



Universitat Politècnica de València  
Departament de Matemàtica Aplicada

PhD. THESIS

# Building networks of sexual partners. Application for the study of the transmission dynamics of Human Papillomavirus (HPV)

Ph.D. CANDIDATE

Víctor Sánchez Alonso

ADVISORS

Dr. Luis Acedo Rodríguez

Dr. Rafael Jacinto Villanueva Micó

Dr. Francisco Javier Villanueva Oller

Valencia - November 2018

Dr. Luis Acedo Rodríguez and Dr. Rafael Jacinto Villanueva Micó, from the Universitat Politècnica de València and Dr. Francisco Javier Villanueva Oller from the Universidad Rey Juan Carlos,

**CERTIFY** that the present thesis entitled *Building networks of sexual partners. Application for the study of the transmission dynamics of Human Papillomavirus (HPV)* has been performed under our supervision in the Department of Applied Mathematics at the Universitat Politècnica de València by Víctor Sánchez Alonso. It constitutes his thesis dissertation to obtain the PhD degree in Mathematics.

In compliance with the current legislation, we authorize the presentation of this dissertation signing the present certificate.

Valencia, September 7, 2018

Luis  
Acedo Rodríguez

Rafael Jacinto  
Villanueva Micó

Francisco Javier  
Villanueva Oller

# Abstract

Sexually Transmitted Diseases (STD) have been a major public health threat for a long time in human history. Modern concerns about STD began with the pandemic of syphilis which spread over Europe in the early sixteenth century.

The Human Papillomavirus (HPV) is the direct cause of more than half million new cases of cervical cancer, the second most common malignancy among women and a leading cause of cancer death worldwide. It also causes anogenital warts. It is estimated that the probability of transmission of HPV is 40-50% per contact. The networks of sexual contacts in human populations are crucial to the spread of sexually transmitted diseases (STDs).

Working in large networks applied to epidemiological-type models has led us to design a simple but effective computed distributed environment to perform a large amount of model simulations in a reasonable time in order to study the behavior of these models and to calibrate them. Finding the model parameters that best fit the available data in the designed distributed computing environment becomes a challenge and it is necessary to implement reliable algorithms for model calibration.

We are able to simulate our model and carry out vaccination campaigns in order to get conclusions concerning the best strategies. This information can be useful for Health Care, policy makers and pharmacological industries to save time and money.

# Resumen en Castellano

Sexually Transmitted Diseases (STD) have been a major public health threat for a long time in human history. Modern concerns about STD began with the pandemic of syphilis which spread over Europe in the early sixteenth century.

The Human Papillomavirus (HPV) is the direct cause of more than half million new cases of cervical cancer, the second most common malignancy among women and a leading cause of cancer death worldwide. It also causes anogenital warts. It is estimated that the probability of transmission of HPV is 40-50% per contact. The networks of sexual contacts in human populations are crucial to the spread of sexually transmitted diseases (STDs).

Working in large networks applied to epidemiological-type models has led us to design a simple but effective computed distributed environment to perform a large amount of model simulations in a reasonable time in order to study the behavior of these models and to calibrate them. Finding the model parameters that best fit the available data in the designed distributed computing environment becomes a challenge and it is necessary to implement reliable algorithms for model calibration.

We are able to simulate our model and carry out vaccination campaigns in order to get conclusions concerning the best strategies. This information can be useful for Health Care, policy makers and pharmacological industries to save time and money.

# Resum en Valencià

Sexually Transmitted Diseases (STD) have been a major public health threat for a long time in human history. Modern concerns about STD began with the pandemic of syphilis which spread over Europe in the early sixteenth century.

The Human Papillomavirus (HPV) is the direct cause of more than half million new cases of cervical cancer, the second most common malignancy among women and a leading cause of cancer death worldwide. It also causes anogenital warts. It is estimated that the probability of transmission of HPV is 40-50% per contact. The networks of sexual contacts in human populations are crucial to the spread of sexually transmitted diseases (STDs).

Working in large networks applied to epidemiological-type models has led us to design a simple but effective computed distributed environment to perform a large amount of model simulations in a reasonable time in order to study the behavior of these models and to calibrate them. Finding the model parameters that best fit the available data in the designed distributed computing environment becomes a challenge and it is necessary to implement reliable algorithms for model calibration.

We are able to simulate our model and carry out vaccination campaigns in order to get conclusions concerning the best strategies. This information can be useful for Health Care, policy makers and pharmacological industries to save time and money.

# Contents

<b>1</b>	<b>Introduction</b>	<b>12</b>
1.1	The revelation of the virus . . . . .	12
1.2	Vaccines . . . . .	14
1.3	Our proposal . . . . .	16
<b>2</b>	<b>Building and Dynamics of the Sexual Contacts Network</b>	<b>19</b>
2.1	Origin of the Data . . . . .	19
2.2	Network Model . . . . .	20
2.3	Semi-Random Construction . . . . .	22
2.4	The Dynamics of HPV Transmission in the Sexual Network . . . . .	26
2.5	Decline of infections . . . . .	27
<b>3</b>	<b>Validation of the model</b>	<b>29</b>
3.1	How to measure the decline . . . . .	30
3.2	Does our model return similar values to those in [1]? . . . . .	30
3.3	Study of the herd immunity effect over HPV LR infection . . . . .	33
<b>4</b>	<b>Calibrating the Sexual Contacts Network</b>	<b>35</b>
4.1	Description of the model . . . . .	36
4.2	Description of our resources . . . . .	38
4.2.1	Computers . . . . .	38
4.2.2	Distributed computing environment . . . . .	39
4.2.3	Random Particle Swarm Optimization (rPSO) . . . . .	39
4.2.4	Fitness function and some features included in the rPSO . . . . .	41
4.3	Selecting an optimal number of particles to calibrate the HPV large network computational model with rPSO . . . . .	43
4.4	Improving the exploration of the rPSO . . . . .	44
4.5	Calibrating the model . . . . .	45
4.6	Selecting the 30 realizations that best capture the data uncertainty . . . . .	46

<b>5</b>	<b>Making Dynamic Sexual Contacts Network and adding Vaccination Campaigns</b>	<b>51</b>
5.1	Including new features into the model . . . . .	51
5.1.1	Introducing age and LSP dynamics into the LSP model	51
5.1.2	When should we start the vaccination schedule? . . . .	53
5.1.3	Introducing vaccination . . . . .	55
5.1.4	Introducing vaccination loss protection . . . . .	55
5.1.5	Introducing variations in the vaccination coverage . . .	56
<b>6</b>	<b>Conclusions</b>	<b>57</b>

# List of Figures

1.1	Cartoon illustrating the genomic organization of a typical mucosal high-risk HPV. The genome contains early and late regions (E), which relate to the positions of the genes within the genome and their timing of expression during the viral life cycle. The early region carries a number of genes which function at the level of viral replication and transcription, i.e., E2, E1, E6, and E7. E2 encodes a protein which has an auxiliary role in viral replication and also functions at the level of transcriptional regulation of the viral early genes. The E6 and E7 genes encode the major transforming proteins of the oncogenic HPVs. The late region (L) encodes viral structural proteins, with L1 being the major capsid protein and L2 being the minor capsid protein, adopted from [2]. . . . .	13
1.2	Harald zur Hausen (born 11 March 1936) is a German virologist and professor emeritus. He has done research on cancer of the cervix, where he discovered the role of papilloma viruses, for which he received the Nobel Prize in Physiology or Medicine 2008. Adopted from [3] . . . . .	14
1.3	Age-adjusted prevalence of cervical human papillomavirus DNA in sexually active women aged 15-69 years. Data are from IARC Prevalence Surveys, 1990-2012.4, adopted from [4]. . . . .	15
2.1	Example Network of 1,000 nodes, MSM in yellow, women in red and men in blue. . . . .	20
2.2	Flow diagram for the algorithm corresponding to the assignment of a number of LSPs to every male and female in the network. . . . .	24

- 3.1 Percentage of women (a), men (b) and MSM (c) aged 14-26 infected of LR HPV 6 and/or 11 after the implementation of the vaccination program. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. We can see the fast decrease for women and men in both scenarios from the very beginning. However, there is no effect on MSM. . . . . 31
- 3.2 Percentage of decline of women (a) and men (b) aged 14-26 infected of LR HPV 6 and/or 11 (and consequently of genital warts) after the implementation of the vaccination program. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. After 2 years, the model shows a decline of 72% for Scenario 1 and 54.8% for Scenario 2, in average, for women and 38.9% for Scenario 1 and 27.7% for Scenario 2, in average, for men. . . . . 32
- 3.3 Percentage of decline of women (a) and men (b) aged 27-64 infected of LR HPV 6 and/or 11 (and consequently of genital warts) after the implementation of the vaccination program in both scenarios. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. Notice that, in average, no significant decline appears in the 5 years after the implementation of the vaccination program. . . . . 33
- 3.4 Percentage of decline of women (a), men (b) and MSM (c) aged 14-64 for the vaccination program in Australia. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. In the figure (a), green and orange lines correspond to women vaccination percentage for Scenario 1 and 2 respectively. Notice that the herd immunity effect contributes to the decline in the number of infections in men and the decline in the number of infections for unvaccinated women. This latter can be seen when the decline lines are over the vaccination line. However, any herd immunity effect can be seen in MSM. . . . . 34
- 4.1 Introduction of the asynchronous rPSO algorithm in the Sisifo environment. rPSO manages the generation of new problems and put them in the *Problem files* folder and the reading and processing of the solution files located in the *Result files* folder. 41

4.2	Histogram of the computation time of each one of the 10,100 realizations of the model using networks of 100,000 nodes. The average time is 1374.5 seconds, around 23 minutes. . . . .	46
4.3	Drawing of the 100 model realizations with error less than 0.01 from month 400 to 499. . . . .	47
4.4	Drawing of the 1282 model realizations with error less than 0.05 from month 400 to 499. These realizations cover the 95% confidence intervals of the data, represented by the blue horizontal dashed lines. . . . .	48
4.5	Drawing of the 30 selected model realizations from month 400 to 499. These realizations are the ones whose means and 95% confidence intervals resemble the most the data in Table 4.2, represented by the blue horizontal dashed lines. . . . .	49
4.6	Means and 95% confidence intervals of the 30 selected realizations (in red) compared to the means and 95% confidence intervals of the data (blue). . . . .	50
5.1	Static LSP network. The ages remain constant over the time. The time only affects the HPV contagion/clearing dynamics. . . . .	52
5.2	Dynamic LSP network. The ages evolve over the time. The age group 40-65 and their sexual behavior disappear when people grow. . . . .	53
5.3	Mean and 95% confidence intervals of the prevalence for 18-64 men and women for HR and LR. Vertical dashed lines indicate milestones of interest. Note the oscillations of the prevalence levels. The blue and green vertical lines correspond to the peaks and valleys of the oscillations. The magenta line, points the month 500 when dynamic network starts. . . . .	54

# List of Tables

2.1	Proportion of males and females per number of life sexual partners LSP per age group. . . . .	19
2.2	Prevalence of HR- and LR-infected women per age groups from the CLEOPATRE study [5]. Co-infections are included in both HR- and LR-infected, mean and 95% confidence intervals. . . . .	27
4.1	Proportion of males and females per number of LSP per age group [6]. Note that the sum of the rows are 1. . . . .	36
4.2	Prevalence of HR- and LR-infected women per age groups [5]. Mean and 95% confidence intervals. Co-infection is considered in both, HR and LR. . . . .	38
4.3	Analysis of the quality and execution times of different rPSO configurations. Results of 1200 evaluations for different number of particles on the rPSO process (# Part.). . . . .	44
4.4	Upper figures: Time of clearance of HPV high risk and low risk, both for men and women. Lower figures: on the left, probability to transmit low risk HPV if the infected is a woman; on the right, probability to transmit high risk HPV if the infected is a man. Red dashed lines correspond to the bounds of each parameter. We can see that most of the search space is explored in all the cases. . . . .	45
4.5	Mean and the 95% confidence interval of the model parameters corresponding to the 30 selected realizations from (4.1). . . . .	50

# Chapter 1

## Introduction

### 1.1 The revelation of the virus

One of the biggest scientific discoveries in the past 30 years was the connection between human papillomavirus infection of the cervix and cervical cancer. This achievement resulted from the original seminal findings by Harald zur Hausen and his group, they found that human papillomavirus genotype 16 can be detected in cervical cancer tissue.

The finding was followed by epidemiologists, molecular biologists, vaccinologists, and clinicians culminating in 2006 with the development of effective prophylactic vaccines for human papillomavirus, which have the means to prevent 70-80% of cervical cancer. Zur Hausen was awarded the Nobel Prize in Physiology or Medicine in 2008, in recognition of his discovery.

HPV types are often referred to as low risk (LR) wart causing or high risk (HR) cancer causing [7], based on whether they put a person at risk for cancer. The types of HPV that can cause genital warts are not the same as the types that can cause cancer. Persistent human papillomavirus (HPV) infections with genotypes 16 and 18 are responsible for about 70% of all cervical cancer, with 40-85% of other anogenital cancers: anal, penile, vaginal, and vulvar cancer, and also 16-28% of the head and neck cancers [8].

HPV is a cause of other non malignant diseases, to mention genotypes 6 and 11 cause about 90% of anogenital warts, and secondarily juvenile onset of recurrent respiratory papillomatosis [9].

Genital human papillomavirus (HPV) is the most common sexually transmitted infection in the United States. More than 40 HPV types can infect the genital areas of men and women, including the skin of the penis, vulva, and anus, and the linings of the vagina, cervix, and rectum. These types can also infect the lining of the mouth and throat.

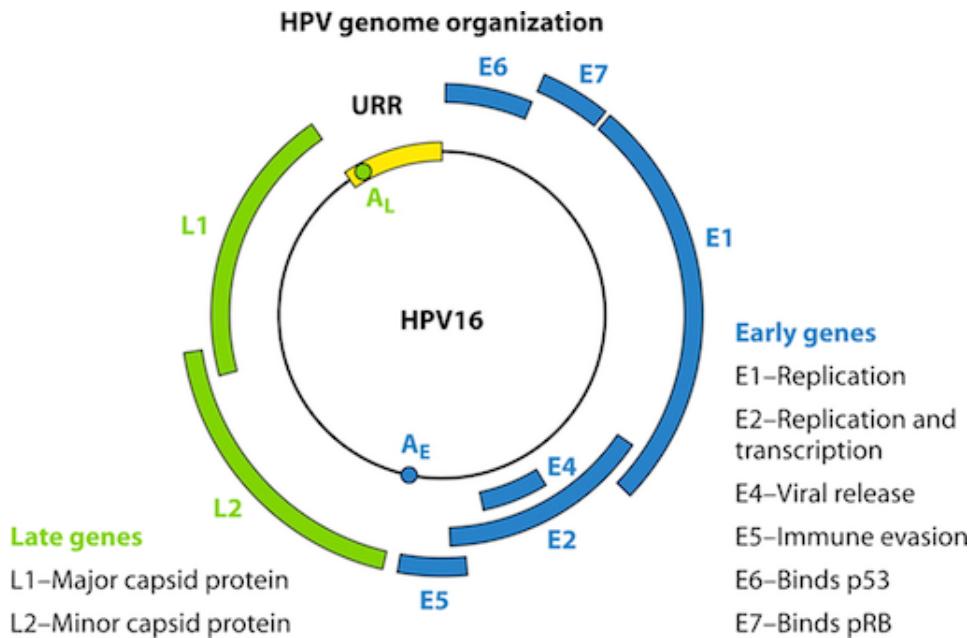


Figure 1.1: Cartoon illustrating the genomic organization of a typical mucosal high-risk HPV. The genome contains early and late regions (E), which relate to the positions of the genes within the genome and their timing of expression during the viral life cycle. The early region carries a number of genes which function at the level of viral replication and transcription, i.e., E2, E1, E6, and E7. E2 encodes a protein which has an auxiliary role in viral replication and also functions at the level of transcriptional regulation of the viral early genes. The E6 and E7 genes encode the major transforming proteins of the oncogenic HPVs. The late region (L) encodes viral structural proteins, with L1 being the major capsid protein and L2 being the minor capsid protein, adopted from [2].

