NYILATKOZAT

Név: Bognár Edit

ELTE Természettudományi Kar, szak: Alkalmazott Matematikus Msc.

NEPTUN azonosító: ZK61DT

Szakdolgozat címe:
Spreading models on random graphs

A **szakdolgozat** szerzőjeként fegyelmi felelősségem tudatában kijelentem, hogy a dolgozatom önálló szellemi alkotásom, abban a hivatkozások és idézések standard szabályait következetesen alkalmaztam, mások által írt részeket a megfelelő idézés nélkül nem használtam fel.

Budapest, 2020.05.30.

Bognár Edit

a hallgató aláírása

Eötvös Loránd University

Faculty of Science

Edit Bognár

Spreading models on random graphs

Thesis submitted for the degree of
MSc in Applied Mathematics

Supervisor:
Ágnes Backhausz
Department of Probability Theory and Statistics

Budapest, 2020

Acknowledgement

First of all, I would like to express my gratitude to my supervisor, Ágnes Backhausz for her valuable guidance as well as amount of time and effort spent on consultations throughout many years.
I am also very grateful to my family and friends for their additional support and motivation.

Contents

Introduction

Spreading models are a subclass of interacting particle systems, creating precise mathematical models with the help of probability theory to numerous real life processes arising in physics, biology and economics. In this thesis we examine two similar stochastic processes, where individuals spread their opinions or a virus between one another depending on relationships or social contacts between them. The main difference between them is that virus transmission is only possible through direct contacts, while opinion can spread between any two individuals. The main interest in this thesis is epidemic spread; as we can experience that in the past as well as nowadays, dealing with virus spread is still difficult. Mathematical models aim to describe the likely outcome of an epidemic and calculate the effect of different public health interventions such as mass vaccination campaigns or restriction measurements. There exist several approaches to model mathematically epidemics, such as using differential equations [18], branching processes [1] and the theory of random graphs [23]. Real life networks possess rather complex structure, as role, degree and type (according to some aspect) of a node can be varying and thus having a great influence in the outcome of the process. For this reason, in this thesis we give a possible implementation to the discretized processes on an underlying random graph, describing social contacts of individuals. It is widely studied that the structure of the graph can have an important impact on the course of the epidemic, thus properties such as degree distribution are essential to understand the dynamics. In our implementation of the model, age of an individual is also taken into consideration, since contact behaviour as well as reaction to a specific disease often depends on age. We use collected statistics and basic assumptions both in the creation of random graphs and in the setting of parameters of the disease.

A variant of the voter model is also studied in this thesis. Voter model is typically used for population dynamics, but also related to epidemics as well, since Durett and Neuhauser [12] applied the voter model to study virus spread. The two processes can be connected by the following idea: We can see virus spread as a special case of the voter model with two different opinions (healthy and infected), but only one of the opinions (infected) can be transmitted, while any individuals with infected opinion switch to healthy opinion after a period of time. Also the virus can spread only through direct contacts of individuals (edges of the graphs), while in the voter model it is possible for the particles to influence one another without being neighbors in the graph. The voter model is usually examined

on the d-dimensional integer lattice, however as in case of epidemics, we also simulate the process on random graphs. Codes of random graphs and processes were written in Matlab and can be found at webpage [24].

The outline of the thesis is as follows. In the first three chapters we examine epidemics described by compartment models: In the first chapter we define $SEIR$ epidemics, where disease transmission is possible between any two individuals of the population. As proved in this chapter is possible to couple the epidemic model with a branching process. The analytical study of virus spread is often achieved by branching processes, thus here we review the most important properties and theorems about dicrete and continuous time branching processes, as in the first two chapters we will strongly rely on these results. In the first chapter in Theorem 1.3.4 we also give an asymptotic property stating the branching process and the epidemic agree up to a time point tending to infinity in probability, as we consider a population growing to infinity. Thus later we can also give the probability of a major outbreak.

In the second chapter we study a special case of $SEIR$ epidemic with no incubation period, the SIR model on random graphs of different structures created by different methods: the complete graph, configuration model graphs and stochastic block model graphs are investigated. Again we strongly rely on theorems and properties of branching processes, since by generation counting infected individuals, we can find a suitable branching process to couple the virus spread with. Through this chapter we consider the same disease with Exponential infectious period of parameter γ, and disease transmitting independently only on the edges of the graph with Exponential rate λ. For each graph structure we compute the mean number of infections caused by an infecting individual through its infectious period in the beginning of the process. This quantity is usually referred as the R_0 of the disease, here proved to be strongly dependent not only on the characteristic of the disease but also measures describing underlying graphs, too.

In the third chapter we carry out simulations of the $SEIR$ process on age-grouped configuration model random graphs of different degree distributions. The underlying graphs are created in a way that they match the real-world data from [18] considering number of individuals in - and number of contacts between age groups. Parameters are chosen to replicate $Covid-19$ virus. We expand a possible implementation of the classical time-discretized $SEIR$ process with possible self-isolation of individuals and also introduction of restricting measurements corresponding to examine the process on an edge-reduced subgraph of the original network. We investigate overall attack rate of the virus and max-

imum number of infected individuals at the same time on graphs under different scenarios.

In the last chapter voter model is studied, however we also define generally interacting particle systems to connect spread of diseases with the voter model. We describe a possible implementation of the discretized process on graphs: analytical study as well as simulations on Erdős–Rényi and Barabási–Albert graphs are also carried out in this model.

Chapter 1

The stochastic $SEIR$ epidemic model in a closed homogeneous community

In this chapter firstly we define $SEIR$ process, and introduce some basic facts about stochastic epidemic models. Then in the second section we introduce the branching process and review the most important results without proofs in order to be able to examine the epidemic model with the help of branching processes according to Part I of book [1]. In the last section of the chapter we prove that in homogeneous mixing population suitable branching process approximations can be used to completely review the connection of the two processes. General results of branching processes also lay the foundation of the second chapter, since $SEIR$ on different random graphs will be originated in a suitable branching process.

1.1 Definition of the model

Compartmental models simplify the mathematical modelling of infectious diseases. In these models individuals are assigned to compartments with labels and might progress between them in time. These compartments corresponds to stages of a disease. The order of the labels usually shows the flow patterns between the compartments as seen in the following. In this section we introduce $SEIR$ (Susceptible - Exposed - Infected -Recovered) compartment model, one of the most frequently used epidemic model for diseases with the characteristic that the individuals who recover from the illness are immune, and no longer prone to the virus. Recovery confers lasting resistance in case of many diseases

such as measles, mumps and rubella. In the $SEIR$ category of models, individuals usually experience a long incubation duration (in the acronym E for "exposed" category), such that the individual is infected but not yet infectious (for example chicken pox). In this sense, SIR models can be interpreted as a special case of $SEIR$ with no incubation period, as examined later in the second chapter.

Through the document we consider the case of a closed community: without influx or born of new susceptibles, mortality and migration, so the size of the population remains fixed for the process.

Firstly we give a precise definition to the $SEIR$ process:

Definition 1.1.1 *The stochastic SEIR epidemic model in a closed homogeneous community. (Definition 1.1.1 in [1])*
Consider a closed population of $N+1$ individuals. At any point in time t each individual is either susceptible, exposed, infectious or recovered. Let $S(t)$, $E(t)$, $I(t)$ and $R(t)$ denote the numbers of individuals in the different states at time t ($S(t)+E(t)+I(t)+R(t) = N+1$ for all t). The epidemic starts at $t = 0$ in a specified state called the index case (often the state with one infectious individual), which corresponds to some individuals being externally infected, and the rest being susceptible. Then the process operates the following way: While infectious, an individual has infectious contacts according to a Poisson process with rate λ. Each contact is with an individual chosen uniformly at random from the rest of the population, and if the contacted individual is susceptible it becomes infected, otherwise the infectious contact has no effect. Individuals that become infected are first latent (called exposed) for a random duration (incubation period) L with distribution F_L, then they become infectious for a duration (infectious period) I with distribution F_I, after which they become recovered and also immune for the remaining time. All Poisson processes, uniform contact choices, latent periods and infectious periods of all individuals are defined to be mutually independent.

The epidemic goes on until the first time τ when there are no exposed or infectious individuals, $E(\tau) + I(\tau) = 0$. At this time no further individuals can get infected so the epidemic stops. The final state hence consists of susceptible and recovered individuals, and we let Z denote the final size, the number of individuals who got infected by the end of the epidemic excluding the index case(s): $Z = R(\tau) - I(0) = N - S(\tau)$.

The rate of infectious contacts is λ, so the rate at which one infectious has contact with a specific other individual is $\frac{\lambda}{N}$ since each contact is with a uniformly chosen other indi-

vidual. This way only the incubation and infectious period is independent of N, however the whole epidemic is not, thus we usually emphasise dependence of N with an index.

Some special cases of the model have received special attention in the literature. The model is called Markovian $SEIR$, if both L and I are exponentially distributed, while with $L = 0$ and exponential I choice the SIR process is known as General stochastic epidemic model examined later in the second chapter.

Only regarding the infectious period the two most studied special cases are when it is exponentially distributed and when it is nonrandom. To model real infections usually neither of these properties hold, and the approximating distribution is strongly dependent on the exact disease to be modelled. Later in the third Chapter when simulating the process on random graphs, we will also choose exponentially distributed latent period, and recovery period with a small variance distribution. To capture more reality, λ parameter can be defined as not a constant, but a function of time elapsed from the infection, however in this paper only time-independent λ is examined in analytical result as well as in simulations.

One of the most important quantities in epidemic models is the R_0 basic reproduction number, which denotes the mean number of infections caused by an infected individual before its recovery during the early stage of an outbreak. So it measures the average number of infections an infectious individual causes during its infectious period in a population of only susceptible individuals (without vaccination or previous immunity). It characterizes the intensity of the epidemic. R_0 has an important threshold value of 1 (critical case), restricting measures and vaccination strategies usually target the decrease of R_0 below this. If $R_0 > 1$, it suggests that an infected is "replaced" with more than one infectee after its recovery in the beginning of the epidemic, possibly resulting in the infection of a significant fraction of the population through the process, later referred as a major outbreak. With the same reasoning in case of $R_0 < 1$ we expect the virus to disappear after reaching negligible ratio of the community. The presumption above later will be engrossed as a theorem and proved with the help of branching processes in Corollary 1.3.5.

1.2 Branching Process

The stochastic and mathematical analysis of the spread of infectious diseases in large populations often relies on the theory of branching processes. Kendall in book [17] was the first in 1956 to suggest a branching process approximation for epidemics.

Branching processes are introduced as a model to describe family trees, where the simplifying assumption is that all individuals have the same probability of having k descendants, and the numbers of descendants of different individuals are independent. The classical model ignores important properties of real populations, such as changing circumstances, however the model has proved to be useful in many situations. The process with varying offspring distribution is also studied in book [16].

With branching process we can model the reproduction of organisms such as human beings, cells or neutrons, in addition in [11] Durrett used it to study the emergence of a giant component in Erdős–Rényi random graphs.

Branching processes are also useful to describe the spread of $SEIR$ epidemics, where an infection can be seen as a birth, with the infector being the ancestor and the infectee the descendant. In this model competition for resources is obvious, since once a susceptible individual gets infected, it cannot be infected ever again. However, if the population size N is large and the number of no longer-susceptible individuals is of $o(\sqrt{N})$, then suitable branching process approximations are very good according to paper [2] in different types of networks such as homogeneous mixing populations, configuration model network populations and multi-type population models. In the third section we prove this result in case of homogeneous mixing populations. When we later introduce epidemics models on graphs we will also rely on these results. In article [8] branching process is used to study epidemics with different vaccination strategies.

Firstly, I compass the reason why it is a natural idea and possible to describe the spread of $SEIR$ epidemics with branching process, then introduce the branching process and some important results without proof. In the next section then an approximation and also a totally equivalent definition is given to $SEIR$ with the help of branching process.

We consider the model with a large population N at the "beginning" of the process, when less than $k = k(N)$ individuals have been infected since $t = 0$. According to Definition 1.1.1 each infectious individual transmits the disease independently with rate λ, however dependence appears due to the fact that individuals become immune to the virus after recovering, so an individual can get infected at most once. Thus making an infectious contact with a previously infected individual (either in E, I or R) results in no further infection. Although, within a large community at the beginning of the outbreak two infectious individuals make contact with the same individuals only with a negligible probability. Therefore, we can approximate the number of infected individuals at the beginning of the outbreak by a branching process: In this branching process "being born"

means to get infected, and correspondingly "dying" is identical to a recovery of individual in the $SEIR$ model.

To examine analytic properties of the process, such as the probability that the branching process does not go extinct corresponding to the disappearance of the disease, it is practical to associate with the process its discrete-time skeleton. If in the $SEIR$ model the latent period is long in comparison to the infectious period, then we can divide infected individuals into generations: the first generation consists of the index cases, the second generation those who were infected by the index case(s), and so on. Since the number of offspring of an individual depends on the duration of the reproductive stage, this dependence has to appear in the discrete-time process: Conditional upon reproductive stage $I = y$, the number of descendants has Poisson distribution with parameter y, thus the unconditional distribution of number of offspring is mixed-Poisson. Therefore if we simply study the number of individuals born in each generation, then our branching process is a Bienaymé–Galton–Watson process with offspring distribution being $MixPoi(\lambda I)$.

In the following I give a precise definition first to discrete and continuous time branching process in general, and review some of the most important results without proof according to $A.1$ part of book [1], which will serve as a theoretical backbone to theorems about epidemics in the next section.

Definition 1.2.1 *Discrete time branching process.*
We consider that initially (at generation 0) an ancestor has ξ_0 children, according to some probability distribution $\mathbb{P}(\xi_0 = k) = q_k$. Each child of the ancestor belongs to generation 1. The i-th of those children has himself $\xi_{1,i}$ children, where the random variables $\{\xi_{k,i}, k \geq 0, i \geq 1\}$ are i.i.d., and all having the same distribution as ξ_0. We define X_n as the number of individuals in generation n: $X_{n+1} = \sum_{i=1}^{X_n} \xi_{n,i}$.

In the following we give the most widely-known result about the connection between probability of extinction and $m = \mathbb{E}(\xi_0)$ according to [1] book Proposition A.1.1. A proof can be easily constructed with the help of the generating function of X_n.

Theorem 1.2.2 *Let $g(z)$ denote the generating function of ξ_0, and $z_\infty = \mathbb{P}(X_n = 0)$ for some n.*

- *If $m \leq 1$, then $\mathbb{P}(X_n = 0) \to 1$ as $n \to \infty$, and $z_\infty = 1$.*
- *If $m > 1$, $\mathbb{P}(X_n = 0) \to z_\infty = q$ as $n \to \infty$, where q is the smallest solution of the equation $z = g(z)$.*

Now we consider the process in continuous time:

Definition 1.2.3 *Binary continuous time branching process.*
The process starts with a single ancestor born at time $t = 0$. This ancestor is characterized by its life length L_0 and the number of its children born on time interval $[0, t]$ denoted by $N_0(t)$, thus the pair $(L_0, \{N_0(t), t \geq 0\})$ describes him. We assume of course that the ancestor does not give birth after his death, so $N_0(\infty) = N_0(L_0)$. We now assume that the individuals are numbered in the order of their birth. To each individual i is attached a pair $(L_i, \{N_i(t)\})$, such that the sequence of pairs $\{(L_i, \{N_i(t)\})\}_{i\geq 0}$ is independent and identically distributed. If the individual i is born at time B_i, the offspring of individual i are born at the jump times of the process $\{N_i(t-B_i), B_i \leq t \leq B_i+L_i\}$. Since B_i depends only upon the pairs $\{(L_j, \{N_j(t)\})\}_{0\leq j<i}$, B_i and $(L_i, \{N_i(t)\})$ are also independent, because $\{(L_i, \{N_i(t)\})\}_{i\geq 0}$ is i.i.d.

