Mutating agents on complex networks

Author: Pol Pastells
Facultat de Física, Universitat de Barcelona, Diagonal 645, 08028 Barcelona, Catalonia, Spain.

Advisor: Marián Boguñá

Abstract: In this work we consider a variation of the SIR model where mutations of the pathogen are allowed. We discuss the appearance of an endemic phase using a mean-field approximation and solve the stochastic model using Monte Carlo simulations. We find that our model has three distinct phases and serves as an interpolation between the SIR and SIS models.

I. INTRODUCTION

An epidemic disease is a contagious disease caused by a biological pathogen that spreads from person to person, e.g. smallpox, polio, rubella, chickenpox, influenza, measles, and sexually transmitted diseases. Dynamical population models are used in the study of epidemic diseases [1] to form hypotheses, predict average behavior or investigate mechanisms across many fields of science such as epidemiology, immunology, and ecology. The course of an epidemic spreading is not only determined by the properties of the pathogen carrying it, but also by network structures within the population it is affecting [2]. In a social network, each person is assigned a node in a graph, and there is an edge between two nodes if two people are capable of transmitting the disease to each other. The disease spreads from the affected to the unaffected via contact infection, and therefore the network will be modified for different types of pathogens. For example, the networks describing the propagation of influenza and HIV will be different. The second one is much sparser, with fewer pairs connected by a node.

Epidemic modeling also translates into other domains. Phenomena described via population dynamics qualitatively similar to that in the spreading of infections include political and religious beliefs, innovations, fads, or fanatic behaviors; dynamics of traits such as the diffusion of ideas [3], the spread of computer viruses [4] or the propagation of rumors. They involve states heavily influenced by peer interaction, where a contagion analogy can be made. The transmission of information is particularly interesting, due to the intentional acts by both the sender and the receiver, as opposed to disease spreading. These interactions can lead to epidemic-like outbreaks including cascades or memes going viral online [5].

II. STANDARD EPIDEMIC MODELS

Most epidemic models belong to either one of the two main families of compartmental models [6], the SIR and SIS models. In these models the population is divided into compartments representing different states, and rules to govern the transitions between states are given. All the properties of the pathogen are translated into mathematical parameters, with the possibility of including a wide range of factors [7], e.g. immunity, vaccination, vital dynamics, and age.

The *Susceptible-Infectious-Recovered* (or Removed), SIR model, describes the behavior of a pathogen that the individuals contract at most once. Given an undirected graph representing a certain population structure, an individual node goes through three potential stages, compartments, during the course of its sickness:

- *Susceptible.* Every node, except some needed to initialize the process, is assumed to be susceptible to the disease at the beginning.
- *Infectious.* If the node gets infected, it becomes infectious and has some probability of infecting each of its susceptible neighbors, given by a rate of infection λ.
- *Recovered.* Each infected individual recovers spontaneously at a constant rate δ. Once recovered from a single infection, complete immunity is granted.

The SIR model has a further interpretation, instead of recovering, the individuals die, henceforth are removed. This means that the SIR model can represent two different kinds of situations. Nevertheless, we will stick to the recovery interpretation for our work.

The *Susceptible-Infectious-Susceptible*, SIS model, allows the individuals to contract the disease more than once, instead of granting them immunity after a single contagion. In contrast to the SIR model, the SIS model may reach a stationary state, making it ideal for the application of many theoretical approaches.

Both models have a phase transition with the ratio of the rates of infection and recovery λ/δ as the order parameter, as shown in Fig. 1. For the SIS model, the two phases distinguish the case where the virus dies out from the one where it stays alive and a stationary state is reached, with a constant prevalence ρ_{st} over time (in the limit of $N \to \infty$, with N the number of nodes), the latter is also called the *endemic phase*. Instead, for the SIR model, no stationary state is reached. In one phase, a macroscopic number of recovered individuals is reached, those who have been infected at some point, while below a critical value the agent dies out before reaching this level.

Several modifications to both models exist [8, 9], like the *SIRS model*, where a stage of immunity happens be-

fore going back to being susceptible, leading to interesting dynamics such as periodic prevalence oscillations [10]; the *SEIR/SEIS model*, with an exposed (E), or latent period before the infected one; or the *MSIR/MSIS model*, where to model passive, or maternally (M) derived immunity, an extra initial compartment is included.