HPV is the most common sexually transmitted infection (STI) in the world. It is estimated that about 1 million new cases of GW are reported each year and the cost of treatment is increased by the tendency of these warts to recur after initial clearance. The cost of the treatment of genital warts was estimated to exceed \$ 3.8 billion in the U.S. in 1997 [10]. In Spain there were 35,000 cases in women in 2007 with an overall annual cost of € 47 millions [5].

Two vaccines consisting of non-infectious HPV virus-like particles (VLP) [11] containing the capsid protein L1 of the virus but not the viral DNA. They induce a high immune response and prevent premalignant lesions. Both vaccines contain genotypes 16 and 18, and one of them adds VLP against



Figure 1.2: Harald zur Hausen (born 11 March 1936) is a German virologist and professor emeritus. He has done research on cancer of the cervix, where he discovered the role of papilloma viruses, for which he received the Nobel Prize in Physiology or Medicine 2008. Adopted from [3]

genotypes 6 and 11 (HPV4).

In the Region of Valencia, Spain, the vaccine is administered to < 15 girls. Similar vaccination strategies of this kind were modelled by Elbasha et al. [12, 13] by means of a compartmental model with 17 age groups for each gender. This model focuses mainly on the development of cervical intraepithelial neoplasia (CIN) and its progression from CIN1 to CIN3. According to these authors vaccination must be implemented for adolescent girls aged between 12 and 14 years. Elbasha et al. also found some evidences that the vaccination of boys could also be cost-effective [12]. By vaccinating girls alone an 83% reduction in the incidence of GW is expected but this reduction is increased to 97% if boys are also vaccinated.

## 1.2 Vaccines

Since the release of the first vaccines in 2006, nowadays there are three available: a quadrivalent (including HPV genotypes 16, 18, 6 and 11) and a bivalent vaccine (including genotypes 16 and 18) and a nonavalent (including genotypes 6, 11, 16, 18, 31, 33, 45, 52 and 58). All vaccines are efficacious to protect against precancerous lesions in the cervix, vulva or vagina; in addition, the quadrivalent and nonavalent prevent precancerous anal lesions,

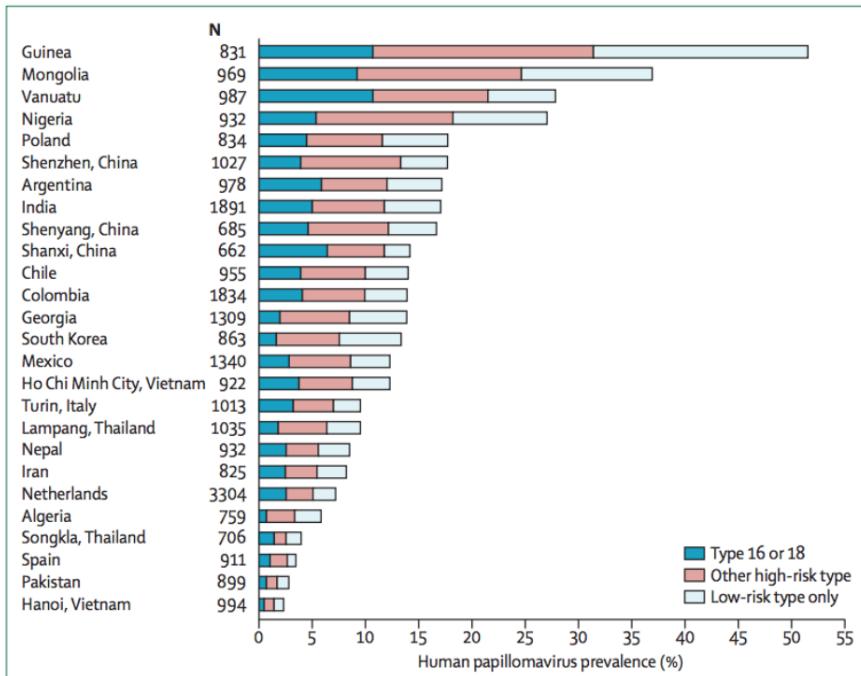


Figure 1.3: Age-adjusted prevalence of cervical human papillomavirus DNA in sexually active women aged 15-69 years. Data are from IARC Prevalence Surveys, 1990-2012.4, adopted from [4].

anal cancer and anogenital warts.

According to the Advisory Committee on Immunization Practices (ACIP) from the Centers for Disease Control (CDC) and Prevention, new recommendations are given for use of a 2-dose schedule for girls and boys who initiate the vaccination series at ages 9 through 14 years. Three doses remain recommended for persons who initiate the vaccination series at ages 15 through 26 years and for immunocompromised persons.

In Spain HPV vaccine is given to adolescent girls as part of the national immunization programme, and is recommended at different age groups in different Autonomous Communities. Numerous cost-effectiveness studies of HPV-vaccination have been published in other countries. However, few studies include the prophylactic effect of all HPV-associated diseases, or the impact of vaccinating men.

There are countries that recommend also vaccinating boys in order to decrease the burden of disease in them. Some models have shown that the female vaccination program has some herd immunity and the impact of implementing the vaccination in males may not be cost effective, however not

vaccinating males leaves them at risk of cancers, especially the groups that do not benefit of the herd immunity, as the males who have sex with males. Besides, there is no economic analysis of the nonavalent program in Spain, and it is important from the decision making perspective.

Even with the high prevalence of sexually transmitted diseases there are few studies devoted to ascertain the structure of sexual networks and its role in disease transmission. Most studies are restricted to small communities such as the Jefferson High Schools project [14] or that of Likoma Island [15].

Random network mathematical models may simulate the interactions and propagations of all these viruses through sexual contacts among a population of more than one million people (including heterosexual and homosexual populations). As this model is based upon a network instead of traditional continuous model approaches [12] we will be able to determine with higher accuracy the effect of vaccination in a short and large periods of time.

The Human papillomaviruses HPV vaccine induces a herd immunity effect in genital warts when a large number of the population is vaccinated. This aspect should be taken into account when devising new vaccine strategies, like vaccination at older ages or male vaccination.

### 1.3 Our proposal

In recent studies [16, 1] a decrease on the number of infected persons and the number of persons with GW is already reported for Australia after two years of administering vaccinations to young girls. It showed both direct and indirect protection in males. These results were more impressive than the predictions of the continuous models. A Bayesian model for HPV vaccination was then proposed by Bogaards et al. [17] and focused on the herd immunity effect of the female vaccination on the male population in a static picture. A dynamic understanding for the short and long term effects of vaccination policies is, however, still necessary and even more so with HPV vaccines because their benefit to the whole population is to be observed in the time span of several decades.

New vaccination schedules, specially vaccination in boys, should take into account the herd immunity effect vaccination in girls (in mainly heterosexual societies), therefore models that simulate STI would be important. Network models are those models in which the relations of the individuals (represented by network sites) are taken into account by means of the network bonds. Random network models have been successfully applied to other infectious diseases such as the Respiratory Syncytial Virus pandemic [18] as well as other social pandemics such as obesity [19].

STI are more likely to produce large-scale infections than other transmissible diseases, such as respiratory transmitted diseases, because the efficacy of sexual contacts for the infection is large and the infectious agent has long latency periods as in the case of HIV or HPV. Moreover, neither the carrier nor his/her partner are aware of their exposure. For example, it has been estimated that around 40–50% of contacts are capable of transmitting HPV [20]. On the other hand, some STIs are caused by oncoviruses such as Hepatitis B or HPV, which increase the death rate of people who develop the disease.

In order to predict the evolution of these diseases we need a reliable model of the underlying social network in which the infection builds up. Individuals who change partners or have several partners simultaneously, are the hubs favouring the spread of STI. The distribution of degrees of the nodes in the network and the average chemical path from an infected individual to a susceptible one, are important parameters controlling the final extension of a new STI in a population and the speed at which it spreads. However, most models are based on some assumptions, which could not be valid for certain populations. Some studies claimed that the web of human sexual contacts is a scale-free network characterized by a power-law distribution for the number of individuals with a certain degree of connectivity,  $k$ :  $P(k) \propto 1/k^\alpha$  with a value of  $\alpha$  in the range  $2 < \alpha < 3$ , and slightly smaller for males than for females [21]. Although  $P(k)$  provides some valuable information about the network, it is not a sufficient prescription on how to build it for a given population size. Moreover, a power-law distribution of contacts could not be representative of some populations, or could vary from country to country. For example, it has been found that a densely connected core appears without the need of a high connectivity degree in the Jefferson High School's network [14].

Some field studies have ascertained the structure of moderate size real networks of sexual contacts. In 2004, Bearman et al. published the results for a set of 800 adolescents in a midsized town of the United States [14]. They showed that the structure of this network is a big cluster with a ring and extended filaments which contained most of the adolescents implying that, potentially, the infection of an individual could spread to the whole of the population, given sufficient time and infectivity. A similar study was performed in 2007 at the Likoma Island in Malawi with the idea of predicting and explaining the expansion of HIV in sub-Saharan populations [15]. That study disclosed that the sexual network contained many cycles, in contrast with the single cycle at Jefferson High School. For that reason, it was speculated that superimposed cycles could be the explanation of the high prevalence of HIV infection in small populations of Africa.

Some recent studies reveal that the evolution of partnerships is also an important factor in the transmission of STIs. In particular, they pointed out that the following items should be considered: (i) the cumulative distribution of the lifetime number of partners, (ii) the distribution of partnership durations, (iii) the distribution of gap lengths between partnerships, and (iv) the number of recent partners. A method for building up networks considering these items has been developed by Schmid and Kretzschmar [22]. However, this information is not available in most surveys, and we therefore face the problem of developing reasonable models for STIs in many countries where information about sexual behaviour is scarce. For example, in the case of Spain, there is only available data about the number of sexual contacts in a lifetime from surveys. This is sufficient for building a sexual network for the transmission of HPV or other diseases with lifetime consequences and progression. In these cases, the important fact is whether the individual has had a contact with risk of infection. The remaining aspects of the network such as the duration of partnership and the time intervals among them can be incorporated effectively into a probability of transmission parameter.

In this dissertation we show how to build a network model for sexual contacts from the usual statistical data in surveys concerning the number of partners in a lifetime. We consider both heterosexual men and men who have sex with men (MSM) populations and the connections among them. We perform simulations over this network substrate on the HPV infections by different genogroups including both LR and HR infections. In particular, we show that for the case of Australia the strategy of a vaccination for 12 – 13 year-old girls plus catch-up lead to a considerable reduction in the number of cases of infection by HPV 6 and/or 11 (which are the main cause of GW). For women in the 14 – 26 age-group we obtain a decrease of 59% after 3.6 – 4.6 years and 39% in men after 3 – 3.75 years. These results agree with the conclusions of the study by Ali et al. [1].

In the following chapters we will explain first how to build the network of sexual contacts, then how to validate our model followed by its calibration. Therefore we will show cases that have been demanded from Public Health representatives running our VPH model. Final chapter is dedicated to conclusions.

# Chapter 2

## Building and Dynamics of the Sexual Contacts Network

### 2.1 Origin of the Data

Demographic data from the region of Valencia (Spain) was collected from the Valencian Institute of Statistics (2013) [23]. Lifetime sexual partners for an individual (LSP) was obtained from the Health and Sexual Habits Survey of 2003 [6], and summarized in Table 2.1.

MALES						
Age	0 LSP	1 LSP	2 LSP	3–4 LSP	5–9 LSP	10 or More LSP
14–29	0.107	0.207	0.131	0.225	0.168	0.162
30–39	0.027	0.225	0.128	0.21	0.17	0.24
40–65	0.019	0.268	0.14	0.193	0.163	0.217

FEMALES						
Age	0 LSP	1 LSP	2 LSP	3–4 LSP	5–9 LSP	10 or more LSP
14–29	0.138	0.43	0.186	0.158	0.056	0.032
30–39	0.029	0.501	0.168	0.177	0.077	0.048
40–65	0.017	0.652	0.138	0.118	0.039	0.036

Table 2.1: Proportion of males and females per number of life sexual partners LSP per age group.

Some features of the distribution of contacts were: (i) the percentage of males and females with no partners is very similar in each age-group; (ii) the

proportion of women with a single partner is, approximately, two times larger than men with only one partner; and (iii) the percentage of men with two or more partners is always larger than that of women except for women in the age-groups 14–29, and 30–39 in the case of two partners. The asymmetry in the behaviour of males and females should be taken into account in the construction of the network.

## 2.2 Network Model

Before proceeding with the definition of our model, it is convenient to give a brief and general perspective on the emergent field of network research for the readers not familiar with these techniques and their application in epidemiology. A network is, basically, a model which derives from the abstract mathematical concept of a graph composed by a set of points (the so-called nodes) connected among them by some lines or edges (known as “links” in the case of networks).

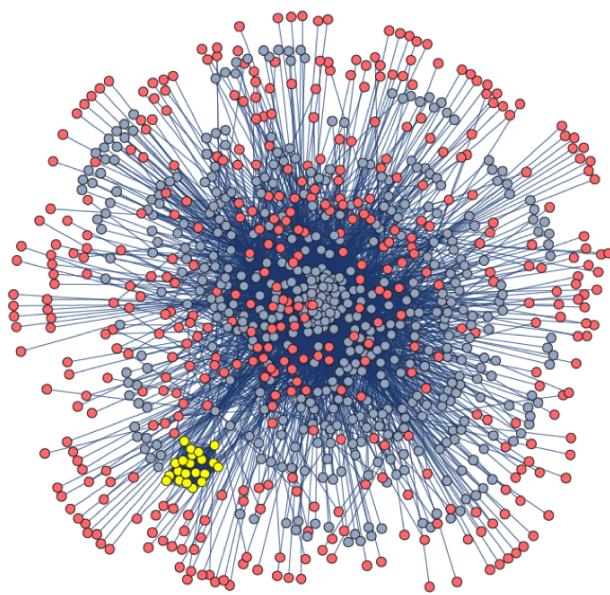


Figure 2.1: Example Network of 1,000 nodes, MSM in yellow, women in red and men in blue.

