple stages of infection

The deterministic model can also be extended by considering not only the infection stages S and I, but more infection stages $S, I_1, I_2, \ldots$. Four different stages were introduced in [11]: S, I_1, I_2 and T. It is assumed that individuals can only go through the infection stages with sequential order. Therefore when an individual gets infected it enters stage I_1, from stage I_1 the individual can enter I_2, and finally from I_2 it can enter the treatment stage T and remain in this stage for the rest of its life. An individual in the treatment phase T is assumed to be no longer infectious. We assume the same for the individual-based model.

The rate at which an individual in stage $i \in \{I_1, I_2\}$ transmits the infection to a susceptible individual is β_i. Given that an individual remains alive for the period under consideration, an individual in stage i enters stage $i+1$ with rate α_i.

Transmission
The assumption that the two infectious stages I_1 and I_2 have two different transmission rates β_1 and β_2 causes the distinction between two transmission events. First the transmission between an infectious individual in stage I_1 and a susceptible individual with the total transmission rate $\beta_1 \cdot [S, I_1]$. Second the transmission between an infectious individual in stage I_2 and a susceptible individual with the total transmission rate $\beta_2 \cdot [S, I_2]$.

Stage transitions
The extension of the model also causes two new possible events: moving from stage I_1 to stage I_2 and moving from stage I_2 to stage T. Upon transmission, an individual enters stage I_1. Let N_i be the number of individuals in infection stage $i \in \{S, I_1, I_2, T\}$. The total rate of entering stage I_2 is then $\alpha_1 N_{I_1}$. The total rate of entering stage T is $\alpha_2 N_{I_2}$.

So we have the following events and their total rates:

Event	**rate**
Birth	$\lambda_1 = \mu N_0$
Death	$\lambda_2 = \mu N$
Separation	$\lambda_3 = \sigma Y$
Forming a partnership	$\lambda_4 = \frac{1}{2}\rho\frac{X^2}{Nn}$
Transmission caused by I_1	$\lambda_5 = \beta_1 \cdot [S, I_1]$
Transmission caused by I_2	$\lambda_{\tilde{5}} = \beta_2 \cdot [S, I_2]$
$I_1 \to I_2$	$\lambda_6 = \alpha_1 N_{I_1}$
$I_2 \to T$	$\lambda_7 = \alpha_2 N_{I_2}$

3.5 Disease-induced death rate

We consider an additional disease-induced death rate. When considering an infection, such as HIV, it is reasonable to consider a higher death rate when an individual is infected. Therefore we will introduce a separate death rate for the infection stages S and I. Let $\mu(i)$ be the death rate for an individual in infection stage $i \in \{S, I\}$, and let $\mu(S) \leq \mu(I)$. Therefore the death of a susceptible individual and the death of an infected individual will be considered as two separate events.

Event	**rate**
Birth	$\lambda_1 = \mu(S)N_0$
Death of S	$\lambda_2 = \mu(S)N_S$
Death of I	$\lambda_{\tilde{2}} = \mu(I)N_I$
Separation	$\lambda_3 = \sigma Y$
Forming a partnership	$\lambda_4 = \frac{1}{2}\rho\frac{X^2}{Nn}$
Transmission	$\lambda_5 = \beta \cdot [S, I]$

Note that by increasing the death rate for infected individuals, we no longer have the stable in- and outflow of the population as we had before, i.e. the birth rate and average death rate are no longer equal.

3.6 Combining the generalizations

We will now combine all the separately discussed generalizations from sections 3.1 - 3.4 and see how this affect the individual-based model.

Model parameters

Many model parameters, as we have previously seen, depend on certain properties of individuals (gender, infection stage) or partnerships (casual, steady). Let us consider the gender $g \in \{m, f\}$, the infection stage $i \in \{S, I_1, I_2, T\}$ and the type of partnership $h \in \{c, s\}$. We have the following parameters and their dependency on g, h and i.

$Z(g,h)$	random variable denoting partnership capacity	4 distributions
$\rho(h)$	parameter partnership formation	2 parameters
$\sigma(h)$	parameter separation	2 parameters
$\beta(g,h,i)$	parameter transmission ($i \in \{I_1, I_2\}$)	8 parameters
$\alpha(i)$	transition rate stage $i+1$ from stage i	
$\mu(i)$	rate death for infection stage i	

The following parameters do not depend on the properties g, h and i.

x	ratio male : female is $1 : x$
N_0	initial population size

Note that we want the ratio male : female in the initial population of size N_0 to be $1 : x$. So the number of male individuals in the initial population is $N_0(m) = \frac{1}{x+1}N_0$ and the number of female individuals is $N_0(f) = \frac{x}{1+x}N_0$.

Characteristics

As explained before each individual in the population has its own characteristics. Combining all generalizations leads to the following characteristics for an individual

$i \in \{S, I_1, I_2, T\}$	infection stage
$g \in \{m, f\}$	gender
$n_s \in \mathbb{N}$	number of steady binding sites
$n_c \in \mathbb{N}$	number of casual binding sites

Note that n_s and n_c are distributed according to the random variables $Z(g,c)$ and $Z(g,s)$, dependent on the gender of the individual. Many other parameters are also dependent on properties as gender, but do not vary individually. Therefore we do not have to keep these parameters as individual characteristics.

The events and their corresponding rates

In this section we will list all events and their rates, that occur in a simulation model combing all above generalizations. But first let us define

$X(g,h)$	the number of free binding sites of type h of individuals of gender g
$Y(h)$	the number of partnerships of type h
$N(i)$	the number of individuals in infection stage i
$[i_m, i_f]_h$	the number of partnerships of type h where the male is in infection stage i_m, the female is in infection stage i_f

We distinguish the events and their corresponding rates shown in table 6.

Event	**rate**
Birth male	$\frac{1}{1+x}\mu N_0$
Birth female	$\frac{x}{1+x}\mu N_0$
Death of stage S	$\mu(S)N(S)$
Death of stage I_1	$\mu(I_1)N(I_1)$
Death of stage I_2	$\mu(I_2)N(I_2)$
Death of stage T	$\mu(T)N(T)$
Separation casual partnerships	$\sigma(c)Y(c)$
Separation steady partnerships	$\sigma(s)Y(s)$
Partnership formation casual	$\tilde{\rho}(c)X(m,c)X(f,c)$
Partnership formation steady	$\tilde{\rho}(s)X(m,s)X(f,s)$
Transmission casual: male I_1 - female S	$\beta(I_1,m,c)[I_1,S]_c$
Transmission casual: male I_2 - female S	$\beta(I_2,m,c)[I_2,S]_c$
Transmission casual: male S - female I_1	$\beta(I_1,f,c)[S,I_1]_c$
Transmission casual: male S - female I_2	$\beta(I_2,f,c)[S,I_2]_c$
Transmission steady: male I_1 - female S	$\beta(I_1,m,s)[I_1,S]_s$
Transmission steady: male I_2 - female S	$\beta(I_2,m,s)[I_2,S]_s$
Transmission steady: male S - female I_1	$\beta(I_1,f,s)[S,I_1]_s$
Transmission steady: male S - female I_2	$\beta(I_2,f,s)[S,I_2]_s$
$I_1 \to I_2$	$\alpha(I_1)N(I_1)$
$I_2 \to T$	$\alpha(I_2)N(I_2)$

Table 6: Transition rates of events in the simulation model with generalizations.

4 The program

In this section we will describe how the individual-based model from sections 2 and 3 is converted to a simulation program in Python (version 3.2). The program will roughly consist of the following steps:

- (Step 1) The rates of all events are calculated.
- (Step 2) The time till the next event is calculated.
- (Step 3) An event is chosen according to the calculated rates.
- (Step 4) The event takes place.
- The event has changed the rates: return to (Step 1).

The general idea of the program is already clear from the explanations of the individual-based model. Therefore in this main text we will not go into full detail on the program structure. A full description of the the code and the process of programming can be found in the appendices A - C. In appendix A the 'blue print' used to program the first version of the program is found. In appendix B the optimization steps are shown that were taken during programming. In appendix C the program structure of the final version of the code is shown.

Despite not fully explaining the code in the main text, there are a few aspects of the code we do want to discuss. In section 4.1 we introduce the 'spin-up' and explain why we make use of it. In section 4.2 we explain what we mean with the terms 'simulation' and 'run'. In section 4.3 we explain the stopping criteria of the simulation. In section 4.4 we clarify some of the chosen input parameters of the program and show how to calculate the model parameters used in section 2 and section 3 from these input parameters. Lastly in section 4.5 we give an explanation of the output of the program.

4.1 Spin-up

We want to introduce the infection in a population that is stable without the infection, i.e. a stable network. Therefore we run the program first without infection, following the usage in climate modelling, we call this phase the *spin-up* of the simulation. This spin-up is stopped when the network 'stabilizes':

Let (g, x_c, x_s) denote the number of individuals in the population with gender g, x_c number of casual partners and x_s number of steady partners. For example $(male, 0, 0)$ denotes the number of single male individuals in the population. The distribution of the individuals over all possible (g, x_c, x_s) is called the degree distribution. If each number of individuals (g, x_c, x_s) stays within certain bounds for a certain amount of time, we say that the population without infection is stable. This is parameter dependent, for which the values will be given as input parameters of the program. An example of a spin-up is given in fig. 1. We can see that for the used parameter set the network appears to stabilize very quickly, it only takes a few years.

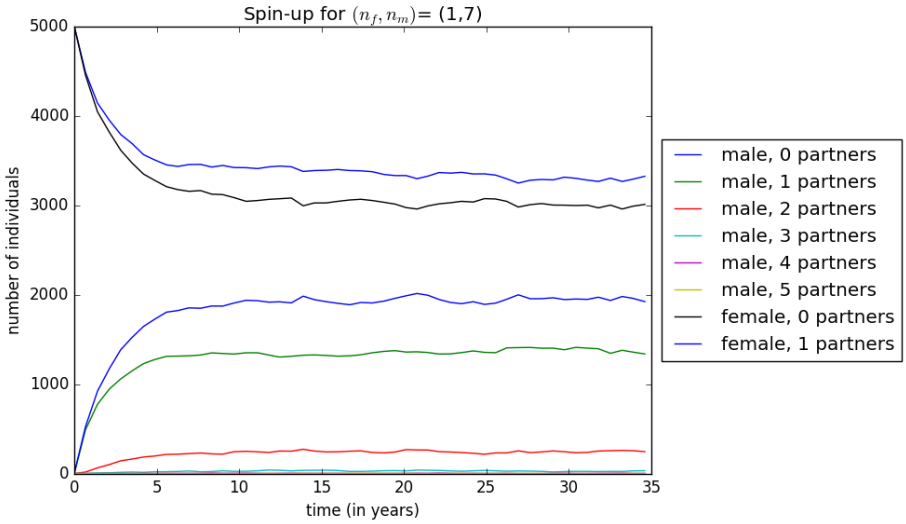

Figure 1: The spin-up of a heterosexual population of size 10.000, where only one type of partnership is allowed. Males have 7 binding sites and females only one. The rest of the parameter values are equal to the values we will use in the next section in table 7.

4.2 Simulation vs. run

We need to keep in mind that every time we perform a simulation, we will get different results. We will need to run the program many times for the same parameter set to get statistically significant results. Therefore we say that one *simulation* consists of a multiple number of *runs*. We introduce a new input parameter *nr_runs* that decides how many times the program will be run to form one simulation, which will be user defined.

4.3 Stopping criteria

After the spin-up, in other words when the network has stabilized, we introduce the infection on the stable network. Since we make use of a simulation program we want to perform multiple runs with the same input-parameter set. Since each time the same input-parameter set is used, the population after the spin-up can be re-used for all the runs. In each run we introduce the infection and want to run the program until the infection is 'stabilized'. Let us clarify the meaning of a stable infection. Let $N(g, i)$ denote the number of individuals of gender g in infection stage i. If each number of individuals (g, i) stay within certain bounds for a certain amount of time, we say that the infection is stable. This is parameter dependent, for which the values will be given as input parameters of the program.

However having multiple runs stop at different times (because the time until the infection is stabilized will vary between the runs) is not the most convenient way. It is much easier to simply give a maximum time and have all runs stop at that time. However we do need to know what approximate time the infection needs to stabilize for that input-parameter set. Therefore we give two options

in the code: either give a maximum time (useful when doing a simulation with multiple runs) or perform the run until the infection stabilizes (useful for only one run, to find the approximate time needed for stabilization of infection).

4.4 From input parameters to model parameters

Not all model parameters in section 3.6 will be used as input parameters. Some of the model parameters, such as 'the rate of separation' are hard to estimate or find in literature. Therefore we will not use these model parameters as input parameters, but express them in new input parameters, such as 'the mean partnership duration' [14]. The following model parameters will not be used as input parameters:

Model parameters	
$\mu(i)$	death rate
$\rho(h)$	parameter formation of partnership
$\sigma(h)$	separation rate
$\alpha(i)$	rate of transition fom the infection stage i to $i+1$ ($i \in \{I_1, I_2\}$)

We would like to express these model parameters in the following new input parameters:

Input parameters	
$L(i)$	the mean sexually active lifetime of individuals after entering infection stage i
$d_I(i)$	the mean duration of the infection stage i (for $i \in \{I_1, I_2\}$)
$d_P(h)$	the mean duration of a partnership of type h (without infection)
$\theta(m, h)$	the mean lifetime number of partners of type h (for male individuals, without infection)

Note that $d_P(h)$ and $\theta(h)$ are defined for the network without infection! Since this gives us enough information to calculate all the model parameters from the input parameters.

4.4.1 Calculations heterosexual population

In this section we will show how to calculate the model parameters from the input parameters in a heterosexual population.

Calculating the death rate $\mu(i)$ from the mean sexual active lifetime after entering stage i is simply

$$\mu(i) = \frac{1}{L(i)} \tag{15}$$

The expected duration of a partnership, without infection, can be expressed as

$$d_P(h) = \frac{1}{\sigma(h) + 2\mu(S)}$$

Therefore,

$$\sigma(h) = \frac{1}{d_P(h)} - 2\mu(S) \tag{16}$$

and since $\sigma(h) \geq 0$ for all h, it must hold that

$$d_P(h) \leq \frac{1}{2\mu(S)} = \frac{1}{2} \cdot L(S) \tag{17}$$

Which means that the mean partnership duration can not be more than half the mean sexual active lifetime of a susceptible individual. The mean duration of the infection phase i is

$$d_I(i) = \frac{1}{\mu(i) + \alpha(i)}$$

Therefore,

$$\alpha(i) = \frac{1}{d_I(i)} - \mu(i) \tag{18}$$

and since $\alpha(i) \geq 0$ for all i, it must hold that

$$d_I(i) \leq \frac{1}{\mu(i)} = L(i) \tag{19}$$

The last model parameter, $\rho(h)$, is much harder to express in the input parameters. Let $\theta(g, h)$ be the mean lifetime number of partnerships of type h of an individual of gender g, without infection. To express $\theta(g, h)$ in the model-parameters, let us first look at the transitions of the two possible states of a binding site of type h belonging to an individual of gender g. A binding site can be in two states: it can be either 'free' or 'occupied'. In figure 2 an overview of the transitions of these two states are illustrated. Note that the rates of the transitions are expressed in the deterministic model parameters. We will convert them later to the simulation parameters.

However we have to keep in mind that in the deterministic model it was assumed that the fraction of free binding sites $F(t) = \bar{F}$ is constant (eq. (5)). We assume the same for the simulation model, or at least that $F(t)$ stabilizes around a mean value $\bar{F}$. We will later check the correctness of this assumption in section 5.

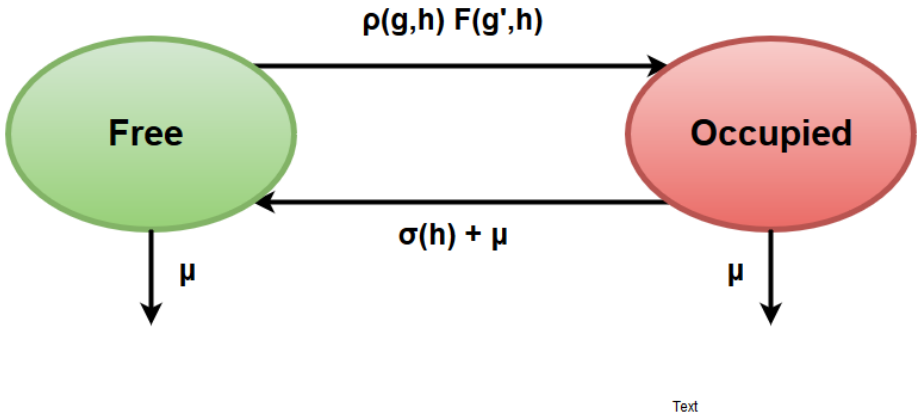

Figure 2: Overview of the transitions of the two possible states of a binding site of type t belonging to an individual of gender g, without infection. The binding site can either be in the state 'Free' or 'Occupied'.

Let $Occ(g,h)$ be the expected future number of times that a free binding site (of type h belonging to an individual of gender g) becomes occupied. The probability that the free binding site becomes occupied is $\frac{\rho(g,h)F(g',h)}{\rho(g,h)F(g',h)+\mu(S)}$. When the free binding site becomes occupied we have to add one to $Occ(g,h)$. The probability that the occupied binding site becomes free again is $\frac{\sigma(h)+\mu(S)}{\sigma(h)+2\mu(S)}$. When the occupied binding site becomes free again, we are back at the beginning, and the expected future number of times the free binding site becomes occupied is again $Occ(g,h)$.