Let X_t denote the number of individuals in the population alive at time t. We say the process to be Markovian if and only if the following properties hold:

- *the pair $(L_i, N_i(t))$ is such that L_i and $\{N_i(t), t \geq 0\}$ are independent.*
- *L_i is an exponential random variable with parameter d.*
- *$N_i(t)$ is a Poisson process with rate b.*

Now we define the two most commonly used values of branching processes describing the properties and intensity of the process in some sense. These parameters of the model play an important part in understanding and determining the behaviour of birth-death process.

Definition 1.2.4 *Malthusian parameter.*
The Malthusian parameter or the epidemic growth rate α of a branching process is defined as the mean number of births minus the mean number of death per unit time. In the super-critical case $(1 < m < \infty)$ α is the unique solution of

$$\int_0^\infty e^{-\alpha t} F(dt) = 1, \tag{1.1}$$

where $F(t) = \mathbb{E}(N(t))$.

We shall denote by X_t^k the number of descendants at time t of k ancestors at time 0. The branching property implies that $\{X_t^k, t \geq 0\}$ is the sum of k independent copies of $\{X_t, t \geq 0\}$.

Definition 1.2.5 *Mean number of offspring.*
Another important quantity is the mean number of offspring of each individual, which is equal to $m = \mathbb{E}(N_0(L_0))$*. The process* X_t^k *is said to be*

- *subcritical, if* $m < 1$ *or* $\alpha < 0$*.*
- *critical, if* $m = 1$ *or* $\alpha = 0$*.*
- *supercritical, if* $m > 1$ *or* $\alpha > 0$

In case of a Markovian process $m = \frac{b}{d}$ and $\alpha = b - d$. If $N(t)$ is Poisson process with rate λ, then $F(t) = \lambda(t)$ is a function of time in Equation (1.1).

Now we try to investigate the dynamics of the process characterized and significantly determined by the parameters above. We put emphasis on Markovian branching processes, since in the second chapter we study the Markovian SIR epidemics on graphs. However, we also consider non-Markovian or so called Crump–Mode–Jagers processes. We have the following results for X_t and X_t^k considering a Markovian branching process:

Theorem 1.2.6 *(Corollary A.1.3. in [1])*
With Markovian branching process we have that $\mathbb{E}(X_t^k) = ke^{\alpha t}$*, where* $\alpha = b - d$*.*

From the theorem above we can see that in the subcritical case $X_t^k \to 0$ in $L^1(\Omega)$, and it is also easy to show with the help of Markov inequality that $\lim_{t\to\infty} X_t^k = 0$ almost surely. (This also holds in the $m = 1$ critical case.)

We now study behaviour of Markovian X_t in the supercritical case, as $t \to \infty$. The next theorem is a key theorem in branching processes. It states that if the number of individuals grows large, then it roughly grows at a rate proportional to $e^{\alpha t}$, where t is the time since the population began, and α is the Malthusian parameter.

Theorem 1.2.7 *(Proposition A.1.4. from [1])*
If $m > 1$*, or equivalently* $\alpha > 0$*, there exists a non-negative random variable* W *such that* $X_t \sim We^{\alpha t}$ *almost surely, as* $t \to \infty$*. Moreover* $\{W = 0\} = \{\exists t > 0 : X_t = 0\}$ *and*

$$\mathbb{P}(W = 0) = \mathbb{P}(\{\exists t > 0 : X_t = 0\}) = \frac{d}{b} \tag{1.2}$$

Thus the branching process goes extinct with a probability of $\frac{d}{b}$*, which is the same as* $\mathbb{P}(\lim_{t\to\infty} X_t e^{-\alpha t} = 0)$*.*

In the non-Markovian case we also have similar, however weaker results considered the growth rate of X_t with $m > 1$.

Theorem 1.2.8 *(Proposition A.1.5. in [1])*

Let $F(t)$ denote $F(t) = \mathbb{E}(N(t))$ as before and $G(t) = \mathbb{P}(L \leq t)$. If $1 < m < \infty$, then there exists a unique $\alpha > 0$ such that

$$\int_0^\infty e^{-\alpha t} F(dt) = 1, \text{ and}$$

$$\mathbb{E}(X_t) \sim a e^{\alpha t}, \text{ where } 0 < a = \frac{\int_0^\infty (1 - G(t)) e^{-\alpha t} dt}{\int_0^\infty t e^{-\alpha t} F(dt)} < \infty.$$

1.3 The early stage of an outbreak

After possessing the necessary concept and properties of branching processes, I return to build up the connection between $SEIR$ models and the previous process along the first chapter of book [1]. We now define the branching process in question to show that the epidemic and branching process have similar distributions in the beginning.

The following notions will be used to construct the approximating branching process as well as the stochastic $SEIR$ epidemic for each N as follows:

Let $L_0, L_1, \ldots$ be i.i.d. latent periods having distribution F_L, and similarly let $I_0, I_1, \ldots$ be i.i.d. infectious periods having distribution F_I. Further, let $\xi_0(\cdot), \xi_1(\cdot), \ldots$ be i.i.d. Poisson processes having intensity λ, and let $U_1, U_2, \ldots$ be i.i.d. $U([0,1])$ random variables. All random variables and Poisson processes above are assumed to be mutually independent.

Definition 1.3.1 *The approximating branching process.(Definition 1.2.2.)*
At time $t = 0$ we start with one new born ancestor having label 0. Let the ancestor have childhood length L_0 and reproductive stage for a duration I_0 (so the ancestor dies at time $L_0 + I_0$), during which the ancestor gives birth at the time points of the Poisson process $\xi_0(\cdot)$. If the jump times of this Poisson process are denoted $T_{0,1} < T_{0,2} < \ldots$ and X_0 denotes the number of jumps prior to I_0, then the ancestor gives birth at the time points $L_0 + T_{0,1}, \ldots, L_0 + T_{0,X_0}$ (the set is empty if $X_0 = 0$). The first born individual is given label 1, and having childhood period L_1, reproductive period I_1 and birth process $\xi_1(\cdot)$ defined above. This individual gives birth according to the same rules (starting the

latency period at time $L_0 + T_{0,1}$), and the next individual born, either to individual 0 or 1, is given label 2 and variables L_2, I_2 and birth process $\xi_2(\cdot)$, and so on. This defines the branching process, and we let $L(t), I(t), R(t)$ respectively denote the numbers of individuals in the childhood state, in the reproductive state and dead, respectively, at time t. The total number of individuals born up to time t, excluding the ancestor/index case, is denoted by $Z(t) = L(t) + I(t) + R(t) - 1$ in the branching process, and the ultimate number ever born, excluding the ancestor, is denoted by Z which may be finite or infinite.

We now define the epidemic for any fixed N with the help of branching process (in the epidemic childhood corresponds to latent and reproductive stage to being infectious). This is done similarly as before with the exception that we now keep track of which individuals get infected using the uniform random variables $U_1, U_2, \ldots$.

Definition 1.3.2 *The stochastic SEIR epidemic with N initial susceptibles. (Definition 1.2.3.)*
We label the $N + 1$ individuals $0, 1, \ldots, N$, with the index case having label 0 and the others being labelled arbitrarily. As for the branching process, the index case is given latency period L_0, infectious period I_0 and contact process $\xi_0(\cdot)$ and the epidemic is started at time $t = 0$. The infectious contacts of the index case occur at the time points $L_0 + T_{0,1}, \ldots, L_0 + T_{0,X_0}$. The first infectious contact is with individuals $[U_1 N] + 1$. This individual, k say, then becomes infected (and latent) and is given latent period, infectious period and contact process L_1, I_1 and $\xi_1(\cdot)$. The next infectious contact (from either the index case or individual k) will be with individual $[U_2 N] + 1$. If the contacted person is individual k then nothing happens, but otherwise this new individual gets infected (and latent), and so on.

Infectious contacts only result in infection if the contacted individual is still susceptible. When a contact is with an already infected individual the branching process has a birth whereas there is no infection in the epidemic: we say a "ghost" was infected when comparing with the branching process. Descendants of all ghosts are also ignored in the epidemic. The epidemic goes on until there are no latent or infectious individuals. This will happen within a finite time, bounded by $\sum_{j=0}^{N}(L_j + I_j)$. The final number of infected individuals excluding the index case is as before denoted $Z^N \in \{0, ..., N\}$. Similar to before we let $L^N(t), I^N(t), R^N(t)$ denote the numbers of latent, infectious and recovered individuals at time t, and now we can also define the number of susceptibles $S^N(t) = N + 1 - L^N(t) - I^N(t) - R^N(t)$.

We now state two important results for these constructions of the branching process and epidemics.

Theorem 1.3.3 *(Theorem 1.2.4. in [1])*
Definition 1.3.2 above agrees with the earlier Definition 1.1.1 of the Stochastic $SEIR$ epidemic in a homogeneous community.

Proof. The equivalence of the two definitions can be easily seen by checking that the second definition satisfies all properties of the first: The latent and infectious periods have the desired distributions, and an infective has infectious contacts with others at overall rate λ, and each time such a contact is with a uniformly selected individual as desired.

□

We can not only give an alternative definition to the stochastic $SEIR$ model with the branching process, but also state an asymptotic property: The following theorem shows that the epidemic behaves like the branching process up to a time point T^N tending to infinity in probability as $N \to \infty$. This implies that we can use theory of branching processes to obtain results for the early part of the epidemic.

In the proofs of the following theorem we will strongly rely on propositions and theorems stated in the previous chapter about branching processes is general. However, in general Definition 1.2.1 and 1.2.3 of discrete and continuous time branching processes we mention and consider no childhood period of individuals. The idea of investigating the number of individuals born in each generation instead of continuous time was already introduced: In this representation of the process the length or even existence of childhood period corresponding to infected but not yet infectious period in the epidemics model makes no difference. Therefore we can lean on theorems of the second chapter without any restriction.

Theorem 1.3.4 *(Theorem 1.2.5. in [1])*
We let M^N denote the number of infections prior to the first ghost, so the number of uniformly selected individuals, before someone was reselected. (If this never happens we set $M^N = \infty$.) Also let T^N denote the time at which the first ghost appears (and correspondingly if this never happens we set $T^N = \infty$).

The branching process and N-epidemic agree up until T^N: $(L^N(t), I^N(t), R^N(t)) = (L(t), I(t), R(t))$ for all $t \in [0, T^N)$. Secondly, $T^N \to \infty$ and $M^N \to \infty$ in probability as $N \to \infty$. So processes defined in 1.3.2 and 1.3.1 are identical up to a time point T^N, which tends to infinity in probability as $N \to \infty$.

Proof. The first statement of the proof is obvious because of the alternative definition of SEIR epidemic with the help of branching process 1.3.2. The only difference between the epidemic and the branching process in our construction occurs after the first time when some individual is contacted again, since this results in new individual born in the branching process but no new infection in the epidemic. Thus the tho processes must agree until the appearance of the first ghost, noted by T^N.

As for the second part of the theorem we first compute the probability that M^N will tend to infinity, and then that the time T^N until the first ghost appears also tends to infinity.

Since $\mathbb{P}(M^N > k)$ can happen if and only if the first k contacts are with k distinct individuals, we can easily compute its value:

$$\mathbb{P}(M^N > k) = 1 \cdot \frac{N-1}{N} \cdot \frac{N-2}{N} \cdots \frac{N-k}{N} = \prod_{j=0}^{k} \left(1 - \frac{j}{N}\right)$$

For fixed k we see that this probability tends to 1 as $N \to \infty$. We can in fact say more. We have the following lower bound:

$$\mathbb{P}(M^N > k) = \prod_{j=0}^{k} \left(1 - \frac{j}{N}\right) \geq 1 - \sum_{j=0}^{k} \left(\frac{j}{N}\right) = 1 - \frac{(k+1)k}{2N}$$

As a consequence, we see that $\mathbb{P}(M^N > k(N)) \to 1$ as long as $k = k(N) = o(\sqrt{N})$. In particular $M^N \to \infty$ in probability as $N \to \infty$. In the following we use w.l.p abbreviation corresponding to "with large probability", meaning with a probability tending to 1 as $N \to \infty$.

This implies that all infectious contacts up to $k(N)$ will w.l.p be with distinct individuals, and therefore we are able to approximate the epidemic with the branching process for any $k(N) = o(\sqrt{N})$.

Let $Z(t)$ denote the number of individuals born before t in the branching process (excluding the ancestor) and $Z^N(t) = N - S^N(t)$ the number of individuals that have been infected before t (excluding the index case) in the N-epidemic. Since the epidemic and branching process agree up until T^N it follows that $Z(t) = Z^N(t)$ for $t < T^N$.

But, since $k(N) < M^N$ w.l.p. it follows that $\inf\{t : Z(t) = k(N)\} \leq T^N$ w.l.p. If the branching process is (sub)critical, then $Z(t)$ remains bounded as $t \to \infty$, so $T^N = \infty$ w.l.p. according to its definition.

Consider now the supercritical case. According to Theorem 1.2.8 the expected number of individuals alive at time t denoted as X_t can be approximated by $ae^{\alpha t}$, where $0 < a < \infty$

is a constant and where α as described in 1.1 solves the equation

$$\int_0^\infty e^{-\alpha s} F(ds) = \int_0^\infty e^{-\alpha s} \lambda(s) ds = 1, \tag{1.3}$$

since $F(ds) = \mathbb{E}(N(s))$, which we can calculate in the following: The function $\lambda(s)$ is determined by F_L and F_I, since $\lambda(s) = \lambda\mathbb{P}(\text{individual infectious at } s) = \lambda\mathbb{P}(L < s < L+I)$ is the rate at which an individual gives birth s units after being born.

However, we need to compute $Z(t)$ the number of individuals born ever before t in the branching process. We know that $Z(t) \leq X_t$, but we can estimate the expected value of $Z(t)$ by integrating X_s for $0 \leq s \leq t$. Thus we get

$$\mathbb{E}(Z(t)) \leq \int_0^t \mathbb{E}(X_s) ds \leq K e^{\alpha t},$$

for some suitable K constant. Since $Z(t) \geq 0$ and $Z(t) \in L^1(\Omega)$ we can use Markov inequality $\forall K_2 > 0$,

$$\mathbb{P}(Z(t) \geq K_2) \leq \frac{\mathbb{E}(Z(t))}{K_2} \leq \frac{K e^{\alpha t}}{K_2}$$

We can conclude $Z(t) = O_P(e^{\alpha t})$, meaning $\mathbb{P}(Z(t) > K_3 e^{rt}) \to 0$, as $K_3 \to \infty$. Because of the reasons above and since $\inf\{t : Z(t) = k(N)\} \leq T^N$ w.l.p.

$$k(N) \leq Z(T^N) \leq c e^{\alpha T^N} \text{ with large probability}$$

which implies that

$$\mathbb{P}\Big(T^N \geq \frac{\log(k(N))}{\alpha} - \log(c)\Big) \to 1, \text{ as } N \to \infty.$$

As a consequence by choosing $k(N) = N^{\frac{1}{3}}$ it follows that $T^N \to \infty$ in probability, since it satisfies $k(N) = o(\sqrt{N})$.