There are several types of networks of interest for the applied sciences. If we classify them according to the degree of a node, i.e., to the number of links for a given node (or, more properly, to the distribution of these number of links), we have two main categories in the literature: (i) random

networks, in which the links or edges occur with a fixed probability and the statistical distribution of this number of links follows a Poisson's law; and (ii) scale-free networks whose distribution of degrees follows a power-law with an algebraic tail of the form  $P(k) \simeq 1/k^\gamma$  with  $2 < \gamma < 3$ . This means that the nodes with very large degrees are more likely to appear in scale-free networks than those in random networks. Random networks have been used in epidemiology [18] and also as an elementary model of the brain [24]. On the other hand, scale-free networks have been successfully applied to the Internet and biological networks in which some nodes with a very large number of links are determinant in the control of the dynamics (see [25]). Small-world networks are also an important paradigm in the science of networks. This concept refers to the average length of a path connecting two typical nodes in the network. It was found that some sparse networks may, on the other hand, present the small-world phenomenon, i.e., we have short paths connecting every pair of nodes through the links with other nodes. A mechanism to generate these networks was discovered by Watts and Strogatz [26]. Many networks in sociology exhibit this small-world property [19, 21, 14].

In this dissertation, we use the random network model as a basis to simulate the network of sexual contacts among individuals, but, in this model, the average number of connections depend upon the age-group as deduced from Table 2.1. A basic property of the network we are going to discuss is that the total number of lifetime sexual partners (LSP) for the male population (M) must coincide with the total number of LSP for the female population (F). This is so because (in a purely heterosexual network) every link starting on a male must end in a female and viceversa. In mathematical terms:

$$\sum_{i=1}^M LSP_i = \sum_{j=1}^F LSP_j . \quad (2.1)$$

The estimation of sexual partners: there are some approaches to the number of LSP in males and females [27, 28] that are difficult to match. Generally speaking, males tend to overestimate the number of their sexual partners and females tend to underestimate it. Therefore, we considered the average LSP male value,  $\langle k \rangle_m$ , and calibrated the network so that results were consistent with data of Table 2.1, and estimated that the number of LSP in males in Spain was at least 4.5. Networks with 250,000, 500,000 and 750,000 have been necessary to perform the present study. It required a substantial computational power.

## 2.3 Semi-Random Construction

From Table 2.1 (proportion of male LSP aged 14–29), we have the following list: (0.107, 0.314, 0.445, 0.67, 0.838, 1) for the accumulated proportion of males less than or equal to a given LSP number. Now, we randomly generate a number  $r$  between 0 and 1 and assign the number of contacts to every male node in the 14 – 29 age group, in the network as follows:

- $r \leq 0.107$  say that the corresponding male does not have an LSP,
- $0.107 < r \leq 0.314$  say that the corresponding male has one LSP,
- $0.314 < r \leq 0.445$  say that the corresponding male has two LSPs,
- $0.445 < r \leq 0.67$  say that the corresponding male has three or four LSPs uniformly distributed.
- $0.67 < r \leq 0.838$  say that the corresponding male has five to nine LSPs uniformly distributed.
- $0.838 < r \leq 1$  say that the corresponding male has 10 or more LSPs.

Every node in the network is labelled by its gender and age randomly assigned according to the population histogram. The assignment of the number of bonds, as another label of the node, is not so straightforward since we must guarantee that the condition in Equation (2.1) is verified. In order to fulfill this condition, we take advantage of the uncertainty of statistics reports concerning individuals with 10 or more LSPs. Starting with the males, we assign the number of LSPs up to nine partners and, for 10 or more partners proceed as follows: let  $i_{\max}$  be the number of males with nine or less partners. The unassigned males should be  $M - i_{\max}$  and the number of bonds that should be distributed among them is  $M \langle k \rangle_m - \sum_{i=1}^{i_{\max}} LSP_i$ . By Euclidian division, this quantity can be expressed as  $(M - i_{\max})n_m + r_m$ , where  $n_m \geq 10$ . In our procedure, we assign a random number of bonds, uniformly distributed, in the interval  $[10, 2n_m - 10]$  to every male with 10 or more LSPs, i.e., to the  $M - i_{\max}$  unassigned males.

Now, we denote as  $p_m$  the total number of bonds of the male population. We must take into account, as expressed in Equation (2.1), that the total number of bonds of the female population should be the same. To impose that condition, we proceed as follows: (i) assign the number of bonds to the females with nine or less partners following the statistical data in Table 2.1; and (ii) the sum of all female LSPs in this group of  $j_{\max}$  members will be denoted by  $s_f$ . Then,  $n_f = F - j_{\max}$  is the number of females with 10

or more LSPs; and (iii) the number of bonds starting in the males and still unassigned to a female is  $p_m - s_f = q_f n_f + r_f$ , where  $0 \leq r_f < n_f$  and  $n_f \geq 10$ . (iv) We assign  $q_f + 1$  bonds to  $r_f$  females still unassigned and  $q_f$  bonds to the rest of  $n_f - r_f$  females. The steps of this algorithm are also enumerated in the flow diagram in Figure 2.2.

Notice that, for men and women with more than 10 LSPs, we assign their LSP in the most equitable way, assuming that all of them have, more or less, the same number of LSPs. Thus, we have a lot of hubs with a low number of contacts instead of a few hubs with a lot of contacts.

From the point of view of STI transmission, the latter situation leads to a faster transmission if the hub is infected, and, also, if the hub is vaccinated, the transmission is cut faster. Therefore, due to the lack of data about people with 10 or more LSPs, we make the decision of being conservative in the transmission of the disease and in the effect of the vaccination campaigns.

Notice that this procedure implies that the condition in Equation (2.1) is verified. After this procedure, we have obtained the following lists:

- $AgeMale[i]$  is the age of the  $i$ -th male,  $i = 1, \dots, M$ ,
- $AgeFemale[i]$  is the age of the  $i$ -th female,  $i = 1, \dots, F$ ,
- $kMale[i]$  is the number of LSP for the  $i$ -th male,  $i = 1, \dots, M$ ,
- $kFemale[i]$  is the number of LSP for the  $i$ -th female,  $i = 1, \dots, F$ .

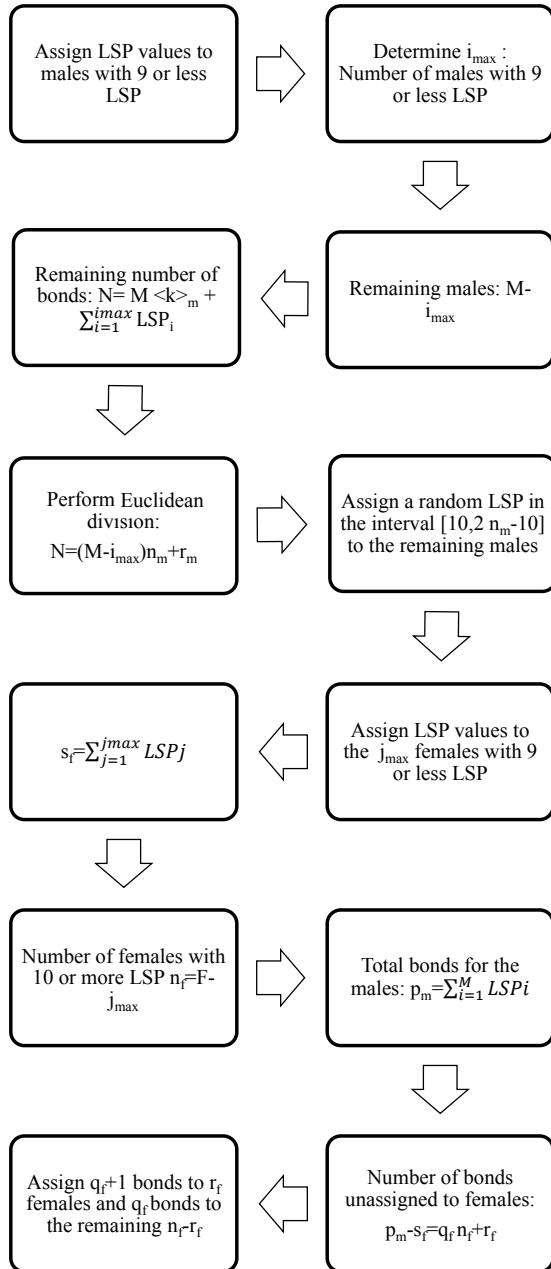


Figure 2.2: Flow diagram for the algorithm corresponding to the assignment of a number of LSPs to every male and female in the network.

These lists will be used to perform the connections of males and females and build the network. Note that, in Table 2.1, there are more females than males with few LSPs (comparing male and female percentages). It implies that there will be few women with a very large number of LSPs. This fact suggests us to start the assignment procedure with women with the largest LSPs. Otherwise, it would be possible that, when we have to assign LSPs of men to a female with a large number of LSPs, there will not be enough men with free sexual partners to be assigned and, for this female, it would be impossible to satisfy the condition that the degree of each node was the number of its LSP.

The assignment of partners was carried out by considering a principle of psychological similarity [29] or assortativity. Hence, we are going to define a weight function assuming that: women with few LSPs usually match men with few LSPs; people with four or more LSPs use to join with people with four or more LSPs; and couples where one of them has a large number of LSPs and the other few LSPs will be uncommon. Then, for the woman  $i$  and the man  $j$ , we define the following weight function:

$$\pi(i, j) = \begin{cases} |kFemale[i] - kMale[j]| & kFemale[i], kMale[j] \leq 4 \\ 0 & kFemale[i], kMale[j] > 4 \\ 100 & \text{otherwise} \end{cases},$$

$$+|AgeFemale[i] - AgeMale[j] - 1.8|. \quad (2.2)$$

The combined weight function, which takes into account the age difference of the partners,  $|AgeFemale[i] - AgeMale[j] - 1.8|$  is defined in this way because some studies show that the average age difference among the members of a couple in Spain is 1.8 years [30].

The MSM population (around a 3.88 % of the total male population in Spain) can also be incorporated into the model, but, in this subpopulation, the connectivity would be larger than the heterosexual network. The MSM population would also be connected with the heterosexual one by links with women in such a way that every MSM individual has a link with a woman with five or more contacts [31]. The assignment of links is then performed by the a Greedy Randomized Adaptive Search Procedure (GRASP) algorithm [32, 33]. Details about the construction of the whole network have been provided in previous studies [31, 34].

## 2.4 The Dynamics of HPV Transmission in the Sexual Network

To implement the simulation of the transmission of HPV among the individuals in the network, a standard epidemiological model with susceptible and infected states and three types of infections (HR, LR and co-infection) was carried out. The epidemiological model is defined by the following parameters:

- We need some probabilities to determine if a sexual partner is going to produce a contagion of another partner in a given time stage. These parameters are different for each age group: 14–17, 18–29, 30–39 and 40–65. Notice that this means that the probability of contagion depends upon the age group of the members of the relationship. Moreover, the probability of connection of these members in the network is also age-dependent as proposed in Equation (2.2). The values of these probabilities are determined in the process of the model fitting.
- Average time an individual infected by a HR HPV clears the infection and recovers.
- A similar parameter for clearing the LR HPV infection.
- If a partner produces the contagion of his/her partner, we need another four parameters to determine if the high or low risk HPV infection is transmitted from man to woman and vice versa.

Another additional parameter is necessary to generate the network. This is the average number of LSPs for men (parameter  $k$ ). Simulations are run by generating a network and carrying out a large number of epidemic evolution time-steps starting with a number of individuals infected by both HPV types as given by the CLEOPATRE study [5]. After the warm-up period, we obtain a stable situation and we can proceed with the calibration by comparing the model predictions with real data and deducing the most probable values of the set of parameters.

We have used a calibration procedure using the Particle Swarm Optimization (PSO) algorithm. The prevalence data for each age group is listed in Table 2.2:

Table 2.2: Prevalence of HR- and LR-infected women per age groups from the CLEOPATRE study [5]. Co-infections are included in both HR- and LR-infected, mean and 95% confidence intervals.

<b>Women</b>	<b>HR-Infected</b>	<b>LR-Infected</b>
18–29 y.o.	24.10%, [21.33%, 26.98%]	6.36%, [4.71%, 8.07%]
30–39 y.o.	11.01%, [7.54%, 15.09%]	1.26%, [0.0%, 3.14%]
40–64 y.o.	5.96%, [4.29%, 7.8%]	2.37%, [1.22%, 3.68%]
18–64 y.o.	16.23%, [14.52%, 17.97%]	4.41%, [3.42%, 5.45%]

Note that the network building and the transmission parameters involve randomness and uncertainty due to the random processes used in the network building and the transmission dynamics of the HPV. This fact is going to be taken into account in the calibration and simulation.

Due to the randomness of the simulations, we will show the average and confidence intervals among the 30 simulations per every scenario that we have considered above. This number was chosen as a compromise among efficiency of the method and computational feasibility. We have found that, with these runs, we obtain reasonable parameters in many of the simulations. Of course, it would be useful to increase this number, but this cannot be achieved, in a sensible computing time, with the computational resources that we devoted to the task. For example, a single run takes 162 hours per run for a network of 500,000 nodes and 256 hours in the case of 750,000 nodes (in a single processor of a Sandy Bridge platform). Simulations were run on a multi core platform with 64 processors and 500 GB RAM and every processor was assigned with the computation of a simulation for a given set of parameters.

## 2.5 Decline of infections

We call  $I$  the number of infected women of LR HPV 6 and/or 11 just before the starting of the vaccination campaign; we call  $V = (v_1, \dots, v_N)$  to the number of infected women of LR HPV 6 and/or 11 every month from the starting of the vaccination program until the end of the simulation. Then, the vector

$$100 \times \left(1 - \frac{v_1}{I}, \dots, 1 - \frac{v_N}{I}\right) \quad (2.3)$$

is a measure of the percentage of decline of number of infected women of LR HPV 6 and/or 11 after the beginning of the vaccination campaign. This will also be applied to men and MSM.

In order to compare GW data given in [1] with our model, results referred to infected women of LR HPV 6 and/or 11, we should take into account that, whether a fixed proportion of HPV 6 and/or 11 infected individuals develops warts, the percentage of decline in warts and in infected women of LR HPV 6 and/or 11 will be comparable.

Another important issue for the natural history of the disease is the persistence of the infection [35]. Our model does not consider the persistence “*a priori*”, but we derive the cases of genital warts from the number of cases of infected individuals by taking this data into account.

# Chapter 3

## Validation of the model

To check the reliability of the model, we simulated the HPV vaccination campaign carried out in Australia [1], and compared them with the actual impact published [1]. In 2007, Australian health authorities started a vaccination program for 12–13 year-old girls with a coverage of 73% (83% in the first dose, 80% in the second dose and 73% in the third dose). In addition, from 2007 to 2009, there was a catch-up vaccination program for women aged 13–26 with a decreasing coverage with age until 52% in women aged 20–26. Their results can be summarized as follows [1]:

- Two years after the vaccine was introduced, the proportion of genital warts diagnosed declined by a 59% in vaccine eligible young women aged 12–26 years in 2007, and by 39% in men of the same age.
- No significant decline was observed in women or men older than 26 years old, non-resident young women, or men who have sex with men.