Therefore, to summarize in one equation, we have that

$$Occ(g,h) = \frac{\rho(g,h)F(g',h)}{\rho(g,h)F(g',h)+\mu(S)}\left(1+\frac{\sigma(h)+\mu(S)}{\sigma(h)+2\mu(S)}Occ(g,h)\right)$$

Which gives us

$$Occ(g,h) = \frac{\rho(g,h)F(g',h)\Big(\sigma(h)+2\mu(S)\Big)}{\mu(S)\Big(\rho(g,h)F(g',h)\sigma(h)+2\mu(S)\Big)} \tag{20}$$

Since individuals of gender g have an expected number of $E[n(g,h)]$ binding sites of type h, the mean lifetime number of partnerships of type h of an individual of gender g can be expressed as

$$\theta(g,h) = Occ(g,h)\cdot E[n(g,h)] = \frac{\rho(g,h)F(g',h)\Big(\sigma(h)+2\mu(S)\Big)E[n(g,h)]}{\mu(S)\Big(\rho(g,h)F(g',h)\sigma(h)+2\mu(S)\Big)} \tag{21}$$

Note that consistency requires that

$$\begin{aligned}\theta(m,h) &= \frac{\rho(m,h)F(f,h)\Big(\sigma(h)+2\mu(S)\Big)E[n(m,h)]}{\mu(S)\Big(\rho(m,h)F(f,h)\sigma(h)+2\mu(S)\Big)} \\ &= x\cdot\frac{\rho(f,h)F(m,h)\Big(\sigma(h)+2\mu(S)\Big)E[n(f,h)]}{\mu(S)\Big(\rho(f,h)F(m,h)\sigma(h)+2\mu(S)\Big)} \\ &= x\cdot\theta(f,h)\end{aligned}$$

If we know the mean number of lifetime partners of one gender, we also know the mean number of lifetime partners for the opposite gender. Therefore we only have the input parameter $\theta(m,h)$ as the mean number of lifetime partners for male individuals.

Let us not forget that the reason for these calculations is to express $\rho(g,h)$ in the input parameters. Therefore let us express the fraction of free binding sites (of type h belonging to an individual of gender g') in the input parameters as

$$F(g',h) = 1 - \frac{d_P(h)\theta(g',h)}{1/\mu(S)\cdot E[n(g',h)]} \tag{22}$$

Since $0 \leq F(g',h) \leq 1$ for all g and h, it must hold that

$$0 \leq d_P(h)\theta(g',h) \leq \frac{E[n(g',h)]}{\mu(S)} = E[n(g',h)]\cdot L(S) \tag{23}$$

Insert equation 16 and equation 22 into the equation 21. This allows us to express $\rho(g,h)$ in the input parameters

$$\rho(g,h) = \frac{E[n(g',h)]\theta(g,h)\mu(S)}{\Big(E[n(g',h)] - d_P(h)\theta(g',h)\mu(S)\Big)\Big(E[n(g,h)] - d_P(h)\theta(g,h)\mu(S)\Big)} \tag{24}$$

We see that

$$\begin{aligned}
\rho(m,h) &= \frac{E[n(f,h)]\theta(m,h)\mu(S)}{\Big(E[n(f,h)] - d_P(h)\theta(f,h)\mu(S)\Big)\Big(E[n(m,h)] - d_P(h)\theta(m,h)\mu(S)\Big)} \\
&= \frac{E[n(m,h)]}{E[n(m,h)]} \cdot \frac{E[n(f,h)]x\theta(f,h)\mu(S)}{\Big(E[n(f,h)] - d_P(h)\theta(f,h)\mu(S)\Big)\Big(E[n(m,h)] - d_P(h)\theta(m,h)\mu(S)\Big)} \\
&= x \cdot \frac{E[n(f,h)]}{E[n(m,h)]} \cdot \frac{E[n(m,h)]\theta(f,h)\mu(S)}{\Big(E[n(f,h)] - d_P(h)\theta(f,h)\mu(S)\Big)\Big(E[n(m,h)] - d_P(h)\theta(m,h)\mu(S)\Big)} \\
&= x \cdot \frac{E[n(f,h)]}{E[n(m,h)]} \cdot \rho(f,h)
\end{aligned}$$

Therefore it must hold that

$$E[n(m,h)]\rho(m,h) = x \cdot E[n(f,h)] \cdot \rho(f,h) \tag{25}$$

We calculate the simulation parameters $\tilde{\rho}(h)$ (see section 3.2) from the deterministic model parameters $\rho(g,h)$,

$$\begin{aligned}
\tilde{\rho}(h) &= \frac{1}{N_m E[n(m,h)]}\rho(f,h) \\
&= \frac{1}{N_m E[n(m,h)]}\rho(m,h)\frac{E[n(m,h)]}{E[n(f,h)]x} \\
&= \frac{1}{N_m E[n(f,h)]x}\rho(m,h) \\
&= \frac{1}{N_f \frac{1}{x} E[n(f,h)]x}\rho(m,h) \\
&= \frac{1}{N_f E[n(f,h)]}\rho(m,h)
\end{aligned}$$

where we have used the equations 25 and

$$N_m = \frac{1}{x}N_f.$$

Therefore we can express $\tilde{\rho}(h)$ as

$$\tilde{\rho}(h) = \frac{1}{N_{g'}} \frac{\theta(g,h)\mu(S)}{\Big(E[n(g',h)] - d_P(h)\theta(g',h)\mu(S)\Big)\Big(E[n(g,h)] - d_P(h)\theta(g,h)\mu(S)\Big)} \tag{26}$$

4.4.2 Overview Calculations

To calculate the model parameters from the input parameters for a homosexual population, we follow the same steps as in section 4.4.1. An overview of the model parameters expressed in the new input-parameters is shown below:

Model parameters expressed in input parameters		
	heterosexual	homosexual
$\mu(i)$	$\frac{1}{L(i)}$	$\frac{1}{L(i)}$
$\rho(h)$	$\frac{1}{N_f} \frac{\theta(m,h)\mu(S)}{\left(E[n(f,h)]-d_P(h)\frac{1}{x}\theta(m,h)\mu(S)\right)\left(E[n(m,h)]-d_P(h)\theta(m,h)\mu(S)\right)}$	$\frac{1}{N} \frac{\theta(h)\mu(S)}{\left(E[n(h)]-d_P(h)\theta(h)\mu(S)\right)^2}$
$\sigma(h)$	$\frac{1}{d_P(h)} - 2\mu(S)$	$\frac{1}{d_P(h)} - 2\mu(S)$
$\alpha(i)$	$\frac{1}{d_I(i)} - \mu(i)$	$\frac{1}{d_I(i)} - \mu(i)$

4.5 Output

Another imporant aspect of the simulatiom program we have not yet discussed is the output. Think for example of:

- $N(g, i)$: the number of individuals of gender g in infection stage i.
- $N(g, x_c, x_s)$: the number of individuals of gender g with x_c number of casual partners and x_s number of steady partners, i.e. the degree distribution.

But of course it would be even more interesting to save the entire network structure: to save for each individual in the population its name, gender, infection stage, partners, partnership capacity etc. Saving the entire network structure is however a lot of output, especially for large populations. Therefore saving the entire network structure is not something we want to do too often in a simulation run. However saving output such as $N(g, i)$ and $N(g, x_c, x_s)$ very often is hardly any problem.

Therefore we implemented to have two different input parameters: *nr_save* and *nr_save_pop*. Where *nr_save_pop* denotes the number of times to save the entire network structure in one run and *nr_save* denotes the number of times to save other, 'smaller' output in one run.

5 Results

This section will be devoted to the results of the simulation model for a certain set of baseline parameter values. In the next section (section 6) we will study the variation around some of these baseline values in a sensitivity analysis.

First we will discuss the chosen parameter values in section 5.1. Next, in section 5.2, we will give a short explanation of confidence intervals, which we will use to illustrate the results of the simulation model. In section 5.3 we will compare the results of the deterministic model with the results of the simulations. Finally, in section 5.4, we will further investigate the network structures of the simulation model, since in the simulations it is known exactly who has a relationship whit whom.

5.1 Parametrization

We have chosen a set of input parameter values used in [14] so that we can compare the results of the simulations to the results of the deterministic model found in this paper, to check the agreement between the two models and the correctness of the mean field at distance one assumption

We consider at a heterosexual population with only one type of partnership and an one to one sex ratio. A simulation will consist of a 100 runs, and we will look at a population of either mean size 1.000 or 10.000. In this population we will follow the spread of a SI_1I_2 infection and we will introduce this infection in the population by putting 100 randomly chosen individuals of the initial population in infection stage I_1. Note that this means that in the case of a population of 1.000 we start with 10% intially infected and in the case of a population size of 10.000 we start with 1% intially infected.

It would be desired to begin with less than 100 initially infected individuals, because then we include the early beginning of the spread of the infection in our simulations. But then the chances of the infection dying out in the early stages of the infection are higher, and since we are interested in the endemic prevalence of the infection, this is a situation we would like to prevent. Starting with 100 initially infected has no influence on the endemic state, only on the beginning of the spread of the infection. Therefore please note that the 'endemic state' and the 'beginning of the epidemic' are considered to be two separate phases of the epidemic. In this section we will focus on the infection when it has reached its endemic state, since we start the simulations with 100 initially infected individuals.

An overview of the input parameters is shown in table 7. We will from now on refer to these input parameter values as the *baseline parameter values*.

	Parameter	Description	Value
Network	N_0	mean population size	$\begin{cases} 1.000 \\ 10.000 \end{cases}$
	x	the ratio male : female (1 : x)	1
	homo	heterosexual (0) or homosexual (1) population	0
	L	mean life expectancy	35 years
	d_P	mean partnership duration	3.92 years
	θ	mean lifetime number of partners	3.5
	$Z(g)$	partnership capacity distributions	$\begin{cases} (n_f, n_m) = (1, 7) \\ (n_f, n_m) = (2, 6) \\ (n_f, n_m) = (3, 5) \\ (n_f, n_m) = (4, 4) \end{cases}$
Infection	β_1	transmission rate of infection stage I_1	2.76/year
	β_2	transmission rate of infection stage I_2	0.106/year
	d_{I_1}	mean duration infection stage I_1	2.9/12 years
	nr_infected(I_1)	the number of initial individuals in infection stage I_1	100
	nr_infected(I_2)	the number of initial individuals in infection stage I_2	0
Rest	*nr_runs*	the number of simulation runs	100
	max_time	the maximum time per simulation run	2000 years

Table 7: Baseline parameter values used in the simulations. The time unit is one year. The parameter values are based one the input parameter set used in [14].

In table 7 we see that we consider multiple values for the following two parameters:

- **The population size** N_0
  We run the simulations for both $N_0 = 1.000$ and $N_0 = 10.000$, to see the effect of the population size on the simulation results. For example it will be interesting to see the influence of the population size to the level of agreement with the deterministic model.

- **The partnership capacity distribution** $Z(g)$
  These distributions will, just like N_0, vary between simulations. Let the males in the population have partnership capacity n_m and the females partnership capacity n_f. Note that the partnership capacity can vary between individuals of different genders, but can not vary between individuals of the same gender. Therefore the 'distributions' are simply constant values. We will use the partnership capacity values $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$ and refer to them from now on as *scenarios*. These four scenarios were used for the deterministic model in [14]. These scenarios have been chosen in such a way that the mean partnership capacity is always equal to 4.

5.2 Confidence Interval

To illustrate the results from the simulations, we will calculate the mean of certain results of the 100 runs. We also derive a *confidence interval* for the mean. A confidence interval for the mean is an interval, calculated from the simulation results, that contains the true mean with some specified probability. If the coverage probability is $1-\alpha$, the interval is called a $100 \cdot (1-\alpha)\%$ confidence interval. In this masterthesis we make use of 95% confidence intervals, which is one of the most common choices in literature. Therefore the confidence intervals calculated in this masterthesis are intervals that contain the true mean with probability 0.95.

To explain the theory behind the confidence interval we follow [15]. Before explaining how to calculate the confidence interval from the simulation results, let us introduce the *Central Limit Theorem*. Let us consider a random sample $\{X_1, X_2, \ldots, X_n\}$ of size n, where $X_1, X_2, \ldots, X_n$ is a sequence of independent and identically distributed random variables with population mean μ and population variance σ^2. Let us define the *sample mean* as

$$\overline{X}_n := \frac{\sum_{i=1}^{n} X_i}{n}.$$

The probability distribution of the sample mean $\overline{X}_n$ is called its *sampling distribution of the sample mean*, with mean $\mu_{\overline{X}}$ and variance $\sigma^2_{\overline{X}}$. The sampling distribution of the sample mean describes how accurate the sample mean $\overline{X}_n$ estimates the population mean μ.

We know from the *Law of Large Numbers* that the sample mean $\overline{X}_n$ converges to μ in probability as $n \to \infty$. The central limit theorem tells us that, as n becomes large, the sampling distribution of $\overline{X}_n$ approaches a normal distribution with mean $\mu_{\overline{X}} = \mu$ and variance $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}$. Let us standardize:

$$Z_n = \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$$

Then Z_n converges to the standard normal distribution $N(0,1)$ as $n \to \infty$.

We can now calculate a size of $100 \cdot (1-\alpha)\%$ confidence interval for the population mean μ. Let $z(\alpha)$ be the value such that the area under the standard normal density function to the right of $z(\alpha)$ is equal to α, for $0 \leq \alpha \leq 1$. Note that, because of the symmetry of the standard normal density function around zero, it must holds that $z(1-\alpha) = -z(\alpha)$. Therefore it holds for $Z \sim N(0,1)$ that

$$P(-z(\alpha/2) \leq Z \leq z(\alpha/2)) = 1 - \alpha$$

Since $\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}}$ approximates a standard normal distribution,

$$P(-z(\alpha/2) \leq \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq z(\alpha/2)) \approx 1 - \alpha$$

As mentioned before, we will make use of a 95% confidence interval. Therefore

$$\begin{aligned} 0.95 &\approx P\left(-1.96 \leq \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq 1.96\right) \\ &= P\left(\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}} \leq \mu \leq \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}\right) \end{aligned}$$

So the 95% confidence interval is $[\overline{X}_n - 1.96\frac{\sigma}{\sqrt{n}}, \overline{X}_n + 1.96\frac{\sigma}{\sqrt{n}}]$.

However the population standard deviation σ is not known. But it can be estimated by calculating the unbiased *sample standard deviation* of the sample, also called its *standard error* s. Using $n-1$ in stead of n in the calculation of the sample standard deviation, gives us an unbiased estimator of the population standard deviation. Therefore,

$$\sigma \approx s := \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})_n^2}$$

Let the sample have mean $\overline{x}$ and standard deviation s. Then the 95% confidence intervals that we will calculate for the simulation results, can be expressed as

$$[\overline{x} - 1.96\frac{s}{\sqrt{n}}, \overline{x} + 1.96\frac{s}{\sqrt{n}}]$$

Please keep in mind that the assumption was made that the sample size $n \to \infty$ (see central limit theorem). Therefore the sample size needs to be large for the confidence interval to be meaningful.

	Population	**Sample**	**Sample distribution**
mean	μ	$\overline{x}$	$\mu_{\overline{X}}$
standard deviation	σ	s	$\sigma_{\overline{X}}$

Table 8: The different symbols used for population parameters and sample statistics.

5.3 Comparison with deterministic model

We will now compare the results of the deterministic model with the results of the simulation model for the baseline parameter values explained in section 5.1. We will compare the following results for both models:

- **The network**
  - **The degree distribution:**
    The degree distribution $P_g(k)$ of the network is defined as the fraction of individuals of gender $g \in \{male, female\}$ with $k \leq n_g$ partners in the total pool of individuals of gender g.
  - **The concurrency index:**
    The partnership-based concurrency index κ_P was introduced in [9] for the homosexual deterministic model with the following operational definition: choose a partnership at random from the pool of all partnerships, focus on one of the two partners and count how many partners it has besides the 'known' partner.

    In the simulation model we will use this interpretation to calculate the concurrency index. We go through all the partnerships in the network at a certain time, and for each partner in the partnership we calculate its number of partners, besides the 'known' partner. We take the mean of all these values to calculate the expected number κ_P.

    But since we are dealing with a heterosexual population in stead of a homosexual population we will calculate a seperate concurrency index for both men and women. Let $\kappa_P(g)$ have the following operational definition: choose a partnership at random from the pool of all partnerships, focus on the partner of gender g of the two partners and count how many partners it has besides the 'known' partner.

- **The infection**
  - **The endemic prevalence**
    The endemic prevalence means the proportion of infected individuals in the population, when the infection has reached endemic stability. We will calculate a seperate endemic prevalence for both male and female individuals and for each infection stage, meaning we calculate the proportion of infected individuals of gender g in infection stage i in the total pool of individuals of gender g for $g \in \{male, female\}$ and $i \in \{S, I_1, I_2, T\}$.
  - **The mean field at distance one**
    The mean field at distance one assumption, used in the deterministic model, has already been discussed in section 1.3 (eq. (10) and eq. (11)). $\Lambda_\pm$ was introduced as the expected number of infectious partners of a susceptible partner of an individual in infection state $\pm$. With the mean field at distance one assumption it is assumed that in the deterministic model these properties concerning partners of partners can be obtained by averaging over the population.

    One of the main reasons that has motivated this thesis is to check the correctness of this assumption, by comparing the values of Λ of the deterministic model (calculated

with the mean field at distance one assumption) with the values of Λ of the simulation model (in which no such assumption is made).

Since we now look at a heterosexual population with an SI_1I_2 infection, let us introduce $\Lambda(i, i_p, g)$ as the expected number of partners in infection stage $i_p \in \{I_1, I_2\}$ of a susceptible partner of an individual of gender g in infection stage $i \in \{S, I_1, I_2\}$. An illustration to clarify the meaning of $\Lambda(i, i_p, g)$ is shown in fig. 3.

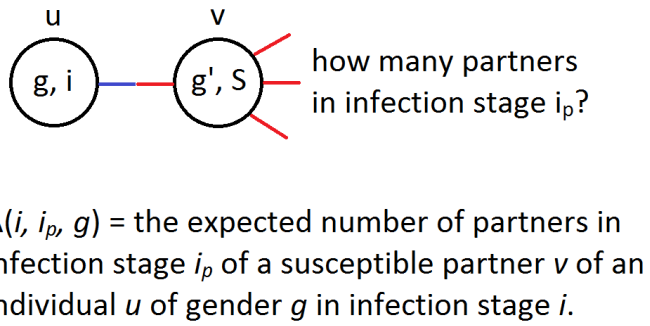

Figure 3: Illustration of the meaning of the mean field at distance one value $\Lambda(i, i_p, g)$.

Network

In this subsection we will compare the network of the deterministic model and the network of the simulations by studying the degree distribution and the concurrency index. As explained in section 4.1 before starting the simulations, we start the program without infection, and introduce the infection after the network has stabilized. With the chosen baseline parameter values (table 7), the infection has no influence on the partnership dynamics. Therefore during the entire simulation we will expect to be dealing with a stable network. Thus concerning the network properties, it should not differ when the means and confidence intervals are calculated, whether it be at the beginning, middle or end of the simulations. To check this, we have calculated the means and confidence intervals on different times in the simulations, and have seen that network properties indeed do not differ significantly. For all future results we have calculated the network properties at the end of the simulation, 2.000 year after introducing the infection.