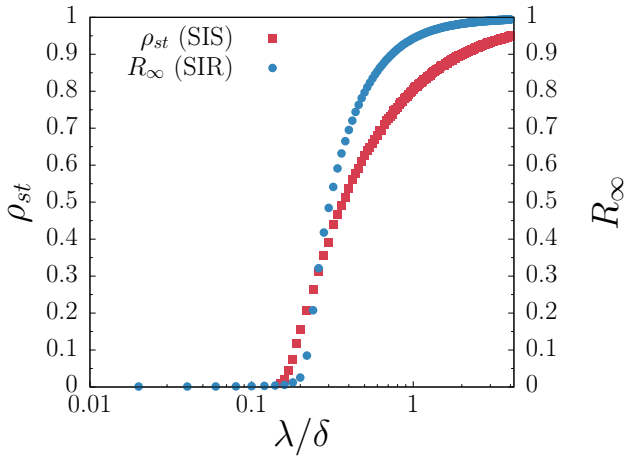

Fig. 1: Averages over 100 realizations for the stationary prevalence (SIS) and total ratio of recovered individuals (SIR) with respect to λ/δ. For the SIS model ρ becomes zero below $\lambda/\delta = 1/\langle k\rangle$, while for the SIR model $R_\infty \simeq 0$ below $\lambda/\delta = 1/(\langle k\rangle - 1)$.

III. SIR MODEL

We have based our work on the SIR model (Fig. 2). We define the rate of recovery $\delta \equiv 1$, so from now on we will not worry about it, and the rate of infection becomes $\lambda \equiv \lambda/\delta$.

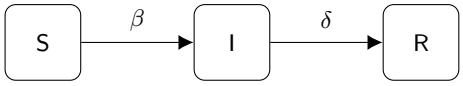

Fig. 2: Flow chart for on node in the SIR model. In the mean-field approximation $\beta = \langle k\rangle\lambda$.

Given that the dynamics of an epidemic, for example influenza [11], are usually much faster than the dynamics of birth and death, these will be omitted hereafter.

For an uncorrelated homogeneous network, or in the mean-field approximation, the equations that describe it are:

$$\begin{aligned}\frac{dS(t)}{dt} &= -\lambda\langle k\rangle\rho(t)S(t)\\ \frac{d\rho(t)}{dt} &= -\rho(t) + \lambda\langle k\rangle\rho(t)S(t)\\ \frac{dR(t)}{dt} &= \rho(t)\end{aligned} \tag{1}$$

$S(t)$, $R(t)$ and $\rho(t)$ are the ratios of susceptible, recuperated and infectious individuals, the latter also called prevalence. t is the time in units of $1/\delta$ and $\langle k\rangle$ is the average number of connections per node.

If the number of nodes is constant, there is no vital dynamics, then

$$S(t) + \rho(t) + R(t) = 1, \tag{2}$$

The system is non-linear, without a general analytical solution. However, significant results can be analytically derived or simulated with Monte Carlo methods, such as the Gillespie algorithm (see VI).

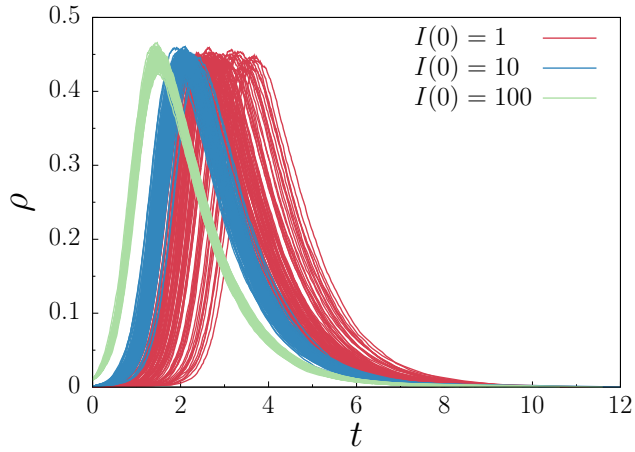

Fig. 3: Prevalence with respect to time for 100 SIR realizations ($\lambda = 1.0$), starting with 1 infected individual (red), 10 (blue) and 100 (green).