As in [36], two different scenarios were considered to be simulated:

- Scenario 1: vaccination of 83% of the 14 year-old girls (or younger girls) plus a catch-up with coverage 73% for 14–26 year-old women.
- Scenario 2: vaccination of 73% of 14 year-old girls (or younger girls) plus a catch-up with a vaccination coverage of 52% for 14–26 year-old women.

These simulations represented the upper and lower bounds of the scenario implemented in Australia.

### 3.1 How to measure the decline

We call  $I$  the number of infected women/men/MSM of HPV just before the starting of the vaccination campaign; we call  $V = (v_1, \dots, v_N)$  to the number of infected women/men/MSM of HPV every month from the starting of the vaccination program until the end of the simulation. Then, the vector

$$100 \times \left(1 - \frac{v_1}{I}, \dots, 1 - \frac{v_N}{I}\right) \quad (3.1)$$

is a measure of the percentage of decline of number of infected women/men/MSM of HPV after the beginning of the vaccination campaign. We can apply this procedure to find out the decline of LR HPV or HR HPV infected people.

Also, we should note that if a fixed proportion of LR HPV infected individuals develops genital warts, then the expression of the decline given by (3.1) may be an estimation of the decline of genital warts cases, taking into account the natural delay between the infection and the apparition of the warts. Analogously, it would be valid for HR HPV infected individuals and the cancer cases.

### 3.2 Does our model return similar values to those in [1]?

In this section, we are going to show figures about prevalence or decline of the percentage of women, men and MSM infected of LR 6/11, the HPV type responsible of 90% of genital warts. Taking into account that genital warts, in average, use to appear 6 months after the infection, the figures about prevalence or decline will be a good estimation of the prevalence and the decline of genital warts.

Figure 3.1 shows the percentage of women, men and MSM aged 1426 infected of LR 6/11 after starting the vaccination program in both simulated scenarios. We can see the fast decrease for women and men in both scenarios from the very beginning. MSM, in average, have a slow decreasing.

In Figure 3.2, we have plotted the same data as in Figure 3.1 but from another point of view: the average percentage of decline of women and men infected of LR HPV 6 and/or 11. As the vaccination program progresses over time, the percentage of decline obviously grows. After 2 years, the model shows a decline of

- Scenario 1: 72.0% with CI 95% [67.7%, 76.5%] for women and 38.9% with CI 95% [32.0%, 45.5%] for men.

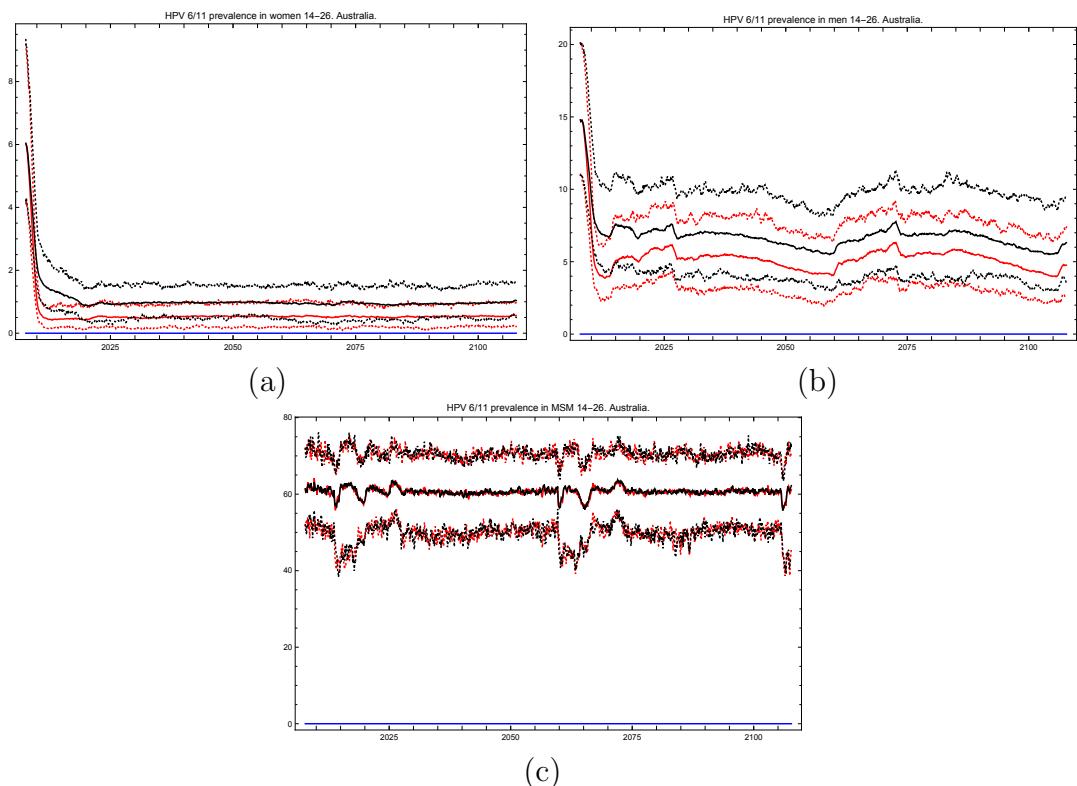


Figure 3.1: Percentage of women (a), men (b) and MSM (c) aged 14–26 infected of LR HPV 6 and/or 11 after the implementation of the vaccination program. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. We can see the fast decrease for women and men in both scenarios from the very beginning. However, there is no effect on MSM.

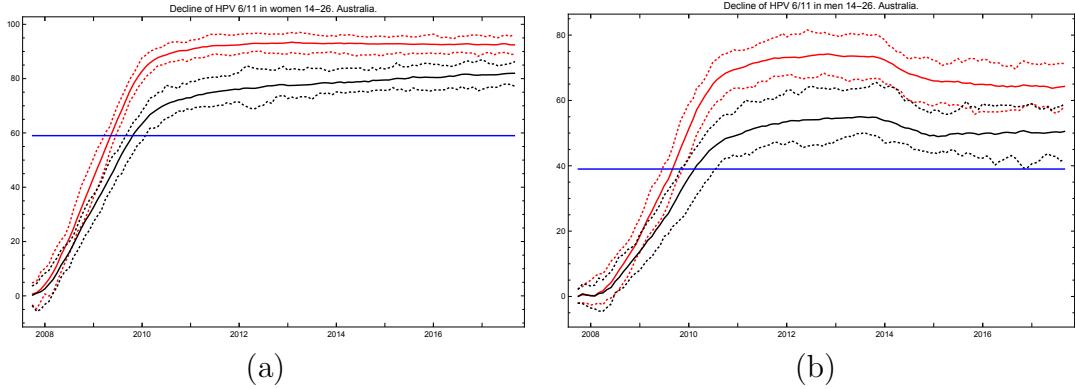


Figure 3.2: Percentage of decline of women (a) and men (b) aged 14-26 infected of LR HPV 6 and/or 11 (and consequently of genital warts) after the implementation of the vaccination program. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. After 2 years, the model shows a decline of 72% for Scenario 1 and 54.8% for Scenario 2, in average, for women and 38.9% for Scenario 1 and 27.7% for Scenario 2, in average.

- Scenario 2: 54.8% with CI 95% [48.5%, 59.0%] for women and 27.7% with CI 95% [21.3%, 34.5%] for men.

Australian reported levels of decline (59% in women and 39% in men aged 14-26) will be reached by the model after

- Viruses, 9(10):30 years with CI 95% [1.5, 1.75] for women and 2.0 years with CI 95% [1.75, 2.16] for men,
- Scenario 2: 2.1 years with CI 95% [2.0, 2.33] for women and 2.42 years with CI 95% [2.08, 2.83] for men.

No significant impact on the rate of infection was observed in men aged 27-64 and women 2 years after the implementation of the vaccination program (Figure 3.3) and the same in women agreeing the observations reported in [1]. It can be explained by the fact that, usually, individuals have sexual intercourses with people more or less the same age.

Then, our model predict figures close to the ones given in [1].

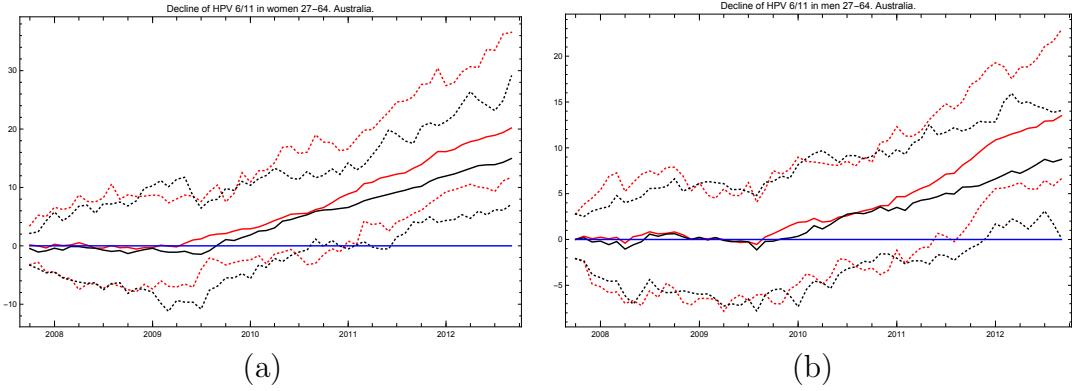


Figure 3.3: Percentage of decline of women (a) and men (b) aged 27-64 infected of LR HPV 6 and/or 11 (and consequently of genital warts) after the implementation of the vaccination program in both scenarios. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. Notice that, in average, no significant decline appears in the 5 years after the implementation of the vaccination program.

### 3.3 Study of the herd immunity effect over HPV LR infection

The herd immunity effect in both scenarios is shown in Figure 3.4 for women, men and MSM. Notice that, in men and MSM, any decline is due to herd immunity. The decline in the whole female population appears when the lines representing their decline are over the vaccination lines (green for Scenario 1 and orange for Scenario 2) also shown in this figure. We see that the herd immunity effect starts after

- for women
  - Scenario 1: 0.58 years with CI95% [0.0, 22.1].
  - Scenario 2: 0.58 years with CI95% [0.0, 22.1].
- for men
  - Scenario 1: 0.0 years with CI95% [0.0, 0.83].
  - Scenario 2: 0.0 years with CI95% [0.0, 0.83].

The herd immunity effect starts very quickly for men. For MSM, there is not a clear herd immunity effect because the decline is stable over the time. For women, we can see that, practically, the CI95% decline lines are over

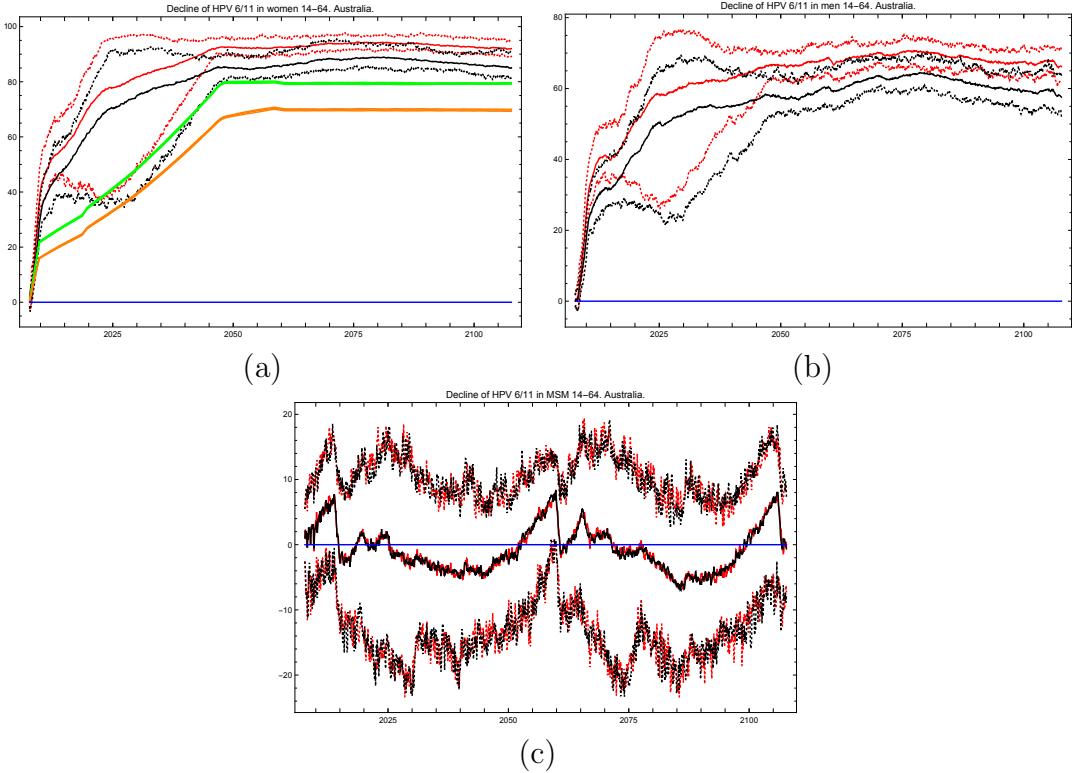


Figure 3.4: Percentage of decline of women (a), men (b) and MSM (c) aged 14–64 for the vaccination program in Australia. The red lines correspond to the average and 95% confidence interval for Scenario 1 and the black lines to Scenario 2. In the figure (a), green and orange lines correspond to women vaccination percentage for Scenario 1 and 2 respectively. Notice that the herd immunity effect contributes to the decline in the number of infections in men and the decline in the number of infections for unvaccinated women. This latter can be seen when the decline lines are over the vaccination line. However, any herd immunity effect can be seen in MSM.

the vaccination lines in both scenarios. This means that, in the worst case, only the vaccinated women will be protected and in the best case, almost all women will be protected by vaccination or by herd immunity effect.

Notice that the herd immunity effect is very clear within the CI95% both for women and men, but it does not appear in the MSM population. In the best case scenario, the MSM subpopulation achieves a constant protection level of 10% – 15%. This could be attributed to the way in which the MSM individuals are connected: with a very large number of LSPs among them and some casual links with women with large LSPs.

## Chapter 4

# Calibrating the Sexual Contacts Network

We already have performed a calibration published in [37]. This calibration allowed to reproduce the Australian scenario [36] where 12-13 years old girls were vaccinated and a catch-up vaccination on women aged 14-26 years old during two years was done. Two years after the vaccine was introduced, the proportion of genital warts diagnosed declined by a 59% in vaccine eligible young women aged 12-26 years in 2007, and by 39% in men of the same age.

