Eötvös Loránd University

Institute of Mathematics

Department of Computer Science

Dynamics of Epidemic Spreading on Hypergraphs

Balázs Csegő Kolok

Applied Mathematics MSc

Supervisor:

Zoltán Király

associate professor

Budapest, 2021

NYILATKOZAT

Név: Kolok Balázs Csegő

ELTE Természettudományi Kar, szak: alkalmazott matematikus, Msc

NEPTUN azonosító: ADUG96

Diplomamunka címe:
Dynamics of Epidemic Spreading on Hypergraphs
(Vírusterjedési dinamikák hipergráfokon)

A **diplomamunka** 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, 2021.05.31

a hallgató aláírása

Acknowledgement

First of all I am very grateful to my supervisor Zoltán Király, who drew my attention to the theme, guided me with a lot of patience and provided me all of his experience and competence. Without his efforts and perseverance this Thesis would not have been finished.

Secondly, I would like to say thank you to my family, who encouraged me a lot and endured with love during this period of my life.

Contents

Introduction

Since the appearance of the novel coronavirus SARS-CoV-2 news and theories about the virus spreading has become part of our daily life. Numerous scientific paper investigated the spreading dynamics and the effects of the epidemic in any sense. There are several mathematical models to describe disease spreading phenomena with the help of differential equations [1] and stochastic processes [2], [3].

The Thesis inspired by the fact, that the field of spreading phenomena on graphs is a widely studied area in the network science [4], but the models that use hypergraph as underlying structure are not well discovered. Why using hypergraph models can be more appropriate? Using networks for modelling the interactions between individuals is a common tool of network science. The virus spreads through social connections, so it is convenient to use dyadic links in social networks, where two people can infect each other if they have a link. However, in the case of airborne diseases like SARS-CoV-2, the virus spreads through the common air of a place. People in one place share their common air. It seems reasonable to describe this relation of the individuals with one multyadic connection rather than many dyadic links. The hypergraphs are decent structures for representing networks with multiadyc connections.

Modelling society with hypergraphs raises many questions. Although social network models using graphs have deep literature, the hypergraph models are not well described. In the aim of building real representations of the society, we extend some well-known graph models like Erdős-Rényi, Barabási-Albert models for hypergraphs. We see the individuals as the nodes and the communities of the society and the events or occasionally meetings as the hyperedges of the hypergraph representation. It can be an interesting question how the epidemic spreading on hypergraphs differs from the well-studied graph models, what we gain from the hyperedge representation of a

community. The main benefit of this approach that we can handle the different types of groups together and specify their characteristics. Like in our study of epidemic spreading, we can create household, workplace and event hyperedges and write down different spreading dynamics inside them. The second important characteristics of representing the society with a hypergraph that we can define the sizes of the communities. This becomes important when we study the recent fight against the coronavirus when there were active restrictions on the size of the gatherings.

In this Thesis, we are focusing on two main topics, the differences of the disease spreading on hypergraphs and graph representations, and the effects of the measures against the virus spreading.

In the first chapter, we introduce some basic definitions and define the extension of random graph models to hypergraphs. We describe the evolving hypergraph extension of Barabási-Albert and other real-world hypergraphs. In real-world networks, the influence of the nodes can be unequally distributed. Some nodes might have fewer and some have a very high number of connections. In the aspect of the virus spreading searching for these highly influential individuals is crucial to slow down the virus. Thus, after the definition of the models, we investigate methods to calculate node centrality or importance measure for hypergraphs.

In the second chapter, we present well-known epidemic models like the Kermack and McKernick model [1] and the stochastic model from Tom Britton [2]. These model definitions serve for a better understanding of our epidemic model.

In Chapter 3, we develop a discrete epidemic on hypergraphs. We make comparisons between the epidemic dynamics on hypergraph and on its clique expansion graph. After that, we investigate how the epidemic dynamics depend on the structure of the underlying hypergraph.

In Chapter 4, our goal is to describe the SARS-CoV-2 virus spreading with our hypergraph model and investigate the effects of precautions like wearing masks, social distancing or measures like tests and quarantine and mass vaccination methods. At the end of the chapter, we fit our model to the last wave of COVID-19 in the Spring of 2021 in Hungary.

Chapter 1

Hypergraphs

1.1 Definition

The theory of finite sets has been studied since the 1930s [5]. These systems were firstly called hypergraphs in [6] by Berge in 1972 and introduced as the extension of the graphs. The hypergraphs are decent tools for describing complex networks. They allow us to interpret not just pair-wise connections, as the graphs could, but also polyadic linkages.

