POLITECNICO DI TORINO

Master Degree Thesis

Multiscale modeling of influenza viral emergence

FABIO MAZZA



POLITECNICO DI TORINO



Master of Science programme in Physics of Complex Systems

Supervisor: Alfredo Braunstein (PoliTO)

Co-supervisor: Chiara Poletto (INSERM)

Academic Year 2018-2019

Influenza is viral disease characterized by a sudden onset of fever, cough, headache, muscle and joint pain, severe malaise, sore throat and a runny nose. While most people recover from these symptoms in about a week without medical attention, there are cases with complications where the illness is severe and can cause death. Influenza spreads around the world in yearly outbreaks, causing about 3 to 5 million cases of severe illness and about 290,000 to 650,000 deaths due to respiratory complications [11]. The World Health organization suggests vaccination to protect the people at higher risk (very young children and elderly people, for example), but the immunity given by the vaccine is not lifelong and vaccinations of one year could be ineffective even the next year.

This is due to the highly mutating nature of the influenza virus, which, following a mutation, can change the antigens that it presents and thus be recognized as of a different "type" by the immune system with respect to its previous configuration. This process, called *viral emergence*, effectively creates a new viral strain, and leads to the huge immunological variability of influenza. The newly created emerging strain can, in some cases, infect individuals that were protected to the previous strain, thus becoming the dominant strain during an epidemic season. This process thus heightens the risk of failure of the influenza vaccine, which, every season, requires the identification of the dominant strain of influenza for each region of the world: in case of a mistake in its prediction, the disease would be able to infect a significantly higher fraction of the population of a country, resulting in greater risks for more susceptible individuals and economic losses due to a drop in productivity. This phenomenon has thus come under strict observation from public health institutions, yet the conditions under which it happens are not fully understood.

This thesis tries to address this problem by developing a model of the spreading of two strains of the influenza virus in a population of hosts. Since this process is influenced by the interplay of within-host factors and the between-host transmission, both these scales have been accounted in the model. Because of the resulting complexity of the model, a numerical approach has been devised in order to study the outcomes of the epidemics, averaging the results over many numerical simulations. This required transforming the within-host model, originally deterministic, into a stochastic simulation, and implementing a set of rules for the transmission of the two strains between individual hosts. The model has been studied first with only one viral strain, exploring the space of the parameters of the model, both at the within-host and between-hosts scales. This first stage was required in order to understand the outcomes produced by the model and reproduce realistic outcomes of influenza epidemics. Afterwards, the second strain was introduced in the simulations, and the effect of parameters of this emerging strain was studied by exploring their possible values. As one of the results of this analysis, it was found that the frequency on which the hosts contacted each other played an important part in the selection of the dominant strain.

CONTENTS

1	INTRODUCTION 1			
	1.1	The immune response 1		
	1.2	An overview of influenza 2		
2	MODELING WITHIN-HOST DYNAMICS 5			
	2.1	Deterministic model 5		
	2.2	Stochastic variant 7		
	2.3	Implementation of the stochastic model 10		
		2.3.1 Plotting the dynamics 10		
3	INCI	LUDING THE BETWEEN-HOST INTERACTION 13		
	3.1	Transmission of the virus 13		
	3.2	Contact selection 14		
	3.3	A first look at the dynamics 15		
4	SINGLE STRAIN CALIBRATION 17			
	4.1	Host-level statistics 17		
	4.2	Mapping the times 18		
	4.3	Tuning the dynamics at the population level 21		
		4.3.1 Experienced hosts 21		
		4.3.2 Contact probability 22		
		4.3.3 Bottleneck factor 22		
5	INTRODUCING THE EMERGING STRAIN 27			
	5.1 Exploring the space of parameters of the emerging			
		5.1.1 Fixed ratio of cross immunity 28		
		5.1.2 Fixed wildtype cross immunity 29		
	5.2	The effect of the contact frequency 30		
		5.2.1 Redrawing the picture 30		

27

v

- 5.2.2 Changing the contact frequency 31
- 6 CONCLUSIONS 35

BIBLIOGRAPHY 37

LIST OF FIGURES

Figure 2.1	Within-host dynamics from integration, naive case 7	
Figure 2.2	Within-host dynamics from integration, experi- enced host 7	
Figure 2.3	Within-host viral dynamics from integration and stochastic simulation, naive host 11	
Figure 3.1	General overview of the model 14	
Figure 3.2	Test of the dynamics on the star graph S_{20} 15	
Figure 4.1	Generation time between two individuals (1 infects 2) 18	
Figure 4.2	Measure of the generation time as function of the value of the threshold V_m , which is plot- ted here divided by <i>K</i> (as explained in section 2.3.1) 18	
Figure 4.3	Viral kinetics of a single host under different values of the parameters (ρ , p_b). The small jump upwards at the end of the infection are not due to the internal dynamics, but to the transmission of viruses from other hosts. As it can be seen, they are not able to "restart" the infection, as by this point the host has already produced a significant immunity. 19	
Figure 4.4	Heat maps of the relevant times with a $G(n, p)$ network of 1000 individuals and $\langle k \rangle = 20$, with $p_c = 0.18$, $V_{th} = 0.1 \cdot K$, $\Delta t_T = 1.0$ 20	
Figure 4.5	Heat map of the generation time over peak time, with a $G(n, p)$ network of 1000 individuals and $\langle k \rangle = 20$, with $p_c = 0.18$, $V_{th} = 0.1 \cdot K$, $\Delta t_T = 3.0$ 21	
Figure 4.6	Plot of the distribution of p_b chosen for the model	24
Figure 4.7	Statistics of a simulation on a $G(n, p)$ random graph with $N = 4000$ nodes and average degree $\langle k \rangle = 20$. Here, the parameters of p_b are $m = 3 \cdot 10^{-8}$ and $s = 1.2$, while $p_I = 0.4$. The results are all for naive hosts, except for R_0 which is for the whole epidemic, and are averaged on 120 simulations for each set of parameters, showing the mean and the variance. 24	