□

We give two corollaries, stating that as long as the branching process stays finite the epidemic and branching process coincide forever. However, if $R_0 > 1$ and the branching process grows beyond all limits, then the epidemic and branching process will not remain identical.

Corollary 1.3.5 *(Corollary 1.2.6. and Corollary 1.2.7. in [1])*

- *If $R_0 \leq 1$, then $(L^N(t), I^N(t), R^N(t)) = (L(t), I(t), R(t))$ for all $t \in [0, \infty)$ w.l.p. As a consequence, $\mathbb{P}(Z^N = k) \to \mathbb{P}(Z = k)$ as $N \to \infty$, and in particular Z^N is bounded in probability.*

- *If $R_0 > 1$, then for finite k: $\mathbb{P}(Z^N = k) \to \mathbb{P}(Z = k)$ as $N \to \infty$. Further, $\{Z^N \to \infty\}$ with the same probability as $\{Z = \infty\}$, which is the complement to the extinction probability, the latter being the smallest solution to the equation $z = g(z)$ as described in Theorem 1.2.2.*

Proof. In both cases we use Theorem 1.3.4. Firstly we consider $R_0 \leq 1$. Since we showed that the epidemic and the branching process agree up until there has been M^N births, where for example $M^N > N^{\frac{1}{3}}$ with large probability. However, we also know according to Theorem 1.2.2 that in (sub)critical case the probability of extinction tends to 1 as $N \to \infty$, thus the probability of the number of births exceeding M^N is tending to 0 as $N \to \infty$. This implies that $\mathbb{P}(T^N = \infty) = 1$ according to definition, therefore the epidemic and the branching process agree forever.

In the supercritical case $R_0 > 1$ we reason as follows: As we saw during the proof of Theorem 1.3.4, if only k births occur, there will be no ghost in the epidemic with large probability. Thus the epidemic and the branching process agree forever w.l.p., and $\mathbb{P}(Z^N = k) \to \mathbb{P}(Z = k)$ as $N \to \infty$ for every finite k.

In the branching process a new individual is born, even when the disease is transmitted to a previously infected individual in the epidemic resulting in no new real infection. Thus we can conclude $Z \geq Z^N \geq M^N$. The coupling construction showed that $M^N \to \infty$ on the other part of the sample space. Thus

$$\mathbb{P}(\lim_{N\to\infty} Z^N = \infty) = \mathbb{P}(Z = \infty).$$

By using the discrete-time skeleton again, we know that the probability of extinction is the smallest solution of equation $z = g(z)$. Thus $\lim_{N\to\infty} \mathbb{P}(Z^N = \infty) = 1 - z$.

□

The two corollaries also show that the final number of infected individuals Z^N will be small with a probability equal to the extinction probability of the approximating branching process, however it will tend to infinity with the remaining probability. According to Section 3.3 in [1] the distribution of Z^N is bimodal with one part close to 0 and the other part being $O(N)$ referred as minor and major outbreaks.

Chapter 2

Epidemic models on random graphs

In this chapter we specialize epidemic models by adding to the process an underlying network of connections between individuals represented by a random (or deterministic) graph. In the last chapter we investigated the process with N individuals, where disease transmission was possible between any two individuals. However, in a real life epidemics, especially with large N, this property seems surreal. In case of most of the viruses, infection spreads with high probability (if not only) between individuals having social relationship corresponding to edges in the random graph. In this chapter we assume that the random graph describing relationships which considered suitable for a disease to spread, is fixed through the process, and only on these edges of the graphs can the virus spread. Explaining a real-life epidemics we should consider a random graph with some edges changing with time corresponding to change in contacts of individuals. Another possible solution is a fixed random graph through time plus some random edges changing as the process runs, for describing the transmission of the disease between two individuals having a random and short interaction, for example infection on public transportation. However, there are diseases such as AIDS or Hepatitis C, where the disease can only spread along a social network of individuals. With a graph representation we can also consider the spread of the disease only by direct contacts neglecting indirect contacts such as airborne transmission or contaminated objects, since the probability of these can be reduced by precautionary actions.

Recently epidemics are commonly studied on graphs, Durrett in [11] had examined SIR epidemics on Newman-Strogatz-Watts random graphs, and the SIS model on scale-free networks. The percolation on NSW random graphs was also studied earlier by Calloway, Newman, Strogatz, and Watts 2000 in [10].

In the following chapter we discuss SIR epidemics model, which can be interpreted as a special case of $SEIR$ model reviewed in the previous chapter with no incubation period $I = 0$. In the first section we define the model and some essential quantities concerning the epidemics. Then according to book [23], we study the process on three types of graphs.

2.1 Definition of SIR

We now describe the spread of SIR diseases on graphs: We consider a population of size N whose individuals are the vertices of a random graph G_N. Similar to $SEIR$ as before, in SIR each individual at every moment is in exactly one of the following groups:

- Susceptible: individuals who can contract the disease
- Infected: individuals who can transmit the disease to susceptibles
- Recovered: individuals who were previously infectious and cannot transmit the disease anymore.

We denote corresponding by S_t, I_t and R_t the sizes of groups at time t, so $S_t + I_t + R_t = N$ since we assume a closed community.

On the graph G_N, the dynamics is as follows. To each individual in I is associated an exponential random clock with rate γ to determine its removal or length of infectious period. (So with the notion of the previous chapter $F_I \sim Exp(\gamma)$.) To each edge with an infectious ego and a susceptible alter, we associate a random exponential clock with rate λ. When it rings, the edge transmits the disease and the susceptible alter becomes infectious.

In this chapter we introduce three graphs of different constructions and properties and study analytically SIR epidemics models on them. As in the previous chapter, we compute the critical value of the basic reproduction number R_0 on each graph: This R_0 takes into consideration not only a characteristics of the disease (values of γ and λ parameters), but also the structure of the graph.

Firstly, we give definition to R_0 in relationship with birth rate in a branching process.

Definition 2.1.1 R_0 *(Definition 2.0.1 in [23])*
The basic reproduction number of the epidemic, denoted by R_0, is the mean offspring

number of the branching process approximating the infectious population in early stages. If we denote by $\beta(t)$ the birth rate at time $t > 0$ in this branching process, then:

$$R_0 = \int_0^\infty \beta(t)dt \tag{2.1}$$

Because of Theorem 1.2.7 and the coupling of the SIR model with branching process, we know that with $R_0 > 1$ if the epidemic grows large, then the number of infectious individuals grows roughly proportionally to $e^{\alpha t}$ during the initial phase of the epidemic, since in this chapter we only consider the Markovian SIR.

The epidemic growth rate α corresponds to the Malthusian parameter for population growth defined before in the first Chapter. In the supercritical case it is a positive constant, which depends on the parameters of the model, through the equation

$$1 = \int_0^\infty e^{-\alpha t}\beta(t)dt \tag{2.2}$$

It is easy to see that the equation below is the same as Equation (1.1), since $F(t) = \mathbb{E}(N(t)) = \beta(t)$ according to its definition.

In the chapter we will use these simplified equations to compute critical values of R_0 in different graphs.

2.2 Complete graph

Firstly, we examine the epidemics on the deterministic complete graph K_N. The process thus remains the same as in the previous chapter (except with no incubation period), and can be approximated with a branching process.

Theorem 2.2.1 *R_0 for homogeneous mixing (Proposition 2.1.1 in [23])*
In the case where $G_N = K_N$ is the complete graph, many results for epidemics in large homogeneous mixing populations can be obtained since the initial phase of the epidemic is well approximated by a branching process.
The reproduction number is given by: $R_0 = \frac{\lambda}{\gamma}$
In the case where $\lambda > \gamma$, then $\alpha = \lambda - \gamma$, and the basic reproduction number can be expressed as $R_0 = 1 + \frac{\alpha}{\gamma}$.

The second expression of R_0 is proved to be helpful in case of the determination of R_0 for a real-life disease, since it is independent of λ which can be complicated to estimate. However, removal rate γ is usually easily measurable and the Malthusian parameter α can be estimated from the dynamics of the emerging epidemics.

Proof. We use Definition 2.1.1 of R_0, but for this first we need to determine the birth rate. In homogeneous mixing population an infected individual makes contacts if it is still infectious at rate λ, while $e^{-\gamma t}$ is the probability that the individual is still infectious t time units after it became infected, so $\beta(t) = \lambda e^{-\gamma t}$.

$$R_0 = \int_0^\infty \beta(t)dt = \int_0^\infty \lambda e^{-\gamma t}dt = \Big[-\frac{\lambda}{\gamma}e^{-\gamma t}\Big]_{t=0}^\infty = 0 - \Big(-\frac{\lambda}{\gamma}\Big) = \frac{\lambda}{\gamma}$$

According to (2.2), we need to solve the following equation:

$$\int_0^\infty e^{-\alpha t}\beta(t)dt = \int_0^\infty e^{-\alpha t}\lambda e^{-\gamma t}dt = \lambda \int_0^\infty e^{-(\alpha+\gamma)t}dt = 1$$

Thus $1 = \frac{\lambda}{\alpha+\gamma}$ resulting in $\alpha = \lambda - \gamma$, which was to be demonstrated. $\square$

2.3 Configuration model

In this section we bring more reality to the epidemics model by describing contacts of individuals by a random graph suitable for satisfying many properties of real life networks.

In 1980 Bollobás was the first to construct graphs with specified degree distributions in [4]. Since then, configuration models play a significant role in the mathematical modelling and study of real-life networks. In paper [8] epidemics with different vaccinating strategies are studied on configuration model graphs, while in [13] the SIR epidemic is examined on an upgraded version of the graph: For each node of the graph not only a single number of degree is given, but also the number of triangles of which the vertice is member. The model is called Clustering Configuration Model (CMC) and was introduced by Miller and Newman in 2009, motivated by the fact that clustering coefficient of graphs plays a crucial role in the description of social networks.

We now introduce the classical Bollobás–Molloy–Reed or Configuration model, which creates a random graph to a given degree sequence or degree distribution.

Definition 2.3.1 *Configuration model.*
Let $p = (p_k, k \in \mathbb{Z}_+)$ be a probability distribution on $\mathbb{Z}_+$. The Configuration model constructs a random graph $G_N(V, E)$ with $N = |V|$ vertices for this probability distribution as

follows. We associate with each vertex $u \in V$ an independent random variable X_u drawn from the distribution p that corresponds to the number of edges attached to u, or the degree of u in G_N random graph. Conditionally on $\sum_{u \in V} X_u$ is even both of the following algorithms create a random graph. (If the sum of degrees is odd, we redraw the degree of the last node u, X_u from probability distribution p until the sum becomes even.) We create X_u half edges or stubs for each u vertex, then we pair them according to one of the following algorithms:

- *During Algorithm 1. We choose two from the remained free stubs uniformly at random, and form them into an edge of G_N graph.*
- *During Algorithm 2. We associate with each stub an independent uniform random variable on $[0, 1]$ and sort the half-edges by decreasing values. Then we pair each odd stub with the following even stub.*

The Configuration model (created by Algorithm 1 or 2) is a multigraph, possibly containing self-loops and multiple edges. However, according to Durrett we have asymptotic result for the number of self-loops and multiple edges, if p_k has a finite second moment.

Theorem 2.3.2 *(Theorem 3.1.2. in [11])*
Let $\mu = \sum_k kp_k$ and $\mu_2 = \sum_k k(k-1)p_k$. As $N \to \infty$, the number of self-loops χ_0 and the number of parallel edges χ_1 are asymptotically independent Poisson$(\mu_2/2\mu)$ and Poisson$((\mu_2/2\mu)^2)$ variables.

Thus the probability that the graph is simple has a positive limit, and $\mathbb{E}(\frac{\chi_0}{N}) \to 0$ and $\mathbb{E}(\frac{\chi_1}{N}) \to 0$ as $N \to \infty$. So, to describe relationships between individuals is reasonable with the help of configuration model, since the number of self-loops and parallel edges is negligible in a large graph. Therefore later in the third section we will be able to simulate the process on configuration model random graphs.

From the aspect of epidemics, there are two definitions in connection with the configuration model, which play a major role in the understanding of disease dynamics on CM graphs.

Definition 2.3.3 *Size-biased degree distribution.*
Lets suppose that $\big(p_k, k \in \mathbb{Z}\big)$ is a graph degree distribution, with mean μ, and variance σ^2. We define $\big(q_k, k \in \mathbb{Z}\big)$ the size-biased degree distribution of p,

$$q_k = \frac{kp_k}{\sum_{l \in Z_+} lp_l} \tag{2.3}$$

Because of the construction, we see that in such a network, given an edge of u, the node v is chosen proportionally to its number of half-edges. Since the epidemic starts by a randomly selected individual being infected from outside, this individual has (approximately) the degree distribution p, while the friends of this individual, (or of any individual), have the size biased degree distribution q.

Another important value is the mean excess degree of a Configuration model graph:

Definition 2.3.4 *Mean excess degree.*
Let us assume that $p = (p_k, k \in \mathbb{Z}_+)$ admits a second order moment, so $\mu = \sum_{k\in\mathbb{Z}_+} kp_k = g^{'}(1)$, and $\sigma^2 = \sum_{k\in\mathbb{Z}_+}(k-\mu)^2 p_k = g^{''}(1) + g^{'}(1) - (g^{'}(1))^2$, where $g(z)$ is the generating function of the degree distribution. The mean excess degree of the configuration model with degree distribution p is defined the following way:

$$\kappa = \sum_{k\geq 0} \frac{k(k-1)p_k}{\sum_{l\in\mathbb{Z}_+ lp_l}} = \frac{\sigma^2}{\mu} + \mu - 1 = \frac{g^{''}(1)}{g^{'}(1)} = \mathbb{E}_q(D-1) \tag{2.4}$$

The mean excess degree κ, is in the context of SIR epidemics spreading on graphs, the mean number of susceptibles that are contaminated by a typical infective (other than his or her own infector).

Now we compute the value of R_0 depending on the parameters of the epidemic model, and p_k degree distribution of the graph.

Theorem 2.3.5 *R_0 for Configuration Model. (Proposition 2.2 in [23])*
Assume that G_N is a configuration model graph whose degree distribution p admits a mean μ and a variance σ^2. Then

$$R_0 = \Big(\frac{\kappa\lambda}{\lambda+\gamma}\Big). \tag{2.5}$$

In the super-critical case, R_0 can also be rewritten as

$$R_0 = \frac{\gamma+\alpha}{\gamma+\alpha/\kappa} = 1 + \frac{\alpha}{\lambda+\gamma} \tag{2.6}$$

Proof. As we did in the first chapter, we try to couple the epidemic model with a suitable branching process to study analytic properties: Let us consider the following continuous time birth-death process $(X_t)_{t\geq 0}$. Individuals live during exponential independent times with expectation $\frac{1}{\gamma}$. To each individual is associated a maximal number of offspring $k-1$, where k (the 'degree' of the individual) is drawn in the size-biased distribution q. We

associate to such an individual $k-1$ independent exponential random variables with expectations $\frac{1}{\lambda}$.
The ages at which the individual gives birth are the exponential random variables that are smaller than the lifetime of the individual. There is an intuitive coupling between $(X_t)_{t\geq 0}$ and $(I_t)_{t\geq 0}$ such as $X_t \geq I_t$ for every t, with the equality as long as no 'ghost' has appeared. We can also associate with the process $(X_t)_{t\geq 0}$ its discrete-time skeleton that is a Galton–Watson process $(Z_n)_{n\geq 0}$ with $Z_0 = 1$.
Now we try to compute the descendant distribution of an individual in the time-discretized process: Conditionally on the degree k and the fact that the chosen individual remains infectious for a duration y, the number of contacts contaminated by this individual follows a binomial distribution with parameters $k-1$ and $1-e^{-\lambda y}$: The chosen individual can transmit the disease to a maximum of $k-1$ individuals independently with $1-e^{\lambda y}$ probability, because an infection occurs if and only if the random variable with $Exp(\lambda)$ distribution corresponding to the time of birth is within the individual's lifetime, and if $\zeta \sim Exp(\lambda)$, then $\mathbb{P}(\zeta < y) = 1 - e^{-\lambda y}$.