Definition 1. *A pair $H = (V, E)$ is a hypergraph if V is a set and E is a set of subsets of V. We call V the nodes or the vertices and E the hyperedges or shortly the edges of the hypergraph H. We also use the notations $V(H)$ for the nodes and $E(H)$ for the hyperedges of H.*

Let $H = (V, E)$ be a hypergraph. If we allow hyperedges only with cardinality two, then H is a graph. Generally, if H has only edges of size k, then H is called k-uniform. The degree of a node v means the number of edges incident to v. We note it with $d_H(v)$ or $d(v)$ if it is clear what is the hypergraph. If every node in the hypergraph has the same degree d, then we call H d-regular. An alternating sequence $P = (u_1, e_1, u_2, e_2, \ldots, e_{k-1}, u_k)$ is a path of length $k-1$ between u_1 and u_k in H, if $u_i \in V$ for $i = 1, \ldots, k$ and $e_i \in E$ and $u_i, u_{i+1} \in e_i$ for $i = 1, \ldots, k-1$. The distance between $u, v \in V$ is the length of the shortest path between them and noted with $dist(u, v)$.

One could make the clique expansion graph of a hypergraph, which can function as a graph representation of the hypergraph on the same nodes. We will use this representation to compare epidemic spreading dynamics on graph and hypergraph.

Definition 2. *Let $H = (V, E)$ be a hypergraph. We gain the clique expansion graph $G^H = (V^H, E^H)$ of H, if we set $V^H = V$ and for all hyperedge h, we create a clique on the nodes contained by h. Here we allow multiedges between vertices, there are k parallel edges between $u, v \in V$, if exactly k hyperedges contain u and v. From now on, for the sake of simplicity we call the clique expansion graph of a hypergraph as its clique graph.*

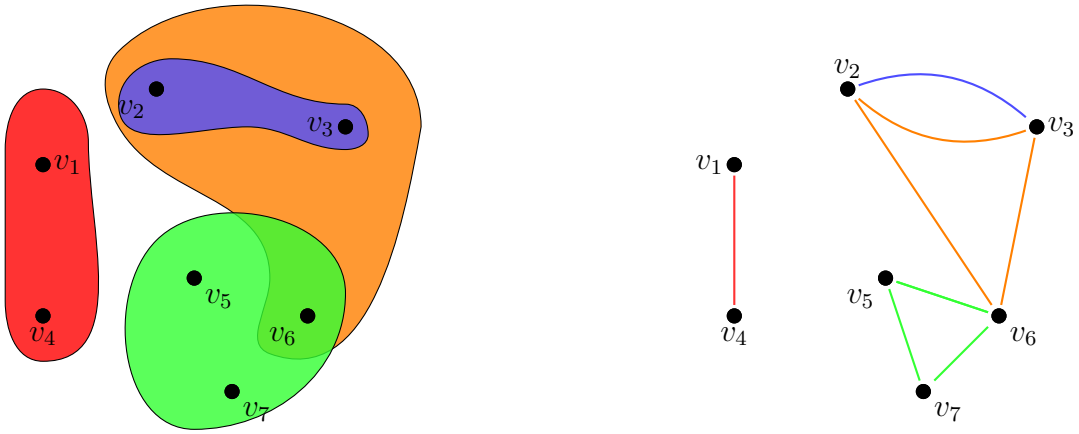


Figure 1.1: We presented a hypergraph on the right and its clique graph on the left side. We coloured the cliques of the graph as they derived from the hyperedges of the hypergraph.

1.2 Random hypergraphs

The introduction of random graphs is originated from Erdős and Rényi [7]. Since then, random graphs have encouraged many mathematicians resulting several interesting pieces of literature in the area and inspired a new field of science called network science. In this section, we would like to describe some well known random graph models and extend them to hypergraphs along [8] which is a decent summarise of the theme.

1.2.1 Erdős-Rényi model

There are alternate models, which appeared to be equivalent in many graph properties. The first was defined in [7] by Erdős and Rényi. Let $G[n, m]$ be a random graph on n vertices and with m edges. The m edges are chosen independently with uniform probabilities from the possible $\binom{n}{2}$ edges. In [9] Gilbert described another random model, we note it with $G[n, p]$. In this model for a given n and $0 \leq p \leq 1$ we generate $G[n, p]$ on n nodes creating an edge between every pair of nodes with probability p. These edge creations are independent from one another. If $p\binom{n}{2} = m$ then $G[n, m]$ and $G[n, p]$ will be almost equivalent.

The extensions of these models are quite straightforward. Using the definitions from [8], a random k-uniform hypergraph $H[n, m, k]$ is a hypergraph uniformly chosen from all possible $\binom{\binom{n}{k}}{m}$ k-uniform hypergraphs having n vertices and m edges. The extension of Gilbert's model is denoted by $H[n, p, k]$. In this scenario for every set of size k in V we either create a hyperedge with probability p $(0 \leq p \leq 1)$ or we do not with probability $1 - p$. These creations are independent from one another. Note that if $k = 2$ then we get a $G[n, m]$ from the first, and a $G[n, p]$ from the second model. More generally, $H[n, \underline{m}, \leq k]$, where $\underline{m} = (m_1, \ldots, m_k)$, denotes the random hypergraph that is a union of the $H[n, m_1, 1], \ldots, H[n, m_k, k]$ hypergraphs. Similarly, one can define the $H[n, \underline{p}, \leq k]$ random hypergraph, where $\underline{p} = (p_1, \ldots, p_n)$.