As explained is section 5.2 we will illustrate the simulation results of the 100 runs by calculating the mean and the 95% confidence interval. We will put the results of the deterministic model with the mean and confidence interval of the simulations together in tables for easy comparison. We have also illustrated the results in plots, which can be found in appendix D.

Degree Distribution $(n_f, n_m) = (1, 7)$			
	Deterministic	**Simulations** mean [95% CI]	
gender, nr partners		$N = 1.000$	$N = 10.000$
male, 0	0.671	0.673 [0.670, 0.677]	0.672 [0.671, 0.674]
male, 1	0.272	0.269 [0.266, 0.273]	0.271 [0.270, 0.272]
male, 2	0.0510	0.0518 [0.0499, 0.0538]	0.0511 [0.0506, 0.0517]
male, 3	0.00543	0.00508 [0.00445, 0.00570]	0.00541 [0.00523, 0.00560]
male, 4	0.000351	0.000217 [7.25 $\cdot 10^{-5}$, 0.000362]	0.000371 [0.000323, 0.000418]
male, 5	1.37 $\cdot 10^{-5}$	0.0 [0.0, 0.0]	6.05 $\cdot 10^{-6}$ [-7.26 $\cdot 10^{-7}$, 1.28 $\cdot 10^{-5}$]
male, 6	2.99 $\cdot 10^{-7}$	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
male, 7	2.81 $\cdot 10^{-9}$	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
female, 0	0.608	0.610 [0.606, 0.614]	0.610 [0.608, 0.611]
female, 1	0.392	0.390 [0.386, 0.394]	0.390 [0.389, 0.392]

Table 9: The degree distribution for $(n_f, n_m) = (1, 7)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years. The degree distribution is the fraction of individuals of gender g with $k \leq n_g$ partners in the total pool of individuals of gender g.

Degree Distribution $(n_f, n_m) = (2, 6)$			
	Deterministic	**Simulations** mean [95% CI]	
gender, nr partners		$N = 1.000$	$N = 10.000$
male, 0	0.670	0.668 [0.664, 0.673]	0.670 [0.669, 0.671]
male, 1	0.275	0.276 [0.272, 0.280]	0.274 [0.273, 0.275]
male, 2	0.0505	0.0501 [0.0480, 0.0522]	0.0505 [0.0498, 0.0511]
male, 3	0.00507	0.00505 [0.00436, 0.00575]	0.00525 [0.00505, 0.00545]
male, 4	0.000290	0.000321 [0.000166, 0.000475]	0.000276 [0.000233, 0.000318]
male, 5	8.91 $\cdot 10^{-6}$	3.88 $\cdot 10^{-5}$ [-1.48 $\cdot 10^{-5}$, 9.24 $\cdot 10^{-5}$]	1.21 $\cdot 10^{-5}$ [1.08 $\cdot 10^{-6}$, 2.32 $\cdot 10^{-5}$]
male, 6	1.14 $\cdot 10^{-7}$	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
female, 0	0.648	0.648 [0.643, 0.653]	0.648 [0.647, 0.650]
female, 1	0.312	0.312 [0.307, 0.316]	0.311 [0.310, 0.313]
female, 2	0.0402	0.0403 [0.0384, 0.0422]	0.0406 [0.0400, 0.0412]

Table 10: The degree distribution for $(n_f, n_m) = (2, 6)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years. The degree distribution is the fraction of individuals of gender g with $k \leq n_g$ partners in the total pool of individuals of gender g.

Degree Distribution $(n_f, n_m) = (3, 5)$			
	Deterministic	**Simulations** mean [95% CI]	
gender, nr partners		$N = 1.000$	$N = 10.000$
male, 0	0.668	0.669 [0.664, 0.673]	0.667 [0.666, 0.668]
male, 1	0.278	0.278 [0.274, 0.282]	0.279 [0.277, 0.280]
male, 2	0.0498	0.0497 [0.0479, 0.0515]	0.0495 [0.0489, 0.0500]
male, 3	0.00457	0.00393 [0.00337, 0.00449]	0.00458 [0.00438, 0.00478]
male, 4	0.000212	0.000204 [$7.07 \cdot 10^{-5}$, 0.000337]	0.000200 [0.000160, 0.000240]
male, 5	$3.96 \cdot 10^{-6}$	0.0 [0.0, 0.0]	$6.00 \cdot 10^{-6}$ [$-7.22 \cdot 10^{-7}$, $1.27 \cdot 10^{-5}$]
female, 0	0.659	0.662 [0.658, 0.667]	0.659 [0.658, 0.660]
female, 1	0.292	0.288 [0.284, 0.292]	0.292 [0.291, 0.293]
female, 2	0.0463	0.0473 [0.0452, 0.0493]	0.0466 [0.0460, 0.0472]
female, 3	0.00251	0.00239 [0.00195, 0.00282]	0.00251 [0.00236, 0.00266]

Table 11: The degree distribution for $(n_f, n_m) = (3, 5)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years. The degree distribution is the fraction of individuals of gender g with $k \leq n_g$ partners in the total pool of individuals of gender g.

Degree Distribution $(n_f, n_m) = (4, 4)$			
	Deterministic	**Simulations** mean [95% CI]	
gender, nr partners		$N = 1.000$	$N = 10.000$
male, 0	0.665	0.664 [0.660, 0.669]	0.664 [0.663, 0.666]
male, 1	0.283	0.282 [0.279, 0.286]	0.284 [0.282, 0.285]
male, 2	0.0486	0.0495 [0.0477, 0.0513]	0.0484 [0.0478, 0.0490]
male, 3	0.00380	0.00385 [0.00329, 0.00441]	0.00373 [0.00356, 0.00389]
male, 4	0.000113	$6.00 \cdot 10^{-5}$ [$-7.22 \cdot 10^{-6}$, 0.000127]	0.000132 [$9.61 \cdot 10^{-5}$, 0.000169]
female, 0	0.665	0.668 [0.664, 0.672]	0.665 [0.663, 0.666]
female, 1	0.283	0.278 [0.274, 0.283]	0.283 [0.281, 0.284]
female, 2	0.0486	0.0492 [0.0472, 0.0513]	0.0487 [0.0481, 0.0493]
female, 3	0.00380	0.00421 [0.00364, 0.00479]	0.00376 [0.00361, 0.00392]
female 4	0.000113	$6.09 \cdot 10^{-5}$ [$-7.42 \cdot 10^{-6}$, 0.000129]	0.0001113 [$8.06 \cdot 10^{-5}$, 0.000141]

Table 12: The degree distribution for $(n_f, n_m) = (4, 4)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years. The degree distribution is the fraction of individuals of gender g with $k \leq n_g$ partners in the total pool of individuals of gender g.

Degree variance and mean			
	Deterministic	**Simulations** mean [95% CI]	
		$N_0 = 1.000$	$N_0 = 10.000$
(4,4)			
mean population	0.392	0.391 [0.386, 0.3968]	0.392 [0.390, 0.394]
male variance	0.360	0.360 [0.355, 0.366]	0.359 [0.357, 0.361]
female variance	0.360	0.362 [0.355, 0.369]	0.360 [0.358, 0.362]
(3,5)			
mean population	0.392	0.389 [0.384, 0.395]	0.392 [0.391, 0.394]
male variance	0.368	0.362 [0.356, 0.369]	0.367 [0.366, 0.369]
female variance	0.346	0.346 [0.340, 0.352]	0.347 [0.345, 0.349]
(2,6)			
mean population	0.392	0.392 [0.387, 0.398]	0.392 [0.390, 0.394]
male variance	0.373	0.373 [0.365, 0.381]	0.374 [0.372, 0.376]
female variance	0.319	0.318 [0.313, 0.323]	0.320 [0.318, 0.321]
(1,7)			
mean population	0.392	0.389 [0.385, 0.393]	0.391 [0.389, 0.392]
male variance	0.377	0.374 [0.367, 0.380]	0.377 [0.376, 0.379]
female variance	0.238	0.237 [0.236, 0.238]	0.238 [0.238, 0.238]

Table 13: Degree Variance and Mean for $(n_f, n_m) = (1,7), (2,6), (3,5)$ and $(4,4)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years. In this table we show the mean of the degree distribution of the population, and the variance of the degree distribution of both male and female individuals.

Concurrency index			
	Deterministic	**Simulations** mean [95% CI]	
		$N = 1.000$	$N = 10.000$
(4,4)			
male	0.310	0.311 [0.301, 0.321]	0.308 [0.305, 0.311]
female	0.310	0.317 [0.304, 0.330]	0.309 [0.306, 0.313]
(3,5)			
male	0.331	0.320 [0.308, 0.332]	0.329 [0.325, 0.332]
female	0.275	0.277 [0.267, 0.287]	0.276 [0.272, 0.279]
(2,6)			
male	0.345	0.341 [0.326, 0.355]	0.347 [0.343, 0.351]
female	0.205	0.204 [0.197, 0.212]	0.207 [0.204, 0.209]
(1,7)			
male	0.355	0.349 [0.337, 0.361]	0.356 [0.353, 0.359]
female	0.0	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]

Table 14: Concurrency Index for $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years.

The results of the degree distribution are shown in Tables 9 - 12, the degree variance and mean in Table 13 and the concurrency index in Table 14. We can see from the results that the network structure of the simulations seems to agree very well with the network structure of the deterministic model. The deterministic value almost always falls within the confidence interval of the simulation model. There are some exceptions for very small values, such as the fraction of male individuals with 5 or more partners in Table 9. This could be caused by the fact that these cases appear very little, or not at all, in the 100 simulation runs. It is not unreasonable for the deterministic results and the simulation results to no longer coincide for such very small values.

This effect can also be seen in the difference in the means and confidence intervals between men and women for the scenario (4,4). In the symmetric scenario (4,4) the results should be the same for men and women. But we can see in the results that this is not the case, simply caused by the variablity between the runs, even when taking a 100 runs.

The mean values of the simulations with $N_0 = 10.000$ do improve slightly compared to the mean values of the simulations with $N_0 = 1.000$. Although the mean values of $N_0 = 1.000$ already seem to be a good approximation of the deterministic model. However the most effect of increasing the population size seems to be on the confidence interval, which becomes much smaller for a larger population size. The variability between the runs is much greater with $N_0 = 1.000$ than $N_0 = 10.000$.

Infection

In this subsection we will compare the infection of the deterministic model and the simulation by looking at the endemic level and the mean field at distance one assumption. We look at the infection when it has reached endemic stability. Therefore all results are calculated at the end of the simulations. Just like the comparison of the network, we will show the deterministic results with the means and confidence intervals of the simulations together in one table. The results illustrated as plots can be found in appendix D.

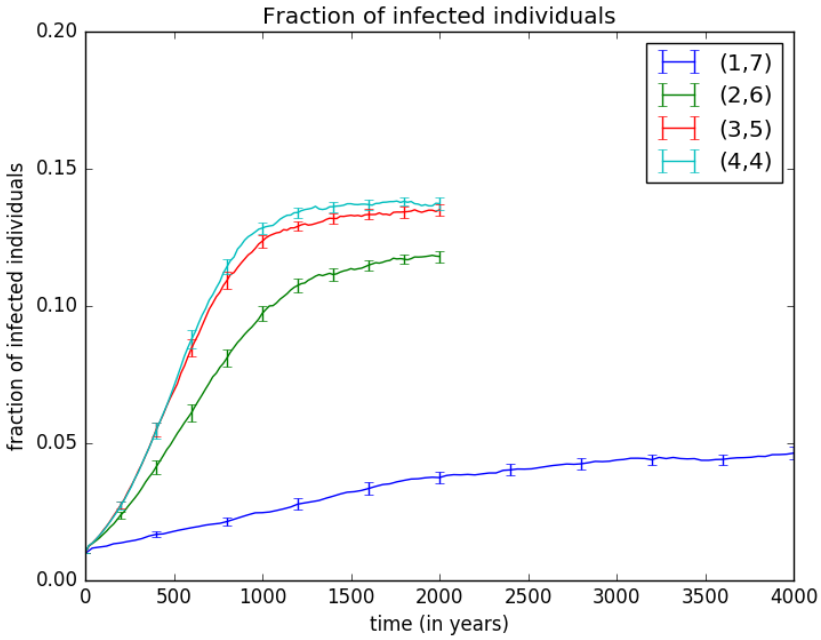

Figure 4: The fraction of infected $(I_1 + I_2)$ (mean and 95% confidence interval) over time for $(n_f, n_m) = (1,7), (2,6), (3,5)$ and $(4,4)$, with $N_0 = 10.000$. Over a period of 2.000 years for $(n_f, n_m) = (2,6), (3,5), (4,4)$ and for a period of 4.000 years for $(n_f, n_m) = (1,7)$.

As explained before we calculate the endemic level at the end of the simulation, 2.000 years after introducing the infection. We have chosen a maximum time of 2.000 years for the parameter baseline values in table 7, based on the stabilization of the infection in the scenario $(n_f, n_m) = (4,4)$ and $N_0 = 10.000$. But when we plot the fraction infected individuals over time for all 4 scenarios (see Figure 4), we see that the time needed for the infection to reach endemic stability differs for the scenarios. Especially for the case $(n_f, n_m) = (1,7)$ it takes much longer to reach endemic stability, and we can see that it has not yet been reached after 2.000 years. This slow spread of the infection for scenario (1,7) will be later explained in section 5.4.

But because of this slow spread of infection, to calculate the endemic level and other properties of the infection in its endemic state, we will take a maximum time of 4.000 years for the scenario $(n_f, n_m) = (1,7)$.

Endemic Level			
	Deterministic	**Simulations** mean [95% CI]	
		$N = 1.000$	$N = 10.000$
(4,4)			
male I_1	0.000956	0.00106 [0.000791, 0.00133]	0.000870 [0.000787 0.000953]
male I_2	0.137	0.130 [0.124, 0.136]	0.137 [0.134, 0.139]
female I_1	0.000956	0.000845 [0.000567, 0.00112]	0.000964 [0.000879, 0.00105]
female I_2	0.137	0.132 [0.125, 0.138]	0.136 [0.134, 0.138]
(3,5)			
male I_1	0.000919	0.000645 [0.000422, 0.000868]	0.000987 [0.000902, 0.00107]
male I_2	0.132	0.126 [0.120, 0.133]	0.132 [0.131, 0.134]
female I_1	0.000939	0.000945 [0.000679, 0.00121]	0.000945 [0.000867, 0.00102]
female I_2	0.135	0.128 [0.121, 0.135]	0.136 [0.133, 0.138]
(2,6)			
male I_1	0.000793	0.000814 [0.000489, 0.00114]	0.000815 [0.000729, 0.000900]
male I_2	0.114	0.100 [0.0944, 0.106]	0.114 [0.112, 0.117]
female I_1	0.000838	0.000637 [0.000439, 0.000835]	0.000809 [0.000732, 0.000887]
female I_2	0.120	0.105 [0.0990, 0.112]	0.120 [0.118, 0.122]
(1,7)			
male I_1	0.000328	0.000321 [0.000154, 0.000487]	0.000328 [0.000275, 0.000380]
male I_2	0.0471	0.0318 [0.0262, 0.0374]	0.0426 [0.0405, 0.0447]
female I_1	0.000377	0.000222 [$6.333 \cdot 10^{-5}$, 0.000381]	0.000326 [0.000272, 0.000381]
female I_2	0.0542	0.0365 [0.0300, 0.0431]	0.0495 [0.0471, 0.0519]

Table 15: Endemic level for $(n_f, n_m) = (1,7), (2,6), (3,5)$ (at $t = 2.000$ years) and $(4,4)$ (at $t = 4.000$ years) with $N_0 = 1.000$ or 10.000.

Let us first compare the endemic level (see Table 15). We see that the endemic levels of the deterministic model often do not lie within the confidence interval of the simulation model for $N_0 = 1.000$. But since in the deterministic model a large population is assumed, it is not strange that the simulation results for a population of size 1.000 do not completely coincide with the results of the deterministic model. The means and confidence intervals of $N_0 = 10.000$ are already a much better approximation and the deterministic values almost always fall within the confidence intervals.

It is notable that from the three cases that the deterministic values do not fall within the confidence interval for $N_0 = 10.000$, two of them are for the scenario (1,7). And that for both the confidence intervals are lower than the endemic levels of the deterministic model. Although we have increased the maximum time for the scenario (1,7), apparently the infection still not has reached endemic stability, even after 4.000 years (see fig. 4). Which have caused the deviating values for (1,7).

Mean field at distance one $(n_f, n_m) = (4, 4)$			
	Deterministic	**Simulations** mean [95% CI]	
		$N = 1.000$	$N = 10.000$
$\Lambda(S, I_1, m)$	0.000143	0.000556 [0.000145, 0.000968]	0.000158 [9.23 $\cdot 10^{-5}$, 0.000224]
$\Lambda(S, I_2, m)$	0.0305	0.0293 [0.0258, 0.0327]	0.0299 [0.0288, 0.0310]
$\Lambda(I_1, I_1, m)$	1.00	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, m)$	0.0305	0.0 [0.0, 0.0]	0.00463 [-0.00444, 0.0137]
$\Lambda(I_2, I_1, m)$	0.000142	0.0 [0.0, 0.0]	6.76 $\cdot 10^{-5}$ [0.0, 0.000200]
$\Lambda(I_2, I_2, m)$	1.03	1.03 [1.02, 1.05]	1.03 [1.03, 1.03]
$\Lambda(S, I_1, f)$	0.000143	6.94 $\cdot 10^{-5}$ [-6.67 $\cdot 1-^{-5}$, 0.000206]	0.000140 [7.22 $\cdot 10^{-5}$, 0.000209]
$\Lambda(S, I_2, f)$	0.0305	0.0298 [0.0263, 0.0333]	0.0302 [0.0291, 0.0312]
$\Lambda(I_1, I_1, f)$	1.00	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, f)$	0.0305	0.200 [-0.192, 0.592]	0.0213 [-0.0204, 0.0630]
$\Lambda(I_2, I_1, f)$	0.000142	0.000625 [-0.000600, 0.00185]	7.94 $\cdot 10^{-5}$ [0.0, 0.000235]
$\Lambda(I_2, I_2, f)$	1.03	1.02 [1.01, 1.03]	1.03 [1.03, 1.03]

Table 16: Mean field at distance one for $(n_f, n_m) = (4, 4)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years.