Figure 4.8	Same plot as the previous figure, but with $\rho_W = 1$ fixed, changing Δt_T , and with $m = 10^{-8}$ and $s = 0.5$ as the parameters for the distribution of p_b . The results are all for naive hosts, except for R_0 which is for the whole epidemic, and are averaged on 120 simulations for each set of parameters, showing the median and the 80% capital damage intermal.
Figuro = 1	Heat map of the relative attack rate, while change
Figure 5.1	ing $\epsilon_{\rm m}$ and keeping the ratio fixed $\epsilon_{\rm E}/\epsilon_{\rm m} = 0.35$
Figure = 2	Heat maps for the case of fixed wildtype cross-
riguite 5.2	immunity parameter $c_{\rm ex} = 0.8$ $c_{\rm ex} = 1$
Figuro = 2	Polative attack rate plot from provious figure with
Figure 5.3	the hourd origin of the set dominance region drawn in
	the boundaries of the co-dominance region drawn in
	black 33
Figure 5.4	Relative attack rate plot from previous figure with
	co-dominance region highlighted 33
Figure 5.5	Relative attack rate plots, with co-dominance region
-	(with r_L and r_U as defined in the text) drawn for
	different transmission intervals 34

The Cambridge English dictionary defines a disease as "(an) illness of people, animals, plants, etc., caused by infection or a failure of health rather than by an accident". This definition encompasses a wide range of conditions that can be very distinct, such as osteoporosis, malaria or smallpox. One important distinction that can be made about diseases is whether they can be transmitted between individuals or not: the ones that can are classified as *infectious*, while the others are called non-infectious. Infectious diseases can be distinguished by the type of infectious pathogen which transmits the illness, thus defining viral, bacterial or parasitic diseases, for example. Also, another important distinction is based on the typical duration of the infection, which separates diseases in essentially two classes, either *acute*, which have rapid recovery, and *chronic*, which last longer and may also be recurrent.

1.1 THE IMMUNE RESPONSE

As the pathogen infects an individual and starts reproducing itself, the infected individual's immune system develops what is generally called an *immune response* to fight the infection. This process is quite articulated, as it involves first the activation of the *innate* immune system, and later of the *adaptive* immune system. While the former is a general response and is not very efficient in eliminating the infectious disease, it has a faster activation time, and is useful to keep the infection at bay while the adaptive immune system develops a more tailored response. The activation of the adaptive immunity involves the recognition of the *antigens*, which are proteins on the surface of the cells, which are presented by the pathogen. The immune system is able to recognize and keep track of the antigens and produces *antibodies*, which can bind only to the targeted antigens. In this way the immune system is able to distinguish the pathogen cells (or the cells infected by it) and destroy them without harming its own healthy ones.

After an infection, the antibodies previously created are kept circulating in the body in reduced numbers: this leads to immunological memory and is the process upon which vaccination is built. A subsequent infection of the same pathogen takes a much shorter time to eradicate, because the antibodies already present are replicated upon encountering the antigens, and the immune system doesn't need to create them again.

A mutation of the pathogen that manages to change its antigens makes it able to escape recognition by preexisting antibodies and thus produce a more sustained infection of a hosts. In this case the immune system will need to produce new antibodies to target the new strain, while it will be able to fight and eliminate efficiently the previous one. In this way, the pathogen is forced to mutate in order to survive.

1.2 AN OVERVIEW OF INFLUENZA

Influenza is viral disease characterized by a sudden onset of fever, cough, headache, muscle and joint pain, severe malaise, sore throat and a runny nose. While most people recover from these symptoms in about a week without medical attention, there are cases with complications where the illness is severe and can cause death. Influenza spreads around the world in yearly outbreaks, causing about 3 to 5 million cases of severe illness and about 290,000 to 650,000 deaths due to respiratory complications [11]. The disease is caused by the influenza virus, which has a very high mutation rate, and this characteristic makes it possible to see a huge variability of the virus in antigenic terms: currently, four types of the virus are known, named alphabetically from A to D. Among these, only types A, B and C are known to infect humans, and type C has not been known to cause epidemics [15, 11]. Therefore, the globally circulating types are only A and B, and these are divided into subtypes (for A) and lineages (for B). Influenza A is the viral type evolving most rapidly among the two, and has already caused numerous pandemics (the 1918 Spanish Flu, for example, or the recent 2009 Avian Flu epidemic, caused by a new strain of H1N1 [10]).

Since the influenza virus mutates rapidly, at any moment during the course of a seasonal outbreak one of the subtypes could change the antigens it presents. This leads, as explained before, to the creation of a variant, or strain, of the virus (called *emerging*) which is novel for the immune system. If the antigens of the new strain do not vary much from the previous viral variant, and are still keeping a similarity to it¹, it can happen that the antibodies for the previous strain (which in this case is called *wildtype*) will still recognize, if present, the emerging strain and begin to fight it. When this happens there is a so-called *cross*reactivity of the antibodies or cross-immunity. This phenomenon has been shown to happen in the case of influenza [12]. The antibodies may be able to counteract the emerging strain as well as the wildtype one, but it could also happen that they are less efficient with the new strain. Considering the second case leads to a phenomenon called *immune escape*, as the new virus can thus elude the immune system and spread to more individuals, potentially infecting a large part of the hosts population and becoming the dominant strain.

This process of creation of new viral strains, called *viral emergence*, has important implications for public health, mainly in the development

¹ The reasons for this similarity are biological. A more detailed explanation is given in [5]

of the vaccine against influenza, where, from the many strains of the subtypes that are circulating in the season, the dominant ones have to be selected for the targeting of the vaccine. Given the previously mentioned variability of the subtypes, one strain has to be chosen for each subtype/lineage in circulation (A/H1N1, A/H3N2, B/Victoria and B/Yamagata are the ones currently in circulation [11]). This step is very important, because if the targeted strain does not become dominant, the vaccine could fail in protecting the population and there would be an increase in the incidence of the disease during the epidemics.

Therefore, it becomes crucial to understand the conditions upon which an emerging strain can overcome the wildtype strain and infect a large part of the population. This is the aim of this thesis, in which a metapopulation model has been designed and implemented to describe the spreading of two strains (wildtype and emerging) of influenza virus in a population of hosts.

The model describes the diffusion of two strains (wildtype, W, and emerging, E) of influenza at different scales, including both the withinhost dynamics and the between-host dynamics, the latter of which accounts for the spreading of the viruses from person to person. The within-host part describes the interaction of the viral strains with the immune system inside each individual.