If we initialize the population with a microscopic number of infected individuals, $\rho(0) \simeq 0$, $S(0) \simeq 0$, $R(0) = 0$, we obtain:

$$S(t) = e^{-\lambda\langle k\rangle R(t)} \tag{3}$$

Together with (2), we find that the total number of infected individuals, $R_\infty = \lim_{t\to\infty} R(t)$, which is the same as the number of recovered ones when there is no infected nodes left, satisfies the self-consistent relation:

$$R_\infty = 1 - e^{-\lambda\langle k\rangle R_\infty} \tag{4}$$

The solution $R_\infty = 0$ always exists, but to get a second one we must have:

$$\frac{d}{dR_\infty}\left(1 - e^{-\lambda\langle k\rangle R_\infty}\right)\Big|_{R_\infty=0} \geq 1 \tag{5}$$

which is equivalent to $\lambda \geq \lambda_c = \frac{1}{\langle k\rangle}$.

λ_c is the epidemic threshold, below which the total number of infected individuals will be microscopic. The *basic reproduction number* $R_0 = \beta/\delta$, with $\beta = \lambda\langle k\rangle$ in the mean-field approximation, is defined as the number of infections one infected individual generates on average over the course of its infectious period, in an otherwise uninfected population. It is useful to determine whether a pathogen can spread through a population. For $R_0 < 1$,

the virus will die out in the long run, while for $R_0 > 1$ the infection will be able to spread. Notice that $R_0 = 1$ implies $\lambda\langle k\rangle = 1 \quad \Rightarrow \quad \lambda = \lambda_c$.

We can do better by noticing that in the SIR model any infected individual, except the initial ones, must have been infected by a neighbor, therefore one less edge than the average should be considered, consequently:

$$\lambda_c = \frac{1}{\langle k\rangle - 1} \tag{6}$$

A similar result can be found for a generic complex network [12], taking into account the degree distribution and possible correlations between nodes.

Depending on the number of initially infected individuals we obtain a slightly different behavior (Fig. 3). If the number is little enough, some realizations will not even reach a macroscopic prevalence. Indeed, starting with a single infectious node there is a phase transition with λ, below a certain value, no realization reaches the macroscopic level of infection, while over it, the probability of a realization generating a substantial epidemic is approximately $1 - 1/R_0$ [1], which grows similarly to Fig. 1. For this work, we have arbitrarily chosen to start all the simulations with 10 infected nodes.

Likewise, from now on we assume an Erdős-Rényi (ER) network with $N = 10^4$ and $\langle k\rangle = 6$. The ER model [13] encompasses all the random networks with a given average degree $\langle k\rangle$. The resulting networks are characterized by a Poisson degree distribution. It is also the *quenched* (all edges are fixed in time) network structure closer to the mean-field description.

IV. MODELING MUTATIONS

One interesting generalization of the SIR model is the allowance of pathogen cohabitation. It is well known that viruses mutate, so several strains of a virus can inhabit the same population. Flu season happens during winter in each hemisphere, yet it is an unsolved puzzle why this is. It has been suggested [14] that the high mutation rate of influenza may promote the rapid evolution of the virus in nature. Temperature changes, rainy seasons or the fact that people spend more time inside could also be pieces of an explanation.

Here we propose an extension of the SIR model to represent a mutating agent, introducing a new parameter, the individual rate of mutation ε. We label the different mutations (strains) with numbers, starting with a single one in the beginning. A mutation event happening after n different strains makes an infected node go from being infected by the strain i to being infected by a new strain $n+1$, recovering from the previous strain, i, in the process. In other words, every new mutation is different from all the previous ones, without restricting the number of different strains. It is not sufficient to say that an individual is recovered anymore, now we have to specify to which strains it is immune, while it stays susceptible to all the other ones. We do restrict the number of strains an individual can be infected of at once, to just one. Otherwise, simulation of the fully stochastic SIR model with mutations is intractable.

Moreover, this model provides a bridge between the standard SIR and SIS models. For when $\varepsilon = 0$ we just have the SIR model, and if the rate of mutations is big enough, once an individual recovers from whichever strain he was infected with, no other individual will have that same strain any longer, so we recover the SIS model (Fig. 4). The model is not independent of the network size anymore. Certainly, a population 10 times larger than another, with the same individual rate of mutation ε, will get about 10 times more different strains, so we will use εN as a parameter instead of ε.