However, the calibration of this kind of random computational models is an open problem and several issues about fitting computational models to data with uncertainty have to be addressed. For instance,

- the fact that, for the same set of parameters, different realizations usually return different outputs, and consequently, one realization may fulfill the fitting requirements and another realization does not,
- the determination of an appropriate measure of goodness-of-fit,
- to find model parameters agreeing the values appearing in the medical literature in a reliable and reproducible way,
- adaptation of the optimization algorithms to these above issues and the available resources,
- the determination of the best parallel implementation of the optimization algorithms, in terms of quality of the solutions and computational efficiency of the program.

In the remainder of this section, we are going to describe thoroughly the procedure we propose to approach the above issues.

## 4.1 Description of the model

In the papers [37, 36], we described how to build lifetime sexual partners (LSP) networks. These networks are built with the main aim to reproduce the demography in Spain, [38] and data about the LSP, for men and women, and for age groups 18 – 29, 30 – 39 and 40 – 65, presented in a survey about sexual habits in Spain [6] and collected in Table 4.1.

MALES						
Age	0 LSP	1 LSP	2 LSP	3 – 4 LSP	5 – 9 LSP	LSP $\geq 10$
14 – 29	0.107	0.207	0.131	0.225	0.168	0.162
30 – 39	0.027	0.225	0.128	0.21	0.17	0.24
40 – 65	0.019	0.268	0.14	0.193	0.163	0.217
FEMALES						
Age	0 LSP	1 LSP	2 LSP	3 – 4 LSP	5 – 9 LSP	LSP $\geq 10$
14 – 29	0.138	0.43	0.186	0.158	0.056	0.032
30 – 39	0.029	0.501	0.168	0.177	0.077	0.048
40 – 65	0.017	0.652	0.138	0.118	0.039	0.036

Table 4.1: Proportion of males and females per number of LSP per age group [6]. Note that the sum of the rows are 1.

There are a lot of possible combinations to create networks fulfilling the above requirements. This fact makes that the built networks contain uncertainty in the sense of the randomness of the building process and the different shapes these networks may have.

Then, over the LSP network, we describe the dynamics of the HPV [37, 36]. We divide the nodes into the age groups 14 – 17, 18 – 29, 30 – 39, 40 – 65. In order to describe properly the transmission dynamics of HPV on the LSP networks, it is necessary to include 11 model parameters that will govern the aforementioned HPV dynamics:

1. Average number of men LSP,
2. Global probabilities in order to determine if the existence of a LSP implies sexual intercourses in the current time step per age group 14 – 17, 18 – 29, 30 – 39 and 40 – 65 (4 parameters),
3. Average time an individual infected by a high (low) risk HPV clears the infection and recovers (2 parameters)
4. Probability that a woman (man) infected of high (low) risk HPV transmits it to his/her partner in a sexual intercourse (4 parameters).

Most of the above parameters have been studied in the literature and there are some estimations we have to take into account:

- The average number of men LSP: it is around 8 [39]. For this parameter, our search will be in the interval [7, 10].
- The time for clearing the infections due to HPV HR, for both men and women: in [12], the authors say that the mean duration of HPV 16/18 infection is 1.2 years; in [40], the mean duration of HPV 16 is 12.19 months (7.1618.17); in [41] the duration clearance varies, in average, between 6.5 months to 11.8 months. Thus, we considered the interval [0.8, 1.2] years.
- The time for clearing the infections due to HPV LR, for both men and women: in [12], the mean duration of HPV 6/11 infection is 0.7 years; in [40], the mean duration of HPV infection is 7.52 months (6.808.61) for any HPV; in [41] the duration clearance varies, in average, between 6.2 months to 11.7 months. Thus, we considered the interval [0.5, 1] years.
- In [12], the authors estimated the probability of HPV infection transmission per partnership and by type and, as in [5], this probability is higher for transmission from males to females (0.8) than that for transmission from females to males (0.7). Given that these data come from estimations (not surveys) and after some runs of our model, we are going to be more flexible and consider the probability interval [0.2, 0.6] for LR transmission and [0.5, 1.0] for HR transmission, given that, women transmit less than men.

Now, we are able to define the dynamics of HPV. Setting the time step in a month, for every month:

- For each node  $i$ 
  - Add a month to the age of  $i$ .
  - If  $i$  is already infected by any HPV, we have to see if this infection clears, depending on the type of infection and the current infection time.
  - We have to see if  $i$  can be infected by a type of HPV he/she is not infected (HPV LR, HPV HR or both).

Hence, we have implemented in C++ a simulator that, given the above described model parameters, it builds a LSP network and performs a realization of the transmission dynamics of HR and LR HPV.

Note that the transmission parameters involve certain probabilities. Then, in order to see if a contagion has been carried out by a sexual intercourse, we simulate this by generating a random number and checking if it is less than the corresponding threshold given by the model transmission parameters. Therefore, the randomness is included into the model in a natural way producing uncertainties on the model output that has to be quantified. Then, there are two sources of uncertainty: the LSP network building and the simulation of HPV dynamics.

Thus, our goal is to find the model parameters in such a way the computational model output is close, in a way we will describe later, to the data in Table 4.2. In this table, we recall the data presented in [5]<sup>1</sup> related to the percentage of women HR- and LR-infected per group ages 18–29 and 18–64.

Women	HR-infected	LR-infected
18 – 29 y.o.	24.10%, [21.33%, 26.98%]	6.36%, [4.71%, 8.07%]
18 – 64 y.o.	16.23%, [14.52%, 17.97%]	4.41%, [3.42%, 5.45%]

Table 4.2: Prevalence of HR- and LR-infected women per age groups [5]. Mean and 95% confidence intervals. Co-infection is considered in both, HR and LR.

## 4.2 Description of our resources

### 4.2.1 Computers

All the realizations are going to be run on two computers with 64 cores on 8 Xeon Sandy Bridge E5-4620 running at 2,2 Ghz, with 16 MB of cache memory and 512 GB RAM memory<sup>2</sup>. The operating system is Ubuntu Server 16.04 LTS.

---

<sup>1</sup>In [5], data related to age groups 30 – 39 and 40 – 65 are available. However, these data has been discarded because there are very few and non-significant reported cases.

<sup>2</sup>Both computers are not exactly the same. There are some minor hardware differences.

### 4.2.2 Distributed computing environment

We also have deployed a distributed computing environment called Sisifo. Sisifo is a client-server based system designed to allow a problem to be solved using distributed computation. Sisifo is able to assign tasks to a set of personal computers (PCs), wait for the tasks to be completed and collect the results for further analysis. Sisifo is made with simplicity as main aim, giving as a result a system that requires almost no maintenance, needs very little configuration time, and can be deployed in just a couple of hours.

The Sisifo Server keeps listening for request of the clients. The Server has stored one or more executors, a set of problems to be solved in the *Problem files* folder, and the solutions sent in the *Result files* folder.

The Sisifo Client is a program stored in one or several PCs that connects to the server, and asks for a *work packet*. This work packet is composed of two elements: a text file containing the model parameter values and the simulator executable file. The Client, once the work packet is received, executes a realization using the model parameters stored in the text file. When the realization finishes, a solution file is generated, returned to the server and dropped in the *Results files* folder. More details about how Sisifo works can be found in [42].

In our case, the Sisifo clients are going to be located in each one of the 64 cores of the Sandy Bridge computer. The Sisifo server is located in a regular PC with Windows 7.

### 4.2.3 Random Particle Swarm Optimization (rPSO)

Using Python3 [43] and *Mathematica* [44], we have implemented an asynchronous version of rPSO adapted to the *Sisifo* computing environment. To do that, first, we recall the random PSO (rPSO) algorithm appearing in [45] applied to the optimization of a function  $F$ .

**Step 1.** Initialization.

- Initialize  $N$  particles  $p_1, \dots, p_N$  chosen randomly in the parameter space.
- Initialize randomly their velocities  $v_1, \dots, v_N$ .
- Evaluate the fitness of all the particles  $F(p_1), \dots, F(p_N)$ .
- Define the individual best fitness as  $p_i^{best} = p_i$ ,  $i = 1, \dots, N$  and the global best fitness  $p_{global}^{best}$  as the  $p_i^{best}$  which fitness is optimum.

**Step 2.** Modify the particle velocities based on the previous individual best and global best positions:

$$v_i^{new} = \omega v_i + \psi_1(p_i^{best} - p_i) + \psi_2(p_{global}^{best} - p_i), \quad i = 1, \dots, N,$$

where  $\omega$  is a random value in  $[\frac{1}{4}, \frac{3}{4}]$ ,  $\psi_1$  is the exploitation rate and  $\psi_2$  is the exploration rate, both randomly generated in each iteration as a number in the interval  $[0, 1.5]$ .

**Step 3.** Update the particle locations:  $p_i = p_i + v_i^{new}$ ,  $i = 1, \dots, N$ .

**Step 4.** Evaluate the fitness of all the particles  $F(p_1), \dots, F(p_N)$ .

**Step 5.** Update the individual best fitness  $p_i^{best} = p_i$ ,  $i = 1, \dots, N$  and the global best fitness  $p_{global}^{best}$ . Go to Step 2.

The above algorithm can be adapted to *Sisifo* computing environment if, using the *Sisifo Server*, the computation of the fitness of the particles is distributed among the *Sisifo Clients*.

However, in a typical PSO procedure, including rPSO, the set of particles is updated once the fitnesses of all the particles have been calculated. This means that, until all the fitnesses have not been evaluated and Step 4 is not completely finished, the particles cannot be updated and new evaluations cannot be performed. Then, in every iteration of rPSO, scenarios where some *Sisifo* clients have finished their evaluations and are idle while other *Sisifo* clients are still performing their evaluations are usual. In these scenarios, we have an under-use of the *Sisifo* system.

In order to avoid the under-use system drawback, we propose the implementation of an asynchronous version of rPSO in such a way that when the fitness of a particle has been evaluated (Step 4), this particle is updated (Steps 5, 2 and 3) without waiting for the evaluation of the remainder particles, considering the current existing global best and its individual best particles. This way, we modify rPSO algorithm parallelizing Steps 2, 3, 4, 5 and sharing the updates of the global best particle.

We show in Figure 4.1 how we set the asynchronous rPSO in the *Sisifo* environment. As we can see, the rPSO procedure generate the problems and write them into the *Problem files* folder to be processed by the *Sisifo Server*. Once the evaluation has been carried out and the solution file is in the *Result files* folder, rPSO reads the content of the solution file, analyses it and calculates the fitness. With this information, rPSO is able (Steps 2 and 3) to generate a new problem file.

When the procedure starts running, with the initialization of the particles (Step 1) we create their corresponding problem files in the *Problem files*

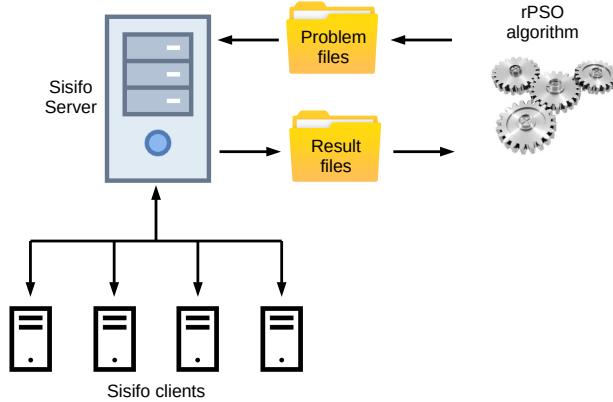


Figure 4.1: Introduction of the asynchronous rPSO algorithm in the Sisifo environment. rPSO manages the generation of new problems and put them in the *Problem files* folder and the reading and processing of the solution files located in the *Result files* folder.

folder. The *Sisifo Server* detects new problem files and distribute them among idle *Sisifo Clients*. These clients carry out the realizations. When a Sisifo client ends its task, a results file is generated and sent to the *Sisifo Server* that drops it on the *Result files* folder. Every time a new results file appears in the *Results files* folder, the asynchronous rPSO, in Step 4, reads the data from the results file and calculates the fitness. Then, updates the velocity taking into account the current existing global best and its individual best particles (Step 2), updates the particle (Step 3) and with the new model parameters creates a new problem file in the *Problem files* folder. And so on.

#### 4.2.4 Fitness function and some features included in the rPSO

There are some features included in our version of the asynchronous rPSO algorithm that we have to describe.

1. In the Step 3, we include the possibility to discard the new particle and generate a new one randomly with 10% probability. If it is not discarded, with 10% probability we apply to the new particle a mutation.
2. The fitness evaluation is made as follows:
  - we start the model and after a warm-up time of 400 simulated months, we get a stabilization of the model output;

- we take the model output from month 401 to 500 for the subpopulations in the data of Table 4.2, i.e., percentage of women HR- and LR-infected per group ages 18 – 29 and 18 – 64;
  - then, for each subpopulation, we calculate the maximum and the minimum of these portions of the model output and we see if these maximum and minimum are inside the corresponding data 95% confidence interval;
  - if this happens, the fitness is zero;
  - otherwise, the fitness is the sum of the distances from these maxima and minima to the corresponding 95% confidence interval of the data.
3. The definition of the fitness function may provide the same fitness values for different particles. Therefore, we are going to store all the particles with the same best fitness and, in Step 2, we chose  $p_{global}^{best}$  randomly among the stored particles.
  4. As we mentioned before, physicians around the world have published estimations of most of the model parameters [39, 12, 40, 41]. It is clear that we have to respect their estimation in our fitting procedure avoiding that some model parameters overpass the estimation intervals. Furthermore, finding model parameters in the range of the estimations, gives credibility to our model. Therefore, when a model parameter is less than 1% closer to an extreme of the interval provided by the physician estimations, we discard this value and it is substituted by a random value inside the interval.

It is worth to note that the above points 1, 3 and 4 will allow us to explore extensively the whole space of parameters avoiding the accumulation of particles close to the borders.

**Remark 4.2.1** *At this point, we want to remark that to evaluate the fitness function, we do not need the model output for all the subpopulations. Only is necessary the output corresponding to HR- and LR-infected women in the group ages 18 – 29 and 18 – 64. Then, when a realization is carried out, we retrieve and process from the results file the data corresponding to the above subpopulations from the month 401 to 500, write them in a row and this row is stored as the result of the realization to be used later.*

*Also, in order to compare properly the model output row with the data, we build the data vector mean, percentile 2.5 and percentile 97.5, repeating 100 times the 4 values of each we have in the Table 4.2 and write them in a row.*

### 4.3 Selecting an optimal number of particles to calibrate the HPV large network computational model with rPSO

As we have mentioned above, on previous implementations of the parallel rPSO, we realized that we can not affirm that *the higher the number of particles, the better the quality of the solution*. Moreover, we also detected that, due to our asynchronous parallel implementation, sometimes the use of more processors does not mean lower execution times. We would need more experiments to detect the correct cause of the delays. However, in this case is more practical to investigate what the optimal combination of particles is to achieve the best quality with the lowest number of processors.