2.1 DETERMINISTIC MODEL

In the model, the dynamic of the infection has been derived from equations 2.1 to 2.4, taken from [13], which describe the evolution of the viral load of the two viruses and the respective immune response of the individual, tailored to each strain.

$$\dot{V}_W = \rho_W V_W (1 - V_W - V_E) - V_W (I_W + \epsilon_W I_E)$$
 (2.1)

$$\dot{I}_W = I_W \left(V_W + \epsilon_E V_E \right) \tag{2.2}$$

$$\dot{V}_E = \rho_E V_E (1 - V_W - V_E) - V_E (I_E + \epsilon_E I_W)$$
 (2.3)

$$\dot{I}_E = I_E \left(V_E + \epsilon_W V_W \right) \tag{2.4}$$

These equations are deterministic and it can be seen that they contain terms of growth and decrease which describe the following processes:

- VIRAL REPLICATION The first terms in equations 2.1 and 2.3 model the replication of the virus, as depending on a parameter ρ , the replication rate, times a factor $(1 - V_W - V_E)$ which limits the total quantity of viruses in the host, imposing the condition that $V_W + V_E < 1$ for the replication of any of the two strains to proceed. Otherwise, this factor is negative and the viral population decreases. Therefore, this term describes the direct competition of the two viruses, as there is a finite amount of cells they can possibly infect.
- VIRAL ELIMINATION The second term in the equations for V_W and V_E accounts for the viruses that are killed by the immune response. While the immune quantity responsible for this is, for the most part, of the same strain as the the virus, there is also the one related to the opposite strain that contributes. This term thus

describes the *cross-immunity* and is mediated by the parameters ϵ_W and ϵ_E , which can have values between 0 and 1.

IMMUNITY BUILDUP The immune response in an individual grows after the encounter of the virus by the immune system of the hosts. This is modeled by equations 2.2 and 2.4, that show a growth term of the immunity which depends not only on the same-strain viral quantity but also on the opposite strain, as one would expect from the *cross-immunity*. Because of this link, this extra growth is mediated by the ϵ parameters.

The model, as previously mentioned, is deterministic, and describes the interaction of the two viral strains in a symmetric way between viral variants. To properly describe the interaction between the emergent and wildtype strains, one has to account for the fact In terms of the model's parameters, this means that since the ϵ_i factor describes the strength of the interaction between V_i and I_i ().

The cross-immunity parameters are present both in the viral and immune equations, but they are exchanged between the equations of *V* and *I* of same strain. Upon close inspection, it can be seen that each ϵ_i is tuning the strength of the interaction of the viral quantity of the same strain V_i with the immunity of the opposite kind I_j ($i \neq j$, $i, j \in \{W, E\}$). Therefore, the higher ϵ_i , the higher the number of viruses of strain *i* killed by the immunity I_j , and the higher the growth of I_j upon encounter of V_i . Since it has been found that the antibodies produced for the emerging strain are more cross-reactive with the wildtype than the opposite [5], the value of the cross-immunity parameters is chosen such that $\epsilon_W > \epsilon_E$.

In order to describe different cases of infection and co-infection, different kinds of individual hosts are defined in [13] based on their initial condition of immune quantities described by the model (I_W^0 , I_E^0): a *naive* individual has very low initial values (10⁻³) of both immune responses, while an *experienced* one has already encountered the wildtype virus, having thus a strong immune response to the wildtype, $I_W^0 = 1$, while $I_F^0 = 10^{-3}$.

The replication rate is an important parameter of the viruses, and while in the aforementioned work the model was studied both with $\rho_W < \rho_E$, where the mutated virus has a replication advantage, and with $\rho_W > \rho_E$, where the emerging has a replication deficiency, only latter case has been considered in this thesis, since the mutations conferring immune escape to the virus are often detrimental to viral replication [2].

The dynamics for the replication deficiency case are shown in Figure 2.1 for a naive host and in Figure 2.2 for an experienced host, with common parameters $\epsilon_W = 0.8$, $\epsilon_E = 0.28$, $\rho_W = 2$, $\rho_E = 1.8$, and initial condition on the viral quantities $V_W^0 = 10^{-2}$, $V_E^0 = 10^{-4}$. In the first case, the wildtype virus is dominating over the emerging,

and it manages to create a sustained infection. The hosts develops immunities for both strains, but in an asymmetrical way because of the cross-immunity parameters. In the experienced host case there is a very short infection from the wildtype, as expected, and a quite long infection of the emerging strain. It is curious to see that, in contrast to the previous case, the host doesn't develop an immunity for the emerging strain, but fights the infection with the wildtype antibodies, and ends up at a very high immune level for that strain. This is a mechanism called *Original Antigenic Sin*, and a detailed discussion can be found in [5].



Figure 2.1: Within-host dynamics from integration, naive case



Figure 2.2: Within-host dynamics from integration, experienced host

2.2 STOCHASTIC VARIANT

The model described in the previous section has been translated into a stochastic simulation for the within-host dynamics. Introducing stochasticity is essential, because, in general, in many biological processes

the number of degrees of freedom are so high that in order to get to a tractable representation it becomes necessary to coarse-grain the description of the system, therefore losing some information about its state. This uncertainty is modeled by adding some degree of stochasticity in the model.

In order to create a stochastic variant of the model, the quantities whose evolution is described in equations 2.1 to 2.4, where they vary from 0 to 1, need to be scaled up in order to find the number of viruses varying from 0 to *K*. By defining $\tilde{V}_i = V_i \cdot K$ and $\tilde{I}_i = I_i \cdot N$ and inserting these equalities into the equations 2.1 to 2.4, we find (removing the tildes):

$$\dot{V}_W = \rho_W V_W \left(1 - \frac{V_W + V_E}{K} \right) - V_W \left(\frac{I_W + \epsilon_W I_E}{N} \right)$$
(2.5)

$$\dot{I}_W = I_W \left(\frac{V_W + \epsilon_E V_E}{K}\right) \tag{2.6}$$

$$\dot{V}_E = \rho_E V_E \left(1 - \frac{V_W + V_E}{K} \right) - V_E \left(\frac{I_E + \epsilon_E I_W}{N} \right)$$
(2.7)

$$\dot{I}_E = I_E \left(\frac{V_E + \epsilon_W V_W}{K}\right) \tag{2.8}$$

where the viral and immune quantities now represent the number of viruses and immune particles.

The dynamical processes described by this system of equations, which are the same as those identified in the previous section, can also be described in probabilistic terms. For each of them, a probability of that phenomenon happening to one viral (or immune) particle can be defined starting from the above system of equations.