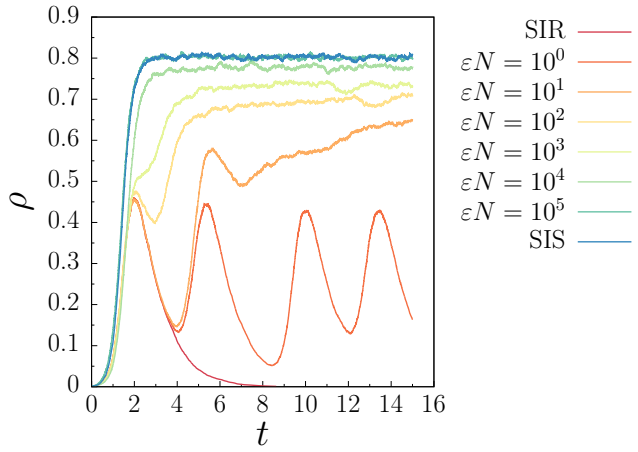

Fig. 4: Prevalence with respect to time for a range of εN, in the SIR model with mutations; SIR and SIS models. All with $\lambda = 1.0$. For $\varepsilon N \gtrapprox 10^5$ we recover the SIS model.

We will have two possible distinguished regimes:

- There is a finite number of mutations and the population ends up recovering from all of them (Fig. 5).
- In the limit of infinite time, an infinite number of mutations develop and the virus becomes endemic (Fig. 6)

We are interested in which of these regimes (phases) the system will be as a function of our parameters, λ and εN. To obtain the phase diagram (Fig. 8) we perform a series of simulations (Fig. 7). For every λ we search for the minimum εN that achieves:

1) at least half of the realizations are still alive after a set time $t = 400$.

2) the same number of alive realizations after $t = 200$ and $t = 400$.

For either criterion, each combination of parameters is simulated 100 times, with different seeds for the Monte Carlo steps. In the $N \to \infty$ limit, we expect them to agree, with all the realizations alive for $\varepsilon \geq \varepsilon_c(\lambda)$ after any set amount of time, and none for $\varepsilon < \varepsilon_c(\lambda)$.

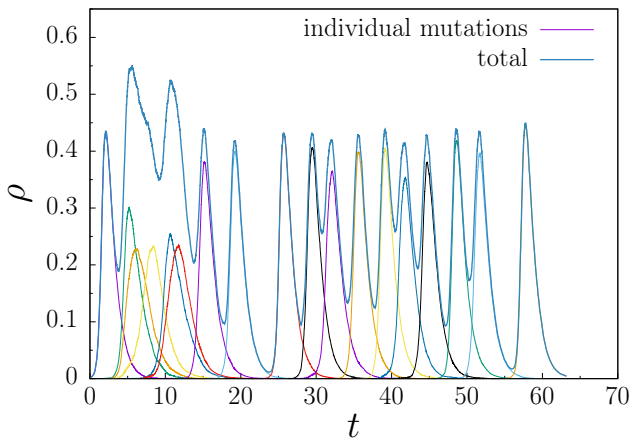

Fig. 5: Total prevalence and prevalence for each strain with respect to time for the SIR model with mutations ($\lambda = 1.0$, $\varepsilon N = 1.5$).

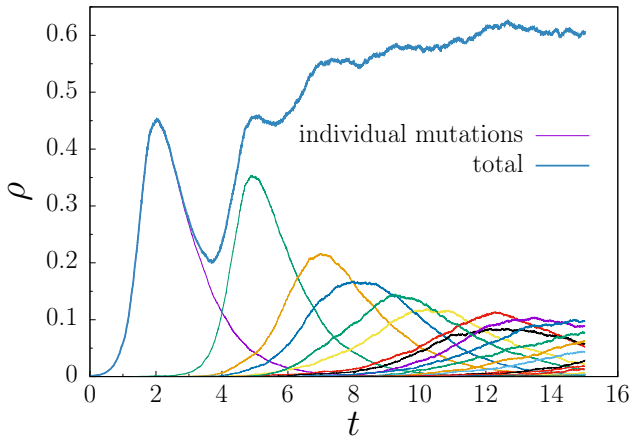

Fig. 6: Total prevalence and prevalence for each strain with respect to time for the SIR model with mutations ($\lambda = 1.0$, $\varepsilon N = 10$).