Let Ψ denote the number of contacts contaminated by a randomly chosen individual and $\psi = \psi(k, y)$ denote the number of contacts contaminated by a randomly chosen individual conditionally on it has degree k and infectious duration of y. To get R_0 we need to calculate $\mathbb{E}(\Psi)$. Using the law of total expectation $\mathbb{E}(\mathbb{E}(\Psi|\mathcal{F})) = \mathbb{E}(\Psi)$ for any σ-algebra $\mathcal{F}$, we only need $\mathbb{E}(\psi(k, y))$. Since $\psi \geq 0$, we use Fubini's Theorem:

$$\mathbb{E}(\psi(k, y)) = \int\int\int \psi(k, y) dP_1 dP_2 dP_3,$$

where P_1, P_2 and P_3 are the probability measures of random variables with Binomial, Exponential distribution and distribution determined by size-biased degree distribution. Since a random variable with $Bin(n, p)$ distribution has expected value of np, and P_2 and P_3 measures are given, we get

$$R_0 = \mathbb{E}(\psi(k, y)) = \sum_{k\geq 0} \frac{k \cdot p_k}{\mu} \int_0^\infty (k-1)(1-e^{-\lambda y})\gamma e^{-\gamma y} dy =$$

$$= \sum_{k\geq 0} \frac{k \cdot p_k}{\mu}(k-1)\frac{\lambda}{\lambda+\gamma} = \Big(\frac{g''(1)}{g'(1)} - 1\Big)\frac{\lambda}{\lambda+\gamma} = \frac{\kappa\lambda}{\lambda+\gamma}$$

From Definition 2.1.1 we can obtain that $\beta(t) = \kappa\lambda e^{-(\lambda+\gamma)t}$. This can also be seen by noting that κ is the expected number of susceptible acquaintances a typical newly infected individual has in the early stages of the epidemic, while $e^{-\lambda t}$ is the probability

that a given susceptible individual is not contacted by the infective over a period of t time units, and $e^{-\gamma t}$ is the probability that the infectious individual is still infectious t time units after he or she became infected.

Now the proof is almost complete, we only need to calculate some alternative expression to R_0 in the super-critical case. If $\alpha > 0$ from Equation (2.2) we can obtain the value of α by solving the following:

$$1 = \int_0^\infty e^{-\alpha t}\kappa\lambda e^{-(\lambda+\gamma)t}dt = \kappa\lambda\int_0^\infty e^{-(\lambda+\gamma+\alpha)t}dt = \frac{\kappa\lambda}{\lambda+\gamma+\alpha}$$

Thus we get $\alpha = \kappa\lambda - \lambda - \gamma$. Rearrange the expression to $\kappa\lambda$ or λ we can bring R_0 to the desired form. By substitution into R_0 obtained before, we get

$$R_0 = \frac{\kappa\frac{\alpha+\gamma}{\kappa-1}}{\frac{\alpha+\gamma}{\kappa-1}+\gamma} = \frac{\kappa\alpha+\kappa\gamma}{\alpha+\kappa\gamma} = \frac{\gamma+\alpha}{\gamma+\alpha/\kappa}, \text{ and}$$

$$R_0 = \frac{\alpha+\lambda+\gamma}{\lambda+\gamma} = 1 + \frac{\alpha}{\lambda+\gamma}.$$

□

The following Theorem can be considered as an interpretation of Theorem 1.2.2 of a branching process, however we now investigate SIR epidemics on CM graphs. The first two parts are consequences of coupling between them, while the third part heuristically says that at the beginning of the epidemics, the population either gets extinct with probability z, or reaches the size ϵn before a t_n order $\log(n)$ time and before extinction with probability $1-z$, since the super-critical process has an exponential growth when it does not go to extinction.

Theorem 2.3.6 *(Proposition 2.2.3 in [23])*
Let us consider the continuous time birth-death process $(X_t)_{t\geq 0}$.

1. *If $R_0 \leq 1$, the process $(X_t)_{t\geq 0}$ dies out almost surely.*

2. *If $R_0 > 1$, the process $(X_t)_{t\geq 0}$ dies with a probability $z \in (0,1)$ that is the smallest solution of*
$$z = \frac{\gamma}{g'(1)}\int_0^\infty g'(z+e^{-\lambda y}(1-z))e^{-\gamma y}dy \tag{2.7}$$

3. *Let us define the times $\tau_0 = \inf\{t \geq 0 : X_t = 0\}$ and $\tau_{\epsilon n} = \inf\{t \geq 0 : X_t \geq \epsilon n\}$. If $R_0 > 1$, then for all sequences $(t_n)_{n\in\mathbb{Z}_+}$ such that $\lim_{n\to\infty}\frac{t_n}{\log n} = \infty$ we have*

$$\lim_{n\to\infty} \mathbb{P}(\tau_0 \leq t_n \wedge \tau_{\epsilon n}) = z \tag{2.8}$$

$$\lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \wedge \tau_0) = 1 - z \tag{2.9}$$

Proof.

The first two points are consequences of the continuous time birth-death process $(X_t)_{t\geq 0}$ defined in the proof of Theorem 2.3.5 that is coupled with $(I_t)_{t\geq 0}$ as long as no ghost has appeared. We can also consider the time-discretized $(Z_t)_{t\geq 0}$ process, obtained from $(X_t)_{t\geq 0}$ by counting individuals in each generation, also defined in the proof mentioned before, since the probability of dying out agrees in the continuous time birth-death process and in its time discretized skeleton Z_n. Thus the first point of the theorem is a straight consequence of Theorem 1.2.2 considered in the sub-critical case.

Concerning the second point of the theorem, we also rely on Theorem 1.2.2. According to it in the super-critical case we need the smallest solution of equation $z = g(z)$ to get the probability of extinction in the branching process, where $g(z)$ is the generating function of the descendant distribution in Z_n. We can easily determine that, since in the proof of Theorem 2.3.5 we found out that conditionally on the degree k and the fact that the chosen individual remains infectious for a duration y, the number of contacts contaminated by this individual $\psi = \psi(k, y) \sim Bin(k-1, 1-e^{-\lambda y})$.

Let $\mathcal{F}$ denote the event that the infected individual has degree k and infectious period y, while υ denotes the descendant distribution of an individual in the discrete time skeleton. We use again the law of total expectation and Fubini Theorem:

$$\mathbb{P}(\upsilon = l) = \mathbb{E}(\mathbb{1}_{\upsilon=l}) = \mathbb{E}(\mathbb{E}(\mathbb{1}_{\upsilon=l})|\mathcal{F}) = \int\int \mathbb{P}\big(\psi(k,y) = l\big) dP_2 dP_3,$$

where P_2 and P_3 are the probability measures of random variables with Exponential and size-biased degree distribution. Summing over k and integrating with respect to y, we can get the probability that in this Galton–Watson process an individual of generation $n \geq 1$ has $\upsilon = l$ offspring. Thus we need to calculate

$$\mathbb{P}(\upsilon = l) = \sum_{k=l+1}^{\infty} \frac{k \cdot p_k}{\mu} \int \mathbb{P}\big(\psi(k,y) = l\big) \gamma e^{-\gamma y} dy$$

$$\mathbb{P}(\upsilon = l) = \sum_{k=l+1}^{\infty} \frac{k \cdot p_k}{\mu} \int_0^{\infty} \binom{k-1}{l} (1 - e^{-\lambda y})^l (e^{-\lambda y})^{k-1-l} \gamma e^{-\gamma y} dy$$

Now instead of trying to calculate manually the integral above, we solve the problem by examining the meaning of it. So far, we concluded that the number of disease transmitting cases by a chosen individual follows $Bin(k-1, 1-e^{-\lambda y})$ distribution, conditionally on the fact that individual is of degree k, and remains infectious for a duration y. Infections caused by the same individual are independent, since we consider a Poisson process, and the probabilities of disease transmissions are the same due to the fact that we examine only constant λ infectious contacts through the process. Therefore, the number of disease transmitting cases by a chosen individual conditionally only on it has k degree also follows Binomial distribution. We now only need to determine the value of the second parameter. (The first parameter $k-1$ is unchanged, as except for the first infectee in the population, an individual is able to infect a maximum of its neighbours minus one individual, the one they got the disease from.) An infection occurs if and only if the lifetime of the individual $\sim Exp(\gamma)$ is longer than the $Exp(\lambda)$ random variable associated to the edge. Thus, we only need to calculate $\mathbb{P}(Z < Y)$, where $Z \sim Exp(\lambda)$, $Y \sim Exp(\gamma)$, and they are independent.

$$\mathbb{P}(Z < Y) = \mathbb{E}\big(\mathbb{P}(Z < Y | Y = y)\big) = \int_0^\infty \mathbb{P}(Z < y)\gamma e^{-\gamma y} dy =$$

$$= \int_0^\infty (1 - e^{-\lambda y})\gamma e^{-\gamma y} dy = \gamma \cdot \Big(\frac{1}{\gamma} - \frac{1}{\gamma + \lambda}\Big) = \frac{\lambda}{\lambda + \gamma}.$$

Thus we determined the parameter of the Binomial distribution in question, and correspondingly got

$$\int_0^\infty (1 - e^{-\lambda y})^l (e^{-\lambda y})^{k-1-l} \gamma e^{-\gamma y} dy = \Big(\frac{\lambda}{\lambda + \gamma}\Big)^l \Big(\frac{\gamma}{\lambda + \gamma}\Big)^{k-1-l} \tag{2.10}$$

The distribution of the number of infectious contacts caused by an individual is given by:

$$\mathbb{P}(\upsilon = l) = \sum_{k=l+1}^{\infty} \frac{k \cdot p_k}{\mu} \binom{k-1}{l} \Big(\frac{\lambda}{\lambda + \gamma}\Big)^l \Big(\frac{\gamma}{\lambda + \gamma}\Big)^{k-1-l}. \tag{2.11}$$

Now we need to calculate the smallest solution of

$$z = \sum_{l=0}^{\infty} z^l \mathbb{P}(\upsilon = l) = \sum_{l=0}^{\infty} z^l \sum_{k=l+1}^{\infty} \frac{k \cdot p_k}{\mu} \binom{k-1}{l} \Big(\frac{\lambda}{\lambda + \gamma}\Big)^l \Big(\frac{\gamma}{\lambda + \gamma}\Big)^{k-1-l}. \tag{2.12}$$

Since we only consider $z \geq 0$ and $\lambda, \gamma \geq 0$ parameters, the expression above is non-negative, and we can use Fubini's Theorem by changing the sums:

$$z = \sum_{k=1}^{\infty} \frac{k \cdot p_k}{\mu} \sum_{l=0}^{k-1} \binom{k-1}{l} z^l \Big(\frac{\lambda}{\lambda + \gamma}\Big)^l \Big(\frac{\gamma}{\lambda + \gamma}\Big)^{k-1-l} \tag{2.13}$$

Using Binomial Theorem this is not other than

$$z = \sum_{k=1}^{\infty} \frac{k \cdot p_k}{\mu} \Big(\frac{z\lambda + \gamma}{\lambda + \gamma} \Big)^{k-1} \tag{2.14}$$

Now we try to obtain the same expression from Equation (2.7) to complete the proof: Since here $g(z)$ is the generating function of the configuration model graph's degree distribution in question, and according to definition $g'(z) = \sum_{k=1}^{\infty} kz^{k-1}p_k$, we need the smallest solution of

$$z = \frac{\gamma}{g'(1)} \int_0^{\infty} \sum_{k=1}^{\infty} kp_k \big(z + e^{-\lambda y}(1-z)\big)^{k-1} e^{-\gamma y} dy =$$

$$= \frac{\gamma}{g'(1)} \int_0^{\infty} \sum_{k=1}^{\infty} kp_k \big(z(1-e^{-\lambda y}) + e^{-\lambda y}\big)^{k-1} e^{-\gamma y} dy.$$

We fall back on Binomial Theorem again, and also use $g'(1) = \mu$:

$$z = \gamma \int_0^{\infty} \sum_{k=1}^{\infty} \frac{k \cdot p_k}{\mu} \sum_{l=0}^{k-1} \binom{k-1}{l} z^l \cdot (1-e^{-\lambda y})^l \cdot (e^{-\lambda y})^{k-1-l} \cdot e^{-\gamma y} dy$$

By using non-negativity again, we get

$$z = \sum_{k=1}^{\infty} \frac{k \cdot p_k}{\mu} \sum_{l=0}^{k-1} \binom{k-1}{l} z^l \int_0^{\infty} (1-e^{-\lambda y})^l \cdot (e^{-\lambda y})^{k-1-l} \cdot \gamma e^{-\gamma y} dy$$

We have already calculated this integral in Equation (2.10), and correspondingly we concluded the probability of dying out to be the smallest solution z of Equation (2.13), and also finished the second part of the proof.

We now prove the third part of the Theorem. We start by proving Equation (2.8): For the birth-death process $(X_t)_{t\geq 0}$ there is no accumulation of birth and death events, so $\lim_{t\to\infty} t_n \wedge \tau_{\epsilon n} = \infty$ almost surely. Since a probability can always be dominated by constant 1, we are able to use Lebesgue's dominated convergence theorem to get the following limit in question:

$$\lim_{n\to\infty} \mathbb{P}(\tau_0 \leq t_n \wedge \tau_{\epsilon n}) = \lim_{n\to\infty} \mathbb{E}(\mathbb{1}_{(\tau_0 \leq t_n \wedge \tau_{\epsilon n})}) = \mathbb{E}(\lim_{n\to\infty} \mathbb{1}_{(\tau_0 \leq t_n \wedge \tau_{\epsilon n})}) = \mathbb{E}(\mathbb{1}_{(\tau_0 \leq \infty)}) = \mathbb{P}(\tau_0 < \infty)$$

Since τ_0 denotes the first time t at which the birth-death process $X_t = 0$, probability $\mathbb{P}(\tau_0 < \infty)$ is nothing else than the extinction probability of the process, which is the smallest solution of Equation (2.7), as proved before denoted by z. Thus $\lim_{n\to\infty} \mathbb{P}(\tau_0 \leq t_n \wedge \tau_{\epsilon n}) = z$.