Mean field at distance one $(n_f, n_m) = (3, 5)$			
	Deterministic	**Simulations** mean [95% CI]	
		$N = 1.000$	$N = 10.000$
$\Lambda(S, I_1, m)$	0.000133	0.000258 [-4.64e $\cdot 10^{-5}$, 0.000562]	0.000118 [6.28 $\cdot 10^{-5}$, 0.000172]
$\Lambda(S, I_2, m)$	0.0315	0.0279 [0.0247, 0.0311]	0.0313 [0.0301, 0.0325]
$\Lambda(I_1, I_1, m)$	1.000	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, m)$	0.0314	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(I_2, I_1, m)$	0.000132	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(I_2, I_2, m)$	1.03	1.02 [1.01, 1.04]	1.03 [1.03, 1.04]
$\Lambda(S, I_1, f)$	0.000130	0.000192 [-8.54 $\cdot 10^{-5}$, 0.000469]	9.83 $\cdot 10^{-5}$ [5.03 $\cdot 10^{-5}$, 0.000146]
$\Lambda(S, I_2, f)$	0.0264	0.0237 [0.0205, 0.0269]	0.0262 [0.0252, 0.0273]
$\Lambda(I_1, I_1, f)$	1.000	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, f)$	0.0263	0.0 [0.0, 0.0]	0.0409 [-0.00177, 0.0836]
$\Lambda(I_2, I_1, f)$	0.000129	0.0 [0.0, 0.0]	0.000249 [8.02 $\cdot 10^{-6}$, 0.000489]
$\Lambda(I_2, I_2, f)$	1.03	1.02 [1.01, 1.03]	1.03 [1.02, 1.03]

Table 17: Mean field at distance one for $(n_f, n_m) = (3, 5)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years.

Mean field at distance one $(n_f, n_m) = (2, 6)$			
	Deterministic	**Simulations** mean [95% CI]	
		$N = 1.000$	$N = 10.000$
$\Lambda(S, I_1, m)$	9.31 $\cdot 10^{-5}$	0.0 [0.0, 0.0]	0.000126 [5.98 $\cdot 10^{-5}$, 0.000192]
$\Lambda(S, I_2, m)$	0.0283	0.0239 [0.0208, 0.0271]	0.0282 [0.0270, 0.0294]
$\Lambda(I_1, I_1, m)$	1.000	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, m)$	0.0283	0.0 [0.0, 0.0]	0.0714 [-0.0103, 0.153]
$\Lambda(I_2, I_1, m)$	9.22 $\cdot 10^{-5}$	0.0 [0.0, 0.0]	0.000204 [-2.48 $\cdot 10^{-5}$, 0.000434]
$\Lambda(I_2, I_2, m)$	1.03	1.03 [1.01, 1.04]	1.03 [1.02, 1.03]
$\Lambda(S, I_1, f)$	8.78 $\cdot 10^{-5}$	0.000138 [-5.23 $\cdot 10^{-5}$, 0.000328]	2.68 $\cdot 10^{-5}$ [9.19 $\cdot 10^{-7}$, 5.26 $\cdot 10^{-5}$]
$\Lambda(S, I_2, f)$	0.0172	0.0166 [0.0143, 0.0190]	0.0178 [0.0171, 0.0185]
$\Lambda(I_1, I_1, f)$	1.00	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, f)$	0.0172	0.0 [0.0, 0.0]	0.0160 [-0.00642, 0.0385]
$\Lambda(I_2, I_1, f)$	8.71 $\cdot 10^{-5}$	0.0 [0.0, 0.0]	0.000134 [-5.07 $\cdot 10^{-5}$, 0.000318]
$\Lambda(I_2, I_2, f)$	1.02	1.02 [1.01, 1.03]	1.02 [1.01, 1.02]

Table 18: Mean field at distance one for $(n_f, n_m) = (2, 6)$ with $N_0 = 1.000$ or 10.000 at $t = 2.000$ years.

Mean field at distance one $(n_f, n_m) = (1, 7)$			
	Deterministic	**Simulations** mean [95% CI]	
		$N = 1.000$	$N = 10.000$
$\Lambda(S, I_1, m)$	3.75 $\cdot 10^{-7}$	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(S, I_2, m)$	0.0121	0.00722 [0.00526, 0.00917]	0.0107 [0.0101, 0.0114]
$\Lambda(I_1, I_1, m)$	1.000	-	-
$\Lambda(I_1, I_2, m)$	0.0115	-	-
$\Lambda(I_2, I_1, m)$	3.76 $\cdot 10^{-7}$	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(I_2, I_2, m)$	1.01	1.02 [0.995, 1.05]	1.01 [1.00, 1.01]
$\Lambda(S, I_1, f)$	0	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(S, I_2, f)$	0	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(I_1, I_1, f)$	1	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]
$\Lambda(I_1, I_2, f)$	0	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(I_2, I_1, f)$	0	0.0 [0.0, 0.0]	0.0 [0.0, 0.0]
$\Lambda(I_2, I_2, f)$	1	1.0 [1.0, 1.0]	1.0 [1.0, 1.0]

Table 19: Mean field at distance one for $(n_f, n_m) = (1, 7)$ with $N_0 = 1.000$ or 10.000 at $t = 4.000$ years.

Let us now take a look at the mean field at distance one assumption in tables 19 - 16. Remember that we have defined $\Lambda(i, i_p, g)$ as the expected number of infectious partners in infection stage

$i_p \in \{I_1, I_2\}$ of a susceptible partner of an individual of gender g in infection stage i (see fig. 3). To calculate $\Lambda(i, i_p, g)$ in the simulations we will go through all the individuals of gender g in infection stage i in the population, and for each susceptible partner we will count its number of partners in infection stage i_p and average these values.

Let us look at the scenario $(n_m, n_f) = (4, 4)$. The mean field at distance one $\Lambda(i, i_p, g)$ should be equal between men and women, but in the case of $\Lambda(I_1, I_2, g)$ the values differ greatly between the two genders. Also the confidence interval of $\Lambda(I_1, I_2, m)$ does not even coincide with the deterministic results. All the other results of $\Lambda(i, i_p, g)$ for the scenario $(n_m, n_f) = (4, 4)$ are approximately equal between men and women and coincide with the deterministic results. Therefore we must wonder what differs in the case of $\Lambda(I_1, I_2, g)$.

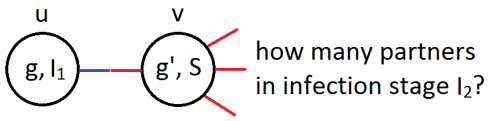

Figure 5: Illustration of the meaning of the mean field at distance one value $\Lambda(I_1, I_2, g)$.

The meaning of $\Lambda(I_1, I_2, g)$ is the expected number of partners in infection stage I_2 of a susceptible partner v of an individual u of gender g in infection stage I_1 (see illustration in fig. 5). The fraction of the population in infection stage I_1 is very small, approximately 0.000956 (see Table 15), which means approximately 1 individual in stage I_1 in case of $N = 1.000$ and approximately 10 individuals in stage I_1 in case of $N = 10.000$. So the number of individuals in infection stage I_1 with a susceptible partner will most likely be very small. Therefore the mean and confidence interval for $\Lambda(I_1, I_2, g)$ is probably calculated with use of a very small 'sample'. Even though during the calculation of the confidence interval the sample size is taken into account (a small sample size causes a large confidence interval), during the calculations a large sample size is assumed (an assumption needed for the Central Limit Theorem). This can cause unprecise confidence intervals, when working with small samples.

These small sample sizes can also be seen in the other four scenarios. Especially in the situation of $(n_m, n_f) = (1, 7)$, where $\Lambda(I_1, I_1, m)$ and $\Lambda(I_1, I_2, m)$ have no value. Apparently no male individual in infection stage I_1 with a susceptible partner has been encountered, therefore no average could be calculated and no value is given for Λ.

Besides the difference between the deterministic results and the simulation results concerning $\Lambda(I_1, I_2, g)$ for all the four scenarios, the mean field at distance one values seem to coincide quite well. In most cases the deterministic value falls within the confidence interval of the simulations.

Conclusion

We have now compared the results of the deterministic model with the results of the simulations for the baseline parameter values, in which we varied the population size N_0 and the partnership capacities (n_f, n_m), all other parameters were held constant. We looked at different aspects of the

network (degree distribution, concurrency index) and infection (endemic prevalence, mean field at distance one assumption).

The network results of both models agree very well (the deterministic values fall within the confidence intervals of the simulations), except for the very small values. But we cannot expect such precision, since we are dealing with a simulation model, even with a larger number of runs. When looking at the infection, we saw a few discrepancies in the endemic prevalence for scenario (1,7). This could be caused by the fact that endemic stability has not yet been reached at the end of each run. But, altogether, the models seem to agree very well.

Next, we have compared the values of $\Lambda(i, i_p, g)$ for both models, to check the correctness of the mean field at distance one assumption made in the deterministic model. A few inconsistencies were encountered, but these could be explained by the very small sample sizes that causes inaccurate results. Any conclusion from these results about the correctness of the mean field at distance one assumption we do not dare to make, but the results are promising.

It would be very interesting to run the simulations for larger population sizes, and see how this affects the mean field at distance one results. This is however not advisable with the code in Python, because of the large running time of the program (see appendix B.9), but could be possible if the code was rewritten in a compiled programming language (see appendix B.8).

Please note that we only compared both models for the baseline parameter values, so no conclusions can be made for other parameter sets.

5.4 Further network structure results

In this subsection we will explore the network and infection further, using the same baseline parameter values that were used for the comparison of the deterministic and simulation model. But we will no longer compare the simulation results to the deterministic model. We will now focus on the study of the network structures and exploit the fact that, compared to the deterministic model, more information about the network is available in the simulations.

Network

Let us first take a closer look at the network structure. A great advantage of the simulation model is that we know the exact network structure, or in other words we know exactly who has a partnership with whom. An important quality of a network is its connected components. A *connected component*, often simpy refered to as a component, is a subnetwork in which any two individuals are connected to each other by partnerships.

In fig. 6 the distribution of the size of connected components in the network is shown. For each run we count the frequency of each component size in the network and divide it by the total number of connected components. We calculate the mean and confidence interval over all the 100 runs, which has led to the distributions shown in fig. 6.

Since connected components of size 1 (single individuals) are most frequent, we do not show these in the plot to better illustrate the rest of the occurring sizes. The maximum occurring size is also

the maximum value on the x-axis.

The figure shows that the distribution of the size of the connected components is very similar for the scenarios $(n_f, n_m) = (2, 6), (3, 5)$ and $(4, 4)$. For the scenario $(n_f, n_m) = (1, 7)$ the maximum size of a connected component is of course 8, but the occurring maximum size in the simulation is only 6. And we can see that connected components of size 6 occur very rarely. The other scenarios have no such limitations on the maximum size, since females are allowed to have more than one partner, which allows larger connected components. Although larger connected components are indeed present for these three scenarios, the frequency with which they occur is very small.

This difference of the component size distribution between scenario (1,7) and the other three, may explain why the infection takes so much longer to reach endemic stability for (1,7) compared to the other scenarios (see fig. 4).

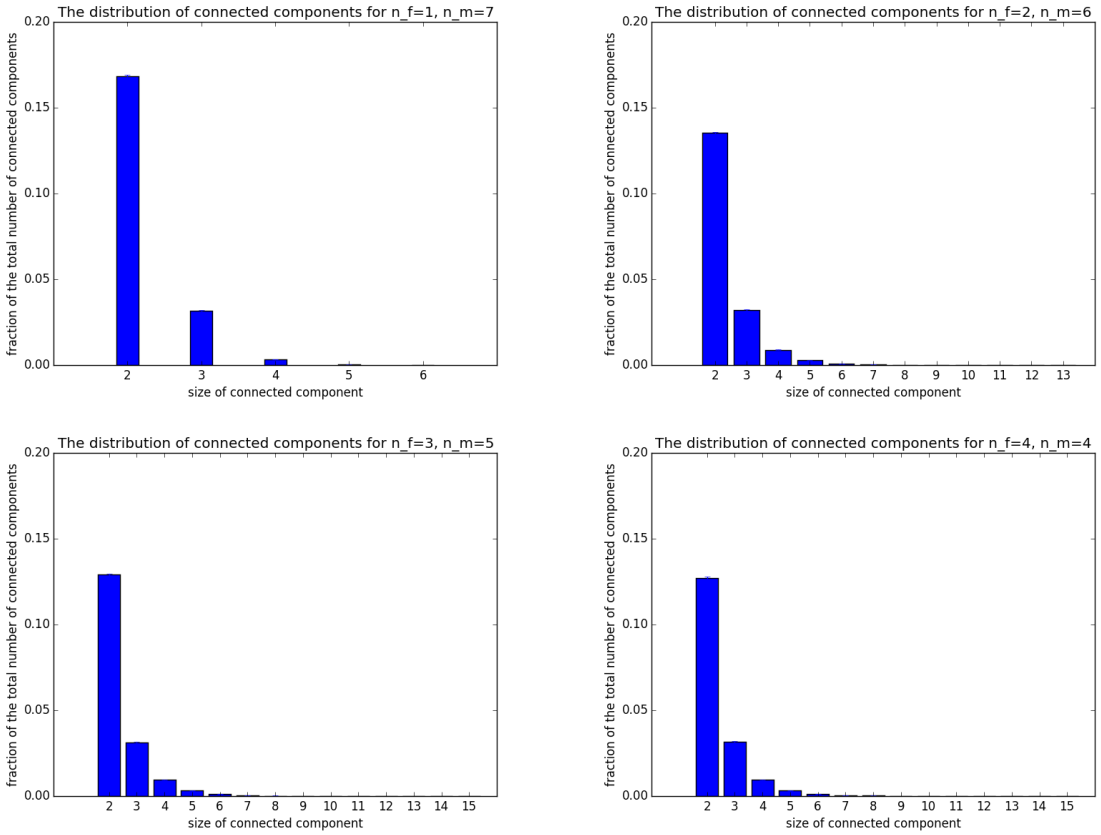

Figure 6: The connected component size distribution of the network (mean and 95% confidence interval) at the end of the simulations with $N_0 = 10.000$, for the scenarios $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$

In fig. 7 we illustrate the mean and variance of the connected component size distributions that are

shown in fig. 6. We can see that the mean size of the connected components is approximatly the same for all scenarios. Even though there do exist large components for the scenarios (2,6), (3,5) and (4,4) it seems that, because they occur so rarely, they have little effect on the mean size.
The variances of the distributions do differ, especially between the scenario $(1, 7)$ and the rest of the scenarios. This could be caused by the previously explained limitation on the maximum size of the connected components when $(n_f, n_m) = (1, 7)$, while the other scenarios have no such limitations. Since for calculating the variance, the difference between the component size and the mean component size is squared, the larger components have more influence on the variance than on the mean.

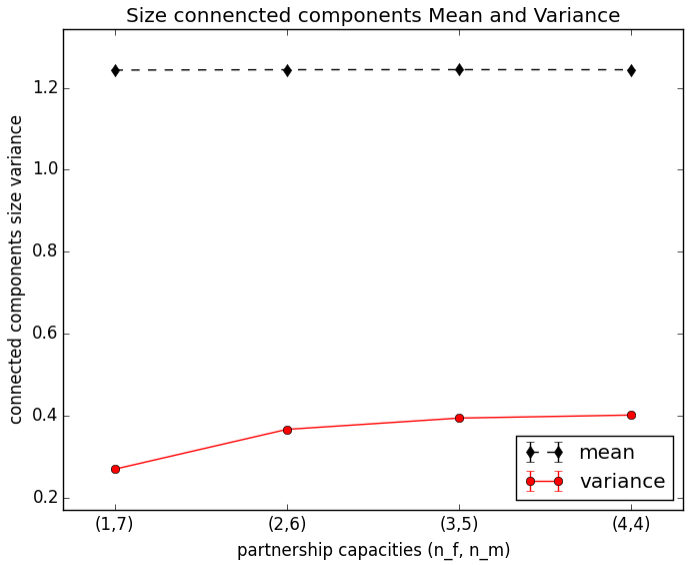

Figure 7: The mean and variance of the connected component size distribution for $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$.

The distributions of the size of the connected components (fig. 6) and its mean and variance (fig. 7) are both calculated at one point in time. However we are also interested in the change of the network structure over time. It is difficult to investigate a dynamic network, when we only take snapshots of the network at certain times. We will try to explore the network changes over time by following a population with a stable network structure for 100 years and look at the distribution of the size of the components every year. We have illustrated the frequency of the size of the connected components as a pixel plot over time in fig. 8 for each scenario. Note that we follow one run for 100 years, therefore the figure does not show means or confidence intervals.

We can see that there is again a large difference between the scenario (1,7) and the other three scenarios, because of the limitation on the maximum component size for (1,7). The other three scenarios are very similar, but we do see that the less variance in the partnership capacities (n_f, n_m), the higher the occurring 'peaks' in existing component sizes.