To perform this test, we are going to build HPV network models with 50,000 nodes. We run 5 repetitions of the same problem with different number of particles. Times are shown in seconds. Thus, the question is, what is the quality of the solutions for different number of particles? Table 4.3 shows relevant information regarding the quality of the solutions for different number of particles and 5 runs of each configuration. Each run carried out 1200 particle evaluations. We measure the quality of the solutions on Table 4.3 by counting the number of fitness values equal to zero in the 5 runs (*Total # 0s*) and on average (*Avg #0s*). In addition, we also recorded the iteration at which the first zero appear on average (*Avg First*), from the last population of particles Fitness of the best, worst and average, averaged over five runs (*Best at end*, *Worst at end* and *Avg at end*). And regarding executions time we show on Table 4.3 information about the average total execution time for 1200 evaluations (*Total Time*), the worst execution time for one particle (*Worst t*) and the best execution time for one particle (*Best t*). Red color indicates the worst configuration, bold letters indicate the best and blue color the second best configuration.

As we can see, 25 and 30 particles are the preferred configurations, since we obtained the higher number of zeros in total and on average and the lower executions times. Since the results of 25 particles shows a very good run with 7 zeros and also a very good execution time, we can conclude that the configuration with 30 particles should be selected bearing in mind both quality and execution time. Results on total number of 0s and total time were statically significant with  $p$  value of 0.1 after an ANOVA analysis.

# Part.	Total #0s	Avg #0s	Avg First	Best at end	Worst at end	Avg at end	Total Time	Worst t	Best
25	18	3.60	531.00	0.015486	0.429090	0.122573	8422.00	148.60	107.
30	28	5.60	594.00	0.003785	0.445587	0.131193	6908.00	186.60	112.
35	8	1.60	834.80	0.010211	0.576688	0.149215	7105.00	250.60	119.
40	16	3.20	692.80	0.003332	0.524417	0.132628	8808.60	389.40	144.
45	14	2.80	790.40	0.006458	0.639775	0.149670	10004.00	502.00	177.
50	11	2.20	796.00	0.005168	0.658020	0.139089	11124.40	648.40	220.
55	2	0.40	1171.20	0.004330	0.518894	0.137088	11905.80	775.00	207.
60	12	2.40	765.00	0.002153	0.546259	0.132464	13787.80	857.00	245.
64	5	1.00	985.40	0.002763	0.551228	0.140977	18770.00	1041.20	379.

Table 4.3: Analysis of the quality and execution times of different rPSO configurations. Results of 1200 evaluations for different number of particles on the rPSO process (# Part.).

## 4.4 Improving the exploration of the rPSO

As we explained on Section 4.2.4, we established a procedure respect to the intervals for the parameter values proposed by physicians. Some algorithms such as differential evolution [46] allows to overpass these limits in the search of a good combination of parameters. However, this is not a good idea when implementing our parallel rPSO. First, remember that our parallel version is asynchronous and the updating of the particles is made once every particle is evaluated. Allowing particles with parameters close to the limits trespass them, could lead to a chaotic search. Second, parameters out the defined bounds have not a medical meaning. Therefore, when a model parameter is less than 1% closer to an extreme of the interval provided by the estimations, we discard this value and it is substituted by a random value inside the interval. This allows also to made a deeper and more efficient exploration of the search space.

Figures in Table 4.4 show the exploration performed in some of the parameters of the model (duration and contagion parameters). Figures represent the values of the parameters during the 20 executions of rPSO algorithm with 500 realizations each one. X axis represents the realization number and Y axis the value of each parameter. We can see that we find points on most of the search space.

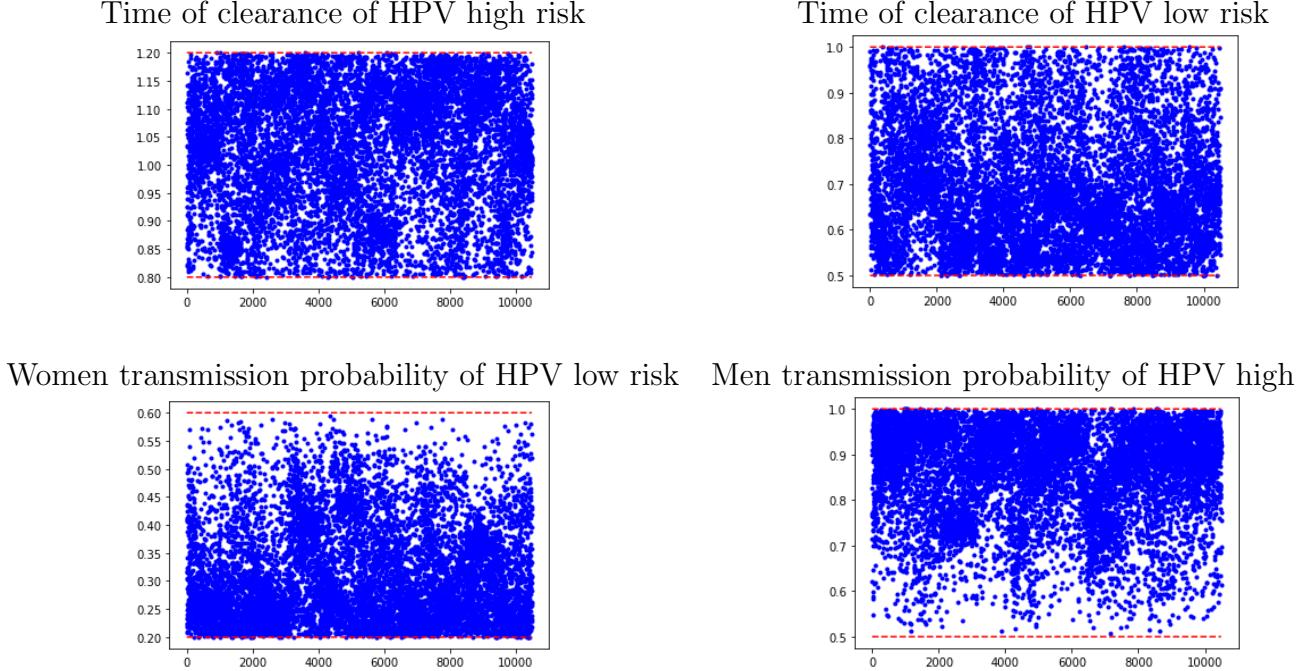


Table 4.4: Upper figures: Time of clearance of HPV high risk and low risk, both for men and women. Lower figures: on the left, probability to transmit low risk HPV if the infected is a woman; on the right, probability to transmit high risk HPV if the infected is a man. Red dashed lines correspond to the bounds of each parameter. We can see that most of the search space is explored in all the cases.

## 4.5 Calibrating the model

Now, we are going to perform the calibration. We use the Sisifo environment with the modified rPSO and 30 particles. The HPV network models will have 100,000 million nodes. In order to guarantee the reproducibility of the realizations, we are going to include in the problem files and store seeds for the generation of the random numbers during the calibration process.

We assume that, initially, data of prevalence from Table 4.2 are not only for women but also for men. Then, we label women and men as infected randomly following these prevalence data. Also, we start running the realization and the first 400 months are used as a warm-up period to stabilize the distribution of infected men and women. Thus, the goal is to calibrate the model parameters in such a way that the model output related to women HR and LR prevalence minimize the fitting function defined in Section 4.2.4.

We have performed 20 different calibrations using rPSO, each one with around 500 realizations and 30 particles. A total of 10,100 realizations of the model were performed with an equivalent sequential total computation time of 161 days.

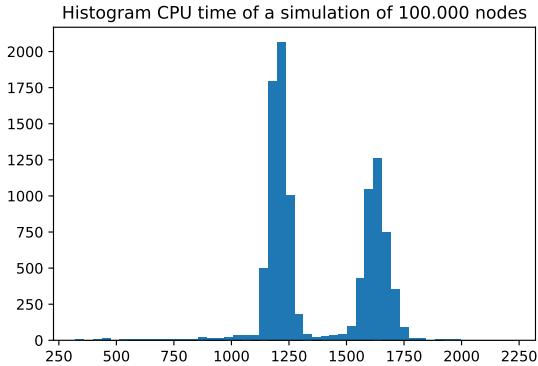


Figure 4.2: Histogram of the computation time of each one of the 10,100 realizations of the model using networks of 100,000 nodes. The average time is 1374.5 seconds, around 23 minutes.

## 4.6 Selecting the 30 realizations that best capture the data uncertainty

Our goal, now, is to find a procedure to select 30 among the 10,100 realizations of the model in such a way that the means and the 95% confidence intervals of these 30 realizations be as much close as possible of the corresponding means and the 95% confidence intervals of the data in Table 4.2. We decided to select 30 because as we saw in Table 4.3, the total computation time is the best for 30 particles running in parallel in the Sandy Bridge computer.

Nevertheless, it would be interesting to reduce the number of eligible realizations to much less than 10,100. In Figure 4.3 we can see the 100 realizations with error less than 0.01, that is, the realizations that almost lie inside the 95% confidence intervals of the data, represented by the blue horizontal dashed lines. If we select 30 among these 100, it is clear that some percentiles of the model will be far from the corresponding percentiles of the data, for instance, the lower parts of the left figures. Therefore, we need to consider realizations with errors greater than 0.01 without forgetting the

objective to reduce the number of eligible realizations.

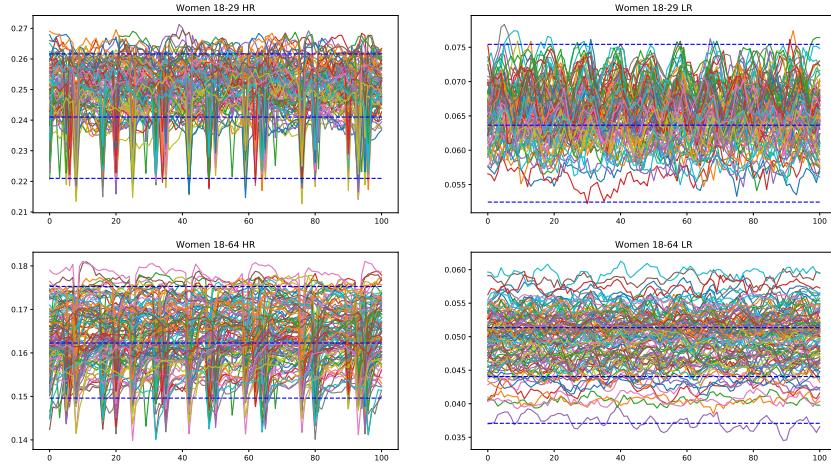


Figure 4.3: Drawing of the 100 model realizations with error less than 0.01 from month 400 to 499.

Thus, we check the number of possible realizations depending on their error. Then, there are 2 realizations with error 0, 100 realizations with error less than 0.01, 392 with error less than 0.025, 1282 with error less than 0.05 and 2607 with error less than 0.075. In Figure 4.4, we draw the 1282 model realizations of with error less than 0.05. Note that the realizations cover completely the 95% confidence intervals of the data.

In order to determine the *best* 30 realizations that capture the mean and the 95% confidence intervals of the data in Table 4.2, we are going to introduce an adapted version of the rPSO algorithm before. Let  $E$  be the realizations considered with  $\text{card}(E) = M$  its number. For instance, for  $E$  the set of realizations with error less than 0.01,  $M = 100$ . Also, for  $E$  the set of realizations with error less than 0.05,  $M = 1282$ . Then  $E(i)$  is the  $i$ -th realization of the set  $E$ . We define the following fitness function  $F$ :

**INPUT:** Set of 30 indexes  $I = \{i_1, \dots, i_{30}\}$ ,  $1 \leq i_j \leq M$ ,  $j = 1, \dots, 30..$

**Step 1.** Select the realizations  $E(i_1), \dots, E(i_{30})$  and calculate the mean, percentile 2.5 and percentile 97.5 of them.

**Step 2.** Calculate the norm of the difference between the mean, percentile 2.5 and percentile 97.5 of the 30 realizations and the data (see Remark 4.2.1), and sum them up.

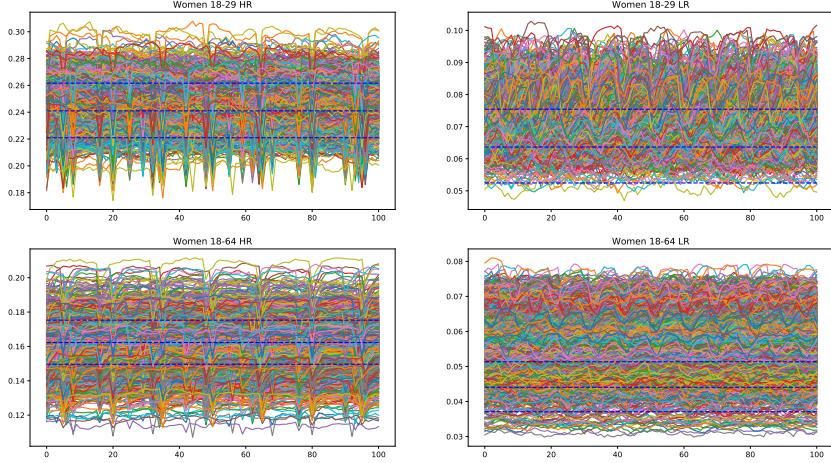


Figure 4.4: Drawing of the 1282 model realizations with error less than 0.05 from month 400 to 499. These realizations cover the 95% confidence intervals of the data, represented by the blue horizontal dashed lines.

Now, the adapted version of rPSO to select the best 30 realizations is

**Step 1.** Initialization.

- We have a set  $E$  with  $M$  realizations.
- Initialize  $N$  index subsets  $S_1, \dots, S_N$  with 30 elements of the set  $\{1, \dots, M\}$  (particles) chosen randomly without repetition.
- Evaluate the fitness of all the particles  $F(S_1), \dots, F(S_N)$ .
- Define the individual best fitness as  $S_i^{best} = S_i$ ,  $i = 1, \dots, N$  and the global best fitness  $S_{global}^{best}$  as the  $S_i^{best}$  which fitness is the minimum.

**Step 2.** For  $i = 1, \dots, N$ , build the new set  $P_i = S_i \cup S_i^{best} \cup S_{global}$ , that is, joining the current particle, its individual best and the global best. After removing index repetitions, the new particle  $S_i$  consists of a random selection without repetition of 30 indexes of  $P_i$ .

**Step 3.** Evaluate the fitness of all the new particles  $F(S_1), \dots, F(S_N)$ .

**Step 4.** Update the individual best fitnesses  $S_i^{best} = p_i$ ,  $i = 1, \dots, N$  and the global best fitness  $S_{global}^{best}$ . Go to Step 2.

Here, we also consider 10% of randomness and 10% of mutation when updating new particles. In this case, mutation consists of changing some of the indexes in the current particle by other randomly chosen indexes, avoiding repetitions.

The above algorithm, in the tests we have run, lasts around 20 minutes for 1 million of evaluations of the fitness function in the Sandy Bridge computer, returning accurate solutions.

We have performed runs for 30, 40, 50 and 60 particles, being  $E$  the realization sets with errors less than 0.025, 0.05 and 0.075. The lowest error has been 0.1166 for the following realizations

$$85, 109, 460, 474, 475, 476, 485, 493, 496, 497, 523, 531, 542, 543, 563, 600, \\ 635, 637, 650, 687, 715, 729, 730, 842, 974, 987, 1058, 1060, 1080, 1238, \quad (4.1)$$

among the 1282 of the set of realizations with error less than 0.05. In Figure 4.5 we draw the selected realizations and in Figure 4.6 we can see the graphical result of the calibration and how resemble the means and 95% confidence intervals.