Now we have to prove the second Equation (2.9). Firstly, we try to determine $\lim_{n\to\infty}\mathbb{P}(\tau_{\epsilon n}\leq t_n\leq\tau_0)$, and then deal with the whole $\lim_{n\to\infty}\mathbb{P}(\tau_{\epsilon n}\leq t_n\wedge\tau_0)=1-z$ expression. We can divide $\tau_{\epsilon n}\leq t_n\leq\tau_0$ into two disjoint events, thus the probability turns into a sum:

$$\mathbb{P}(\tau_{\epsilon n}\leq t_n\leq\tau_0)=\mathbb{P}(\tau_{\epsilon n}\leq t_n \text{ and } \tau_0=\infty)+\mathbb{P}(\tau_{\epsilon n}\leq t_n\leq\tau_0<\infty).$$

We can easily prove the second part of the sum converging to zero, as $n\to\infty$. Since $\mathbb{P}(\tau_{\epsilon n}\leq t_n\leq\tau_0<\infty)\leq\mathbb{P}(t_n\leq\tau_0<\infty)$, and by using dominated convergence theorem again

$$\lim_{n\to\infty}\mathbb{P}(t_n\leq\tau_0<\infty)=\mathbb{P}(\lim_{n\to\infty}t_n\leq\tau_0<\infty)=0$$

as we consider a sequence t_n such that $t_n\to\infty$. Now we only have to concentrate on the first component of the sum. For that we need to deflect a little bit: it can be proved with martingale techniques that

$$\lim_{n\to\infty}\frac{\log X_t}{t}=\alpha,$$

where α is the initial epidemic growth rate defined in 2.2 and that is positive in the super-critical case $R_0>1$.

This is a straight corollary of Theorem 1.2.7, since we consider a Markovian process. According to it, there exists a non-negative random W such that $X_t\sim We^{\alpha t}$ almost surely, as $t\to\infty$, and the branching process dies out with a probability of $\mathbb{P}(W=0)$. However, we examine the event $\tau_{\epsilon n}\leq t_n$ and $\tau_0=\infty$, where $\tau_0=\infty$ means that the branching process does not die out, as stated before. Therefore $W_t=X_te^{-\alpha t}$ is a positive martingale and $\lim_{t\to\infty}W_t=W$ almost everywhere and also in $L^1(\Omega)$.

Now we calculate the first component of the sum: Let us consider $n>\frac{1}{\epsilon}$, so that $\log(\epsilon n)>0$. Since $\mathbb{P}(\lim_{n\to\infty}\tau_{\epsilon n}=\infty)=1$, we have on $\{\tau_0=\infty\}$ that:

$$\lim_{n\to\infty}\frac{\log(\epsilon n)}{\tau_{\epsilon n}}\geq\lim_{n\to\infty}\frac{\log(X_{\tau_{\epsilon n}}-)}{\tau_{\epsilon n}}=\alpha>0,$$

because $\tau_{\epsilon n}$ denoted the first time t such that $X_t\geq\epsilon n$, thus $\epsilon n\geq(X_{\tau_{\epsilon n}}-)$.

$$\text{So inequality } \lim_{n\to\infty}\frac{\log\epsilon n}{\tau_{\epsilon n}}\geq\alpha \text{ corresponds to } \lim_{n\to\infty}\frac{\tau_{\epsilon n}}{\log\epsilon n}\leq\frac{1}{\alpha}.$$

We deduce that:

$$\lim_{n\to\infty}\mathbb{P}(\tau_{\epsilon n}\leq t_n,\tau_0=\infty)=\lim_{n\to\infty}\mathbb{P}(\frac{\tau_{\epsilon n}}{\log\epsilon n}\leq\frac{t_n}{\log\epsilon n},\tau_0=\infty)=\mathbb{P}(\tau_0=\infty),$$

since by our choice of t_n, $\lim_{n\to\infty}\frac{t_n}{\log(n)}=\infty$, and also $\lim_{n\to\infty}\frac{t_n}{\log(\epsilon n)}=\infty$, however $\lim_{n\to\infty}\frac{\tau_{\epsilon n}}{\log\epsilon n}$ remains bounded as we showed before. According to definition $\tau_0=\infty$

event means our process X_t never reaches 0, thus $\mathbb{P}(\tau_0 = \infty)$ is the probability that the continuous time birth-death process does not die out $= 1 - z$, as seen before.

Now we concluded

$$\lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \leq \tau_0) = \lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \text{ and } \tau_0 = \infty) + 0 = 1 - z,$$

however, in Equation (2.9) we need to determine $\lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \wedge \tau_0)$. So far, we know

$$1 - z = \lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \leq \tau_0) \leq \lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \wedge \tau_0) \leq \lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq \tau_0).$$

It remains to show that $\lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \wedge \tau_0) \leq 1 - z$. Since we have already proved Equation (2.8), we got

$$z = \lim_{n\to\infty} \mathbb{P}(\tau_0 \leq t_n \wedge \tau_{\epsilon n}) \leq \lim_{n\to\infty} \mathbb{P}(\tau_0 \leq \tau_{\epsilon n}), \text{ thus } \lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq \tau_0) \leq 1 - z.$$

It is clear that $\lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} = \tau_0) = 0$, resulting $\lim_{n\to\infty} \mathbb{P}(\tau_{\epsilon n} \leq t_n \wedge \tau_0) = 1 - z$ what was to be demonstrated.

□

2.4 Stochastic block model

In this section we introduce the stochastic block model, which can be interpreted as a multi-type generalization of Erdős–Rényi random graphs. SBM is a random graph model with planted clusters, commonly used to study clustering and community detection, but also have proven to be useful in network and data sciences. Because of its wide range of importance the model appeared independently in multiple scientific communities: SBM terminology comes from the machine learning and statistics literature Holland in [15], while in the mathematics literature Bollobás [6] refers to the model as the inhomogeneous random graph model.

Definition 2.4.1 *Stochastic block model graph (SBM).*
A stochastic block model graph is a undirected graph, where each vertex is given a type independently from the others, all with the same probability, and where each pair of vertices is linked independently of the other pairs with a probability depending on the types of the vertices. If there are K types, say $\{1, \dots K\}$, we will denote by $(\rho_i)_{i\in\{1,\dots K\}}$ the probability distribution of the types, and by $\pi_{i,j}$ the probability of linking a vertex of type i with a vertex of type j.

With different number of types, or sets of $\pi_{i,j}$ probabilities, the model can offer a great variety of distinct interpretations: As mentioned before, SBM is a generalized model of Erdős–Rényi graphs, since with $K = 1$ type it constructs an $ER(p)$ graphs, while with $K = 2$ and $\pi_{1,1} = \pi_{2,2}$ choice we get a bipartite graph.

In fact, SBM random graphs are $\mathcal{W}$-random graphs determined by a graphon. According to article [7], a graphon W is a bounded measurable functions $W : [0,1]^2 \to [0,1]$, such that $W(x,y) = W(y,x)\ \forall x, y \in [0,1]$, and a $G(n,W)$ $\mathcal{W}$-random graph with n nodes can be created with the help of W the following way: Let $i, j \in \{1, 2, \ldots, n\}$ denote two nodes, then we connect i and j with an edge in graph G with probability $W(\frac{i}{n}, \frac{j}{n})$. In case of a SBM, the graphon generating random graphs with corresponding structures can be described by $\pi_{i,j}$ and (ρ_i), $i, j \in \{1, \ldots, K\}$ parameters. We divide two sides of the $[0,1] \times [0,1]$ square according to distribution (ρ_i). The graphon associates the appropriate $\pi_{i,j}$ probabilities on the rectangles created as above.

We now focus on the graphs model from the point of view of epidemics: Often a community contains different types of individuals displaying specific roles in contact behaviour. Types can relate to different aspects such as age, geographic situation, profession/ everyday occupation, or social behaviour. If the disease affects individuals of different age significantly distinctly, with a suitable small K and statistically appropriate age- and contact-distribution it is reasonable to construct an age-group graph with SBM. Another possible interpretation of the types is the household structure, in which with a large K $\pi_{i,i} = 1$, while $\pi_{i,j}, i \neq j$ are tiny.

Now we compute the value of R_0 in SBM graphs with the help of multi-type branching process.

Theorem 2.4.2 *R_0 for SBM (Proposition 2.3. in [23])*
Let us assume that there are K types of individuals, labeled $\{1, 2, \ldots, K\}$, and that the infection rate from an ego of type i to an alter of type j is $\frac{\lambda_{i,j}}{N}$. Denote by ρ be the largest eigenvalue of the matrix with elements $\lambda_{i,j}\rho_j$. Then basic reproduction number of the disease on SBM random graphs is given by

$$R_0 = \frac{\rho}{\gamma} = 1 + \frac{\alpha}{\gamma} \tag{2.15}$$

Proof.

We can couple here the infection process with a multi-type branching process. The rate at which a given i individual gives birth to an individual j corresponds to the rate, in

the epidemic process, at which an individual i infects individual j at time t after infection. Let $a_{i,j}$ denote this rate. Since an i individual contacts a given individual of type j at rate $\frac{\lambda_{i,j}}{N}$, while there are $N\rho_j$ individuals of type j, and an i individual is still infectious t time after being infected with probability $e^{-\gamma t}$, we can express

$$a_{i,j}(t) = \frac{\lambda_{i,j}}{N} N\rho_j e^{-\gamma t} = \lambda_{i,j}\rho_j e^{-\gamma t} \tag{2.16}$$

For multi-type branching processes, it is well known according to [21] that the basic reproduction number $R_0 = \rho_M$ is the largest eigenvalue of the matrix M with elements $m_{i,j} = \int_0^\infty a_{i,j}(t)dt = \frac{\lambda_{i,j}\rho_j}{\gamma}$, and the epidemic growth rate α is such that $1 = \int_0^\infty e^{-\alpha t}\rho_{A(t)}dt$, where $\rho_{A(t)}$ is the largest eigenvalue of the matrix $A(t)$ with elements $a_{i,j}(t)$. We also know that $\rho_{A(t)} = \rho e^{-\gamma t}$. Therefore

$$R_0 = \rho_M = \int_0^\infty \rho e^{-\gamma t} = \frac{\rho}{\gamma},$$

and we can also compute

$$1 = \int_0^\infty e^{-\alpha t}\rho_{A(t)}dt = \int_0^\infty e^{-\alpha t}\rho e^{-\gamma t}dt = \frac{\rho}{\alpha+\gamma}.$$

Thus $\rho = \alpha + \gamma$, and we can express R_0 with the help of the parameters of the disease:

$$R_0 = \frac{\rho}{\gamma} = \frac{\alpha+\gamma}{\gamma} = 1 + \frac{\alpha}{\gamma}.$$

□

Chapter 3

Discretized $SEIR$ with isolation

In this chapter we describe a possible implementation of discretized $SEIR$ process enriched with isolation of individuals run on random graphs. From now on, we consider the basic reproduction number only as a characteristic of a disease regardless the structure of the random graph, denoted by R_0' to emphasise the difference. Firstly, we examine the process with different means and proportion of individuals participating in self-isolation.

In the previous semesters according to [18] we also examined discretized virus spread on random graphs with different vaccination campaigns, which were usually constructed based on age groups. However, now we consider a disease with no available vaccine, and assume no previous immunity of individuals, thus isolation of individuals and restriction measurements are introduced to flatten the curve of infectious individuals. Most of the time we try to set parameters of the disease to replicate properties of Covid-19. Later we run the process (with the same transmission rate) not only on a fixed random graph, but on 5 graphs, from which 4 can be obtained as a subgraph of the first one by deleting edges corresponding to decreasing density. This is the equivalent graph representation of restriction measurements applied to individuals.

Firstly, we define the precise implementation of the model, review necessary data and parameter sets, describe examined original random graphs, and means of creating subgraphs from them.

3.1 Implementation of the model

We added a quarantine compartment to the classical model as follows. Here we only describe a general representation of isolating individuals, possible means of execution will

be detailed later.

Definition 3.1.1 *Discretized $SEIR$ with isolation*

Similarly to the classical $SEIR$ model, each individual is in exactly one of the following compartments at each time steps during the virus spread.

- *Susceptible: Individuals are healthy, but can be infected.*
- *Exposed: Individuals are infected but not yet infectious.*
- *Infectious: Individuals are infected and infectious.*
- *Recovered: Individuals are not infectious anymore, and immune (cannot be infected again).Thus R is a terminal point.*
- *Quarantined/Isolated: Individuals regardless infected or infectious cannot transmit the disease to anyone, and also healthy individuals are not able to get infected as long as they are staying in this compartment.*

We could rather consider group of isolated individuals not only as one compartment, but four different sub-compartment in S, E, I and R, since individuals in quarantine also can experience infected and infectious periods, just not able to interact with others.

The rate at which individuals leave compartments are described by probabilities (transmission rates) and the parameters of the model (incubation rate, recovery rate). We can associate different meanings to isolation of individuals (individuals are transmitted to a hospital, or self isolation), later we will consider different implementations to the problem.

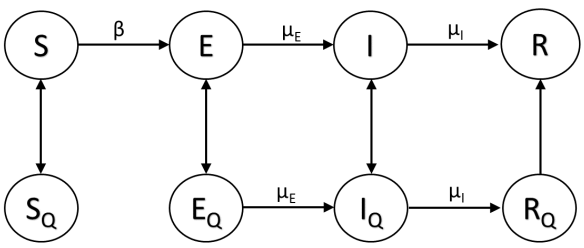

Figure 3.1: Compartments in SEIR with self-isolation

The process can by summarized by the Figure 3.1, describing possible moving of individuals between compartments. Let Q-indexed stages represent isolated individuals within the respective group, so E_Q denotes infected but not infectious individuals isolating themselves. In our implementation of the model recovered individuals do not get isolated. (Individuals can leave a stage only on an edge heading out.)

3.2 Data and parameters

3.2.1 Data

To describe the underlying network, we use real-life data. We distinguish individuals according to their age. In particular, we consider 5 age groups since they have different social contact profiles, and also they are affected by the virus differently. The age groups and the number of individuals in each per $N = 10000$ citizen (distributed as the 2005 European Union population, Eurostat 2006):

- 0-9: $N^1 = 1050$
- 10-19: $N^2 = 1200$
- 20-39: $N^3 = 2850$
- 40-65: $N^4 = 3250$
- 65+: $N^5 = 1650$

To describe the social relationships of the different age groups, we used the contact matrix obtained in [18]:

$$C = \begin{pmatrix} 5,3580 & 1,0865 & 3,0404 & 2,4847 & 0,8150 \\ 0,9507 & 10,2827 & 2,8148 & 3,6215 & 0,7752 \\ 1,1201 & 1,1852 & 6,5220 & 4,1938 & 0,9016 \\ 0,8027 & 1,3372 & 3,6776 & 5,2632 & 1,3977 \\ 0,5187 & 0,5638 & 1,5573 & 2,7531 & 2,0742 \end{pmatrix},$$

where the elements $c_{i,j}$ represent the average number of contacts an individual in age group i has with individuals in age group j.

We create random graphs detailed later which fulfill the distribution of age groups, and edge densities between age groups.

Later we will also consider the process with restricting measurements: these actions aim to reduce contacts between individuals. We examined five different stages of restricting measurements corresponding to five graphs. (All graphs are created from the original random graph by deleting edges, thus reducing contacts in the population.) We tried to set the proportion of contacts decreased to imitate real-life data, thus we have taken into account results of questionnaires organized by University of Szeged, found at page [25].

1. Normal scenario: no restrictions, contact matrix C remains the same.
2. School closings: contacts within the first two age groups reduced drastically by 80%.
3. Home office: contacts within age group 3 and 4, and contacts between them reduced by 60% uniformly.
4. School closing + Home office + mild restrictions: This action contains contact reduction of action 2 and 3, in addition all remained contacts is also reduced by 20% regardless of age groups.
5. School closing + Home office + strict restrictions: In this action all remained contacts reduced by 40% regardless of age groups. (Graph created here is also a subgraph of the previous one.)

3.2.2 Parameters

The model is specified by the following family of parameters. We set parameters of the disease to model $Covid-19$ virus, according to article [3] and page [26].