For example, let's take the viral replication term $\rho_i V_i \left(1 - \frac{V_W + V_E}{K}\right)$ (where *i* indicates one of the strains). The process modeled here is the creation of a new virus from a preexisting virus. There are V_i such viral particles, and we can say that from each one of them, a new one can be created at any instant in time with probability rate p_{ρ} . After looking at the term, we can identify $p_{\rho} = \rho_i \left(1 - \frac{V_W + V_E}{K}\right)$, as on average the number of new viruses created per unit time is exactly $\rho_i V_i \left(1 - \frac{V_W + V_E}{K}\right)$.

The destruction of viral particles, which is modeled by the term $V_i\left(\frac{I_i+\epsilon_i I_j}{N}\right)$, happens after a viral particle of strain *i* encounters an immune particle of either strain $(i \neq j, i, j \in \{W, E\})$. Therefore, the probability rate of this event is proportional to the probability of one viral particle finding an immune particle, that is $\frac{I_i}{N}$ for the same strain and $\frac{I_j}{N}$ for the other strain. In the latter case, the interaction is mediated by the parameter ϵ_i . Therefore, we find that the probability rate for the destruction of viral particles is $\frac{I_i+\epsilon_i I_j}{N}$. Again, there are V_i such viral

particles which have share this probability of encountering the immunity. It is straightforward to apply the same reasoning to the immunity buildup term, finding a probability rate $\frac{V_E + \epsilon_W V_W}{\kappa}$.

The processes thus identified can be treated, for a small interval of time Δt , as homogeneous Poisson point processes for the number of events (of the same type) N_k , with probability rates λ_k (found above), and probability law:

$$\mathbb{P}\left[N_k(\Delta t) = m\right] = \frac{\left(\lambda_k \Delta t\right)^m}{m!} e^{-\lambda_k \Delta t}$$

At first order, there are only two possible outcomes which could happen in a very small Δt : either $N_k = 0$, with $\mathbb{P}(0) = e^{-\lambda_k \Delta t}$, or $N_k = 1$, with $\mathbb{P}(1) = \lambda_k \Delta t e^{-\lambda_k \Delta t}$. Expanding the expressions for the probabilities, we get that

$$\mathbb{P}\left(N_k=0\right) = 1 - \lambda_k \Delta t$$

and

$$\mathbb{P}\left(N_k=1\right) = \lambda_k \Delta t$$

which means that N_k is now a Bernoulli random variable with probability $p = \lambda_k \Delta t$.

Taking into account that each of the possible type of event can happen to any particle (viral or immune, depending on the event), we can say that the total number of particles which undergo the same process follows a binomial distribution. Therefore, we find the following mapping:

$$\begin{split} \rho_i V_i \left(1 - \frac{V_W + V_E}{K} \right) &\longrightarrow \Delta V_i^{(+)} \sim Bin\left(\rho_i, \Delta t \left(1 - \frac{V_W + V_E}{K}\right)\right) \\ V_i \left(\frac{I_i + \epsilon_i I_j}{N}\right) &\longrightarrow \Delta V_i^{(-)} \sim Bin\left(V_i, \Delta t \left(\frac{I_i + \epsilon_i I_j}{N}\right)\right) \\ I_i \left(\frac{V_i + \epsilon_j V_j}{K}\right) &\longrightarrow \Delta I_i \sim Bin\left(I_i, \Delta t \left(\frac{V_i + \epsilon_j V_j}{K}\right)\right) \end{split}$$

where $X \sim Bin(N, p)$ denotes a random variable X that follows a binomial distribution with N as the number of trials and p the probability of success, and $i, j \in \{W, E\}$ correspond to different strains $(i \neq j)$. In the mapping the random variables $\Delta V_i^{(+)}, \Delta V_i^{(-)}$ and ΔI_i have also been defined, and V_i and I_i have become natural numbers which describe the number of viral and immune particles present.

Therefore, during a simulation, at each time step Δt the viral and immune quantities, which are now natural numbers, are updated as follows:

$$V_W \leftarrow V_W + \Delta V_W^{(+)} - \Delta V_W^{(-)}$$
(2.9)

$$I_W \leftarrow I_W + \Delta I_W \tag{2.10}$$

$$V_E \leftarrow V_E + \Delta V_E^{(+)} - \Delta V_E^{(-)}$$
(2.11)

$$I_E \leftarrow I_E + \Delta I_E \tag{2.12}$$

2.3 IMPLEMENTATION OF THE STOCHASTIC MODEL

The simulation of the model has been implemented using the C++ programming language, because of its very high speed of execution. This was the only choice as the simulations have been intended to run with a very high number of individual hosts and a small time step Δt in the update to the within-host dynamics. In order to be able to draw from random variables with constantly changing parameters, the GNU Scientific Library [9] has been used extensively in the simulation.

In order to run the simulations, the scale of the virus counts and the immunities needs to be defined. Since the number of viruses inside an individual can become very high, the values $K = 10^8$ and $N = 10^5$ have been chosen and kept fixed in all successive steps, except when explicitly noted. One example of the dynamics from the simulation, compared to the numerical solution of the deterministic system of equations, is shown in Figure 2.3, where the viral quantities are scaled by $1/\kappa$ in order to make the comparison.

2.3.1 Plotting the dynamics

The convention that, in the viral quantity axis, the value 1 corresponds to having *K* viral particles will be kept in future plots, except when otherwise noted, as it helps understanding the dynamics. It will also be used in future plots of the immunity, where from now on a value of 1 will corresponds to *N* viral particles.

The semi-log plot is essential to observe the whole viral dynamics, since a virus can start infecting an host even from a very small number of particles, and eventually reach *K*, which usually happens with an exponential growth.



Figure 2.3: Within-host viral dynamics from integration and stochastic simulation, naive host

INCLUDING THE BETWEEN-HOST INTERACTION

During the course of an epidemics, pathogens are able to move from one host to another. This happens when a contact between the hosts occurs, and since influenza virus is transmitted through the air, excreted by infected individuals via coughs and sneezes, any kind of contact could potentially spread the virus. Therefore, it is crucial to find a good description of the contacts between individuals in the course of the infection.

In the thesis's model, the description of the contacts between hosts is based on a network, or graph, where the nodes represent the hosts and the edges represent all the possible contacts between the hosts. Following this network, hosts interact with each other, and during said interaction they exchange viral particle. Each of the hosts is modeled as described in the previous chapter, therefore there is a dynamic which involves not only a single individual, but the whole population.

3.1 TRANSMISSION OF THE VIRUS

During contacts between individuals, viral particles might jump from an infected host to another one, where they might start a new infection. In the model, this is represented by a certain probability p_b (called, for historical reasons, the *bottleneck factor*) for each virus to go from its current host to a new one. Thus, in the simulation, since all the viral particles of the individual in question have the same probability p_b and all the receiving hosts are independent, the number of particles transmitted from one infected host to each neighbor follows a multinomial distribution (whose random sampling has been implemented according to [7]).