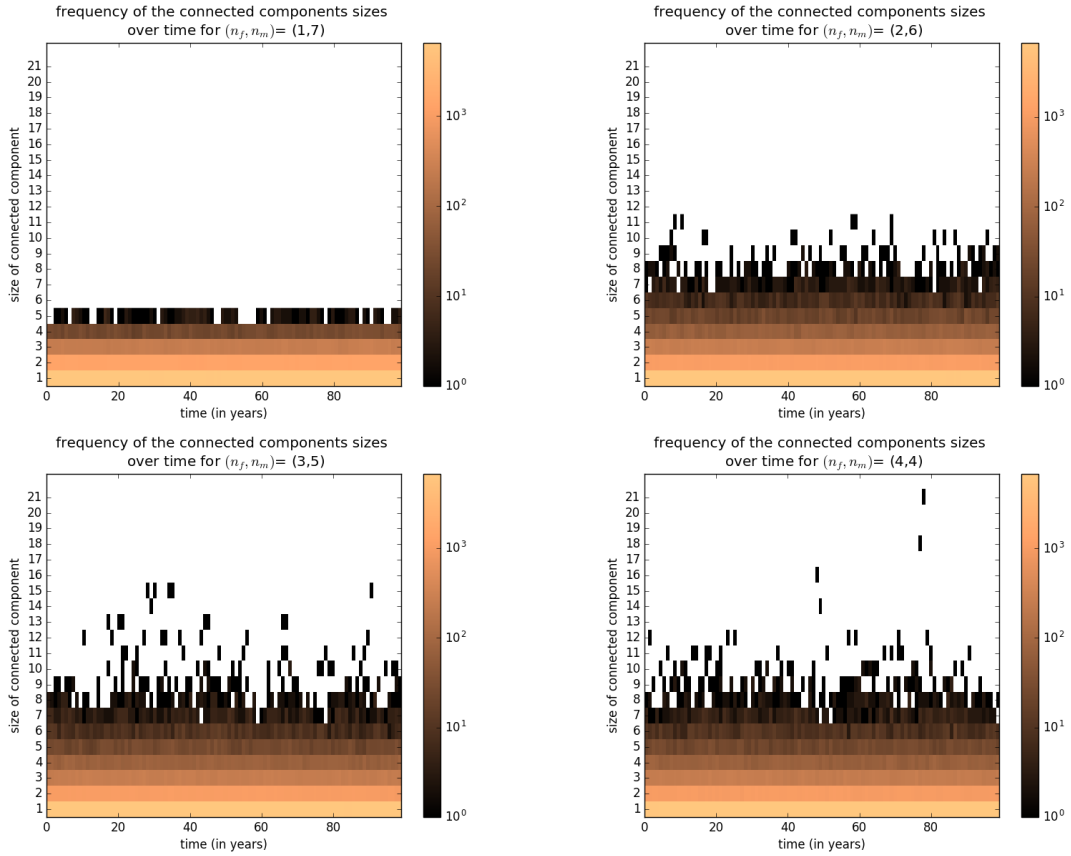

Figure 8: The frequency of the sizes of the connected components over 100 years with time steps of 1 year. With $N_0 = 10.000$, for $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$.

Infection

In Figure 6 we have seen that for the scenarios $(n_f, n_m) = (2, 6), (3, 5)$ and $(4, 4)$ large connected components exists, but not very often. But maybe, even though large connected components do not occur often, they could give the transmission of infection an extra 'boost'. Therefore we would like to investigate if the probability that an individual gets infected is higher when it resides in a large component.

We cannot simply compare the fraction of transmission that occur in a certain component size with the fraction of individuals that reside in a component of that size. Because we also need to take into account that the probability of transmission strongly depends on the individuals' number of partners. And the component size in which the individual resides has some influence on its number of partners. For example if an individual resides in a component of size 2 we already know that it only has one partner. But if an individual resides in a large component size, it could have a large number of partners, making the probablity of getting infected higher. Therefore we will compare

the fraction of transmissions to an individual with k partners in a component of size s with the fraction individuals with k partners residing in a component of size s at the moment of transmission.

We follow a population in endemic state for 100 years. Everytime transmission occurs, we save the entire network structure. Thus we do not only have the information in which component size the transmission has occured, but also the newly infected individuals' number of partners and the component size distribution. This gives us all the information needed to construct the wanted figures. We will only illustrate the results in the figure if transmission (to an individual of a certain number of partners in a component of certain size) has occured more than 10 times during the 100 years of following the population.

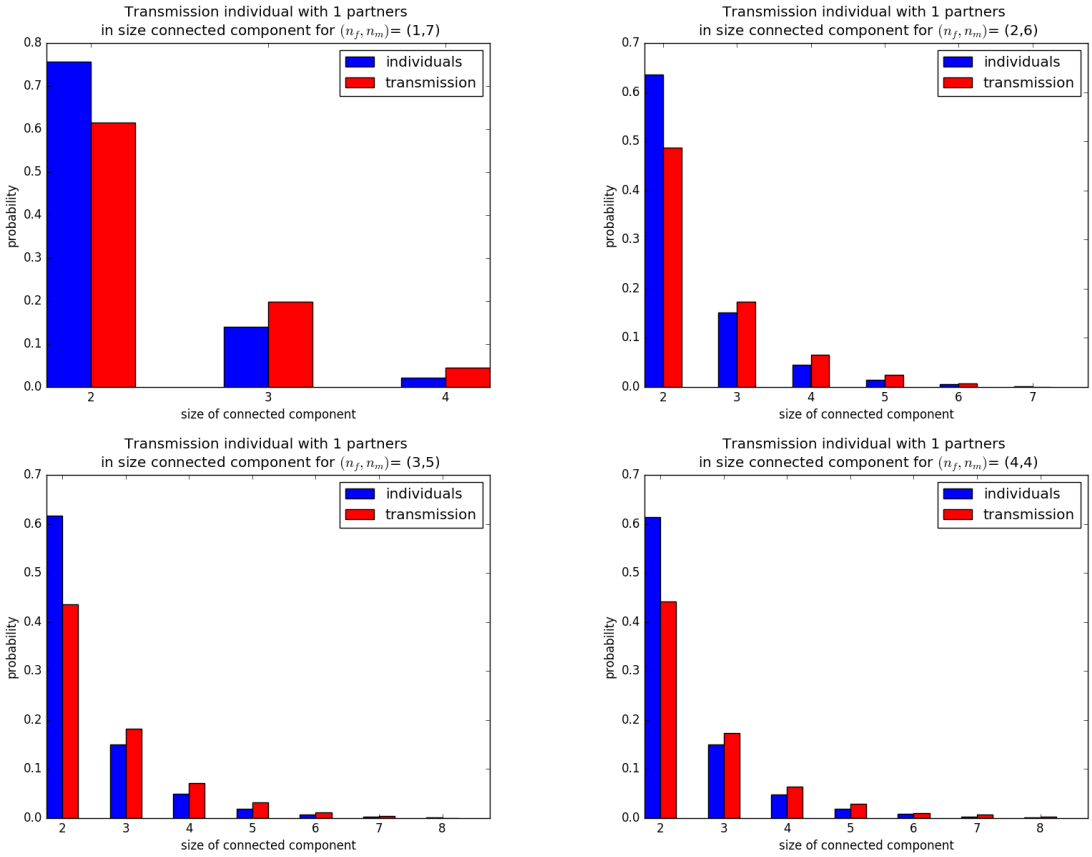

Figure 9: The fraction of transmissions to an individual with one partner in a component of size s vs the fraction of individuals with one partner residing in a component of size s. We look at a population size $N_0 = 10.000$, for the scenarios $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$.

In this paper we only show the fraction of transmissions in component size s versus the probabillity of an individual residing in a component of size s, given that the (newly infected) individual has only **one** partner, which is illustrated in fig. 9. The results for multiple partners are not shown, but the same conclusion could be drawn from those results. We see for all scenarios that the probability

of transmission in a component of size two is smaller than the probablity of an individual residing there. For component size three or more, the probability of transmission is larger. But in fig. 9 it is hard to see if it is true that the larger the connected component, the higher the probablility for an individual to get infected. Therefore we want to take a closer look at the ratio between the two probabilities:

$$\text{ratio} = \frac{\text{fraction transmissions to individual with k partners in component of size s}}{\text{fraction of individuals with k partners residing in component of size s}}$$

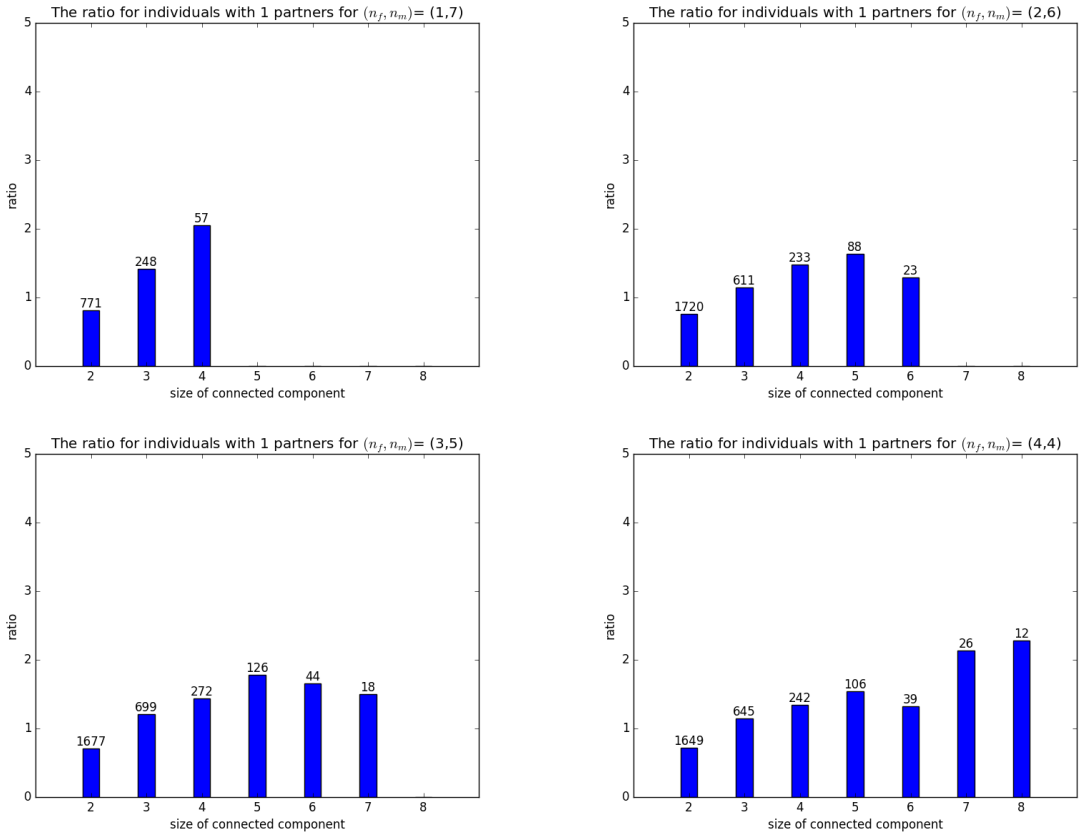

Figure 10: The ratio: (the fraction of transmissions that occur in a component of size s, given that the newly infected has 1 partner) : (the fraction of individuals with 1 partner residing in a component of size s). We follow a population of size $N_0 = 10.000$ for 100 years, for the scenarios $(n_f, n_m) = (1,7), (2,6), (3,5)$ and $(4,4)$. The numbers on the bars indicate the number of times transmission has occured in the 100 years, to an individual with 1 partner in a certain component size.

We have illustrated the ratios in fig. 10. We see that for all scenarios the ratios indeed do increase in value, when the connected component increases in size. Telling us that the probablity for an

individual to get infected is more likely in a larger component. But we also see that with component size 6, the ratio seems to decrease again for all scenarios. These irregularities after component size 6 may be caused by the fact that these larger components are not stable in time, causing the calculations of the ratio to be based on little data.

As we have said before, we only illustrate the results if transmission has occured more than 10 times. But this is still very little data to draw conclusions from. Since the frequency of occurence has a large impact on the reliability of the ratio, we show on every ratio bar the frequency of the transmission that has occured in that component size.

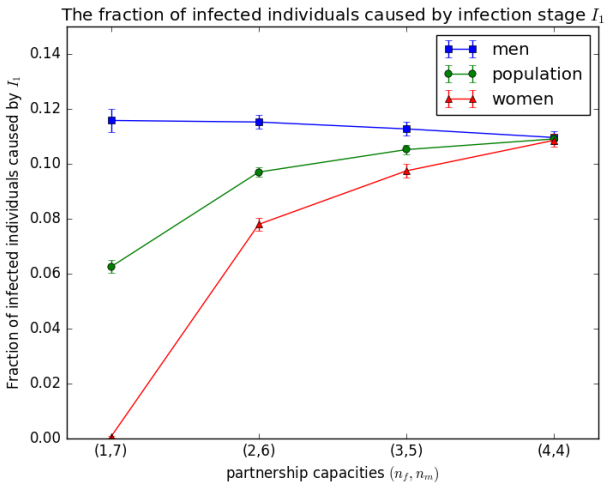

Figure 11: The fraction of the total number of infected individual that are infected by an individual in infection stage I_1.

Next we will look at the contribution of the infection stage I_1 to the transmission of the infection, illustrated in fig. 11. We have plotted the fraction of infected individuals that are infected by an individual of gender g in infection stage I_1, out of the total pool of infections caused by an individual of gender g.

We see that decreasing the variance of the partnership capacities, causes an increase in the contribution of the infections caused by I_1. The acute phase I_1 is a relative short period (2.9/12 years), but during this period the individual is very infectious. Therefore it advantegeous for the transmission of infection if there are larger components in the network, which allows the infection during the acute phase to spread quickly through the component. And as we have seen in fig. 6 the scenarios with less variance between the partnership capacities, allow larger components in the network.

We also see that changing the partnership capacities has very little effect on the contribution of I_1 of men. We can see in the tables 9 - 12 that the degree distribution of men changes a lot less compared to the degree distribution of women, when altering the partnership capacities. Even though the difference in the number of binding sites for the scenarios (1,7) and (4,4) is the same

for men ($7 - 4 = 3$) and women ($4 - 1 = 3$). But since there are approximately the same number of men as there are women in the population, for the scenario (1,7) there are simply not enough female binding sites for the number of male binding sites. Therefore the binding sites of the men do not reach their full 'potential' in scenario (1,7), while the female binding sites do not have that limitation since there are many more male binding sites. All in all, changing the partnership capacities has more effect on women than on men.

6 Sensitivity analysis

In section 5 we have only looked at the baseline parameter values given in table 7. All input parameters were held constant, except for (n_f, n_m) and N_0 for which a few different scenario's were considered. The used parameter values are estimates, and are bound to be uncertain. Therefore we will perform a so called *Sensitivity Analysis*, the study of how uncertainty in the output of a model can be attributed to different sources of uncertainty in the model input [16]. In this section we are going to investigate the influence of variation around these baseline values on the model results.

In section 6.1 we will use the so called univariate *one-at-a-time* (OAT) approach. OAT analysis looks at the impact of a parameter on the model output one parameter at a time, keeping the others constant. OAT is a simple approach and does not explore the interaction between input parameters.

6.1 One-at-a-time sensitivity analysis

In this section we will investigate the influence of some of the input parameters on the model output with the *one-at-a-time* (OAT) approach. We will vary one parameter at a time, while we keep the rest of the parameters constant to the previously used values in table 7. We will not perform sensitivity analysis for all the model parameters, but we will look at the following three parameters: the mean partnership duration d_P, the mean lifetime number of partners θ and the mean duration of the acute phase d_{I_1}. By varying the parameters d_P and θ we change network parameters and investigate how this influences not only the network structure but the spread of the infection. By changing d_{I_1} we change characteristics of the infection itself and will only investigate how this influences the spread of the infection.

We will perform the sensitivity analysis for a population of size $N_0 = 10.000$ and for the scenarios $(n_f, n_m) = (1, 7)$ and $(4, 4)$. We have to keep in mind that not all combinations of parameter values are allowed. Taking the restrictions 17, 19 and 23 (see section 4.4) into account, we will consider the allowed intervals for the OAT approach shown in table 20. These intervals are chosen to surround the baseline parameter values. Note that because of the restrictions, we consider two different intervals of d_P for the scenarios (1,7) and (4,4).

parameter	**description**	**constant**	**interval OAT**
d_P	the mean duration of a partnership	3.92	[1,10], if $(n_f, n_m) = (1, 7)$ [1,17.5], if $(n_f, n_m) = (4, 4)$
θ	the mean lifetime number of partners	3.5	[1.5, 5.5]
d_{I_1}	the mean duration of the infection stage I_1	2.9/12	[1/12, 6/12]

Table 20: The input-parameters and their values that will be used for one-at-a-time (OAT) sensitivity analysis.

Out of the OAT-interval 200 random samples will be chosen for each scenario. Unlike in section 5.3, we will not perform 100 runs for each parameter set, but we will perform one run per parameter set. Despite performing only one run for each chosen sample, the relations between the varying parameter and the results are clear as we will see in sections 6.1.1 - 6.1.3.

We will investigate what effect varying each of these parameters will have on the following model output:

- The network
  - The degree distribution
  - The concurrency index
  - The distribution of the size of the connected components
- The infection
  - The endemic prevalence
  - The fraction of transmissions caused by individuals in I_1

For each of these results we will plot the output of the 200 samples against the varying parameter. We will fit these data points to a mathematical function, with the use of non-linear least squares. Some of these fitted mathematical function could be compared to the functions in the deterministic model. Altough we will not compare them in this paper, it would be interesting to keep in mind for future studies.

To measure the sensitivity of the output to the input parameters, we calculate the *Sensitivity Index (SI)*. This sensitivity measurement *SI* analysis the local sensitivity in the sense that it measures the sensitivity in one point, in our case, the baseline values that we have used in all previous results (see table 20). Let $x_1, \ldots, x_p$ be the input parameters. We calculate the sensitivity index for a small change Δx_i from the constant value x_i for the ith parameter. The rest of the input parameters will be held constant. Then *SI* can be calculated as

$$SI = \frac{\frac{M(x_1,\ldots,x_i+\Delta x_i,\ldots,x_p)-M(x_1,\ldots,x_i,\ldots,x_p)}{M(x_1,\ldots,x_i,\ldots,x_p)}}{\frac{\Delta x_i}{x_i}} \tag{27}$$

where M is the model output variable [17], for example the endemic level. We will calculate the model output M with the use of the fitted function. As you can see, for small Δx_i, eq. (27) approximates the first-order partial derivative of M with respect to the ith parameter, multiplied by the factor $\frac{x_i}{M(x_1,\ldots,x_i,\ldots,x_p)}$. This factor causes the scaling of the effects: *SI* represents the *relative* importance of the ith parameter on the model output M.

To summarize, we: vary one parameter, calculate the model output for the different parameter values, fit the model output to a function and calculate the sensitivity index in the baseline value with the use of the fitted function.

6.1.1 d_P: the mean partnership duration

We will start with the OAT sensitivity analysis on the parameter d_P: the mean duration of a partnership. Let us investigate what kind of influence this parameter has on the network structure and the spread of the infection.

Network - degree distribution

In fig. 12 we have illustrated the relation between the degree distribution and the mean partnership duration. Not only have we plotted the mean and the variance of the degree distribution of the runs, we have also plotted the functions fitted to this data. We will use these fitted functions to calculate the *sensitivity index* from eq. (27) for the baseline value $d_P = 3.92$. An overview of all the calculated SI for $d_P = 3.92$ will later be shown in table 21.

Please remember that for (1,7) we look at $d_P \in [1, 10]$ and for (4,4) we look at $d_P \in [1, 17.5]$, therefore the x-axis in fig. 12 are of different scale.