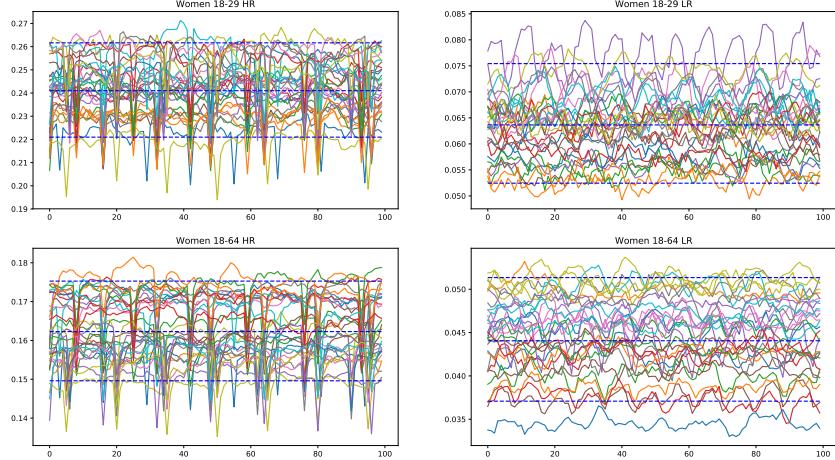


Figure 4.5: Drawing of the 30 selected model realizations from month 400 to 499. These realizations are the ones whose means and 95% confidence intervals resemble the most the data in Table 4.2, represented by the blue horizontal dashed lines.

The mean and the 95% confidence interval of the model parameters corresponding to the 30 selected realizations from (4.1) are given in the Table 4.5.

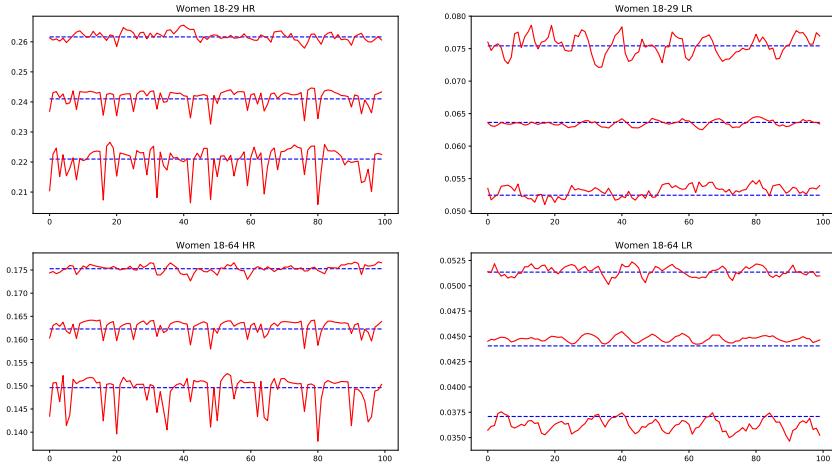


Figure 4.6: Means and 95% confidence intervals of the 30 selected realizations (in red) compared to the means and 95% confidence intervals of the data (blue).

Note that the obtained model parameters are in accordance to the medical parameters appearing in the literature and collected in Section 4.1.

Model parameter	Mean	95% confidence interval
Average LSP men	8.63	[7.15, 9.86]
Average time clearing HR HPV	1.08	[0.88, 1.19]
Average time clearing LR HPV	0.60	[0.52, 0.82]
Probability a woman transmits LR	0.23	[0.21, 0.29]
Probability a man transmits LR	0.28	[0.22, 0.37]
Probability a woman transmits HR	0.81	[0.68, 0.95]
Probability a man transmits HR	0.91	[0.74, 0.97]

Table 4.5: Mean and the 95% confidence interval of the model parameters corresponding to the 30 selected realizations from (4.1).

In the following, we are going to use the sets of parameters corresponding to the realizations (4.1) (and their corresponding seeds) to perform the simulations below.

# Chapter 5

## Making Dynamic Sexual Contacts Network and adding Vaccination Campaigns

### 5.1 Including new features into the model

#### 5.1.1 Introducing age and LSP dynamics into the LSP model

The computational model presented so far is static, that is, once we have assigned the LSPs, they do not change over the time and the nodes do not increase their age. We modulate the possibility of a contagion introducing global probabilities that determine if the existence of a LSP implies sexual intercourses in a given time step per age group 14 – 17, 18 – 29, 30 – 39 and 40 – 65. Also, note that the available data about LSPs come from 2003 and the data of prevalence appears in a study (CLEOPATRA [5]) conducted in 2007 – 2008. The sexual behavior has changed in the last 10 – 15 years and we would like to include it in our model in some way. Nevertheless, newer data about LSPs and prevalence in Spain are not available.

Thus, in order to introduce age and LSP assignation dynamics in a simple way, we are going to *move* the available data over the time. Figure 5.1 shows the constant age distribution that do not varies over the time.

Then, we evolve the ages of the nodes over the time. As long as the nodes do not change the age group, there is nothing to do. However, how do we *recycle* the nodes that turn 65? And how do we transform the nodes in the first age group 14 – 29 when they turn 30?

To answer the first question, we propose to

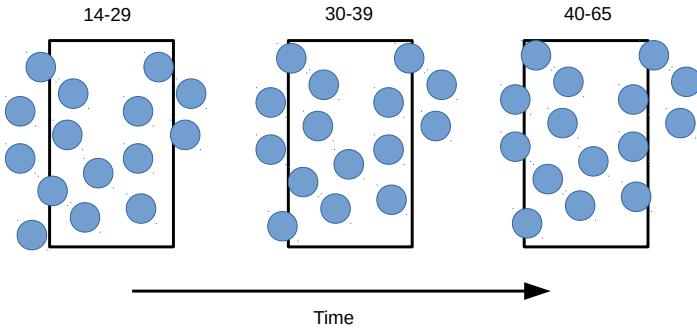


Figure 5.1: Static LSP network. The ages remain constant over the time. The time only affects the HPV contagion/clearing dynamics.

- preserve its sex in order to maintain the proportion males/females,
- assign 14 years old to the node, erase all its LSPs and transform it in susceptible,
- assign new LSPs according to the LSPs of the age group 14 – 29 following the assortativity property.

To answer the second question, we propose to

- preserve its sex and the infectious state,
- erase all its LSPs and assign new LSPs according to the LSPs of the age group 30 – 39 following the assortativity property.

This way, as time goes on, we will have the LSP structure by ages as we can see in the Figure 5.2.

If the node is MSM (male who have sex with males), we perform the above procedure taking into account the division of age groups given in [39], 14 – 19, 20 – 24, 25 – 29, 30 – 39, 40 – 49, 50 – 59 and 60 – 64+.

Taking into account that the global number of LSP in the age groups 40 – 65 is less than in the age group 30 – 39, as the time goes on, the global number of LSP in the whole network will increase and the transmission of HPV will also increase. However, the proposed approach may be considered as very conservative because the sexual behavior change in the last 10 – 15 years seems to result in an increase of the sexual intercourses greater than the number that we can approach with the proposed dynamic network. Nevertheless, the lack of data do not allow us to quantify the mentioned change.

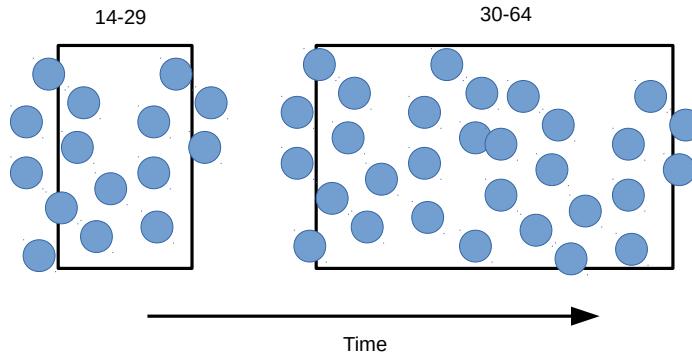


Figure 5.2: Dynamic LSP network. The ages evolve over the time. The age group 40-65 and their sexual behavior disappear when people grow.

### 5.1.2 When should we start the vaccination schedule?

When a realization runs, we use 500 months to stabilize the static network, then, we change to a dynamic network. Thus, a key point arises and it is to decide in which time instant starts the vaccination schedule. We have performed a simulation with the selected 30 sets of parameters with 2300 months (around 191.6 years) where the network turns dynamic from the month 500. In the Figure 5.3 we can see the levels of prevalence predicted by the model for 18-64 years old men and women for HR and LR, over the next years.

The Figure 5.3 shows the mean and 95% confidence intervals of the 30 simulations for the prevalence of men and women for HR and LR. The horizontal axis indicates the month and the vertical axis the percentage of infected. As it can be seen, there are oscillations in the evolution of the prevalence.

The vertical lines correspond to:

- magenta: month 500, the static network turns dynamic;
- blue: months 980, 1580 and 2180, and point out the peaks in the oscillations in the means and the percentiles. Between them, there are 600 months (50 years), the time of a complete generation in the model;
- green: months 1280 and 1880, and point out the valleys. Between the valleys there are also 50 years, and 25 years between a peak and a valley.

In some test runs with vaccination, we have observed that, if we start the vaccination schedule when the prevalence is in the decreasing part of the

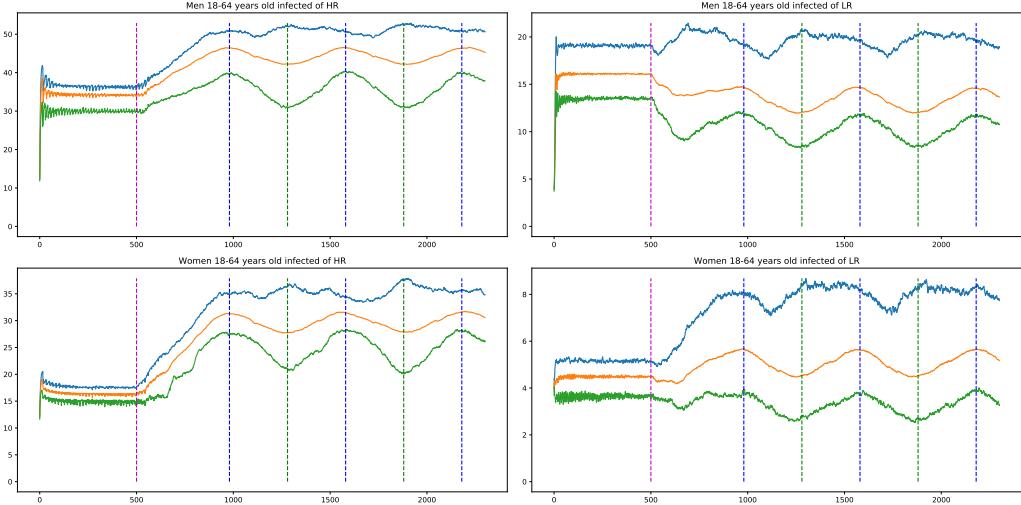


Figure 5.3: Mean and 95% confidence intervals of the prevalence for 18-64 men and women for HR and LR. Vertical dashed lines indicate milestones of interest. Note the oscillations of the prevalence levels. The blue and green vertical lines correspond to the peaks and valleys of the oscillations. The magenta line, points the month 500 when dynamic network starts.

oscillation, the herd immunity appears very much sooner than when we start the vaccination schedule in the increasing part.

If we take into account that in the paper [1], the authors reports extraordinary results in Australia, where two years after the vaccine was introduced, the proportion of genital warts diagnosed declined by a 59% in vaccine eligible young women aged 12–26 years in 2007, and by 39% in men of the same age, we conjecture that, if there are oscillations in the prevalence of HPV over the time, they have have taken advantage of a decreasing part of the oscillation.

Recall that our goal is to determine the appropriate month where start the vaccination schedule, trying to save computations and favoring the apparition of the herd immunity effect as soon as possible, in order to minimize the opposite effect when the oscillation is in the increasing trend if we have vaccinated a large enough number of individuals.

The oscillations in all the cases, men, women, HR and LR, are very similar, as we can see in the Figure 5.3, except, maybe in some upper percentiles. Also, there are similarities from the first peak in valleys and peaks. Therefore, in order to take the maximum advantage of the decreasing trends and to save computation, we are going to select the earliest peak, the month 980, as the starting point for the vaccination.

### 5.1.3 Introducing vaccination

For our simulations, we are going to consider the vaccine GARDASIL9 [47], that protects against the HPV types 6/11/16/18/31/33/45/52/58. The two first (6/11) are LR and the remainder are HR. Thus, GARDASIL9 protects against 90% of genital wart cases and 90% of the cancer cases [48]. Although GARDASIL9 may protect partially against other HPV types, for modelling purposes we assume that it does not happen. Therefore, it would be interesting to introduce changes in order to monitor if a node is infected by HPV types included in GARDASIL9 or not, and then, to simulate accurately the protection effect of the vaccine.

Following the study conducted in [5], if a woman is infected by HPV LR, the probability to be only infected by 6/11 is 34.23%, 63.06% only infected by others than 6/11 and 2.70% to be infected by both. Also, if a woman is infected by HPV HR, the probability to be infected only by 16/18/31/33/45/52/58 is 30.44%, 23.66% only infected by others than 16/18/31/33/45/52/58 and 45.90% by both. Due to the lack of information about men, we will also use the above percentages for men.

Then, before starting the vaccination, we label men and women as infected of HPV LR 6/11 or infected of HPV LR other than 6/11 or infected of both, following the above percentages. Analogously, we label men and women as infected of HPV HR 16/18/31/33/45/52/58 or infected of HPV HR other than 16/18/31/33/45/52/58 or infected of both.

Once these assignments are done, we continue with the HPV transmission dynamics including the vaccination, taking into account the new labels we included. If a node has been vaccinated, it can be infected by the types of HPV different to those that GARDASIL9 protects and, in this case, they will never be infected of 6/11 nor 16/18/31/33/45/52/58. If a node has not been vaccinated, can be infected by any HPV type.

The assumed effectiveness of the vaccine is 96.5%.

### 5.1.4 Introducing vaccination loss protection

In the previous section we assume that the protection of GARDASIL9 is forever. In fact, until now, people vaccinated by GARDASIL (previous version of GADASIL9) do not have experienced any loss in the protection. But this does not mean that it could happen in the future.

In fact, we want to simulate the worst possible scenario, that is, the sudden drop to zero of the protection. To simulate this possibility, we will introduce a new parameter that represents the time after the vaccination where the protection is complete. Therefore, after this time, the vaccinated

individual will behave as a non-vaccinated individual.

### **5.1.5 Introducing variations in the vaccination coverage**

One of our goals is to simulate scenarios where variations in the vaccination coverage have been occurred and we want to study the effect in the global protection against HPV due to these variation of the coverage. To simulate these scenarios, we will include into the model vectors of coverage, indicating the vaccine coverage every month, and vaccinating the people following these variable coverages.