Also, the exchange of viruses is instantaneous, lasting a time step Δt of the simulation. Hosts have been defined as infected if the total viral load is larger than a predefined value V_m , and only infected individuals transmit viral particles to others, therefore it must happen that $V_W + V_E \geq V_m$.

Since there are indications that there is a latent period in the course of influenza infection [6], meaning that the diffusion of viral particles is not immediate after infection of a host, it seemed reasonable at the beginning to add another constrain in the model on the transmission of viral particles, that is the existence of an higher threshold level on the total viral load of an individual. This can be again written as $V_W + V_E \ge V_{th}$, thus defining this particular threshold. In the rest of the thesis, indication of a value of V_{th} means that this threshold has been used



Figure 3.1: General overview of the model

during the simulations. When studying the full model with two strains, the threshold has been removed, because it was deemed unnecessary after the single strains tests.

3.2 CONTACT SELECTION

During the simulations, hosts contact each with a frequency $\frac{1}{\Delta t_T}$, where Δt_T is the contact interval (also called transmission interval). During each transmission step, every individual chooses the person to reach out to with a probability p_c among his/hers neighbors on the interaction network. This randomization of the interaction has been included in order to account for the variation of the contacts during the course of the epidemic.

The network containing all the possible interactions between individuals in the epidemics has been generated using the G(n, p) procedure from the Erdős–Rényi model [8]. Initially, community networks ,generated using a procedure derived from the configuration model, were also used, but because this kind of modeling is novel, it was decided to concentrate on simple homogeneous networks such as Erdős–Rényi.



Figure 3.2: Test of the dynamics on the star graph S_{20}

3.3 A FIRST LOOK AT THE DYNAMICS

The model now contains both a within-host dynamics and the exchange of viruses. This already produces complex patterns of infection, and as the number of individual grows, it becomes very difficult to understand what happens. A simple test case of a graph with 21 hosts can be seen in Figure 3.2. In the plots, the color of the line represents the dynamics of a particular host.

4

SINGLE STRAIN CALIBRATION

The model presented so far is complete, but it is very difficult to understand what happens during an epidemic. In order to simplify the picture, the model has been studied with a single strain (the wildtype) first. At this stage, the influence of the parameters has been observed through the analysis of the epidemics simulated by the model. This required the definition of both host-level and population-level statistical quantities which are of significance in the characterization of infectious diseases. The model parameters have been therefore tuned in order to reproduce the outcomes of influenza epidemics.

4.1 HOST-LEVEL STATISTICS

During the course of an epidemics, there are several important quantities that are used to characterize a disease, among them, the following have been introduced and measured:

- PEAK TIME In viral diseases, the peak time is the time that it takes for the virus to reach the maximum load in the individual. It is an important parameter of models for viral dynamics, as the peak of the disease is associated to the onset of the symptoms [3], meaning that it is available experimentally.
- GENERATION TIME The generation time is defined as the time interval between *primary* case and infection of a *secondary* case caused by the primary case [14]. This definition is clearer when looking at Figure 4.1, where it can be seen that the generation time is measured by looking at the intersections on the green dashed line. The value of this line corresponds to the threshold at which the host becomes infected, which is called V_m . As can be seen in Figure 4.2, the measure of the generation time does not change when changing the value of V_m . This quantity, therefore, has been set to $V_m = 2 \cdot 10^{-4} \cdot K$ to avoid having a too large value, which would have been deceptive of the its role, and a too small value, which would have caused errors in the recognition of the infected status of the host (see the small "reinfections" shown in Figure 4.3).
- ACTIVE TIME The time of activity, or active time, of an infected host is the time that it stays infectious, meaning the time interval between the moment when the total viral load surpasses V_m and the moment when it become lower than V_m . Since the dynamic

of infection is primarily driven by the internal dynamics, when a host's total viral load starts declining and goes below V_{m} , it rarely goes upwards again (if the value of V_m is well-determined in order to avoid spurious reinfections).



Figure 4.1: Generation time between two individuals (1 infects 2)



Figure 4.2: Measure of the generation time as function of the value of the threshold V_m , which is plotted here divided by K (as explained in section 2.3.1)

4.2 MAPPING THE TIMES

The measures defined above play an important role in tuning the model. In order to compare find the outcomes of the model with real-world epidemics, since the time of the simulations is in arbitrary units, it becomes essential to find a "translation" of an unit of simulation time into days or hours. To tackle this problem, we observe the average values of the viral peak time, the generation time and their ratio as the parameters of the simulations are changed. The heat maps in Figure 4.4 and 4.5 have been produced running many simulations with different values of the replication rate ρ_W , the bottleneck factor p_b , the transmission interval Δt_T (in simulation time) and the infectiousness threshold V_{th} .

From these results, a few important considerations can be made. First, the viral peak time, as one could expect, does not depend on



Figure 4.3: Viral kinetics of a single host under different values of the parameters (ρ , p_b). The small jump upwards at the end of the infection are not due to the internal dynamics, but to the transmission of viruses from other hosts. As it can be seen, they are not able to "restart" the infection, as by this point the host has already produced a significant immunity.

the bottleneck factor, and as a matter of fact it depends only on the replication rate, as neither the transmission interval or the threshold level for the transmission of viruses have any effect on it.

Based on these results, and on empirical data which shows that the generation time of influenza is $\approx 2.3-3.4$ days and the average peak time of the virus is ≈ 2 days [4], it can be seen that the region corresponding to realistic dynamics of infection is on the bottom left corner, for $\rho_W \leq 1.0$ and $p_b \leq 10^{-6}$. Using this empirical information we are able to define the length of a day in simulation time as half the viral peak time, thus finding a link between the units of measurements of the time.



Figure 4.4: Heat maps of the relevant times with a G(n,p) network of 1000 individuals and $\langle k \rangle = 20$, with $p_c = 0.18$, $V_{th} = 0.1 \cdot K$, $\Delta t_T = 1.0$



Figure 4.5: Heat map of the generation time over peak time, with a G(n, p) network of 1000 individuals and $\langle k \rangle = 20$, with $p_c = 0.18$, $V_{th} = 0.1 \cdot K$, $\Delta t_T = 3.0$

4.3 TUNING THE DYNAMICS AT THE POPULATION LEVEL

The generation time and viral peak time are important statistics of influenza, but there are other quantities which are also important in the characterization of the disease.