In the same spirit of the basic reproduction number we search for a *basic mutation number* M_0. It should fulfill the conditions that:

$$\begin{cases} \text{For } M_0 < 1, \text{ the virus will die out in the long run,} \\ \text{while for } M_0 > 1, \text{ the infection will become endemic.} \end{cases}$$

For a single SIR realization, there is a total of NR_∞ infected individuals, of which about εNR_∞ will mutate during its infection. Of this, only a fraction $P_\infty \approx 1 - 1/R_0$ will spread the new mutation to a substantial part of the population. Then,

$$M_0 \equiv P_\infty \varepsilon N R_\infty. \tag{7}$$

The εN that fulfills $M_0 = 1$ is:

$$(\varepsilon_c N)_{th} = \frac{1}{R_\infty P_\infty}. \tag{8}$$

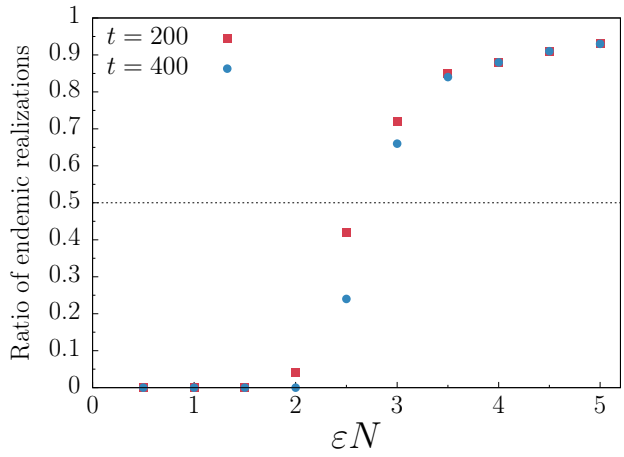

Fig. 7: Ratio of endemic realizations after $t = 200$ and $t = 400$ with respect to εN, with $\lambda = 1.0$.

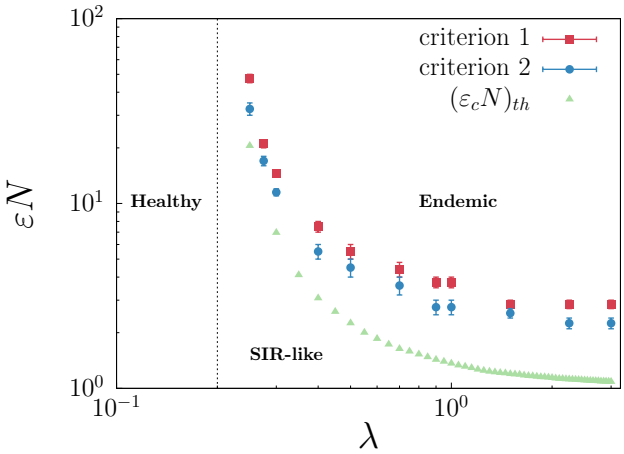

Fig. 8: Phase diagram for the SIR model with mutations. The two criteria explained above are represented. We compute $(\varepsilon_c N)_{th}$ via simulations, averaging R_∞ and P_∞ over 100 realizations. Below λ_c (6) no realization will reach a macroscopic level of infections (healthy phase), over λ_c, but below $\varepsilon_c(\lambda)$ there can be oscillations ending up with the epidemic disappearing (SIR-like phase), on top of both λ_c and $\varepsilon_c(\lambda)$ an (endemic phase) exists.

$(\varepsilon_c N)_{th}$ is the minimum εN that will achieve an endemic phase. It needs to be one that just achieves a principal strain at each time, similar to Fig. 5. Note that it is a function of λ.

As can be seen in Fig. 8, $(\varepsilon_c N)_{th}$ is a crude lower bound estimation of $\varepsilon_c(\lambda)N$. Although it predicts the $(\lambda - \lambda_c)^{-1}$ behavior separating the endemic and SIR-like phases, it does not characterize the tendency for big λ's, where both criteria tend to a constant $\varepsilon_c N$. Thus a finite fixed rate of individual mutations is needed to reach the endemic phase, which could be interesting for epidemiologists.

V. CONCLUSIONS

We have described a SIR model with a single mutating pathogen that encompasses both the SIR and SIS models in the limits of no mutations and infinite mutations, respectively. Instead of two distinct regimes, like for the SIR and SIS models, we have three different phases. One healthy phase where the pathogen does not spread, a SIR-like phase with multiple oscillations and an endemic phase that reaches a stationary prevalence ρ_{st}.