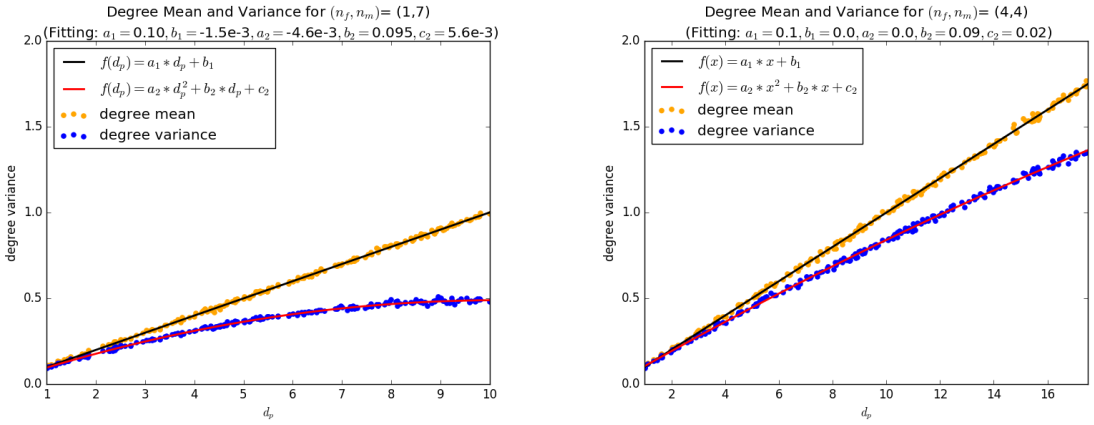

Figure 12: The mean and variance of the degree distribution dependent on the mean partnership duration d_P for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

We see a linear relation between the mean of the degree distribution and the mean partnership duration. This linear relation seems to be very similar for both the scenarios (1,7) and (4,4). However the relation between the variance of the degree distribution and the mean partnership duration differs between the two scenarios. The variance increases slower in the case of (1,7) than (4,4). This causes a lower variance of the degree distribution for (1,7) compared to (4,4), especially for the higher values of d_P.

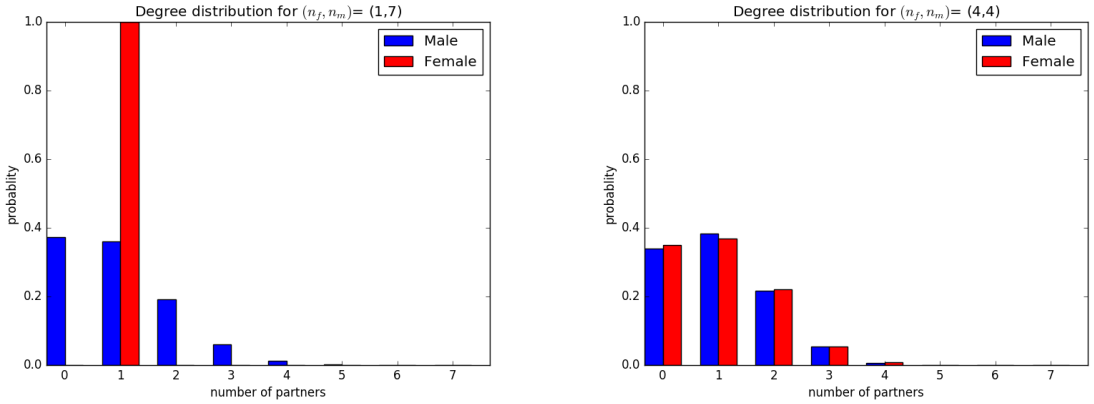

Figure 13: The degree distribution for $d_P \approx 10$ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Let us look closer to the degree distributions of both scenarios. In fig. 13 the degree distributions are plotted for both scenarios, of the run with the d_P value nearest 10. We see that in case of (1,7) all females have one partner. Since the mean is very close to one, the variance of the females (approximately half the population) is very small. When plotting the degree distribution for other values of d_P for scenario (1,7), we see the same situation. Although for lower values of d_P there are more single females, the degree variance of females is much lower than in the male population. This causes the much smaller variance for scenario (1,7).

When we compare the degree distributions for males for scenarios (1,7) and (4,4) in fig. 13, we see little difference. Even though males in scenario (1,7) are allowed to have seven partners at the same time, we see that such high number of partners is rarely reached. We have seen this phenomenon before in section 5.4. It was caused by the fact that there are simply not enough female binding sites compared to the number of male binding sites. Increasing the partnership duration has no influence on this phenomenon.

Network - concurrency index

Next, we will discuss the relation between the partnership-based concurrency index (explained in section 5.3) and the mean partnership duration, illustrated in fig. 14. There seems to be a linear relation between the mean partnership duration d_P and the concurrency index κ_P for both scenario's. However, the slope for (4,4) is higher than in the case of (1,7). In scenario (1,7) all females will have a concurrency index equal to zero, regardless of the mean partnership duration. The males are allowed to have a maximum of seven partnerships at the same time, and therefore could have a very high concurrency index. But as we have seen in fig. 13 even for a high mean partnership duration, males rarely have such a high number of partners. Unlike in scenario (1,7), where only males contribute to a higher concurrency index, in scenario (4,4) both males and females contribute to the concurrency index. This causes the higher slope for (4,4).

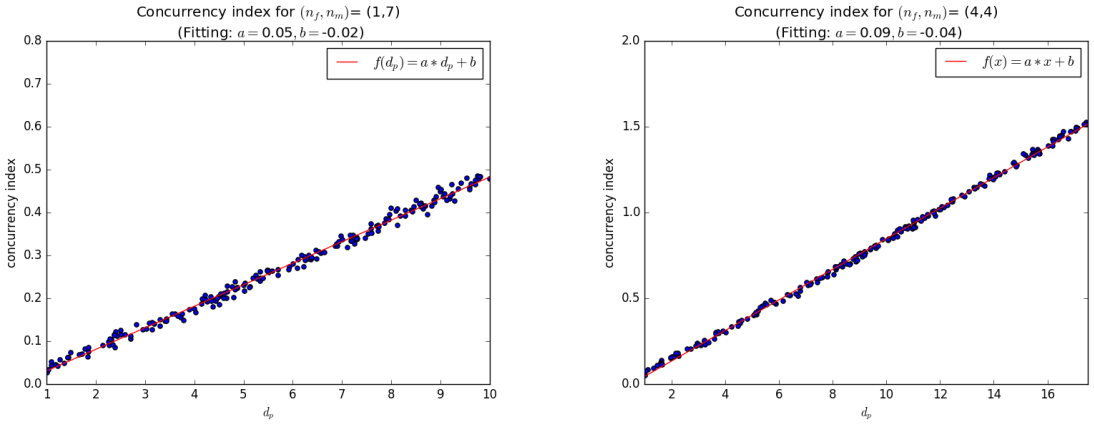

Figure 14: The concurrency index dependent on the mean duration of a partnership d_P for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Network - connected components

In fig. 15 the mean and standard deviation of the distribution of the random variable Q is illustrated, in which Q represents the size of the connected components in which the individual resides at the end of each run. We calculate for that moment in time $P(Q = q)$, the fraction of individuals that resides in a connected component of size q, for all occuring q. Note that we do not calculate the probability that a connected component of size q occurs, but the probability that an individual resides in a connected component of size q. Therefore a connected component of size q 'counts' q times.

We have chosen to illustrate the standard deviation instead of the variance, because the variance reaches very high values. Chosing to plot the standard deviation gives us the advantage that the mean and standard deviation are of the same order of magnitude and therefore can be illustrated very well in one figure.

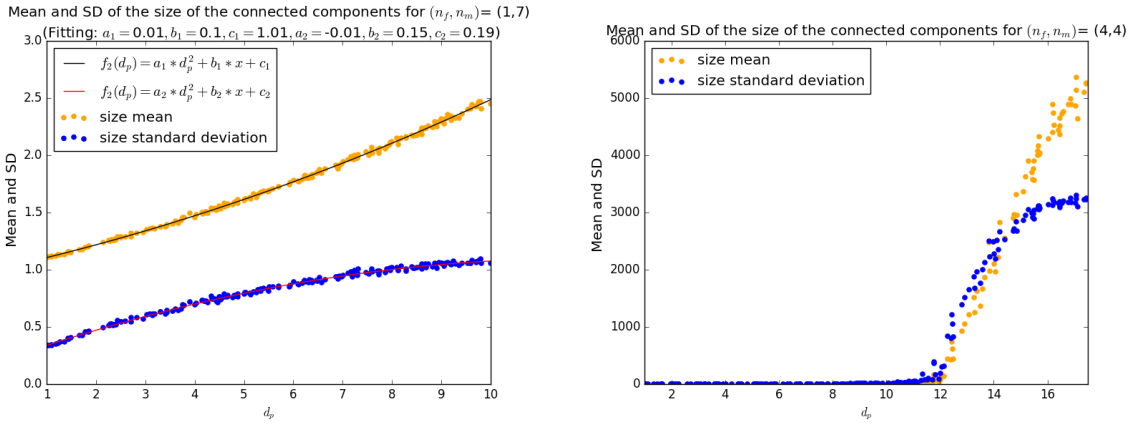

Figure 15: The mean and standard deviation of the distribution of the size of the connected components, dependent on the mean partnership duration d_P for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right). Note the difference in x-scale (for (1,7) we look at the interval $d_P \in [1, 10]$ and for (4,4) we look at the interval $d_P \in [1, 17.5]$) and y-scale!

Note that in fig. 15 both the x and y axis have very different values for the scenario's $(1, 7)$ (left) and $(4, 4)$ (right). We were able to fit the results of scenario $(1, 7)$ to quadratic functions, but were unable to find a well fitting function for scenario $(4, 4)$ due to its strange structure. We can see that for scenario $(4, 4)$ there seems to be a threshold $d_P \approx 12$ after which there is a sudden increase in both the mean and the standard deviation. The mean increases to values around 5000 for a population of approximate 10000 individuals. This means one very large component is formed in which many individuals of the population reside. We speculate that the sudden increase around $d_P \approx 12$ is caused by the appearence of only one very large component. We will later check this statement in fig. 17, where we investigate the relation between the largest occuring component and the mean partnership duration.

Since for scenario $(4, 4)$ the mean and the standard deviation reach such high values for $d_P > 12$ it is difficult to see the graph for the lower values of d_P. The difference in axis-values makes it difficult to compare the two scenario's from fig. 15. Therefore we zoom in for the values $d_P \in [1, 6]$, such that the axis can have similar ranges for both scenario's. The resulting graphs are shown in fig. 16. Previously, for the scenario $(4, 4)$, we were unable to fit the results to functions for the entire inteval $d_P \in [1, 17.5]$. But now we are able to fit the results for just the interval $d_P \in [1, 6]$. We will use the fitted functions shown in fig. 16 to calculate the sensitivity index for $d_P = 3.92$.

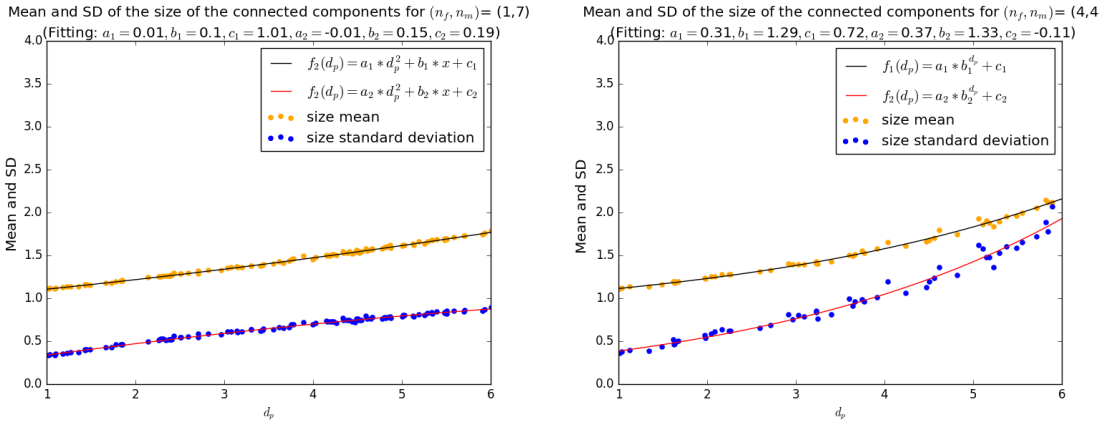

Figure 16: The mean and standard deviation of the distribution of the size of the connected components, dependent on the mean partnership duration d_P for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right) for $d_P \in [1, 6]$.

We can see in fig. 16 that the standard deviation increases much faster for scenario $(4, 4)$ than $(1, 7)$. This can again be explained by the fact that there is a limit on the maximum size of the connected components that can occur for scenario $(1, 7)$, and no such limit exists for scenario $(4, 4)$, giving the possibility of a larger variance between the sizes of the components.

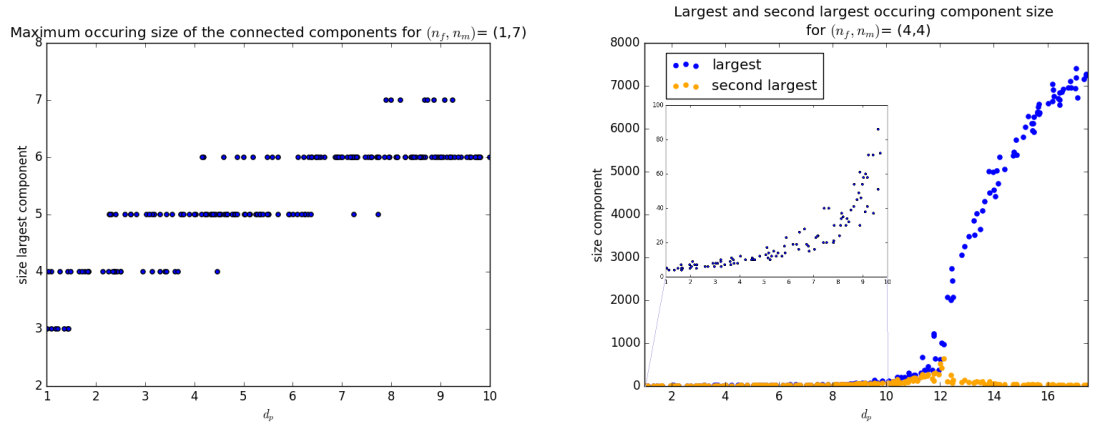

Figure 17: Size of largest connected component dependent on the mean partnership duration $d_P \in [0, 10]$ for $(n_f, n_m) = (1, 7)$ (left). The largest and second largest component size dependent on the mean partnership duration $d_P \in [0, 17.5]$ for $(n_f, n_m) = (4, 4)$ (right). For scenario $(4, 4)$ a zoom is included on $d_P \in [0, 10]$.

We have seen in fig. 15 that in case of $(4, 4)$ the mean component size grows very large from $d_P \approx 12$, we speculated that this is caused by the appearance of one very large component. Since for $d_P \approx 17$

the mean is approximately 5000 in a population of around 10000 individuals, this must indeed be the case. But for smaller values of d_P this is not necessarily true. Therefore for scenario $(4, 4)$ we plotted in fig. 17 not only the largest, but also of the second largest component size, to check whether only one very large component is formed. We can see from the results that this is indeed true. There is a very large gap between the largest and second largest connected component for $d_P > 12$. However for $d_P \approx 12$ there is a peak of the second largest component size. There seem to be two large components coexisting in the network, maybe even more. But when d_P grows larger and the largest component grows in size, the size of the second largest component decreases. This is an interesting phenomenon. Therefore we will investigate the network structures in more detail.

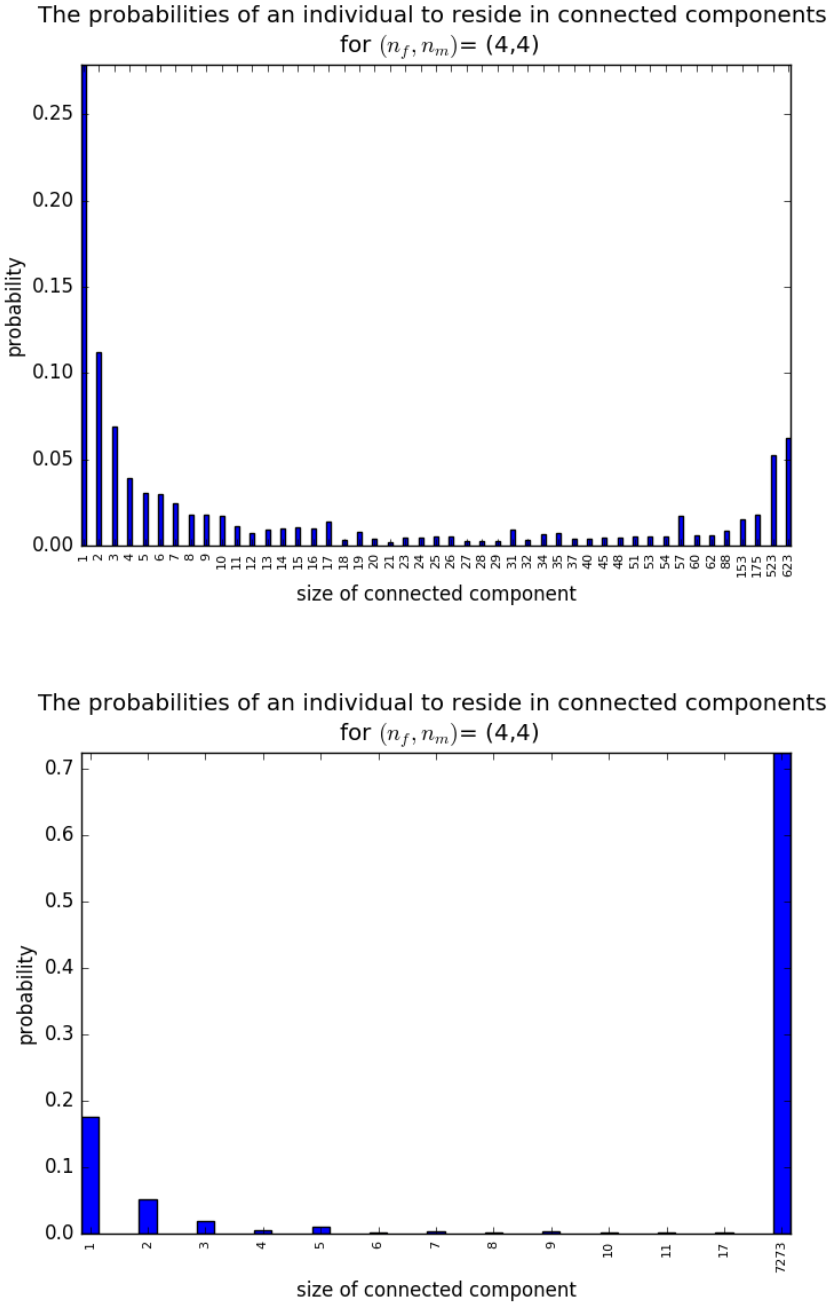

Figure 18: The frequency distribution that an individual resides in a connected component of certain size. For the scenario $(n_f, n_m) = (4, 4)$ for $d_P \approx 12$ (left) and $d_P \approx 17.5$ (right).