The first is the *attack rate*, which is the percentage of individuals of a population infected by the virus during the epidemic. If the total number of individuals is fixed and known, an equivalent statistic is the number of cases in the course of the epidemics. Another important quantity is the *basic reproductive ratio* R_0 , which is the average number of secondary infections per primary infection. In other words, it is the number of hosts who, on average, contract the disease from the same infected host.

Initial analysis of the model showed that the attack rate was too high for influenza, as almost all individuals were infected by the virus. Therefore, a few changes to the model are needed, with the aim of modeling the spreading of the virus in a more realistic way.

4.3.1 Experienced hosts

Up until this point, during the simulation all hosts have been considered naive (as defined in subsection 2.1) to the virus. This is not realistic, since during an epidemic there is always a fraction of the population which has contracted the disease beforehand and is therefore immune. To include this in the model, experienced individuals are introduced by adding a probability p_I of the host being experienced to the wildtype virus since the beginning of the simulation.

4.3.2 *Contact probability*

Having a fixed probability of contact for any kind of network of interaction (used in the simulation) does not seem realistic. In order to find a better estimation of the probability p_c that is dependent of the network, a data-based approach has been devised. The statistics upon which this approach has been built is the number of people contacted in a day by an individual, which will be called *c* in the rest of the thesis. In [1] this quantity has been estimated, for the first time, using a large scale survey in France, finding an average number of contacts $\langle c \rangle = 8$.

In order to include this information in the model, it is necessary to find an expression for $\langle c \rangle$ as a function of the parameters used. Focusing for a moment on a single individual and his/her neighbors in the graph, we can say that, since different neighbors are independent, the probability of reaching *m* individuals during a day will follow a binomial law *Bin* (*k*, *p*_d), where *p*_d is the probability of contacting an individual at least once in a day and *k* is the number of neighbors. From simple probabilistic reasoning, it follows that

$$p_d = 1 - (1 - p_c)^T$$

where *T* is the number of trials of contact per day, and can be expressed, given the duration of the day *D* and the contact interval Δt_T , as $T = \frac{D}{\Delta t_T}$.

Since the neighbors are independent, for a single individual the average number of contacts will be:

$$\langle c \rangle = k \left[1 - \left(1 - p_c \right)^{\frac{D}{\Delta t_T}} \right]$$

Inverting this formula, and considering the whole network of individuals we find the wanted probability of contact:

$$p_{c} = 1 - \left(1 - \frac{\langle c \rangle}{\langle k \rangle}\right)^{\frac{\Delta t_{T}}{D}}$$
(4.1)

where $\langle k \rangle$ is the average degree of the network. This last step is valid only in networks with homogeneous degree of nodes, like the ones used in this model.

Using equation 4.1, it becomes possible to determine a value for p_c in an automated way, as Δt_T , $\langle c \rangle$ and D are input parameters of the simulation, by calculating the average degree of the network at the beginning. The value for D in simulation time can be found using the correspondence drawn in the section 4.2, where the information that the viral peak time of influenza corresponds to 2 days has been used.

4.3.3 Bottleneck factor

In the model so far, the bottleneck factor p_b has been kept fixed for all the events of viral transmission. This has been changed for successive epidemic simulations in order to account for the variability of the contacts between individuals. Thus, at this point we introduce a log-normal distribution for the bottleneck factor p_b (so that $ln(p_b)$ is normally distributed), so that

$$f_{p_b}(x) = \frac{1}{sx\sqrt{2\pi}} \exp\left(-\frac{\left(\ln x - \mu\right)^2}{2s^2}\right)$$

is the probability density function. Here, μ and s are parameters of the distribution which are linked to the mean and variance through these equations:

$$\mathbb{E}[p_b] = \exp\left(\mu + \frac{s^2}{2}\right)$$
$$\mathbb{V}[p_b] = \left[\exp\left(s^2\right) - 1\right]\exp\left(2\mu + s^2\right)$$

Since these relations are quite complicated and cannot be inverted (a numerical attempt to find μ and s given $\mathbb{E}[p_b]$ and $\mathbb{V}[p_b]$ did not converge), the mode m of the distribution has been used as parameter instead of μ :

$$m\left(\mu,s\right) = \exp\left(\mu - s^2\right)$$

Inverting this relationship leads to $\mu(m, s) = \ln m + s^2$, allowing for the parametrization of $f_{p_b}(x)$ in m and s. Since s is the variance of the distribution of $\ln (p_b)$, it will be called log-variance in the rest of this thesis.

Different values of the mode and log-variance have been evaluated by checking the outcomes of the simulations (the ones for one couple of values of *m* and *s* is shown in Figure 4.7). As a result of the exploration of these parameters, the values $m = 10^{-8}$ and s = 0.5 have been determined as final, as they lead to a large variation of the attack rate when the transmission interval Δt_T is changed. The results with the final parameters are shown in Figure 4.8, and the corresponding distribution can be seen in Figure 4.6.



Figure 4.6: Plot of the distribution of p_b chosen for the model



Figure 4.7: Statistics of a simulation on a G(n, p) random graph with N = 4000nodes and average degree $\langle k \rangle = 20$. Here, the parameters of p_b are $m = 3 \cdot 10^{-8}$ and s = 1.2, while $p_I = 0.4$. The results are all for naive hosts, except for R_0 which is for the whole epidemic, and are averaged on 120 simulations for each set of parameters, showing the mean and the variance.



Figure 4.8: Same plot as the previous figure, but with $\rho_W = 1$ fixed, changing Δt_T , and with $m = 10^{-8}$ and s = 0.5 as the parameters for the distribution of p_b . The results are all for naive hosts, except for R_0 which is for the whole epidemic, and are averaged on 120 simulations for each set of parameters, showing the median and the 80% confidence interval.

INTRODUCING THE EMERGING STRAIN

After the calibration stage of the model has been completed, the emerging viral strain has been included in the simulations. As a first scenario, it has been introduced at the start of the epidemic season, in a host chosen with uniform probability among the population. As in previous analysis, there is a fraction of the hosts that are *experienced* ($p_I = 0.4$), the average number of neighbors contacted in a day by an individual $\langle c \rangle$ is set to 8 [1], and the bottleneck factor is extracted for each contact from a log-normal distribution with mode $m = 10^{-8}$ and log-variance s = 0.5. In the following, the replication rate of the wildtype strain ρ_W has been fixed to 1.0, and the replication rate of the emerging strain has been always considered lower then the wildtype ($\rho_E \leq \rho_W$) because of previously mentioned fitness considerations. The network of contacts has been generated once with the Erdős–Rényi model, with number of nodes (individuals) N = 1000 and average degree $\langle k \rangle = 20$, and then used in all the simulations.