The fact that the parameter describing the onset of the endemic phase is εN instead of just ε is of paramount importance, for in the $N \to \infty$ limit, any mutation rate will be enough to produce an endemic infection.

Application to scale-free networks could be interesting. Even though they have an absence of an epidemic threshold, the individual mutation rate ought to be enormous to achieve an endemic phase for small λ's. Further understanding of the phase transition between the SIR-like and endemic phases could be achieved using bigger systems in our simulations. However, a proper theoretical understanding of the model would certainly help guiding the computational work.

VI. APPENDIX: GILLESPIE ALGORITHM

To simulate the behavior of our system of equations we have used two facts:

- The processes are memoryless, with the probability of an event predicted only by the state attained in the previous event. Consequently, they are Markov processes.
- The assumption that all the processes we are interested in are described by independent Poisson distributions. We can then use the following result: If $X_i \sim P(\mu_i)$ for $i = 1, .., n$ are independent, and $\mu = \sum_i^n \mu_i$, then $Y = (\sum_i^n X_i) \sim P(\mu)$.

A system of discrete Markovian stochastic processes can be accurately simulated using the Gillespie algorithm [15, 16].

A Monte Carlo step of the algorithm consists of:

1. Generate a random number for the next time interval from $\tau \sim Exp(\mu) = \mu e^{-\mu\tau}$. So instead of discretizing time with a fixed step, we use the exact interval from the probability distribution.
2. Generate a second random number to choose which event will come about from $i \sim \frac{\mu_i}{\mu}$.
3. Once the event has occurred, the changes in all μ_i have to be accounted for in the next iteration.
4. Iterate.

A. SIR example

We have two independent processes:

- Infection: $\mu_{inf} = \#_{active\ edges} \times \lambda$
- Recovery: $\mu_{rec} = \rho \times N$

Where we have defined an *active edge* as an edge between an infected and a susceptible node.

At every step of the algorithm $\mu = \mu_r + \mu_i$. If a node is recovered, ρ decreases, and the number of active edges may decrease, for every neighbor that is susceptible, by one. Likewise, if a node gets infected, ρ increases, and the number of active edges can either increase, decrease or stay the same. For every susceptible neighbor, they increase by one, while they decrease for infected neighbors. Therefore $\Delta_{active\ edges} = \#S_{nn} - \#I_{nn}$.

Acknowledgments

I would like to thank Dr. Marián Boguñá for his guidance and patience during the last months.

[1] N. T. Bailey, *et al.*, *The mathematical theory of infectious diseases and its applications* (Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975). I, III
[2] D. Easley, J. Kleinberg, *et al.*, *Networks, crowds, and markets*, vol. 8 (Cambridge university press, 2010). I
[3] L. M. Bettencourt, A. Cintrón-Arias, D. I. Kaiser, C. Castillo-Chávez, *Physica A* **364**, 513 (2006). I
[4] R. Pastor-Satorras, A. Vespignani, *Physical review letters* **86**, 3200 (2001). I
[5] J. H. Fowler, N. A. Christakis, *PNAS* **107**, 5334 (2010). I
[6] W. O. Kermack, A. G. McKendrick, *PNAS A* **115**, 700 (1927). II
[7] R. M. Anderson, B. Anderson, R. M. May, *Infectious diseases of humans* (Oxford university press, 1992). II
[8] H. W. Hethcote, *SIAM review* **42**, 599 (2000). II
[9] N. Masuda, N. Konno, *J. Theor. Biol.* **243**, 64 (2006). II
[10] M. Girvan, D. S. Callaway, M. E. Newman, S. H. Strogatz, *Physical Review E* **65**, 031915 (2002). II
[11] N. M. Ferguson, *et al.*, *Nature* **437**, 209 (2005). III
[12] M. Boguñá, R. Pastor-Satorras, A. Vespignani, *Statistical mechanics of complex networks* (Springer, 2003), pp. 127–147. III
[13] P. Erdős, A. Rényi, *Publ. Math. Debrecen* **6**, 290 (1959). III
[14] J. D. Parvin, A. Moscona, W. Pan, J. Leider, P. Palese, *Journal of virology* **59**, 377 (1986). IV
[15] D. T. Gillespie, *J. Comput. Phys.* **22**, 403 (1976). VI
[16] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977). VI