We will take a closer look at the propertions of individuals residing in a component of certain size. In fig. 18 we have illustrated these results for scenario $(4, 4)$ from the two runs with the values d_P nearest 12 and 17.5. We see for $d_P \approx 17.5$ there is one very large component (of size 7273) and a very large gap to the second largest component (of size 17). We see that 73% of the population resides in the largest component and that 18% of the population is single. This leaves a very small percentage of individuals in components of other sizes.

In the network for $d_P \approx 12$ there is not one large component, but a lot of 'medium' sized components. Unlike in the network of $d_P \approx 12$, there is no sudden large gap between occuring sizes.

Infection - Endemic level

In fig. 19 we have illustrated the relationship between the endemic level, when the infection has reached stability, and the mean partnership duration. For small values of d_P the infection dies out. Note that the baseline value $d_P = 3.92$ lies very close to the transition of whether or not the infection survives. Therefore, according to the results of the simulation model, a small decrease of the baseline value of d_P can cause the infection to die out.

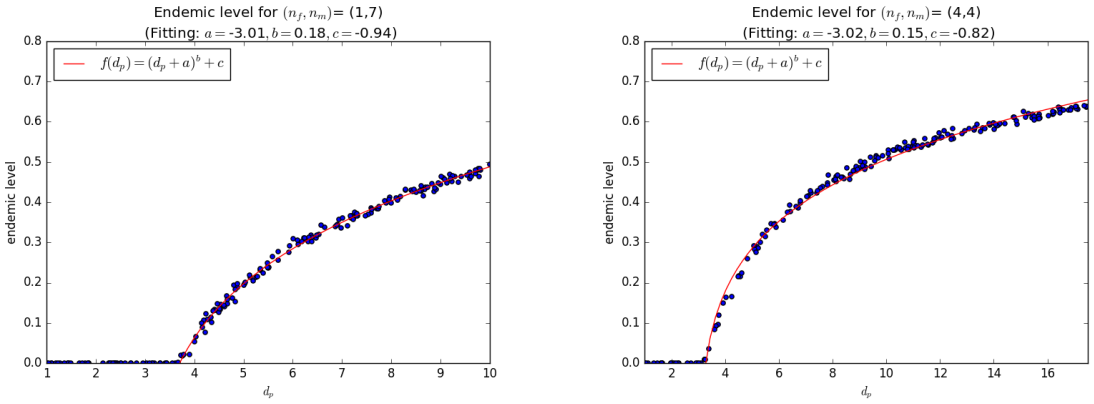

Figure 19: The endemic level dependent on the mean duration of a partnership d_P for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

For the scenario (4,4) the infection can establish itself for smaller values of d_P ($d_P \approx 3.3$) than for scenario (1,7) ($d_P \approx 3.7$). Another observation is that for all values $d_P \in [3.3, 10]$ the endemic level is higher for the scenario (4,4) than for the scenario (1,7). However the difference between the endemic levels of the two scenarios in that interval decreases, when the mean partnership duration increases.

Infection - Contribution of infection stage I_1

In fig. 20 we have illustrated the relation between the contribution of the infection stage I_1 to the transmissions and the mean partnerhsip duration. With the contribution of the infection stage I_1 we mean the fraction of infected individuals, that were infected by an individual which was in infection stage I_1 when the transmission occured, out of the total pool of infected individuals. This contribution is calculated at the end of each run, when the infection has reached endemic stability.

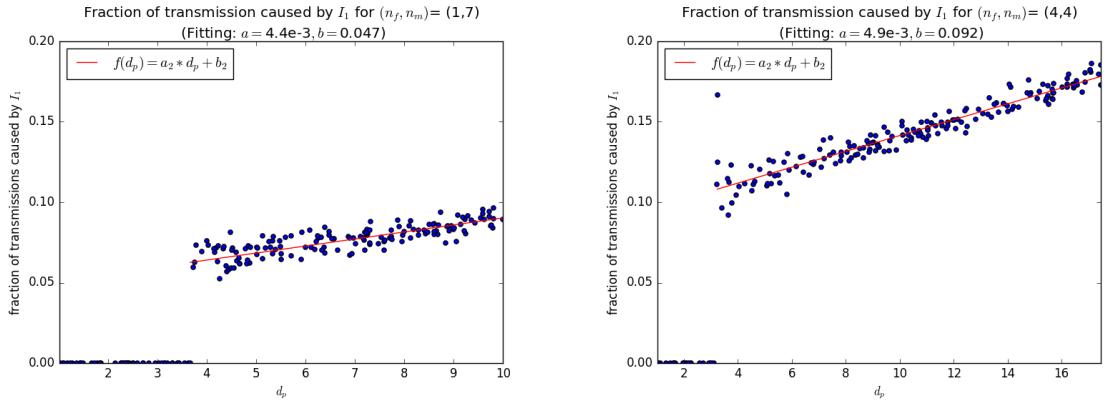

Figure 20: The fraction of transmissions caused by an individual in infection stage I_1 dependent on the mean duration of a partnership d_P for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

When the infection survives, the relation between the contribution of I_1 and the mean partnership duration can be best described as linear. The slope of the linear relation is very similar for both scenarios, but the values are higher for scenario (4,4) than (1,7).

The number of individuals that a newly infected indvididual in infection stage I_1 will infect will be influenced by its current number of partners. However we can also think of an individual acquiring a new partner and infecting this new partner, while being in infection stage I_1. But since the infection survives only when the mean of the exponentially distributed partnership duration is more than 3 years and the acute phase I_1 has a mean of only 2.9/12 years, the probability of this occuring is negligible. Therefore we take only its current partners into account. In fig. 14 we have calculated a measure of concurrency and have seen that the concurrency index is higher for scenario (4,4) than (1,7), which may have caused the higher values of the contribution of I_1 for scenario (4,4) compared to the values of (1,7).

Sensitivity indices
In table 21 the calculated senstivity indices (SIs) for several model results for $d_P = 3.92$ are shown. We see that for both scenarios the mean partnership duration has the most effect on the endemic level. The SIs for the endemic level are very high compared to the other indices, especially for scenario (1,7). In fig. 19 the fitted functions increase indeed very fast around 3.92. Since the endemic levels for (1,7) are so small and the SI is a *relative* measure, this explains why the sensitivity index for the endemic level has such a high value for (1,7).

This high SI for the endemic level tells us that a small error in the estimation of the baseline value $d_P = 3.92$ can cause great changes in the calculated endemic level.

Sensitivity Index for $d_P = 3.92$		
	Scenarios	
	$(1,7)$	$(4,4)$
mean degree	1.0039	1.0002
variance degree	0.7480	0.9256
concurrency index	1.1098	1.1460
mean size components	0.3671	0.5464
SD size components	0.5805	1.2290
endemic level	**16.1796**	**3.7627**
contribution I_1	0.2690	0.1733

Table 21: The sensitivity index values for the different model results for $d_P = 3.92$.

Changing the mean partnership duration around $d_P = 3.92$ seems to have the most effect on scenario $(4,4)$ concerning the network, and the most effect on scenario $(1,7)$ concerning the infection. Scenario $(4,4)$ simply allows the most changes to the network, having none of the limitations of $(1,7)$. Why varying the mean partnership duration has such great effect on the infection for scenario $(1,7)$, could be explained by the fact that the basic reproduction number R_0 is close to 1 (near a bifurcation point).

6.1.2 θ: the mean lifetime number of partners

In this subsection we will perform the OAT sensitivity analysis on the parameter θ: the mean lifetime number of partners. Let us investigate what kind of influence this parameter has on the network structure and the spread of the infection. We will vary the parameter in the interval $\theta \in [1.5, 5.5]$.

Network - degree distribution
In fig. 21 we have plotted the relation between the degree distribution and the mean lifetime number of partners. We see, just like the senstivity analysis of d_P in fig. 12, that there exists a linear relation for the mean of the degree distribution and a quadratic relation for the variance of the degree distribution.

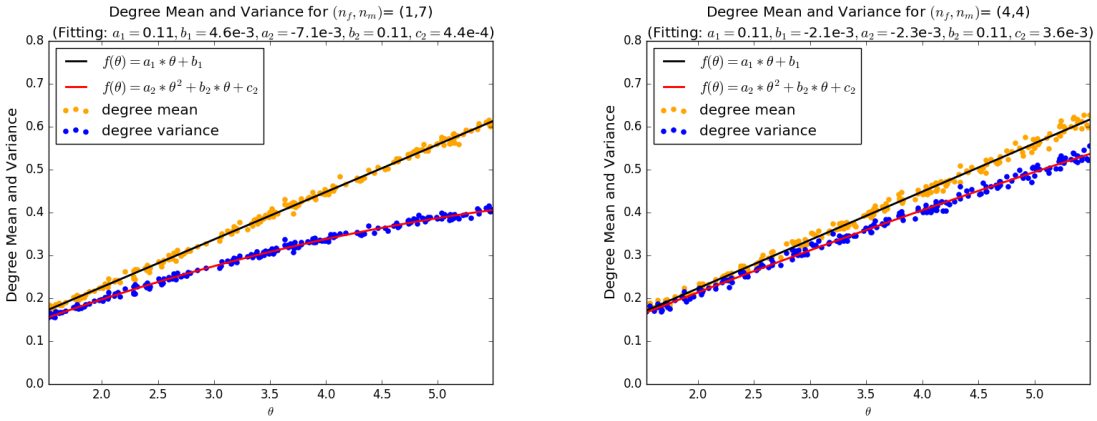

Figure 21: The mean and variance of the degree distribution dependent on the mean lifetime number of partners θ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

The linear relation for the mean of the degree distribution are again very similar for both scenarios and the variance for (4,4) grows much higher than the variance for (1,7). Let us take a closer look at the degree distribution of the run closest to the lowest value $\theta = 1.5$ (fig. 22) and closest to the highest value $\theta = 5.5$ (fig. 23). We see that the degree distribution for a very low mean lifetime number of partners $\theta = 1.5$, is very similar for both scenarios. The network consists mostly of single individuals, some have one partner, but individuals having two partners or more are almost non-existent. Therefore the difference in partnership capacities has little influence on the degree distribution for such low value of θ.

But when looking at a high mean lifetime number of partners $\theta = 5.5$, we see that it causes very different degree distributions for the two scenarios. The same effect was seen and explained in the senstivity analysis of d_P (section 6.1.1).

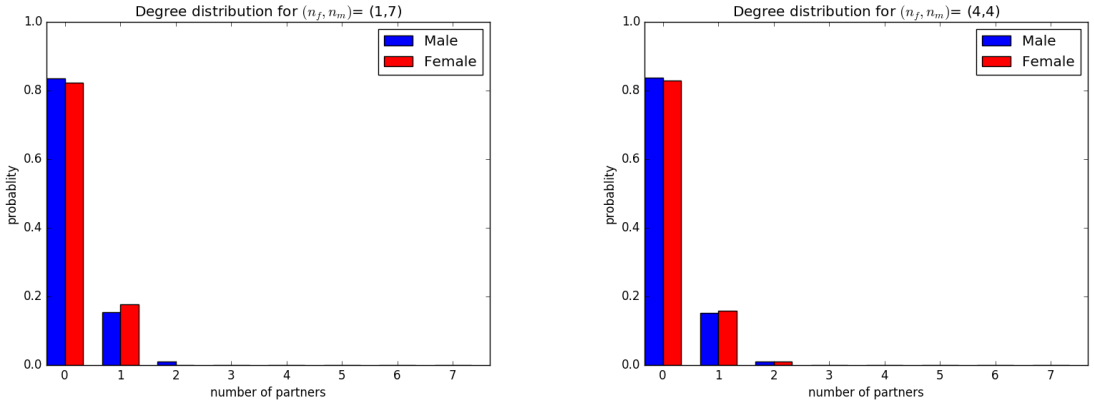

Figure 22: The degree distribution for $\theta \approx 1.5$ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

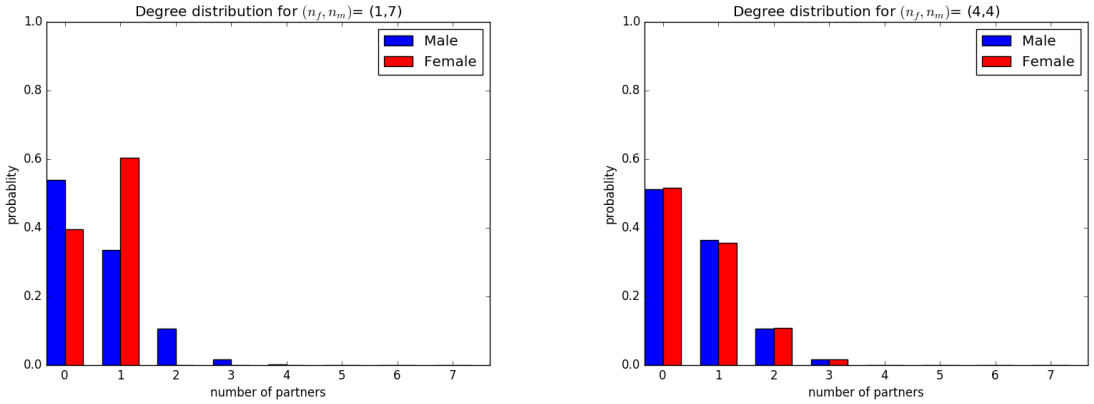

Figure 23: The degree distribution for $\theta \approx 5.5$ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Network - concurrency index

In fig. 24 we have plotted the relation between the concurrency index and the mean lifetime number of partners. We can see that the fitted functions for both scenarios are almost the same as the fitted functions in the senstivity analysis of d_P in fig. 14. It seems that whether you vary the network parameter d_P or θ, the changes to the network are very similar.

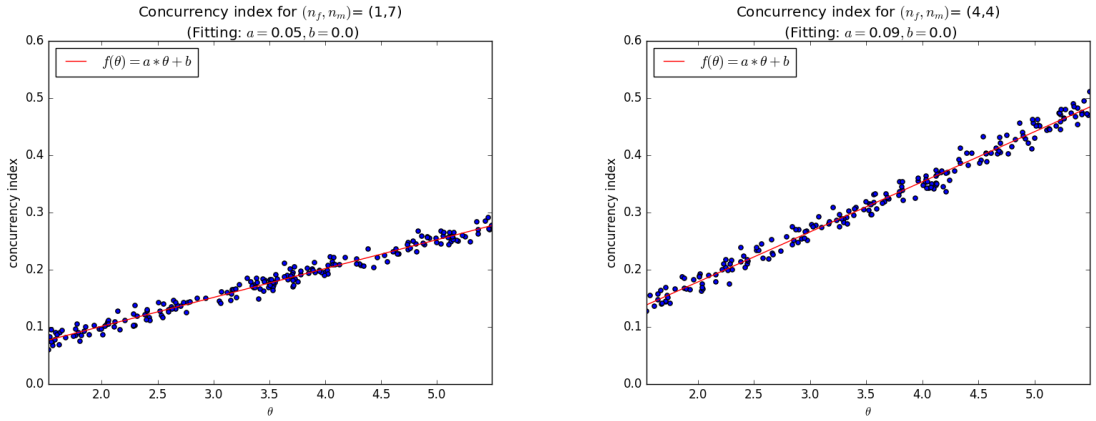

Figure 24: The concurrency index dependent on the mean lifetime number of partners θ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Network - connected components

In fig. 25 we have illustrated the relation between distribution of the size of the connected components in which the individuals reside and the mean lifetime number of partners. Again, the functions are very similar to the functions in the senstivity analysis of d_P in fig. 16.

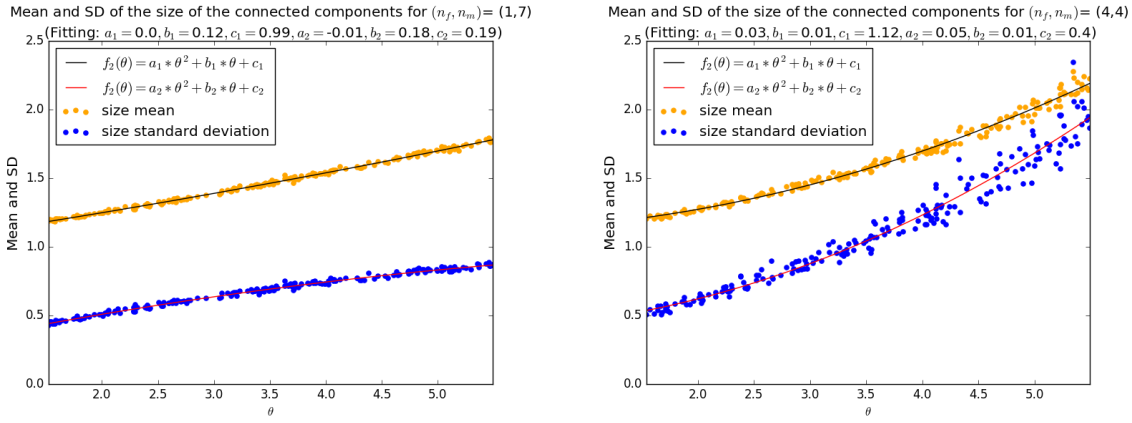

Figure 25: The mean and standard deviation of the distribution of the size of the connected components, dependent on the mean lifetime number of partners θ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

All in all, it seems that changing mean partnership duration d_P or the mean lifetime number of partners θ have very similar results on the network. This can also later be seen in table 22, where we will see that the sentivity indices are also very similar.