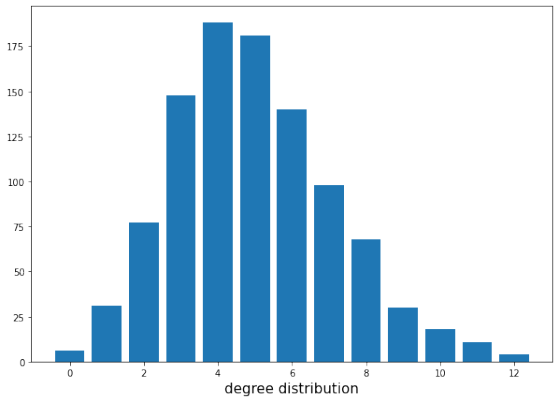


Figure 1.2: The degree distribution of an Erdős-Rényi hypergraph $H[1000, 1000, 5]$.

Some of the famous problems related to the random graphs can be extended to hypergraphs like connectivity, but not all of their solution derives easily from the case $k = 2$. In [8] one can find a detailed discussion on these problems.

1.2.2 A random d-regular hypergraph

We present a way to construct a special type of d-regular hypergraph. Let V be the nodes of the hypergraph H, where $|V| = n$. To create the hyperedges, we generate d random partitions of the nodes, where we sample the size of the subsets in the partition from distributions $F_1, \dots, F_d$. Let us denote the resulting partitions with $\mathcal{W}_1, \dots, \mathcal{W}_d$. After that, the union of these partitions will be the hyperedges of H, i.e. $\bigcup_{i=1}^{d} \mathcal{W}_i = E(H)$. From now we say simply d-regular hypergraph if we want to refer to this type of hypergraph and this will not cause any confusions because we won't work any other d-regular hypergraphs throughout the document.

1.3 Real-world hypergraphs

Several experimental studies have confirmed that complex networks in real life like the WWW network, protein networks or social networks differ in many properties from the random networks [10], [11]. And yet, we have just talked about random networks. We will collect these properties based on the papers [12] and [13].

Small-world phenomenon: The average distance between two nodes is relatively small compared to the size of the network. Erdős and Rényi proved that in their random graph model the average distance depends logarithmically on the size of the graph.

Clustering coefficient: Let $G = (V, E)$ be a graph. The clustering coefficient of a node v measures the density of the vicinity of v. Precisely, let's note the neighbours of v with $N(v)$ and with $G[N(v)]$ the graph induced by them. The clustering coefficient of v is

$$C(v) = \frac{2|E(G[N(v)])|}{|N(v)|(|N(v)| - 1)}.$$

One can compute the clustering coefficient of G by aggregating the clustering coefficients on every node. Let $H = (V, E)$ be a hypergraph and let us note the hypergraph induced by the neighbours of v with $H[N(v)]$. If H is a k-uniform, then we could define the clustering coefficient of v as follows:

$$C(v) = \frac{|E(H(N(v)))|}{\binom{|N(v)|}{k}}.$$

If H is not k-uniform, then

$$C(v) = \frac{|E(H[N(v)])|}{2^{|N(v)|}}.$$

One can compute the clustering coefficient of H analogously to the case of graphs. Watts and Strogatz in their work [12] have pointed out that real networks have rather a high clustering coefficient and the Erdős-Rényi model leads us small coefficient. So they proposed a model, which generates a graph with a high clustering coefficient and small average distance.

Scale-free property: The degree distribution of a real network fits to a power law distribution. Thus, $\mathbb{P}(d)$, i.e. the probability that a randomly chosen node has d neighbours, is proportionate to $cd^{-\alpha}$ independently from the scale of the network, where α is positive and c is the constant that normalises the distribution. We call this graph property as scale-free property and it was demonstrated by Albert and Barabási in [13]. In this paper, they also proposed a new model, which is called Barabási-Albert model or Preferential Attachment model.

1.3.1 Hidden parameter model

In this Subsection we present a simple way to generate a hypergraph with a help of two distributions F_V, F_E which may control the hypergraph's degree and edge size distributions respectively. Let $H = (V, E)$ be a hypergaph with $V = \{v_1, \ldots, v_n\}$ nodes and $E = \emptyset$. First, sample n values $\eta_1, \ldots, \eta_n$ from the distribution F_V. Then assign the weight $w_i = \frac{\eta_i}{\sum_{j=1}^{n} \eta_j}$ to node v_i for $i = 1, \ldots, n$. Then add m hyperedges one by one. To create the hyperedge h first, sample its size k_h from the given distribution F_E, then chose the k_h incident nodes to build up the hyperedge with probabilities proportional to their weights. So we add $h = \{v_{i_1}, \ldots, v_{i_{k_h}}\}$ hyperedge to the hypergraph, where we chose i_j to be l with probability:

$$\mathbb{P}(i_j = l) = w_l = \frac{\eta_l}{\sum_{j=l'}^{n} \eta_{l'}}$$

Let us note if we use the uniform distributions with expected numbers 1 and k for F_V and F_E respectively we gain an $H[n, m, k]$ Erdős-Rényi hypergraph. In Figure 1.3 one can see the degree distributions when we used the power law distribution for F_V with parameters $\alpha = 2$ and $\alpha = 5$ in the other case and uniform distribution with mean 5 and 2 for F_E in both cases.