5.1 EXPLORING THE SPACE OF PARAMETERS OF THE EMERGING STRAIN

As in the previous chapter, a few key measurements of the epidemics have to be defined. Focusing on the infection status of the hosts, the following quantities have been measured: the attack rates of the two strains (along with the total one) and the number of hosts which have been infected by both strains. As previously explained, the attack rate of a strain is the number of individuals infected by the strain during the epidemic. It is important to mention that the last measure does not indicate the cases where the two infections happen simultaneously (called *co-infection*), but it can only be an upper bound to the number of such co-infections of the two strains.

This time, the space of parameters of has been explored by changing the values of ρ_E , the replication rate of the emerging strain, and of the cross-immunity parameters ϵ_W and ϵ_E . In order to avoid having too many parameters, and since ϵ_W and ϵ_E are linked, two different kinds of parametrization have been run, one where the parameters ϵ_W and ϵ_E are varied together as their ratio $\frac{\epsilon_E}{\epsilon_W}$ is kept fixed, and the other where ϵ_W is fixed while ϵ_E is changed, thus changing their ratio.

In both cases the transmission interval has been fixed to $\Delta t_T = 3.0$ (which, after the mapping done in the previous chapter, corresponds to 12 hours). The results have been averaged over 100 simulations. These investigations produced heat maps of the median value measured quantities, such as in Figure 5.1 and 5.2. Showing the median instead

of the mean value makes a big difference for the attack rates, whose empirical distribution appeared skewed in previous plots (see Figure 4.8 for an example of the variability in different simulation runs).

5.1.1 Fixed ratio of cross immunity

A few remarks have to be made for the case of fixed cross immunity rate (Figure 5.1). First, when the cross immunity factors ϵ_W and ϵ_E are very low (≤ 0.2), the two viral strains are almost separate diseases, and this can be seen by noticing that they have approximately the same attack rate, and that the fraction of hosts infected by the two is very high (≈ 0.6). This happens independently of the replication rate of the emerging strain. The generally higher attack rate of emerging strain can be explained by the fact that there is a fraction of the hosts who are immune to the wildtype, where the emerging strain is favored.



(a) Ratio of the attack rates, emerging over wildtype



(b) Fraction of hosts infected by both strains

Figure 5.1: Heat map of the relative attack rate, while changing ϵ_W and keeping the ratio fixed $\epsilon_E/\epsilon_W = 0.35$

5.1.2 *Fixed wildtype cross immunity*

Other epidemics runs have been launched with fixed cross-immunity for the wildtype strain, setting $\epsilon_W = 0.8$. The resulting plots are shown in Figure 5.2. Here, a transition is clearly visible in both the total attack rate and the ratio of attack rates: as the former approaches 1, there is a rapid increase in the latter. Looking at Figure 5.2b, we can see that, for the range of parameters corresponding to the region in the bottom right corner, the emerging strain overcomes the wildtype and becomes dominant. There is also a region in the top left corner where the converse happens, and the wildtype strain is dominant.

By looking at Figure 5.2*a*, we can see that where the emerging strain is dominating the total attack rate reaches very high values ($\approx 0.8 -$

0.9), while in the opposite region it is much lower ($\approx 0.4 - 0.5$). This difference can be explained by reminding the role of the parameters ϵ_i , who are tuning the strength of the within-host interaction between V_i (the same viral variant as the parameter) and I_j (the immunity for the other strain, following the same convention as before $i \neq j, i, j \in \{W, E\}$). In this investigation, ϵ_W has been fixed to a high value while ϵ_E is changed. This means that when an host develops an immunity for the emerging strain, he/she is also protected from the wildtype strain, but when he/she has an immunity for the wildtype (as is the case of experienced hosts), its effect on the emerging strain will depend on the value of ϵ_E . Therefore, when ϵ_E takes high values, it will be much more difficult for the emerging virus to infect experienced hosts, and this will result in a lower attack rate for the emerging.

This reasoning explains the two distinct dominance regions, as once an host has been infected with one strain, a subsequent infection of the other strain is more difficult because of the cross-immunity. By looking at Figure 5.2c we are able to confirm this intuition, as the fraction of individuals who experience both infections is quite low far from the boundary of the two dominance regions.

5.2 THE EFFECT OF THE CONTACT FREQUENCY

After these first two investigations, we started to consider other factors who might have an influence on the transmission of the two strains. Since the case of fixed cross-immunity for the wildtype strain shows, as explained above, a distinct transition between the two regions of dominance of the strains, it has been decided to keep using the same parametrization, while changing other parameters. Among them, the transmission interval Δt_T , or contact frequency $\frac{1}{\Delta t_T}$, has been identified as most promising.

5.2.1 *Redrawing the picture*

Before running other simulations with different contact frequencies, a procedure has been devised to highlight the regions of dominance of the two strains, in order to make the plot of the attack rates clearer. One natural way to define such regions is by using the ratio of the attack rates, which will be called r from now on. As can be seen from Figure 5.2b, this ratio goes from below 1, when the wildtype strain is dominating, to values higher than 10 when it is the emerging virus to dominate. While these two extrema are clear, the value at the boundary isn't so easy to find. To solve this problem, three regions have been defined based on the value of r:

- for $r < r_{L}$, we are in the *wildtype dominant* region (WD)
- for $r \ge r_U$, we are in the *emerging dominant* region (ED)

• for $r_L \le r \le r_U$, we call this the *co-dominance* region.

The values of r_L and r_U , the lower and upper bounds of the co-dominance region, have been chosen to be $r_L = 0.5$ and $r_U = 2$ by visual inspection of plots of the ratio of attack rates. The boundaries defined by these values are shown in Figure 5.3 superimposed to the plot in Figure 5.2b. In successive plots, such as in Figure 5.4, only the co-dominance region is shown, implying that the region to the top-left is the WD and the one to the bottom-right is the ED region.