- $R_0' = 2$: basic reproduction number. It characterizes the intensity of the epidemic. Here we consider R_0' as only a characteristic of the disease regardless the structure of the graphs.
- L: duration of latent period with distribution $F_L \sim Exp(5)$. (Durations are rounded to days, since we investigate the discretized model).
- I: duration of infectious period, $\mathbb{E}(I) = 5$, with distribution F_I:

$$\mathbb{P}(I = 5) = 0.5, \qquad \mathbb{P}(I = 4) = \mathbb{P}(I = 6) = 0.25$$

- β: transmission rate. They control the rate of the infection between a susceptible and infectious individual. Since we assume that characteristic of connections between different age groups is implemented into the structure of our random graph, we consider an universal β transmission rate. We determined β according to the definition of R_0', the mean number of infections caused by an infectious individual during its infectious period in a population of susceptibles. Thus we set the disease transmissions to $\tilde{\beta} = \frac{R_0'}{\mathbb{E}(I) \cdot 12.8113}$. Only the overall average density of the graph (in case of every considered graph this is approximately 12.8113), severity of the disease and average time spent in infectious period can affect the parameters.

Isolation is implemented in the process according to two aspects by the following. Each individual gets a type at the time of its birth, determining its reaction to the virus and behaviour in self isolation during the virus spread:

- Individuals of type 2: They are considered to be the most endangered fraction of the population. After catching the disease, they need treatment in hospitals. Thus, the number of infected individuals of type 2 has crucial importance through the process, restriction measurements are usually put into actions in connection with this measure rather than the whole number of infected individuals, since we can only estimate those. Since the virus has a characteristic that elderly people are more prone to be affected seriously by the disease, the proportion of type 2 individuals is also determined in connection with age groups: Each individual in age group 5 has type 2 independently with probability 0.2, while for other age groups only with probability 0.05. Type 2 individuals go into quarantine after spending 2 days in the infectious period, and also all neighbours of such an individual (regardless of type) engage in self isolation with probability μ_{map}. The self isolation of neighbours can be considered as post-mapping contacts of infected individuals by authorities.

- Individuals of type 1: Independently from age groups individuals are considered to be the "conscious" part of the population. These individuals show symptoms of the disease, however they need no special treatment. After noticing some symptoms (also after the second day of the infectious period) they decide to self isolate. The proportion of individuals of type 1 can be increased by providing information about the possible symptoms and means of spread of the disease. Thus we examine the process with different μ_{con} proportion of the population engaging in self isolation.

- Individuals of type 0: Individuals need no special treatment in hospitals after getting the disease. They self isolate if and only if a neighbour of type 2 gets infected as described above.

Regardless of the cause of self isolation, individuals spend exactly 14 days in quarantine.

3.2.3 Random graphs

We would like to create an underlying network and examine the outcome of virus spread on this given graph.

We generated random graphs of different degree distributions of $N = 100000$ nodes with the help of the configuration model, such that each node has a type corresponding to the age of the individual. In fact, for each node 5 different degrees were drawn from the same distribution with different appropriate means representing the number of neighbours of the individual broken down to age groups. Within an age group the ordinary configuration model created a graph, while between any two age groups we modified the configuration model to create a bipartite graph of the given degrees (paying attention to the total number of degrees be equal in two age groups). The age distributions and number of contacts in the graph between age groups comply with statistic properties detailed above. Since the contact matrix C describes only the average number of contacts, the variances can be different.

As mentioned before, configuration model creates a multigraph containing self-loops and multiple edges. However, according to Theorem 2.3.2 number of self-loops and parallel edges can be neglected. Since we consider a $SEIR$ process, disease transmission on self-loops makes no sense, thus we leave them out of consideration. Contrarily, on k-multiple edges the disease can spread with k-times bigger probability, since we assume a stronger connection of these individuals in some way.

We examined three different degree distributions:

- Poisson distribution: Originally instead of Configuration model random graphs with Poisson distribution we were planning to generate SBM-random graphs. However, degree sequence of Erdős–Rényi random graphs are known to follow Poisson distribution [14]. Since we can observe degrees of a node broken down from the aspect of age groups as degrees in a Erdős–Rényi graph with corresponding density, and

the number of age groups is only $K = 5$, we are able to approximate in SBM the degree distributions of a node in different age groups with Poisson distributions.

From numerical point of view, generating SBM graphs with N nodes is $\mathcal{O}(N^2)$, since we create edges independently. In comparison configuration model is only $\mathcal{O}(|E|)$, which can be significantly faster in case of spare networks. By approximating the structure of SBM graphs with Configuration model random graphs of Poisson degree distribution enabled us to run the process on $N = 100000$ individuals.

- Pareto distribution: Some years ago it was generally believed that real-life and social networks have degree distributions that approximately follow a power law $P(k) \sim k^{-\gamma}$ with some $2 < \gamma < 3$ constant, called scale-free networks. We examine Pareto degree distribution with shape parameter or also called tail index =2.7, however now it is proved that despite the fact that most of the observed networks have fat-tailed degree distributions, only a small percent of them is scale-free according to [9]. (Tail index was chosen to be $\gamma = 2.7$, because variance of Pareto distribution increases significantly, as we decrease γ, with parameter $\gamma < 2$ the variance is not even finite. Thus an outstanding degree made the run of Configuration Model time consuming. In some cases it was not even possible to create a graph, since sum of degrees between different age groups must comply.)

- Geometric distribution: However this distribution is not heavy-tailed, Geometric distributions with the appropriate means correspond bigger variance than Pareto(2.7) distribution.

After describing necessary parameters of the model and possible generation of random graphs, we review the exact steps of our program:

In the discretized process, we start with 10 infectious nodes chosen randomly and independently from the age groups. We usually observe a 200 day period. At a time step firstly the infectious, but not self-isolated nodes can transmit the disease to their neighbours. Only not self-isolated nodes in S can be infected, and they cannot be infected ever again. When a node becomes infected, its position is set immediately to E, and also the number of days it spends in E is generated. Secondly, we check if a node has reached the end of its latent/infectious period, and we set its position to I or R. (As soon as a node becomes infectious, the days it spends in I is also calculated.) Individuals of type 1 and 2 self-isolate after spending two days in infectious period, and some neighbours of

type 2 individuals also are quarantined.

We change the underlying graph at certain time steps previously given, or determined by the describing values of the virus spread at the moment. As described before, when we create a random graph four edge-decreased version of it is also generated to model restrictions. We change the underlying graph (restriction and also easing is possible) at the beginning of an iteration, disease can only spread on the edges of the actual graph. The β transmission rate remains the same on any restricted subgraph. To bring more reality to the model infectious nodes are also able to transmit the disease to any (not necessarily neighboring) node with probability 0.05 at each time steps.

3.3 Basic scenario

Firstly, we describe a so-called basic scenario, serving as a benchmark to different parameter sets of the process. In this scenario we only consider the process on the original graph, so no restriction measurements take place. We set $\mu_{map} = 0.5$, thus expectedly half of the neighbours of an individual with type 2 will self-isolate after its infection, and $\mu_{con} = 0.1$ proportion of the population will self-isolate themselves after experiencing symptoms.

As described before, we only run the process on random graph generated by the configuration model. Since in this section we consider R_0' only as a measure of the disease regardless of the graph structure, and according to Theorem 2.3.5 the real R_0 calculated on the given graph is $R_0 = \frac{\kappa\lambda}{\lambda+\gamma}$, where $\kappa = \frac{\sigma^2}{\mu} + \mu - 1$ with mean and variance of the generating degree distribution, we expect to get different results on random graphs with Poisson, Pareto(2.7) and Geometric degree distributions. In the construction of the graphs we set edge densities within and between age groups to the same, so the mean must comply in case of every graph. However variances are respectively around 21, 36 and 77, thus we would expect a significant virus outbreak in case of random graphs with Geometric degree distribution.

In the following we examine usually studied measures of the intensity of the virus spread such as proportion of individuals recovered from the virus at the end of an examined period called overall attack rate, and maximum number of individuals being infected at the same time, while they are of crucial importance from the aspects of hospital capacity and building herd-immunity.

To study the result, we generated 5 random graphs with $N = 100000$ nodes for each degree distribution, and run the process 20 times on a random graph with independent

initial choice of infected individuals. All in all, these 100 results were averaged. (Generating graphs with 100000 nodes is rather time-consuming, however it is acceptable to run the process on the same graph many times. Types of the individuals and initially infected individuals at the beginning of the process are re-drawn from the appropriate distributions in question, thus we can get very different result on the same graph.) We usually consider a time period of 200 days.

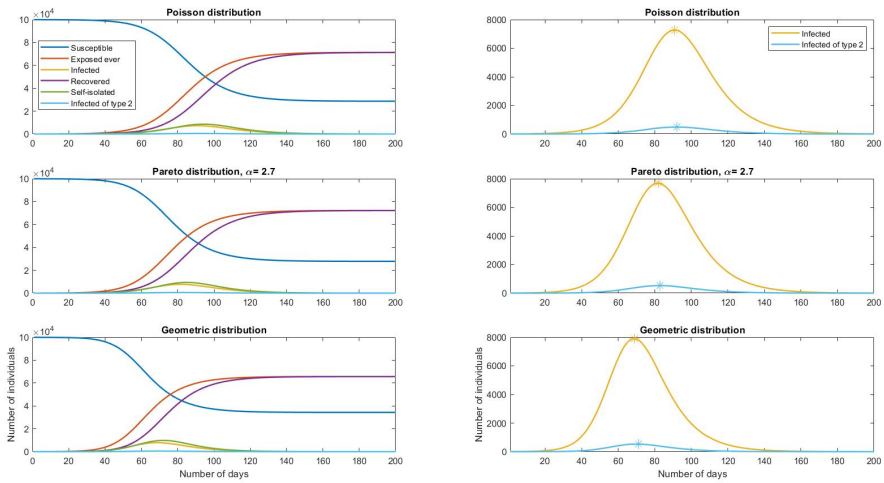

Figure 3.2: Basic scenario virus spread on CM graphs of different degree distribution

We do can see differences between outcomes of the process on random graphs with different degree distribution, however they are not always straightforward.

We can generally say that runs on graphs with generating degree distribution of bigger variance results in swift and peaky curve of infected individuals, while in case of smaller variance infectious individuals tend to be better distributed in time. We can experience a difference of around 10% in the number of infectious individuals at the same time between graphs of different degree distributions, and 7% in the overall attack rates. However, degree distribution of the biggest attack rates do not coincide with the graph model of the highest number of infectious individuals at the same time. Infectees on random graphs with geometric degree distribution reaches its peak only around day 69 with number of individuals approximating 8% of the population in case of $N = 100000$ individuals. However, after the peak, the virus has a short falloff period, because many nodes have only a couple of neighbours. Therefore we get the smallest attack rates after a 200 days period on random graphs with Geometric degree distribution. The number of individuals

in self-isolation reaches its maximum value a few days after the peak of the virus, by affecting 10% of the whole population at that time.

Precisely measurable value is the number of individuals of type two, behaving the same as the total number of infectees only with a couple days of delay. The number of individuals in age group 5 dominates infectious people of type 2, because of the characteristic of the disease. When we investigate the process broken into age groups, we can conclude that the virus reaches its peak at different times, and also reaching different proportion of people, because of the significant differences of the contact matrix.

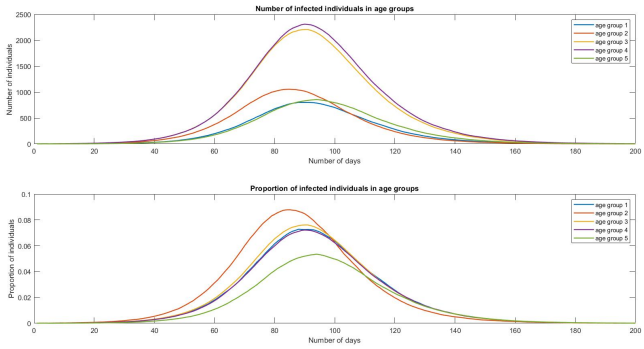

Figure 3.3: Number and proportion of infected individuals in age groups

3.3.1 Change of parameters in the basic scenario

We investigate how and in what sort is it possible to change the outcome of the process only modifying parameters of μ_{con} and μ_{map}, without examining the process on reduced graphs of restrictions. In real life increase of μ_{map} could be achieved by trying to accomplish more successful mapping of possible contacts with infectious individuals. The increase of μ_{con} can be achieved by increasing awareness and knowledge of people about the characteristic of the disease.

Properties				
Scenario	Degree distribution	Poisson	Pareto(2.7)	Geometric
Basic	Overall attack rate	0.71297	0.72216	0.65672
$\mu_{map} = 0.5$	Max individuals in I	7520	8000	8229
$\mu_{con} = 0.1$	Peak day	91	82	69
	Overall attack rate	0.63412	0.64596	0.59406
$\mu_{map} = 0.5$	Max individuals in I	5583	6071	6609
$\mu_{con} = 0.3$	Peak day	106	92	75
	Overall attack rate	0.69204	0.70263	0.63500
$\mu_{map} = 0.7$	Max individuals in I	6988	7455	7569
$\mu_{con} = 0.1$	Peak day	91	83	70

The table above summarizes properties of the process with different sets of parameters: We can experience bigger movement of measures by increasing the value of μ_{con}, since it affects more individuals. It influences significantly both examined values, corresponding to the basic scenario with approximately $R_0' = 1.8$. The increase in μ_{map} also suppresses the disease, however we can see no delay in the peak of the virus, only a reduction of its intensity.

In these cases still only relationships of type 2 individuals are sought out, however contacts of infectious individuals of type 1 can be also taken into consideration as well as random testing of individuals.

To study en masse testing of individuals I modified the model in which neighbours of individuals of type 1 also sent into quarantine, if the individual is proven to be infected. With neighbours of type 2 individuals self-isolated with probability 0.5, and neighbours of type 1 individuals with probability 0.3, with 0.3 proportion of the population being of type 1 could keep down the spread of the virus effectively by decreasing the overall attack rate to ≈ 0.5 in both cases, and flatten the infectious curve significantly as maximum number of individuals being in I at the same time become half of the basic scenario.

We also considered an extreme case of neighbours of type 2 individuals self-isolated with probability 0.7, while neighbours of type 1 individuals with probability 0.5, with half of the population being of type 1.Even with these parameter sets of the model it is not possible to talk about the probability of a minor outbreaks (disappearance of the virus in the early stage of the spread, affecting only a small number of individuals). We could experience this happening only in case of graphs with Poisson degree distribution with an

insignificant probability (3 out of 500 runs).

Thus, with so intense parameters of the virus ($R_0 = 2$) it seems almost impossible to stop the virus completely, even if monitoring and mapping of possible infectious contacts of individuals begins at the first day of appearance of the virus (and of course with the additional assumption that no disease transmission can occur coming from outside of this examined population).

3.4 Restriction measurements

We examined 3 different sets of restriction and releasing actions, from which two are executed at predefined time steps regardless of the actual state of the virus spread. In the third case, actions are taken according to the number of infectious individuals. For both of them we use basic parameter choice of $\mu_{map} = 0.5$ and $\mu_{con} = 0.1$.