1.3.2 Barabási-Albert model

Bollobás and Riordan made the definition of the Barabási-Albert model mathematically precise in their paper [14]. In their work, the evolution of the graph starts from a single node v_1, and it is considered the graph G_m^1. In step t we get G_m^t from G_m^{k-1} by adding a new node and m edges to the graph.

First, we define the case of $m = 1$. In the step t we add the new edge between v_t and v_i choosing i randomly according to the following probabilities:

$$\mathbb{P}(i = l) = \begin{cases} \frac{d_{G_1^{t-1}}(v_l)}{2t-1} & \text{, if } l = 1, \dots, t-1 \\ \frac{1}{2t-1} & \text{, if } l = t \end{cases}.$$

If $m \geq 2$, then we add m edges one by one counting all the edges what we have added to the model and the actual new edge as one-half of it has already attached to v_t. Or we could say equivalently, we run the same process as in the case of $m = 1$ until we get G_1^{tm} with nodes $v'_1, \dots, v'_{tm}$. Then we gain G_m^t by contracting $v'_{i1}, \dots, v'_{im}$ to one node v_i for each $i = 1, \dots, t$. In [14] Bollobás and Riordan have proved that the diameter of G_m^n is asymptotically $\log n$, if $m = 1$, and in case $m \geq 2$ it is in $\Theta(\log \log n)$.

The question is, how can one extend the generative model for real-world hypergraphs? There are several different approaches from the past few years [15], [16]. We present a model, which generates a k-uniform hypergraph based on the Barabási graph model. Unlike in the definition of Bollobási and Riordan, we will not allow loops in the hypergraph, because in our later work we need loop-free hypergraphs.

Let $H_1^1 = (V_1^1, E_1^1)$ be a hypergraph, where $V_1^1 = \{v_1, \dots, v_k\}$ and E_1^1 stands from one hyperedge of size k on the nodes. In the step t we get hypergraph H_1^t from H_1^{t-1} by extending the hypegraph with node v_t and adding a new hyperedge $e = \{v_{t_1} = v_t, v_{t_2} \dots, v_{t_k}\}$. We chose $v_{t_2}, \dots, v_{t_k}$ randomly proportional to the degrees of the nodes. Thus

$$\mathbb{P}(t_i = l) = \frac{d_{H_1^{t-1}}(v_l)}{k(t-1)} \qquad l \in \{1, \dots, t-1\}. \tag{1.1}$$

Note that this model will not lead us to a k-uniform hypergraph. In case after the random choosing, there is a node which appears among $v_{t_2} \dots, v_{t_k}$ more than

once, then we add a hyperedge with size less than k. We can generate k-uniform hypergraph too. If for t_i we chose an index which has appeared already in $t_1, \dots, t_{i-1}$ we sample again from the notes with the same probabilities and repeat this while we get an index that we did not choose before. This repetition will end with a finite expected number of repeats.

In case $m \geq 2$ we proceed analogously to the model of Bollobás and Riordan. Initially, let the hypergraph be $H_m^1 = H_1^1$. Then in the step t we add node v_t to the hypergraph and create hyperedges $e_{t_1}, \dots, e_{t_m}$, where $e_{t_i} = \{v_{t_{i1}} = v_t, \dots, v_{t_{ik}}\}$ for $i = 1, \dots, m$. We chose t_{ij}, according to the following probabilities:

$$\mathbb{P}(t_{ij} = l) = \frac{d_{H_m^{t-1}}(v_l) + d_{li}}{\sum_{l'=1}^{t} d_{H_m^{t-1}}(v_{l'}) + d_{l'i}}, \qquad t \geq 2, j \geq 2, \tag{1.2}$$

where d_{li} is equal to $|\{e_{t_r} | v_l \in e_{t_r}, 1 \leq r < i\}|$.

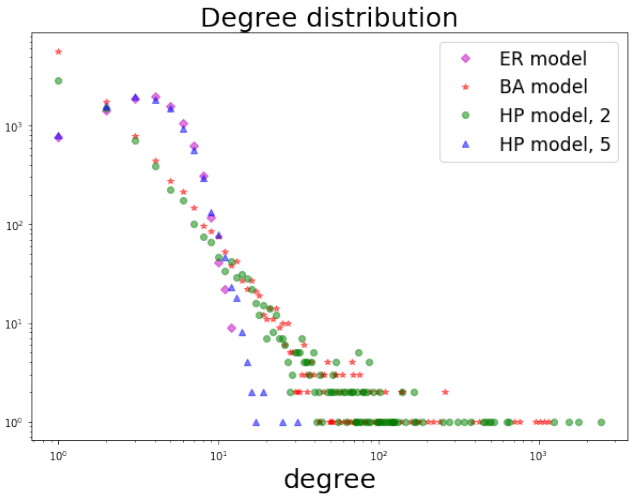


Figure 1.3: The degree distribution of Erdős-Rényi hypergraph is from a $H[10000, 10000, 5]$. For the two Hidden parameter model hypergraphs we chose F_V to be a power law distribution with parameter α, where $\alpha = 5$ for the blue and $\alpha = 2$ for the green one. The Barabási-Albert model is with $n = 10000, m = 1, k = 5$.