5.2.2 *Changing the contact frequency*

The effect of different contact frequencies, or contact intervals, has been investigated by repeating the procedure of section 5.1.2 for different values of the contact intervals, then drawing the co-dominance region as explained in the previous subsection. In this way, the plots in Figure 5.5 have been produced. From these figures, it is clear that changing the frequency of contact between the hosts modifies the boundary of the co-dominance region, which becomes larger with higher frequencies. Since the lower limit of this region is related to where the emerging strain becomes dominant, it follows that raising the contact frequency makes the ED region shrink, yet it doesn't change very much the WD boundary. Therefore, we can say that the interval of contact between hosts can influence the selection of the viral strain.













Figure 5.2: Heat maps for the case of fixed wildtype cross-immunity parameter ϵ_W = 0.8, ρ_W = 1



Figure 5.3: Relative attack rate plot from previous figure with the boundaries of the co-dominance region drawn in black



Figure 5.4: Relative attack rate plot from previous figure with co-dominance region highlighted



(a) With $\Delta t_T \approx 3$ hours in green and $\Delta t_T \approx 12$ hours in red



(b) With $\Delta t_T \approx 6$ hours in orange, $\Delta t_T \approx 12$ hours in red

Figure 5.5: Relative attack rate plots, with co–dominance region (with r_L and r_U as defined in the text) drawn for different transmission intervals

CONCLUSIONS

This thesis has been centered on modeling the emergence of a new viral strain of influenza at both the within-host scale and the between-host scale. In order to explore the phenomenon and find possible outcomes, a computational model based on stochastic simulations has been developed. This involved first adapting the within-host dynamic, transforming a set of dynamical equations into a simulation with stochastic update rules, than designing and implementing the between-host transmission dynamic.

This expansion of the simple simulation which had been previously built for an individual host required the use of networks for the description of the contacts of the population, and the definition of new rules for the exchange of viral particles between individuals. At this stage, the bottleneck factor, the transmission interval and the probability of contact were introduced in order to describe the transmission of the virus.

Afterwards, the model was characterized with a single virus. A great time and effort have been devoted to this stage, to both understand the dynamics of the model and tune it in order to recreate influenza dynamics. To achieve this goal, a correspondence between simulation time and real time needed to be found, as it was essential to understand the significance of single-host statistical quantities that were measured during the simulations. On the other hand, the measurement of population level statistics showed that it necessary to introduce new features of the model: a log-normally distributed bottleneck factor in the betweenhost transmission, a probability of contact derived from the average degree of an individual, and a fraction of the hosts immune to the wildtype at the beginning of the simulation. At the end of this stage, the parameters values were chosen in order to yield a realistic epidemic of influenza.

Finally, the emerging strain was introduced into the model. After an initial twofold exploration with different kinds of parametrization, we have seen that changing together the replication rate of the emerging strain and the ratio of the cross-immunity parameters leads to the model exhibiting two main regions, one where the wildtype strain dominates, and one where the emerging is dominating. We have showed that changing the contact frequency leads to a change in the emerging dominance region. This indicates that it is important to model the interaction between individuals in a realistic way in order to understand the mechanisms of viral emergence at the population level.

- [1] Guillaume Béraud et al. "The French Connection: The First Large Population-Based Contact Survey in France Relevant for the Spread of Infectious Diseases". In: *PLOS ONE* 10.7 (July 15, 2015), e0133203. ISSN: 1932-6203. DOI: 10.1371/journal.pone.0133203.
- [2] E. G. M. Berkhoff et al. "Fitness costs limit escape from cytotoxic T lymphocytes by influenza A viruses". In: *Vaccine*. Proceedings of the Second European Influenza Conference 24.44 (Nov. 10, 2006), pp. 6594–6596. ISSN: 0264-410X. DOI: 10.1016/j.vaccine. 2006.05.051.
- [3] Laetitia Canini et al. "Heterogeneous shedding of influenza by human subjects and its implications for epidemiology and control". In: *Scientific Reports* 6 (Dec. 14, 2016), p. 38749. ISSN: 2045-2322. DOI: 10.1038/srep38749.
- [4] Fabrice Carrat et al. "Time Lines of Infection and Disease in Human Influenza: A Review of Volunteer Challenge Studies". In: *American Journal of Epidemiology* 167.7 (Apr. 1, 2008), pp. 775–785. ISSN: 0002-9262. DOI: 10.1093/aje/kwm375.
- [5] Sarah Cobey and Scott E Hensley. "Immune history and influenza virus susceptibility". In: *Current Opinion in Virology*. Emerging viruses: intraspecies transmission Viral immunology 22 (Feb. 1, 2017), pp. 105–111. ISSN: 1879-6257. DOI: 10.1016/j.coviro.2016. 12.004.
- [6] A. Cori et al. "Estimating influenza latency and infectious period durations using viral excretion data". In: *Epidemics* 4.3 (Aug. 1, 2012), pp. 132–138. ISSN: 1755-4365. DOI: 10.1016/j.epidem.2012.06.001.
- [7] C. S. David. "The computer generation of multinomial random variates". In: *Computational Statistics and Data Analysis* 16 (1993), pp. 205–27.
- [8] P. Erdős and A. Rényi. "On Random Graphs I". In: Publicationes Mathematicae 6 (1959), pp. 290–297.
- [9] GSL GNU Scientific Library. https://www.gnu.org/software/ gsl/. Last Accessed: 1/07/2019.
- [10] Influenza. In: Wikipedia. Page Version ID: 904423468. July 2, 2019.
- [11] Influenza (Seasonal) World Health Organization. https://www. who.int/en/news - room/fact - sheets/detail/influenza -(seasonal). Last Accessed: 3/07/2019.

- [12] Michal Mandelboim et al. "Significant cross reactive antibodies to influenza virus in adults and children during a period of marked antigenic drift". In: *BMC Infectious Diseases* 14.1 (June 20, 2014), p. 346. ISSN: 1471-2334. DOI: 10.1186/1471-2334-14-346.
- K. M. Pepin et al. "Phenotypic differences in viral immune escape explained by linking within-host dynamics to host-population immunity". In: *Journal of Theoretical Biology* 265.4 (Aug. 21, 2010), pp. 501–510. ISSN: 0022-5193. DOI: 10.1016/j.jtbi.2010.05.036.
- [14] Åke Svensson. "A note on generation times in epidemic models". In: *Mathematical Biosciences* 208.1 (July 1, 2007), pp. 300–311. ISSN: 0025-5564. DOI: 10.1016/j.mbs.2006.10.010.
- [15] Types of Influenza Viruses. https://www.cdc.gov/flu/about/ viruses/types.htm. Last Accessed: 3/07/2019.