Let G be the original graph of CM with any degree distribution we examine. We denote by $G_1 = G, G_2, G_3, G_4, G_5$ the appropriate edge reduced graphs created from G as described before. We examine the following set of actions:

1. Late introduction of restrictions: Responses to the virus are overdue, and thus an intensive restriction is needed. We change to graph G_4 only after 60 days of the appearance of the virus, then to G_5 at day 90. Then we start easing by switching back to G_4 graph at day 110, and all restrictions are dissolved on day 150 by using again the original graph $G = G_1$.

2. Early introduction of restrictions: In this campaign we use a balanced sequence of actions taken. Restrictions begin as early as day 10 by closing schools (graph G_2), then we use graph G_4 from day 45, and graph G_5 from day 90. Easing actions occur at the same days and pace as before: Graph G_4 is used from day 110, and G_1 from day 150.

3. Restrictions made according to number of infectious individuals: In this case it is possible to keep switching between restriction and easing measurements, however after an action is taken the same graph must be used for at least 14 or 21 days. Measurements are introduced if the number of infected individuals I reach a certain

threshold. We use

$$\begin{cases} G_1, & \text{if } I < 1000 \\ G_3, & \text{if } 1000 \leq I < 2000 \\ G_4, & \text{if } 2000 \leq I < 3500 \\ G_5, & \text{if } I \geq 3500. \end{cases} \tag{3.1}$$

In the previous sections we could see differences in intensity and values on random graphs of different degree distributions, however the process overall behaved similarly. With restrictions taken place curves are distorted at certain time steps, thus emphasizing importance of a few days in the delay/ introduction of an action.

Now we examine the process on a 250 days period, since in some cases restrictions flatten the curve so much that significant proportion of individuals are still infected or get infected after day 200. As before, we also generate 5 random graphs with $N = 100000$ nodes and run the process 20 times on each graph (or sequence of graphs), then average results. (In the third case it makes no sense to average different trajectories of the virus spread, since actions can take place at different time steps.)

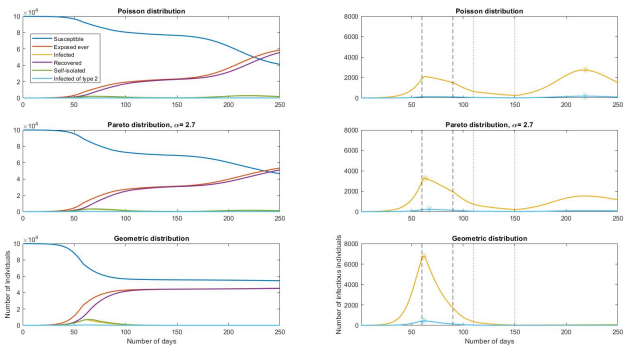

Figure 3.4: Late introduction of restrictions

We can generally conclude that effect of a restriction can be noticed only five days after introduction of actions.

A delay of only a few days in actions can have a huge impact of the process, since with late introduction of restrictions in case of Geometric degree distributed graphs the peak is around day 69 makes almost no help of restrictions introduced on day 60. It is only resulting in faster falloff of the virus, and thus smaller overall attack rate, since many nodes have small degree. In case of the two other random graphs, we do can see help of

the actions taken in flattening the curve. However, with too strict measurements taking place at the beginning the curve was pressed down to a minimal level, and thus easing actions resulted in another peak. The difference in the size of the peaks is depending on the proportion of susceptible individuals. We can also observe that it was possible to reach significantly smaller overall attack rate, even if at the end of the process we run the process on the original $G = G_1$ graph with no restrictions. With measurements taking place in the

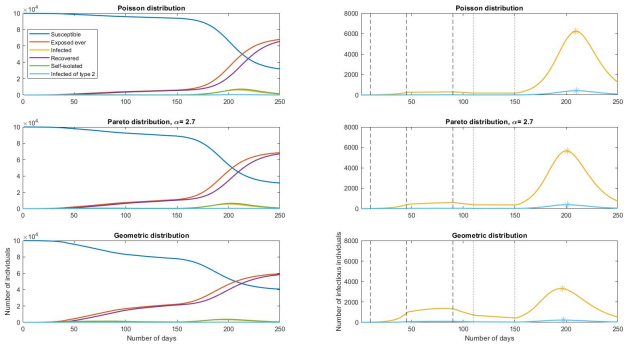

Figure 3.5: Early introduction of restrictions

early days of the virus spread, on most of the graphs it only postponed almost the same effect in volume of virus. On the first two graphs even introduction of home office could tame the rate of increase of infected individuals, while we can say that on graphs G_4 the number of infected individuals remained almost constant in any graph. However, in case of Geometric degree distribution early reaction was proved to be efficient, but also here graph G_5 seems to be too radical, since it is even able to decrease infected individuals. But this success is only temporary, since after lifting all restrictions even on this graph we can experience another peak.

With actions taken according to the numbers of infectious individuals we still can experience difference in volume between graphs of different degree distributions. Not only number of infectious individuals should be taken into consideration but also the steep of the curve, which can be measured by the Malthusian parameter of the disease. By alternating restricting and easing measurements the number of infectious individuals were better distributed in time, a peak was created only at the beginning of the process. However, this peak did not exceed significantly the chosen threshold of 4000 individuals. In Figure 3.6 we illustrated a typical trajectory of infectious individuals evolving in time with measurements taken according to their number. (The numbers on dotted or scattered

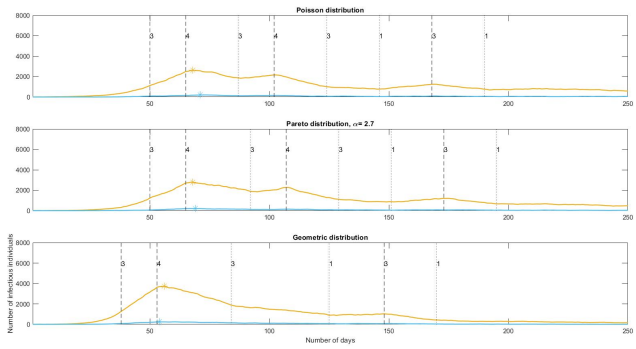

Figure 3.6: Typical trajectory of infected individuals with restrictions

line denote the number of subgraphs we run the process on, after the certain time step.) As we can see, very strict restrictions corresponding to graph G_5 were never even introduced, thus aligned actions taken at the right time could not only flatten the curve, but also allow the use of less severe restrictions.

Chapter 4

Voter model

Interacting particle systems is a large and growing field of probability theory motivated by models arising in statistical physics, biology, economics, and other fields. They are often used as a simplified model for stochastic phenomena that involve a structure in space, for us in this paper graph structures will serve as a spatial structure. Voter models and contact processes are among the most studied in the field.

In this chapter firstly we introduce interacting particle systems generally, then define and review a few results about the voter model according to book [19]. Contact processes also will be presented to make a connection between the voter model and epidemics processes. In the second section we show a possible implementation of the time-discretized voter model on a random graph, and by making assumptions about the structure of the graph and the allocation of opinions we show an analytical property of the proportion of nodes representing one of the opinions. Therefore this enables us to run the discretized process defined on random graphs in the next chapter.

4.1 Voter model in continuous time

Interacting particle systems are usually examined, as in books [19] and [20], as an η_t continuous time strong Markov process, whose transition measures are weakly continuous in the initial state, also known as Feller process, on the compact configuration space $\{0,1\}^S$, where S is an arbitrary countable set. The process η_t is usually described by the rates at which transition occurs, so the system changes from one configuration to another. Here configuration refers to a state of a node according to some aspect. In case of SIS epidemics the possible configurations are susceptible and infectious, while in the voter

model configuration of a node is its opinion. If S is finite, saying that transition $\eta \to \psi$ occurs at rate c means that for $\eta_t \neq \psi$: $\mathbb{P}^\eta(\eta_t = \psi) = ct + o(t)$ as $t \to 0$.
These changes are generally local, in that only one or two sites change state at any given time.

To describe the transition rates, the following notations are used for interacting particle systems: If $\eta \in \{0,1\}^S$, and $x, y \in S$, then $\eta_x, \eta_{x,y} \in \{0,1\}^S$ are defined by

$$\eta_x(u) = \begin{cases} \eta(u), & \text{if } u \neq x \\ 1 - \eta(u), & \text{if } u = x \end{cases} \tag{4.1}$$

$$\eta_{x,y}(u) = \begin{cases} \eta(u), & \text{if } u \neq x, y \\ \eta(y), & \text{if } u = x \\ \eta(x), & \text{if } u = y \end{cases} \tag{4.2}$$

So η_x is obtained from η by changing only its value at x, while $\eta_{x,y}$ is obtained from η by interchanging the values at x and y. The transition $\eta \to \eta_{x,y}$ can be interpreted as moving a particle from x to y or vica versa.

Definition 4.1.1 *Voter model.*

In the voter model η_t process can be described by the following: $p(x,y)$ are the transition probabilities for a Markov chain on S, such that $p(x,y) \geq 0$ and $\sum_y p(x,y) = 1$. The transition rates are given by

$$\eta \to \eta_x \text{ at rate } \sum_{y:\eta(y)\neq\eta(x)} p(x,y).$$

We can give the following interpretation to the model: the sites are individuals who can have one of two opinions (denoted by 0 and 1) at any time. At exponential times of rate 1, the individual at x chooses a y with probability $p(x,y)$ and adopts y's opinion.

In books [19] and [22] only the canonical choice of the d-dimensional integer lattice $S = \mathbb{Z}^d$ and the case with $p(x,y) = p(0, y-x)$ transition probabilities for an irreducible random walk are considered, however we can also consider the model on a graph. Firstly, we put an emphasis on the connection of voter model and virus spread models, then I review some analytical result with $S = \mathbb{Z}^d$ case according to Liggett [19]. However, from the beginning of next section we consider the time-discretized voter model on random graphs.

Now, with the help of 4.1 notion we define contact process, a model which is usually used to describe spread of infection, however it is also related to Reggeon Field Theory in high energy physics, and it is a building block for more complex models in biology.

Definition 4.1.2 *Contact process.*
In the contact process S is a graph whose vertices have bounded degree, and λ is a positive parameter. The notation $x \sim y$ means that vertices x and y are connected by an edge. Then for each $x \in S$

$$\eta \to \eta_x \text{ at rate} \begin{cases} 1, & \text{if } \eta(x) = 1 \\ \lambda \cdot |\{u \sim x : \eta(u) = 1\}|, & \text{if } \eta(x) = 0 \end{cases} \tag{4.3}$$

The process describes a virus spread: sites with $\eta(x) = 1$ are the infected individuals, while sites with $\eta(x) = 0$ are healthy. Infected sites recover from the infection after an exponential time of rate 1, while healthy sites become infected at a rate proportional to the number of infected neighbours.

From the definition above we can easily see that contact process in terms of virus spread corresponds to the SIS epidemics model in which recovery ensures no immunity to further infection.

Now we concentrate on properties of the voter model with the special $S = \mathbb{Z}^d$ and $p(x, y) = p(0, y - x)$ transition probabilities case.

According to [19] the behavior of the process depends heavily on the recurrence or transience properties of the symmetrized random walk $X(t) - Y(t)$, where $X(t)$ and $Y(t)$ are independent random walks with unit exponential jump times and transition probabilities $p(\cdot, \cdot)$.

Firstly, we give an informal insight to build up this property: Suppose we want to determine the opinion $\eta_t(x)$ of the voter at x at a large time t. His opinion must came from some other voter at x_1 at some earlier time t_1. Taking steps backward this way, we find that $\eta_t(x) = \eta_0(X(t))$ for some random $X(t)$, where the process $X(t)$ is simply a random walk with transition probabilities $p(\cdot, \cdot)$ and initial point $X(0) = x$. We can similarly construct the opinion of some $y \neq x$ at time t by $\eta_t(y) = \eta_0(Y(t))$ for some other $Y(t)$ random walk, $Y(0) = y$. However, $X(t)$ and $Y(t)$ random walks are not independent: they are independent only up to the first τ time when $X(\tau) = Y(\tau)$, after that they evolve together. Thus $\eta_t(x)$ and $\eta_t(y)$ can agree for two different reasons:

- $t > \tau$
- $t < \tau$ and $\eta_0(X(t)) = \eta_0(Y(t))$.

We say that a random walk is recurrent if it visits its starting position infinitely often with probability one and transient if it visits its starting position finitely often with probability one. If for the independent random walks, $X(s) - Y(s)$ is recurrent, then the coalescing random walks will agree eventually with probability 1, and hence $\eta_t(x) = \eta_t(y)$ with large probability for large t.

However, from this property it cannot be straightforwardly concluded that for each x and y, $\eta_t(x) = \eta_t(y)$ from some time on since, changing the t in the argument changes the random walks $X(s)$ and $Y(s)$.

The following theorem reviews results about the stationary distributions of the process in the recurrent case:

Theorem 4.1.3 *(Theorem 3. in book [19])*

1. *For every $\eta \in \{0,1\}^S$ and every $x, y \in S$,*

$$\lim_{t\to\infty} P^\eta(\eta_t(x) \neq \eta_t(y)) = 0 \tag{4.4}$$

2. *We use the notion $\mathcal{I}$ stationary distributions for the process. This set $\mathcal{I}$ is convex, and we will denote its extreme points by $\mathcal{I}_e$. It is easy to see that the trivial stationary distributions for the linear voter model are the pointmasses δ_0 ($\eta \equiv 0$) and δ_1 ($\eta \equiv 1$). Moreover, $\mathcal{I}_e = \{\delta_0, \delta_1\}$.*

3. *Let $\mu S(t)$ denote the distribution of the process at time t, if the initial distribution was μ. If for the initial distribution $\mu\{\eta : \eta(x) = 1\} = \lambda$ for all $x \in S$, then*

$$\lim_{t\to\infty} \mu S(t) = \lambda\delta_1 + (1-\lambda)\delta_0.$$

4.2 Discretized voter model

In this section a possible implement of the voter model in discrete time is defined according to Definition 4.1.1 of the process in continuous time. As mentioned before, from now on we examine a model not only on $S = \mathbb{Z}^d$ d-dimensional integer lattice, but more general on an arbitrary undirected graph.

Definition 4.2.1 *Voter model in discrete time.*

An undirected graph $G(V, E)$ is given. The individuals are represented by the n nodes of the graph, contacts between them are the edges. Initially each node has an opinion represented by a number in $\{0, 1\}$: Each node independently chooses opinion 1 with probability v, and opinion 0 with $1 - v$. Then individuals can change their opinion randomly in time, under the influence of other vertices. Unlike in the case of epidemic spread, in the general voter model, interaction is possible between any pair of vertices. However, the frequency of the event that vertex x convinces vertex y depends on the distance of x and y, denoted by $d(x, y)$, just like in case of the studied continuous time. Usually $d(x, y)$ represents the classic graph distance, however on weighted graphs it denotes the smallest sum of weights of a path between x and y.

Since the process in continuous time can be modelled with a family of independent Poisson processes, for each pair of vertices (x, y) we have a Poisson process of rate $q(x, y)$, which describes the moments x convincing y. The rate $q(x, y)$ increases as the distance $d(x, y)$ decreases. In the continuous case, every time a vertex is influenced by another one, it changes its opinion immediately.

In our discretized voter process, there are two phases at each time step:

- *First, nodes try to share their opinions and influence each other, which is successful with probabilities $q(x, y)$ depending on the distance of the two vertices. Still, every vertex can "hear" different opinions from many other vertices within a time step.*

- *In the second phase, if a node v receives the message of m_0 nodes with opinion 0, and m_1 nodes with opinion 1 during a time step, then v will represent opinion 0 with probability $\frac{m_0}{m_0+m_1}$ during the next step, and 0 otherwise. If a node v does not receive any opinions from others at a time step, then its opinion remains the same.*

This way, the order of influencing message in the first phase can be arbitrary, and it is also possible that two nodes exchange opinions.