1.3.3 Bianconi-Barabási model

If we have a bias on the degree distribution, of the nodes, like in the example of social networks we may assume that mid-age people have more connections than other age groups, then we can fit our model according to this assumption.

We generate the hypergraph H_m^t in almost the same way as we did in the case of the Barabási model. The initial step is the same, but in whenever we add node v_t to the graph (including the initial step), we assign a positive fitness η_t to the node. In step t we add m hyperedges to the hypergraph to gain H_m^t. Let us note these hyperedges with $e_{t_1}, \ldots, e_{t_m}$, where $e_{t_i} = \{v_{t_{i1}} = v_t, \ldots, v_{t_{ik}}\}$ for $i = 1, \ldots, m$. We chose t_{ij}, according to the probability distribution:

$$\mathbb{P}(t_{ij} = l) = \frac{\eta_l \left(d_{H_m^{t-1}}(v_l) + d_{li}\right)}{\sum_{l'=1}^{t} \eta_{l'} \left(d_{H_m^{t-1}}(v_{l'}) + d_{l'i}\right)}, \qquad t \geq 2, j \geq 2, \tag{1.3}$$

where d_{li} is equal to $|\{e_{t_r} | v_l \in e_{t_r}, 1 \leq r < i\}|$. Note that, if we set every $\eta_t = 1$ for all nodes we get back the Barabási model.

Our aim in Chapter 4 to represent social networks where not every community has the same size, so we want to define models where not every hyperedge has the same size. We need a little manipulation to our existing Bianconi-Barabási hypergraph model. In each step before we generate a new hyperedge h we determine the size of h, k_h. We can sample this k_h from any given distribution and we chose the consisting k_h nodes just after we determined the size. The choosing method is the same as before but instead of k we use k_h. This generalisation can be applied on every model what we have mentioned before. For requiring the exact k_h size, we can use the method we used for Barabási-Albert hypergraphs to reach k-uniform hypergraphs.

1.4 Important nodes in hypergraphs

Our goal is to define the relative importance of the nodes in hypergraphs, so we could use that for defining the vaccination order in our SARS-CoV-2 hypergraph model in Chapter 4. Thus, we aim to find the influence of the nodes in a spreading phenomenon. Let $H = (V, E)$ be a hypergraph with $V = \{v_1, \ldots, v_n\}$ and $E = \{e_1, \ldots, e_m\}$. The incidence matrix of H is $B \in \mathbb{R}^{n \times m}$, where

$$(B)_{ij} = \begin{cases} 1, \text{ if } v_i \in h_j \\ 0, \text{ if } v_i \notin h_j \end{cases} \qquad \forall v \in V, \forall h \in E$$

If there is a weight function w_V on the nodes, then we represent it with an $n \times n$ real diagonal matrix $W_V = diag\left(w_V(v_1), \ldots, w_V(v_n)\right)$. If there is a weight function

w_E on the hyperedges, then we represent it with an $m \times m$ real diagonal matrix $W_E = diag\,(w_E(h_1), \ldots, w_E(h_m))$.

1.4.1 Degree

A simple way to define a node's importance is by its degree. This approach is based on a local measurement and does not concern the global structure of the hypergraph. For example, two nodes with the same degree have the same measure no matter how important their neighbours are. If there is a weight function w_E defined on the edges, we can use the weighted degree as importance function.

$$\mu_d(v_i) = (BW_E\mathbb{1})_i$$

Where $\mathbb{1}$ is an all 1, m dimensional vector.

1.4.2 Eigenvector centrality

For graphs, the Perron eigenvector of the adjacency matrix assigns a value for each node in a graph. This value called the eigenvector centrality of the node. We may use this node centrality for measure the importance of the nodes. There are two ways to extend the node centrality for hypergraphs. One is when we just simply calculate the spectral centrality of the hypergraph using some of its expansion graphs like the clique graph or the bipartite representation. The other is based on using the higher dimension adjacency tensor of the hypergraph, but in this case, the hypergraph must be k-uniform for some k.

One can find these extensions for hypergraphs well described in a recent study [17]. They also presented a new method for calculating the node centrality of hypergraphs using four nonlinear functions. We will have a brief overlook of these methods.

First, we define the eigenvector centrality for graphs. Let $G = (V, E)$ be a graph on node $V = \{v_1, \ldots, v_n\}$ with adjacency matrix A_G. For a node, we define its centrality with the help of its neighbours. Precisely, we calculate the centrality of a node with a positive linear combination of their neighbours' centrality. This can be written down as

$$A_G\mathbf{x} = \lambda\mathbf{x} \qquad \lambda > 0,\ \mathbf{x}_i > 0 \ \ \forall i = 1, \ldots, n.$$

Due to the Perron-Frobenius theorem, this eigenvector problem has a unique solution if A_G is irreducible, i.e. G is connected. This way we gain the centrality of a node v_i by taking $\mathbf{x}_i$. This idea leads us to measure the importance of the hypergraph's nodes by taking its clique graph or its bipartite representation. Using the clique graph is quite straightforward. We just have to calculate the adjacency matrix of the clique graph and determine the Perron eigenvector which provides us with the required centrality of the nodes.