For the mean lifetime numer of partners we have looked at a smaller interval ($\theta \in [1.5, 5.5]$) than the

mean partnership duration ($d_P \in [1, 17.5]$). This larger interval for d_P allowed us to see something very interesting: a certain threshold after which the network allows the existence of a very large component. We have not seen such an occurence in the senstivity analysis of θ. Therefore it would be very interesting for future research to check if increasing θ further would also allow us to observe such a threshold.

Infection - endemic level

Let us now look whether varying the lifetime number of partners also has similar effects on the endemic prevalence, compared to varying the mean partnership duration. The infection can establish itself in the population in scenario (1,7) for $\theta \geq 3.3$ and in scenario (4,4) for $\theta \geq 3.0$. In both the senstivity analysis of d_P and θ, it is the scenario (1,7) that needs the highest value for the parameter in question, to cause the infection to survive.

The baseline value of θ is, just like the baseline value of d_P, very close to the threshold of the survival of the infection, especially for the scenario (1,7). Only a small decrease in the mean lifetime number of partners will cause the infection to die out in the model population.

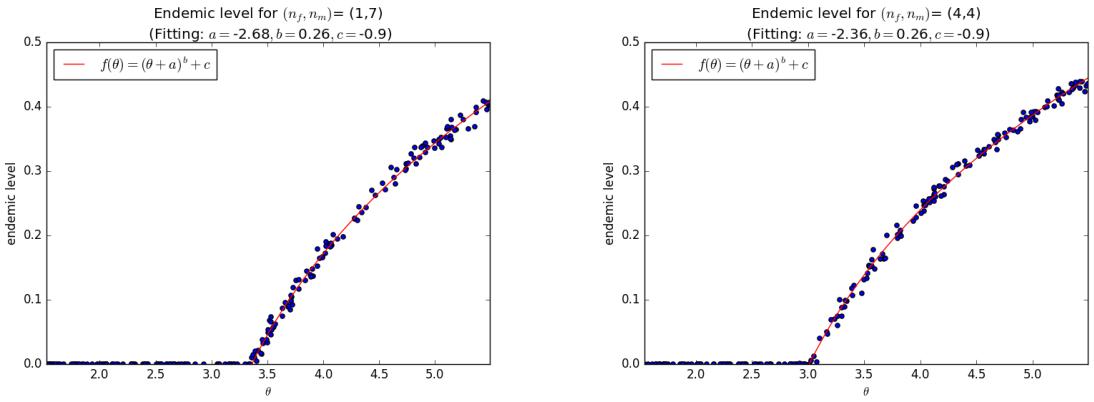

Figure 26: The endemic level dependent on the mean lifetime number of partners θ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Infection - contribution I_1

In fig. 27 we have illustrated the relation between the contribution of the acute phase I_1 and the mean lifetime number of partners. We see that varying θ has hardly any influences on the contribution of I_1. This is surprising, since increasing θ does increase the concurrency index (fig. 24). And because of a higher measure of concurrency, we would expect a higher number of transmission during the acute phase. Although the contribution is measured in the fraction of total transmissions, and the concurrency index influences all transmissions. Therefore the fraction may indeed change little.

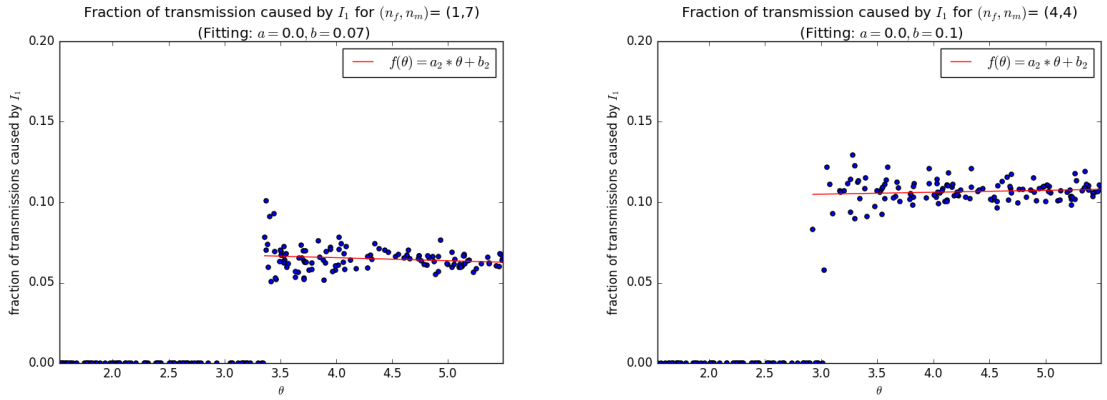

Figure 27: The fraction of transmissions caused by an individual in infection stage I_1 dependent on the mean number of lifetime partners θ for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Sensitivity indices for $\theta = 3.5$

In table 22 we have shown the sensitivity indices for the results for $\theta = 3.5$. When comparing the sensitivity indices of θ with the indices of d_P (in table 21), we see that the sensitivity indices concerning the network are very similar. This coincides with all the results found in this subsection: θ and d_P seem to have approximately the same effect on the network. When looking at the infection we see that varying the parameter has again the largest effect on the endemic level. However θ seems to be hardly any effect on the contribution of I_1.

Sensitivity Index for $\theta = 3.5$		
	Scenarios	
	$(1, 7)$	$(4, 4)$
mean degree	0.9883	1.0055
variance degree	0.7173	0.9115
concurrency index	0.9999	0.9887
mean size components	0.3614	0.5432
SD size components	0.5462	1.1905
endemic level	23.5030	5.8698
contribution I_1	0.0992	0.0396

Table 22: The Sensitivity Index values for different model results for $\theta = 3.5$.

6.1.3 d_{I_1}: the mean duration of the acute infection stage

Lastly we will perform the OAT sensitivity analysis on the parameter d_{I_1}: the mean duration of the acute infection stage I_1. Since we hold the other parameters constant the infection will not influence the network dynamics, and the network will remain stable when varying d_{I_1}. Therefore we will only investigate the effects this parameter has on the infection. We will look at the values $d_{I_1} \in [1/12, 6/12]$.

Infection - Endemic level

Let us take a look at the relation between the endemic level and the mean duration of the acute phase, shown in fig. 28. Increasing or decreasing the length of the very infectious acute phase seems to have hardly any influence on the endemic level. Even though increasing the duration of the acute phase causes a higher endemic level, its effect is much smaller than the increase of the network parameters d_P and θ.

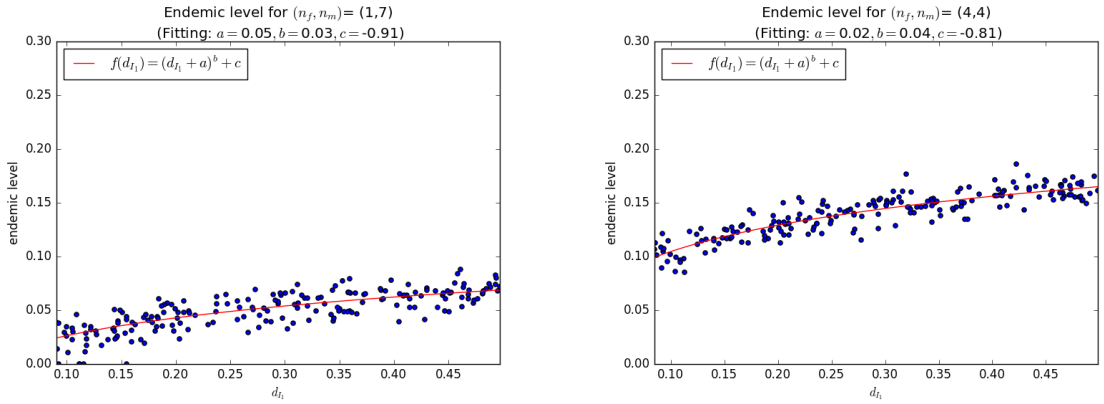

Figure 28: The endemic level dependent on the mean duration of the acute phase d_{I_1} for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Infection - Contribution I_1

In fig. 29 we illustrate the effect changing the mean duration of the acute phase has on the contribution of that phase. Increasing the duration of the acute phase seems to have the most effect in scenario $(4, 4)$. This is however hardly surprising when looking at the network: scenario $(4, 4)$ has a much higher concurrency index than scenario $(1, 7)$. Therefore scenario $(4, 4)$ can take more 'advantage' of the longer acute phase than scenario $(1, 7)$.

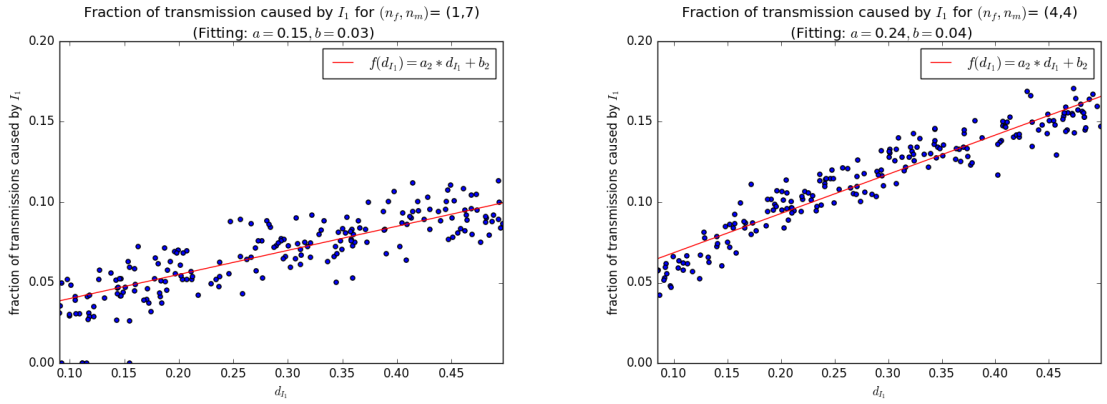

Figure 29: The fraction of transmissions caused by an individual in infection stage I_1 dependent on the mean duration of the acute phase d_{I_1} for $(n_f, n_m) = (1, 7)$ (left) and $(n_f, n_m) = (4, 4)$ (right).

Sensitivity indices for $d_{I_1} = 2.9/12$

In table 23 we have shown the calculated senstivity indices for d_{I_1}. As we have seen in this subsection, varying the mean duration of the acute phase has very little effect. Resulting in the very low SIs in table 23. Comparing these SIs to the indices calculated for d_P and θ we see that especially the effect on the endemic level is surprisingly small. It seems that changing parameters concerning the network, in stead of the infection itself, can have very large influences on the infection.

Sensitivity Index for $d_{I_1} = 2.9/12$		
	Scenarios	
	$(1, 7)$	$(4, 4)$
Endemic Level	0.5671	0.2802
Contribution I_1	0.5901	0.5714

Table 23: The sensitivity index values for different model results for $d_{I_1} = 2.9/12$.

7 Conclusion and Discussion

In this thesis we have investigated how good of an approximation the 'mean field at distance one' assumption is that was made in the deterministic model of Leung et al. [9], [10]. We have used an individual-based simulation model to verify this assumption. First we compared the network (the degree distribution, concurrency index) and the infection (endemic level) for the deterministic model for a set of baseline parameter values with results of the simulation model for these parameter values, to check agreement of the two models. In the set of baseline parameters most values were held constant, but four different scenarios were considered for the partnership capacities for men and women: $(n_f, n_m) = (1, 7), (2, 6), (3, 5)$ and $(4, 4)$, and two different population sizes were considered: $N_0 = 1.000$ and 10.000. We found that the network structure (degree distribution, concurrency index) and the infection (endemic level) agree well between the two models.

Next we compared results concerning the mean field at distance one assumption. Mainly the results concur with each other, but there were a few discrepancies. These however could be explained by the very small sample sizes that can cause inaccurate results. All in all, the results are very promissing and the mean field at distance one assumption seems to be a good approximation. Also, please note that we have only compared the two models for the baseline value parameters, so no conclusions can be drawn for other parameter sets.

Besides checking the mean field at distance one assumption, we have also further investigated the network structure of the simulation model. A great advantage of the simulation model is that the entire network structure is known: we know exactly who has a relationship with whom. This knowledge is not incorporated in the deterministic model. Therefore we were able to investigate for example the connected component size distribution.

We found that for the baseline parameter values, the connected component size distributions for the four scenarios were not that different. For the scenarios $(n_f, n_m) = (2, 6), (3, 5)$ and $(4, 4)$ the distributions were even very similar. Components of size six and larger, hardly ever existed in the networks. In the 100 runs we performed, no greater component was found than of size 15. The distribution of scenario $(n_f, n_m) = (1, 7)$ did differ from the others. This scenario has a limitation on the maximum size of the components, because of the monogamous women. Therefore the variance of the distribution is smaller than the variance for the other three scenarios.

We have also investigated if the probability of getting infected is higher when an individual resides in a large component. We found that this is indeed true, to a certain extent. Since components of size six and larger hardly ever occur in the network, the conclusion is not reliable for these component sizes.

We have performed a senstivity analysis with the one-at-a-time (OAT) approach. We have investigated the effects of varying the mean partnership duration, the mean lifetime number of partners and the mean duration of the acute phase. We have found that both the network parameters (the mean partnership duration and the mean lifetime number of partners) have similar effects on the results, and both have a very large influence on the endemic level around their baseline values. An interesting observation was made during the sensitivity analysis of the mean partnership duration: above a mean partnership duration of 12 years, the network allows the existence of a very large

component. We also found that the variation of the mean duration of the acute phase had very little effect on the results.

There are a lot of possible investigations with the simulation model we have not yet performed in this thesis. Therefore we will now give a few possible future plans. First we have seen that the infection takes a long time to reach endemic stability in the population, especially in the scenario (1,7). It would be very interesting to see if this coincides with the deterministic model.

Secondly, we have not yet investigated the beginning of the epidemic. We have started our simulations with 100 initially infected, to prevent the infection from dying out in the beginning. Therefore we have skipped the early stages of the infection. It would be intriguing to see how the infection spreads if the simulation is started with only one, or at least very few, initially infected.

Thirdly, we have not yet fully exploited all the generalizations we have implemented in the simulation program. For example we could consider two types of partnerships, introducing an infection-induced death rate or not taking the partnership capacities n_f and n_m constant.

Fourthly, we have not yet investigated the entire sample space during the senstivity analysis. We have used the one-at-a-time (OAT) approach, it is a simple approach that does not explore the interaction between input parameters. Therefore varying multiple parameter at the same time, would be an interesting next step.

We have seen during the sensitivity analysis of the mean partnership duration that from a certain value there arises one very large component in the network. The results of the sensitivity analysis for the mean lifetime number of partners were very similar to the analysis of the mean partnership duration, but such an occurence was not observed. However for the senstivity analysis of the mean lifetime number of partners, we looked at a smaller interval. Therefore we could investigate if such existence of a very large component also happens if we extend the OAT interval for the mean lifetime number of partners.

References

[1] AVERT 1986 - 2016, "Global information and advice on HIV & AIDS." `http://www.avert.org/`. Accessed: Augustus 2015.

[2] WHO, "HIV/AIDS - Fact sheet." `http://www.who.int/mediacentre/factsheets/fs360/en/`, July 2015.

[3] M. N. Lurie, B. G. Williams, K. Zuma, D. Mkaya-Mwamburi, G. P. Garnett, A. W. Sturm, M. D. Sweat, J. Gittelsohn, and S. S. A. Karim, "The impact of migration on hiv-1 transmission in south africa: a study of migrant and nonmigrant men and their partners," *Sexually transmitted diseases*, vol. 30, no. 2, pp. 149–156, 2003.

[4] C. H. Watts and R. M. May, "The influence of concurrent partnerships on the dynamics of hiv/aids," *Mathematical biosciences*, vol. 108, no. 1, pp. 89–104, 1992.

[5] M. Morris and M. Kretzschmar, "Concurrent partnerships and transmission dynamics in networks," *Social Networks*, vol. 17, no. 3, pp. 299–318, 1995.

[6] K. M.E.E., D. Y.T.H.P., and S. A.J., "Modeling prevention strategies for gonorrhea and chlamydia using stochastic network simulations," *American Journal of Epidemology*, vol. 144.3, 1996.

[7] M. Morris and M. Kretzschmar, "Concurrent partnerships and the spread of hiv," *Aids*, vol. 11, no. 5, pp. 641–648, 1997.

[8] M. N. Lurie and S. Rosenthal, "Concurrent partnerships as a driver of the hiv epidemic in sub-saharan africa? the evidence is limited," *AIDS and Behavior*, vol. 14, no. 1, pp. 17–24, 2010.

[9] K. Leung, M. Kretzschmar, and O. Diekmann, "Dynamic concurrent partnership networks incorporating demography," *Theoretical population biology*, vol. 82, no. 3, pp. 229–239, 2012.

[10] K. Leung, M. Kretzschmar, and O. Diekmann, "SI infection on a dynamic partnership network: characterization of R_0," *Journal of mathematical biology*, vol. 71, no. 1, pp. 1–56, 2015.

[11] K. Leung, "Easy generalizations (unpublished)." 2015.

[12] S. F. Railsback and V. Grimm, *Agent-based and individual-based modeling: a practical introduction*. Princeton university press, 2011.

[13] R. G. Gallager, *Stochastic processes: theory for applications*. Cambridge University Press, 2013.

[14] K. Leung, K. Powers, and M. Kretzschmar, "Gender asymmetry in concurrent partnerships and hiv prevalence (unpublished, but submitted)." 2015.

[15] J. Rice, *Mathematical statistics and data analysis*. Nelson Education, 2006.

[16] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola, *Global sensitivity analysis: the primer*. John Wiley & Sons, 2008.

[17] A. Van Griensven, T. Meixner, S. Grunwald, T. Bishop, M. Diluzio, and R. Srinivasan, "A global sensitivity analysis tool for the parameters of multi-variable catchment models," *Journal of hydrology*, vol. 324, no. 1, pp. 10–23, 2006.

[18] P. Software Foundation, "The Python Profilers." `https://docs.python.org/2/library/profile.html`, c1990-2015. Accessed: June 2015.

[19] T. pandas development team, "Python Data Analysis Library," c2008-2014.

[20] S. Behnel, R. Bradshaw, D. S. Seljebotn, G. Ewing, W. Stein, and G. Gellner, "Cython C-extensions for Python." `http://cython.org/`, c2015. Accessed: Augustus 2015.

[21] "Time complexity." `https://wiki.python.org/moin/TimeComplexity`, June 2015.

[22] M. Harrison, *Treading on Python Volume 1: Foundations of Python*. CreateSpace Independent Publishing Platform, 2013.