Now we specify the probability that a vertex x manages to share its opinion to vertex y in the first phase, in a way that vertices situated closer to each other have higher chance that their opinion "reaches" the other one. We transform graph distances $d(x, y)$ into a matrix of transmission probabilities with choice $q(x, y) = e^{-c \cdot d(x,y)}$, where c is a constant. This is not a direct analogue of the continuous case, but it is still a natural choice of a

decreasing function of d. (Usually we use $c = 2$, however later we also investigate cases $c \in \{0.5, 1, 2, 3\}$. Decreasing c escalates the process.)

4.2.1 Analytical study

In the following we analytically study the time-discretized voter model in our implementation defined as above. We examine the proportion of nodes with opinion 1 on a random graph of simplified structure: Let X_n denote the number of nodes representing opinion 1 after n iteration of the discretized voter model, while N denotes the number of nodes in the graph. The degree sequence of the random graph is given by $d = (d_i)_{i=1}^N$. A standard way to turn a difficult model into a much easier model is to take the mean-field limit. In book [22] the voter model is studied on the complete graph with N vertices instead of $\mathbb{Z}^d$, and it is also common to approx values by averaging.

Thus we make the following assumptions:

- There exists some k, such that each subgraph obtained from the k-radius of a node, is a tree. Then, for a given node u, the $\mathcal{N}_u^j = \{v : d(u, v) = j\}$ denotes the neighbours of distance j of node u. With the help of d degree sequence we can compute $|\mathcal{N}_u^j| = s_u^{(j)}$.
- For each node u the proportion of neighbours of distance j with opinion 1 is equal to the proportion of nodes with opinion 1 in the whole graph. So each node u has $s_u^{(j),1} = s_u^{(j)} \frac{X_n}{N}$ number of neighbours with opinion 1, and $s_u^{(j),0} = s_u^{(j)} \frac{(N-X_n)}{N}$ number of neighbours with opinion 0.

With the assumptions above, each node u independently with probability e^{-2j} (in case of $c = 2$) is convinced by other nodes being j distance from him. Then each node u is convinced expectedly

$$s_u^{(1)} \cdot e^{-2} + s_u^{(2)} \cdot e^{-4} + \cdots + s_u^{(k)} \cdot e^{-2 \cdot k} = \sum_{j=1}^{k} s_u^{(j)} \cdot e^{-2 \cdot j} \tag{4.5}$$

times, from which it is convinced with opinion 1

$$\sum_{j=1}^{k} s_u^{(j),1} \cdot e^{-2 \cdot j} \tag{4.6}$$

times.

Then a node u will represent opinion 1 with probability

$$\frac{\sum_{j=1}^{k} s_u^{(j),1} \cdot e^{-2\cdot j}}{\sum_{j=1}^{k} s_u^{(j)} \cdot e^{-2\cdot j}} = \frac{X_n}{N} \tag{4.7}$$

and opinion 0 with probability $\frac{(N-X_n)}{N}$, regardless of the degree of node u according to the mean-filed assumptions. In the discrete-time model a node influences others independently, so each node will represent opinion 1 independently with probability $\frac{X_n}{N}$.

Thus after $n+1$ iterations of the voter model, we conclude the following about the number of nodes with opinion 1:

$$\mathbb{E}(X_{n+1}|X_1, X_2, \ldots, X_n) = \mathbb{E}(X_{n+1}|X_n) = X_n + (N - X_n) \cdot \frac{X_n}{N} - X_n \cdot \frac{(N - X_n)}{N} = X_n. \tag{4.8}$$

Because of this martingale property of the process, with the assumptions above the expected number of individuals representing opinion 1 after any iteration of the voter model is the same as in the beginning of the process.

Generally we cannot ensure the process to satisfy the second assumption. Most of the random graphs also fail to possess the desired structure, however according to book [5] in case of a d-regular random graph created by the configuration model, an uniformly drawn node satisfies to have a tree for some radius j with probability converging to 1, as the size of the graph converges to infinity.

4.3 Simulations on graphs

In this section we run a simplified version of the discrete time voter model on random graphs. Firstly, we describe the implementation of creating Erdős–Rényi and Barabási–Albert graphs. Later we examine the probability of disappearing type 1 after some given *viter* iteration of the model, and investigate how it is possible to explain differences in outcome in connection with the underlying graph structure. At last, we study the process on Barabási–Albert graphs with different choices of nodes representing type 1, since in this graph the role of nodes is rather asymmetrical.

In the discretized voter model described in Definition 4.2.1, on a graph on n nodes, at every time step our algorithm consists of $\mathcal{O}(n^2)$ steps, which can be problematic for bigger graphs if our aim is to make sample with $viter = 100$ or 200 iterations of the voter model (*viter* denotes the number of steps of the voter model). However, with $c = 2$ a node x

convinces vertices y with $d(x, y) = 3$ only with a probability of $e^{-6} = 0,0025$. Thus we used the following simplified model: When we created a graph, we stored the list of edges and also calculated for each node the neighbours of distance 2. The simplified voter model spreads opinions only on these reduced number of edges with the proper probabilities. Thus we were able to run the original discretized model only on graphs with $n = 100$, while the simplified version can deal with $n = 1000$ nodes. We made the assumption that neglecting those tiny probabilities cannot significantly change the outcome of the process. From now on we only model the simplified version of the process.

4.3.1 Graph models

We study the voter model on Erdős–Rényi(n, p) and Barabási–Albert(n, m) random graphs.

Erdős–Rényi random graphs $ER(n, p)$**:** The $ER(n, p)$ random graph consists of n nodes, and every possible pair $x, y \in V$ is connected by an edge independently with probability $0 \leq p \leq 1$.

The Barabási–Albert graph is a dynamic model: we keep adding nodes to the graph one by one and connect them to the previous nodes with exactly m edges in relationship with the degrees of the old nodes. Thus, the network will consist of a few nodes with huge degree. Instead of giving a precise definition, we focus only on the implementation of the graph model:

Barabási–Albert random graphs $BA(n, m)$**:** Initially we start with a graph G_0. At every time step we add a new node v to the graph and attach it exactly with m edges to the old nodes with preferential attachment probabilities. Let D denote the sum of degrees in the graph before adding the new node, then we attach an edge independently to u with probability $\frac{d(u)}{D}$.

We generated graphs starting from $G_0 = ER\big(50, \frac{m}{(50-1)}\big)$ graph of complying density. Multiple edges can be created by the algorithm, however loops cannot occur. Attachment probabilities are not updated during a time step. Multiple edges do matter in the voter model, since they somehow represent a stronger relationship between individuals: opinion on a k-multiple edge transmits with a k-times bigger probability.

4.3.2 Probability of disappearing

We examine the voter model on graphs above to understand the differences of the process resulting from the structure. We compare graphs with the same density, $BA(1000, m)$

graphs with $m = \{4, 5, \ldots, 10\}$ and $ER(1000, p)$, where $p \in [0.004, 0.01]$. Initial probability of opinion 1 is set to 0.05 in both graphs. We compare the probability of disappearing the opinion with $viter = 50$ iteration of the voter model. We generated 10 different graphs from each structure and ran voter model on each 20 times with independent initial opinions. Altogether the results of 200 trials were averaged. Figure 4.1 shows the results.

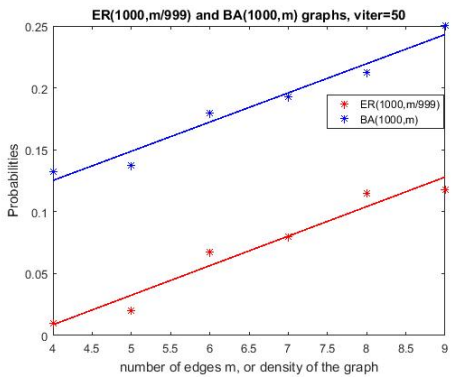

Figure 4.1: Probabilities of disappearing opinion 1 with variable graph density

Before the phase transition of Erdős–Rényi graphs, that is, with $p < \frac{\ln n}{n} \approx 0.007$ with $n = 1000$ nodes (BA graphs of the same density are belonging to $m \leq 7$) the graph consists of several components with high probability. It is possible that nodes being in a tiny component of the graph, or even isolated nodes get opinion 1 initially, resulting in the co-existence of the two opinions. (In a small component of the graph within a few time steps any opinion can disappear easily, remaining the same for the rest of the process since no other nodes can influence them.) This could be one of the reasons why disappearing probabilities in ER graphs are significantly less. The another reason is in connection with the next part: Since some nodes in BA graphs can possess outstanding number of degrees, while most of them have only a few, the process can be influenced by the properties of nodes chosen to represent opinion 1 initially. On the other hand, on BA graphs not only the probability of the vanishing of opinion 1 is higher, but it is also more likely to get extreme results. Proportions of opinion in ER graphs are more stable. We can generally say that the increase of the density of graphs escalates voter model, since in expected value more convictions happen at every time step, resulting in a more volatile proportion of opinions, and thus in higher probabilities of the disappearance of the underrepresented opinion.

Different choices of L_0

As mentioned before, in this sequel we investigate extreme outcomes of the process caused by one of the most important properties of Barabási–Albert graphs. Since nodes do not play a symmetrical role in Barabási–Albert graphs, fixing the proportion of nodes representing opinion 1 (we usually use $v = 0.05$, so 50 nodes represent opinion 1 in expected value), but changing the position of these nodes in the graph can lead to different results. We examined the following three ways of initial opinion setting:

- randomly: Each individual chooses opinion 1 with probability v.
- "oldest nodes": We deterministically set the first 50 nodes of the graph to represent opinion 1. These nodes usually have the largest number of degrees, thus they play a crucial part in the process. Not only have they large degrees, but they are also very likely to be connected to each other (this is the densest part of the graph).
- "newest nodes": We deterministically set the last 50 nodes of the graph to represent opinion 1. These nodes usually have only m edges, and they are not connected to each other with a high probability.

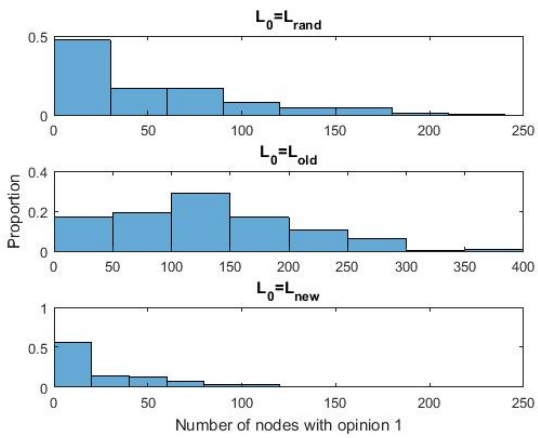

Figure 4.2: Distribution of nodes with opinion 1, different L_0

The histogram on Figure 4.2 shows the distribution of nodes with opinion 1 with the three different choices of L_0 vectors after $viter = 50$ iterations of the voter model on $BA(1000, 5)$ graphs. We experience differences in terms of probabilities of disappearing opinion 1: with random opinion distribution 11%, with L_{new} almost one third of the cases

resulted in extinction of opinion 1, while for L_{old} this probability was negligible (0.005%). Actually, for L_{old} after only one iteration of the voter model it is impossible to see any structure in the distribution of individuals with opinion 1. Vector of opinions became totally random, but with a probability of 0.12. Indeed only with one step of the voter model individuals with opinion 1 could double in number, however opinion 1 cannot take advantage of any special positions in the graph anymore. All in all, giving a certain opinion to individuals who are more likely to be connected in the graph, reduces the probability of disappearing, since they can keep their opinion with a high probability, while with opinion 1 scattered across the graph (in case of L_{new} as well as L_{rand}) with a dynamic parameter setting of c number of individuals with opinion 1 can reduce drastically even in a few time steps.

Bibliography

[1] Frank Ball,Tom Britton, Etienne Pardoux, *Stochastic Epidemic Models with Inference*, Springer, 2019

[2] Frank Ball, Peter Donnelly *Strong approximations for epidemic models*, Stochastic Process. Appl., 55 (1995), pp. 1–21

[3] Boldog P, Tekeli T, Vizi Zs, Dénes A, Bartha F, Röst G, *Risk assessment of novel coronavirus COVID-19 outbreaks outside China*, Journal of Clinical Medicine, 9(2), 571 (2020)

[4] Béla Bollobás, *A Probabilistic Proof of an Asymptotic Formula for the Number of Labelled Regular Graphs*, Europ. J. Combinatorics (1980) 1, 311-316

[5] Béla Bollobás, *Random Graphs Second Edition*, Cambridge University Press (2001)

[6] Bála Bollobás, Svante Janson, and Oliver Riordan. (2007) *The phase transition in inhomogeneous random graphs.* Random Struct. Algorithms, 31(1):3–122

[7] C. Borgs, J. Chayes, L. Lovász, V.T. Sós and K. Vesztergomi, *Limits of randomly grown graph sequences*,Eur. J. Combin. 32 (2011), 985-999.

[8] Britton, T., Janson, S. and Martin-Löf, A. (2007). *Graphs with specified degree distributions, simple epidemics, and local vaccination strategies*, Adv. in Appl. Probab. 39(4): 922-948.

[9] Anna D. Broido, Aaron Clauset, *Scale-free networks are rare*, arXiv preprint, 1801.03400.

[10] Callaway, D.S., Newman, M.E.J., Strogatz, S.H., and Watts, D.J. (2000) *Network robustness and fragility: percolation on random graphs.* Phys. Rev. Letters. 85

[11] Durrett R., *Random graph dynamics*, Vol. 20 of Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2007.

[12] Durrett R., Claudia Neuhauser, *Coexistence results for some competition models*, Ann. Probab. 7 (1), (1997), 10–45.

[13] Fransson, C. and Trapman, P. (2019). *SIR epidemics and vaccination on random graphs with clustering*

[14] Remco Van Der Hofstad, *Random graphs and complex networks* Volume 1. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, Cambridge, 2017.

[15] P. W. Holland, K. Laskey, and S. Leinhardt. (1983) *Stochastic blockmodels: First steps. Social Networks*, 5(2):109–137

[16] P. Jagers, *Branching Processes with Biological Applications*, Wiley, 1975.

[17] Kendall, D.G. *Deterministic and stochastic epidemics in closed populations*, Berekely Symp. Staisti. Prob. Vol. 4, 1956, 149–165.

[18] Diána H. Knipl, Gergely Röst *Modelling the strategies for age specific vaccination scheduling during influenza pandemic outbreaks*, Math. Biosci. Eng. 8 (1) (2011)

[19] Thomas M. Liggett, *Interacting Particle Systems*, Springer, New York, 1985.

[20] T.M. Liggett, *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes*, Springer, Berlin, 1999.

[21] Sophie Pénisson, *Conditional limit theorems for multitype branching processes and illustration in epidemiological risk analysis* https://tel.archives-ouvertes.fr/tel-00570458

[22] J.M. Swart, *A Course in Interacting Particle Systems*, 2015.09.26., arXiv:1703.10007

[23] Viet Chi Tran, *Stochastic epidemics in a heterogeneous community*, Springer, 2019

[24] https://github.com/bognaredith/spreading_models

[25] *MASZK* https://www.facebook.com/szte.sed/

[26] https://www.worldometers.info/coronavirus/