However, if we are talking about hypergraphs we may determine the centrality of not just nodes but of hyperedges too. To follow the previous reasoning, we could say that the centrality of the nodes incident to the hyperedge may determine the centrality of the hyperedge. Like in the example of social networks, if important people are in one community, that must be a prominent community and vice versa. Thus, if a person has important companies he or she must be an important people. So the centrality of the incident hyperedges determines the centrality of the node. Given this we could write down this relationship with equations for hypergraph $H = (V = \{v_1, \ldots, v_n\}, E = \{h_1, \ldots, h_m\})$:

$$\begin{cases} \sum_{h_j \in E: v_i \in h_j} \mathbf{y}_j & = \lambda \mathbf{x}_i \qquad \forall i = 1 \ldots n \\ \sum_{v_j \in V: v_i \in h_j} \mathbf{x}_i & = \lambda \mathbf{y}_j \qquad \forall j = 1 \ldots m \end{cases}, \lambda > 0,\ \mathbf{y}_j, \mathbf{x}_i > 0. \tag{1.4}$$

The centrality of the node v_i is $\mathbf{x}_i$ and the centrality of the hyperedge h_j is $\mathbf{y}_j$. To write the equation system in matrix form we use B, the incidence matrix of H.

$$\begin{cases} B\mathbf{y} & = \lambda \mathbf{x} \\ B^T\mathbf{x} & = \lambda \mathbf{y} \end{cases} \qquad \lambda > 0,\ \mathbf{x}, \mathbf{y} > \mathbf{0}. \tag{1.5}$$

If we want to write down this with one matrix then we get the matrix of the bipartite graph's adjacency matrix which represents the hypergraph.

$$\begin{pmatrix} 0 & B \\ B^T & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \qquad \lambda > 0,\ \mathbf{x}, \mathbf{y} > 0 \tag{1.6}$$

This means that the Perron eigenvector of the bipartite representation of a hypergraph can define the centrality of the nodes and the hyperedges.

If we have weights on the nodes and the edges we may compute the importance this way:

$$\begin{pmatrix} 0 & BW_E \\ B^T W_V & 0 \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} = \lambda \begin{pmatrix} \mathbf{x} \\ \mathbf{y} \end{pmatrix} \qquad \lambda > 0, \mathbf{x}, \mathbf{y} > 0. \tag{1.7}$$

Until now, we worked with graph representations of hypergraphs. In the spectral theory of hypergraph, the tensors are the most commonly used objects to represent hypergraphs. Of course, in this case the hypergraph $H = (V, E)$ what we represent must be k-uniform. We can describe it with the following k dimensional adjacency tensor $\mathcal{A}$:

$$\mathcal{A}_{i_1,\ldots,i_k} = \begin{cases} w_E(h) & \text{if } (v_{i_1}, \ldots, v_{i_k}) = h \in E, \\ 0 & \text{otherwise,} \end{cases} \tag{1.8}$$

where w_E is a weight function on the hyperedges. There are several ways to define the eigenvalues of such tensor, but we are going to use what appears in [17].

$$\sum_{i_2,\ldots,i_k} \mathcal{A}_{i_1,\ldots,i_k} \mathbf{x}_{i_2} \mathbf{x}_{i_3} \ldots \mathbf{x}_{i_k} = \lambda \mathbf{x}_{i_1} \tag{1.9}$$

Instead of trying to solve this, we will use the theorem and the algorithm presented in the article [17] which uses four nonlinear functions to calculate the eigenvector problem (1.9). They generalised equations (1.4) to nonlinear equations to calculate the centrality of the nodes and the hyperedges:

$$\begin{cases} g\left(BW_E f(\mathbf{y})\right) = \lambda \mathbf{x} \\ \psi\left(B^T W_V \phi(\mathbf{x})\right) = \mu \mathbf{y} \end{cases} \qquad \mathbf{x}, \mathbf{y} > \mathbf{0}, \ \mu, \lambda > 0, \tag{1.10}$$

where f, g, ϕ, ψ are real functions operating element-wise on vectors. The following theorem is Theorem 3.1 from [17] .

Theorem 1. *([17] Theorem 3.1) Let $H = (V, E)$ be a k-uniform hypergraph with $w_V(v) = 1$ for all $v \in V$. If $\mathbf{x}$ is a positive solution of (1.10) with $f(\mathbf{x}) = \mathbf{x}$, $g(\mathbf{x}) = \mathbf{x}^{\frac{1}{2}}$, $\psi(\mathbf{x}) = e^{\mathbf{x}}$ and $\phi(\mathbf{x}) = \ln(\mathbf{x})$, then it is an eigenvector centrality solution of the tensor eigenvalue problem (1.9).*

After this theorem they also proved that under the right preconditions, there is a unique solution for eigenvalue problem (1.10). They derived their proof from the nonlinear Perron–Frobenius theorem for multihomogeneous mappings. They also mention that if we use this centrality for nonuniform hypergraphs then it will generalizes the ZEC eigenvector centrality introduced in [18]. They also presented a power algorithm that can compute the solution for (1.10) with arbitrary low error. We included this to Appendix A.