# Chapter 6

## Conclusions

The random network of sexually transmitted HPV including up to 100,000 nodes, was developed to fit the data of surveys concerning the number of sexual partners throughout life [37, 36]. Standard continuous models are insufficient to accurately predict transmission because they do not account for the individual to individual transmission of the infection, the role of hubs in disseminating the virus through the rest of the population and neither the vaccination campaigns targeting specific groups of individuals.

This network has successfully been applied to the stable state of infections by LR and HR HPV genotypes in Spain [37]. In this study we mimicked the results found in the HPV vaccination campaign in Australia [1], and showed very reliable results.

Models based upon continuous differential equations predict a slower decrease in the number of infected individuals after implementing similar vaccination campaigns [12]. Hence, the case of the HPV vaccination in Australia provides one of the best real scenarios for testing new network models in mathematical epidemiology. There is an on-going debate on the pertinence of an approach based upon networks on epidemiology [49] and this work contributes to show the necessity of such an approach in many cases, in particular, in those corresponding to STI.

To validate the model, we used the Australian experience, with two different vaccination coverages: routinely vaccination campaign for 12–13 year-old girls with a coverage of 73% and 83% and a catch-up program in the 14 – 26 age group with an average coverage of 52% and 73%. This program revealed an important herd effect [1], so that vaccination decreased the incidence of genital warts (GW) even in the non-vaccinated men because of the protection of infection conferred by the vaccine, and the decreased transmission of the virus.

The model predicted a fast decline in the number of infections parallel

to the decline in the number of GW in Australia with very similar values. However, this model was built with Spanish data on sexual behavior [6] and prevalence of HPV infection [5], that might differ to the Australian one, and may explain the minor differences found between the model and the actual data published. Herd immunity in this model of STI is predicted much sooner than in other highly transmitted aerial transported infectious diseases as influenza or RSV, due to the structure of the network. This supports the need to build appropriate LSP networks.

Other models have also predicted the protection of males by vaccinating girls and women, but only for men, as the model used by Bogaards et al. [17]. This model uses Bayesian techniques to study the herd immunity effect. However, in contrast with our model, it does not take into account the dynamics of the HPV transmission, the importance of age-groups and the different roles they play in the propagation of these viruses or the links among the MSM subpopulation and the heterosexual network. In this sense, a network model is required to study the impact of the vaccination strategies in short, medium and long time scales.

Vaccination strategies should seek an optimal effectiveness and efficiency. In this case, it can be seen the quick apparition of the herd immunity effect on males and females only vaccinating women. However, the herd immunity effect does not appear in MSM. This can be the consequence of the large LSP numbers for MSM and their casual connections with women with large LSP numbers in the heterosexual subnetwork.

The model considers a quiet close community, where there is not much contact with other communities. This may not be the case in Spain which in 2016 received over 75 million tourists [50], representing almost the double of the number of Spanish inhabitants, and when sexual contacts are frequent. This may bias the results, as the herd immunity in Spain may not be so clear as in countries with less tourism.

Another issue that we must take into account, is the modelling of the population with a high number of contacts because these individuals are hubs in the network whose vaccination may induce a faster decline of the virus prevalence. Our approach is rather conservative in the assignment of LSP for men and women with 10 or more links because we assume that all of them have similar LSP. However, it is expected that individuals with extreme values of LSP are favouring the transmission of HPV in such a way that a targeted vaccination can show its benefit in a very short time.

# Bibliography

- [1] H. Ali, B. Donovan, H. Wand, T. R. Read, D. G. Regan, A. E. Grulich, C. K. Fairley, and R. J. Guy, “Genital warts in young australians five years into national human papillomavirus vaccination programme: national surveillance data,” *Bmj*, vol. 346, p. f2032, 2013.
- [2] M. A. Stanley, “Epithelial cell responses to infection with human papillomavirus,” *Clinical Microbiology Reviews*, vol. 25, no. 2, pp. 215–222, 2012.
- [3] W. Commons, “Harald zur hausen,” 2010. [Online; accessed 6 September 2018].
- [4] J. L. Marx, “Human papilloma virus and cervical cancer.,” *The Medical journal of Australia*, vol. 144 3, p. 164, 1986.
- [5] X. Castellsagué, T. Iftner, E. Roura, J. A. Vidart, S. K. Kjaer, F. X. Bosch, N. Muñoz, S. Palacios, M. S. M. Rodriguez, L. Serradell, L. Torcel-Pagnon, and J. C. and, “Prevalence and genotype distribution of human papillomavirus infection of the cervix in Spain: The CLEOPATRE study,” *Journal of Medical Virology*, vol. 84, pp. 947–956, apr 2012.
- [6] Instituto Nacional de Estadística, “Encuesta de salud y hábitos sexuales. (health and sexual habits survey),” 2003. [Online; accessed 6 March 2017].
- [7] G. M. Clifford, R. K. Rana, S. Franceschi, J. S. Smith, G. Gough, and J. M. Pimenta, “Human papillomavirus genotype distribution in low-grade cervical lesions: Comparison by geographic region and with cervical cancer,” *Cancer Epidemiology and Prevention Biomarkers*, vol. 14, no. 5, pp. 1157–1164, 2005.
- [8] J. Olsen and T. R. Jørgensen, “Revisiting the cost-effectiveness of universal hpv-vaccination in denmark accounting for all potentially vaccine

preventable hpv-related diseases in males and females,” *Cost Effectiveness and Resource Allocation*, vol. 13, no. 1, p. 1, 2015.

- [9] C. J. Lacey, C. M. Lowndes, and K. V. Shah, “Burden and management of non-cancerous hpv-related conditions: Hpv-6/11 disease,” *Vaccine*, vol. 24, pp. S35–S41, 2006.
- [10] J. S. C. Roberts, T. O'Neill, and C. Lacey, “Vaccine for genital warts,” *Vaccines for Human Papillomavirus Infections and Anogenital disease, Medical Intellegence Unit*, vol. 14, 1999.
- [11] C. McNeil, “Who invented the vlp cervical cancer vaccines?,” *Journal of the National Cancer Institute*, vol. 98, no. 7, pp. 433–433, 2006.
- [12] E. H. Elbasha, E. J. Dasbach, and R. P. Insinga, “Model for assessing human papillomavirus vaccination strategies,” *Emerging Infectious Diseases*, vol. 13, pp. 28–41, jan 2007.
- [13] E. H. Elbasha and A. P. Galvani, “Vaccination against multiple hpv types,” *Mathematical biosciences*, vol. 197, no. 1, pp. 88–117, 2005.
- [14] P. S. Bearman, J. Moody, and K. Stovel, “Chains of affection: The structure of adolescent romantic and sexual networks,” *American journal of sociology*, vol. 110, no. 1, pp. 44–91, 2004.
- [15] S. Helleringer and H.-P. Kohler, “Sexual network structure and the spread of hiv in africa: evidence from likoma island, malawi,” *Aids*, vol. 21, no. 17, pp. 2323–2332, 2007.
- [16] C. Fairley, J. Hocking, M. Chen, B. Donovan, and C. Bradshaw, “Rapid decline in warts after national quadrivalent hpv vaccine program,” *Sex Transm Infect*, vol. 85, pp. 499–502, 2009.
- [17] J. A. Bogaards, J. Wallinga, R. H. Brakenhoff, C. J. Meijer, and J. Berkhof, “Direct benefit of vaccinating boys along with girls against oncogenic human papillomavirus: bayesian evidence synthesis,” *Bmj*, vol. 350, p. h2016, 2015.
- [18] L. Acedo, J.-A. Moraño, R.-J. Villanueva, J. Villanueva-Oller, and J. Díez-Domingo, “Using random networks to study the dynamics of respiratory syncytial virus (RSV) in the Spanish region of Valencia,” *Mathematical and Computer Modelling*, vol. 54, pp. 1650–1654, oct 2011.

- [19] N. A. Christakis and J. H. Fowler, “The spread of obesity in a large social network over 32 years,” *New England journal of medicine*, vol. 357, no. 4, pp. 370–379, 2007.
- [20] A. N. Burchell, H. Richardson, S. M. Mahmud, H. Trottier, P. P. Tellier, J. Hanley, F. Coutlée, and E. L. Franco, “Modeling the sexual transmissibility of human papillomavirus infection using stochastic computer simulation and empirical data from a cohort study of young women in montreal, canada,” *American journal of epidemiology*, vol. 163, no. 6, pp. 534–543, 2006.
- [21] F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, and Y. Åberg, “The web of human sexual contacts,” *Nature*, vol. 411, no. 6840, p. 907, 2001.
- [22] B. V. Schmid and M. Kretzschmar, “Determinants of sexual network structure and their impact on cumulative network measures,” *PLoS computational biology*, vol. 8, no. 4, p. e1002470, 2012.
- [23] Valencian Institute of Statistics, “Portal estadístico de la generalitat valenciana (statistical portal of the government of the community of valencia),” 2013. [Online; accessed 6 March 2017].
- [24] L. Acedo and J.-A. Moraño, “Brain oscillations in a random neural network,” *Mathematical and Computer Modelling*, vol. 57, no. 7-8, pp. 1768–1772, 2013.
- [25] S. N. Dorogovtsev and J. F. Mendes, *Evolution of networks: From biological nets to the Internet and WWW*. OUP Oxford, 2013.
- [26] D. J. Watts and S. H. Strogatz, “Collective dynamics of small-worldnetworks,” *nature*, vol. 393, no. 6684, p. 440, 1998.
- [27] A. Chandra, C. E. Copen, and W. D. Mosher, “Sexual behavior, sexual attraction, and sexual identity in the united states: Data from the 2006–2010 national survey of family growth,” in *International handbook on the demography of sexuality*, pp. 45–66, Springer, 2013.
- [28] W. D. Mosher, A. Chandra, J. Jones, *et al.*, “Sexual behavior and selected health measures: men and women 15-44 years of age, united states, 2002,” 2005.
- [29] D. Gentner and A. B. Markman, “Structure mapping in analogy and similarity.,” *American psychologist*, vol. 52, no. 1, p. 45, 1997.

- [30] P. Miret, “La similitud entre los componentes de las parejas jóvenes en España en la primera década del siglo XXI, cada vez más iguales?,” *Revista de Estudios de Juventud*, no. 90, pp. 225–255, 2010.
- [31] L. Acedo, C. Burgos, J.-I. Hidalgo, V. Sánchez-Alonso, R.-J. Villanueva, and J. Villanueva-Oller, “Calibrating a large network model describing the transmission dynamics of the human papillomavirus using a particle swarm optimization algorithm in a distributed computing environment,” *The International Journal of High Performance Computing Applications*, p. 1094342017697862, 2017.
- [32] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*. MIT press, 2009.
- [33] T. A. Feo and M. G. Resende, “Greedy randomized adaptive search procedures,” *Journal of global optimization*, vol. 6, no. 2, pp. 109–133, 1995.
- [34] L. Acedo, R. Martí, F. Palma, V. Sánchez-Alonso, F. Santonja, R. Villanueva, and J. Villanueva-Oller, “Building lifetime heterosexual partner networks,” in *Modeling Human Behavior: Individuals and Organizations* (d. l. P. E. A. L. E. Jódar, L., ed.), Psychology of Emotions, Motivations and Actions, ch. 19, pp. 235–251, Nova Publishers, 2017.
- [35] N. G. Campos, E. A. Burger, S. Sy, M. Sharma, M. Schiffman, A. C. Rodriguez, A. Hildesheim, R. Herrero, and J. J. Kim, “An updated natural history model of cervical cancer: derivation of model parameters,” *American journal of epidemiology*, vol. 180, no. 5, pp. 545–555, 2014.
- [36] J. Díez-Domingo, V. Sánchez-Alonso, R.-J. Villanueva, L. Acedo, J.-A. Moraño, and J. Villanueva-Oller, “Random network models to predict the long-term impact of HPV vaccination on genital warts,” *Viruses*, vol. 9, p. 300, oct 2017.
- [37] L. Acedo, C. Burgos, J.-I. Hidalgo, V. Sánchez-Alonso, R.-J. Villanueva, and J. Villanueva-Oller, “Calibrating a large network model describing the transmission dynamics of the human papillomavirus using a particle swarm optimization algorithm in a distributed computing environment,” *The International Journal of High Performance Computing Applications*, p. 109434201769786, apr 2017.
- [38] “Portal estadístico de la Generalitat Valenciana (Statistical portal of the government of the Community of Valencia),” 2013.

- [39] “Estudio de conducta sexual entre homosexuales (Study of sexual behavior among homosexuals),” tech. rep., Durex, 2002.
- [40] A. R. Giuliano, J.-H. Lee, W. Fulp, L. L. Villa, E. Lazcano, M. R. Papenfuss, M. Abrahamsen, J. Salmeron, G. M. Anic, D. E. Rollison, and D. Smith, “Incidence and clearance of genital human papillomavirus infection in men (HIM): a cohort study,” *The Lancet*, vol. 377, pp. 932–940, mar 2011.
- [41] A. G. Nyitray, M. Chang, L. L. Villa, R. J. C. da Silva, M. L. Baggio, M. Abrahamsen, M. Papenfuss, M. Quiterio, J. Salmerón, E. Lazcano-Ponce, and A. R. Giuliano, “The natural history of genital human papillomavirus among HIV-negative men having sex with men and men having sex with women,” *Journal of Infectious Diseases*, vol. 212, pp. 202–212, feb 2015.
- [42] J. Villanueva-Oller, L. Acedo, J. A. Moraño, and A. Sánchez-Sánchez, “Epidemic random network simulations in a distributed computing environment,” *Abstract and Applied Analysis*, vol. 2013, pp. 1–10, 2013.
- [43] “Pyhton Software Foundation.”
- [44] “Wolfram Mathematica: Modern Technical Computing.”
- [45] N. Khemka and C. Jacob, “Exploratory toolkit for evolutionary and swarm-based optimization,” *The Mathematica Journal*, vol. 11, pp. 376–391, feb 2010.
- [46] R. Storn and K. Price, “Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces,” *Journal of global optimization*, vol. 11, no. 4, pp. 341–359, 1997.
- [47] “Official Site for GARDASIL9.”
- [48] S. Hartwig, J.-J. Baldauf, G. Dominiak-Felden, F. Simondon, L. Alemany, S. de Sanjosé, and X. Castellsagué, “Estimation of the epidemiological burden of HPV-related anogenital cancers, precancerous lesions, and genital warts in women and men in europe: Potential additional benefit of a nine-valent second generation HPV vaccine compared to first generation HPV vaccines,” *Papillomavirus Research*, vol. 1, pp. 90–100, dec 2015.
- [49] S. Eubank, S. Kumar, M. Marathe, A. Srinivasan, and N. Wang, “Structure of social contact networks and their impact on epidemics,” vol. 70, 01 2006.

- [50] “Estadística de movimientos tursticos en fronteras 2017 (statistics of the turistic movements in the spanish borders 2017. press release from the Instituto Nacional de Estadística),” Jan 31st, 2017.