1.4.3 PageRank

The PageRank algorithm is mainly used for ranking websites on the WWW network. It is a powerful method to categorise the nodes by their influence on the network. The main idea comes from a process on the WWW network. Let us imagine a random surfer on the internet, who moves randomly from a website to another through the links in the current website where he stays. Sometimes he randomly jumps to a random website from all the websites. Whenever he reaches one site he notes it. The importance of websites comes from how many times the surfer visited the page. The algorithm runs on graphs and we use it to calculate the importance of the nodes in hypergraphs. We will use the bipartite expansion of our hypergraph and run the PageRank algorithm on that. This way the algorithm ranks both the nodes and the hyperedges of the hypergraph. See more detailed in [19] or [20].

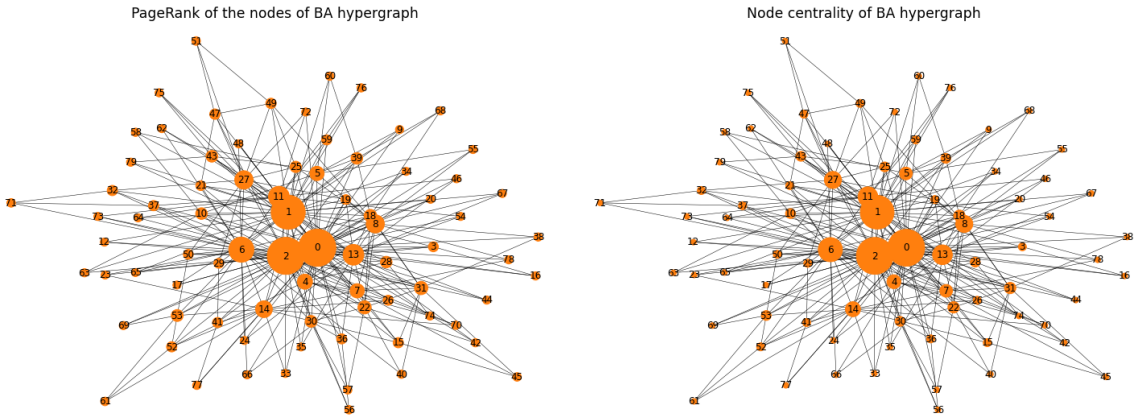


Figure 1.4: Node importance measures of a 4-uniform Barabás-Albert hypergraph with $n = 80$, $m = 1$ on its clique graph. left: PageRank; right: Node centrality from Tudisco's method

Chapter 2

Epidemic models

In this chapter, we take an overview of the SEIR epidemic model. However, we will use discrete and stochastic models to simulate virus spreading on hypergraphs later on. For better understanding, we present continuous and deterministic models in the first section of this chapter. After that, we will move on to the discrete and stochastic models in the second section. One can find a decent monograph from Daley and Gani [21] and a detailed survey from Tom Britton [2] of the theme.
The first epidemic models were proposed by Kermack and McKendrick in [1]. They grouped the individuals into compartments according to their state of the disease. Their model is still the base of the modern epidemic modelling systems. For a more realistic approach, we will investigate some extensions of their model.

2.1 The deterministic SEIR model

Firstly, we present the basic SEIR model on closed population, i.e. the number of people will be constant N. In this model the population consists of four compartments: susceptible (S), exposed (E), infectious (I), and recovered (R). We will use the following notations through the document:

S(t): The number of susceptible individuals at the time t, who have not been infected by the disease. If an infectious individual makes an infectious contact with a susceptible one then the susceptible contracts the disease and transitions to exposed.

E(**t**): The number of exposed individuals at time t, who have been infected by the disease, but not yet infectious. We call the time latent period, that an individual spends in the exposed state.

I(**t**): The number of infectious individuals at the time t, who have been infected by the disease and capable to infect susceptibles.

R(**t**): The number of recovered individuals at the time t, who have been recovered from the disease and gained immunity. Thus they can not be infected by others anymore.

Therefore,

$$S(t) + E(t) + I(t) + R(t) = N(t) = N. \tag{2.1}$$

$$s(t) + e(t) + i(t) + r(t) = 1 \tag{2.2}$$

The lowercase letters denote the rate of individuals being in the compartments respectively. The model rests on the assumption that the population is mixing homogeneously. That means all individuals have identical rates of making infectious contacts. And we also assume that the disease has the same process in every individual, i.e. they have equal susceptibility and infectivity. With these assumptions, the spreading process can be considered as a mass action mechanism. The following system of differential equations describes the dynamics of the disease spreading.

$$\begin{cases} \frac{ds(t)}{dt} &= -\beta s(t)i(t) \\ \frac{de(t)}{dt} &= \beta s(t)i(t) - \lambda e(t) \\ \frac{di(t)}{dt} &= \lambda e(t) - \gamma i(t) \\ \frac{dr(t)}{dt} &= \gamma i(t) \end{cases} \tag{2.3}$$

Where the rate of infection is β and the rate of recovery is γ, and λ defines the rate of the exposed transitions to infectious. It is easy to see, that $(1, 0, 0, 0)$ is the trivial disease-free equilibrium point of the system, and it is unstable.

More interesting case, when there is a small number of infectious ϵ in the population, and the rest of it is susceptible. So let us investigate the model with the initial problem, $s(0) = 1 - \epsilon, e(0) = 0, i(0) = \epsilon, r(0) = 0$.

From reordering (2.2) we get $r(t) = s(t) + e(t) + i(t) - 1$. Thus $s(t), e(t)$ and $i(t)$ determine the value of $r(t)$. Therefore, the system (2.3) reduces to

$$\begin{cases} \frac{ds(t)}{dt} &= -\beta s(t) i(t) \\ \frac{de(t)}{dt} &= \beta s(t) i(t) - \lambda e(t) \\ \frac{di(t)}{dt} &= \lambda e(t) - \gamma i(t) \end{cases} \tag{2.4}$$

If we add the last two equation, we get

$$\frac{d(e+i)(t)}{dt} = (\beta s(t) - \gamma) i(t). \tag{2.5}$$

The basic reproductive number is $R_0 = \frac{\beta}{\gamma}$. This means that in average how many people get infected by one infectious in the early stages of the spread when $s(t) \approx 1$ and $i(t) > 0$. If $R_0 < 1$ then the number of exposed and infectious people is decreasing from the beginning and the virus simply vanishes from the population. If $R_0 > 1$, then there will be more infection than recovery and the virus will emerge to an epidemic outbreak.

2.2 The stochastic SEIR model

According to Håkan Andersson and Tom Britton [3], there are several reasons to model a spreading of a disease as a stochastic process. Firstly, the natural way to describe an infection between two individual is stochastic. They either make infectious contact with probability p or not with probability $1-p$. The deterministic models are relying on the law of large numbers. If the population size is small, then a deterministic model will not describe accurately the process. Another investigation is that in a large population a disease either causes a minor outbreak or emerges to an epidemic. The probability of the two outcomes can be computed using the stochastic model.

One of the first complete stochastic models which have gained significant attention was introduced by Reed and Frost. They described the spreading process as a binomial chain. Although, they never published their results, just presented on a series of lectures. The definition of the model that we included below is from Tom Britton [2].

The definition

Let us group the population into four compartments as we did in section 2.1. We assume that in the population is closed, so $S(t) + E(t) + I(t) + R(t) = N$ for all t. The process starts at $t = 0$ from a specified initial vector $(S(0), E(0), I(0), R(0)) = (n_s, n_e, n_i, n_r)$, often $(S(0), E(0), I(0), R(0)) = (N - m, 0, m, 0)$ where $m << N$. Then the epidemic evolves as follows: 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 he or she becomes infected, otherwise, the infectious contact has no effect. Individuals that become infected are first latent (called exposed) for a random duration L with distribution F_L, then they become infectious for a duration I with distribution F_I, after which they become recovered and 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 process ends at the first τ, when there is no exposed or infectious individual $(E(\tau) + I(\tau) = 0)$, therefore no infection can occur furthermore.
Let Z denote the number of individuals who have been infected during the epidemic, also known as the final size of the epidemic. Then $Z = N - S(\tau) = R(\tau) - I(0)$.

Chapter 3

Discrete epidemic model on Hypergraphs

In this chapter we investigate the epidemic spreading on hypergraphs. The definition of our model is inspired by the work of Bodó, Katona and Simon [22], who have defined epidemic spreading on hypergraphs using SIS model. The main differences between the models that we use SIERD model and that we define the way of the infections differently.

Firstly, we present a definition of the process based on the previous chapter. After that, we compare simulations running on the model on hypergraph and on its clique graph. We generate hypergraphs according to the models described in Chapter 1. Finally, we make some investigation what is the effect of the underlying hypergraph structure on the epidemic spreading.

3.1 The definition of the model

In this section, we develop the definition of the epidemic model on hypergraphs. For a given hypergraph $H = (V, E)$ the nodes correspond to individuals and the hyperedges represent the units of the population such as workplaces, households or some kinds of communities and gatherings. In this model unlike in the previous chapter, we take into account the deaths caused by the disease, but we still ignore the natural deaths.

We group the population into S, E, I, R, D compartments we also call these states.

The S, E, I, R compartments cover the same group as in the previous models, and $D(t)$ notes the number of deaths caused by the disease until the time t. Let us note that here we assume that if an individual dies, he or she must be infectious before that. There are no births and natural deaths or migration in the model, thus $S(t)+E(t)+I(t)+R(t)+D(t) = N$ for all t. We start from $t = 0$ with an initial state distribution $(S(0), E(0), I(0), R(0), D(0))$ on the nodes. We usually have a small positive number for the initial infectious individuals and the rest is susceptible. This case, we chose the initial infectious people uniformly random from the population. After we defined the initial states of the nodes, the epidemic model operates as follows: In timestep t each susceptible individual may contract the disease through an incident hyperedge h with probability $w_h f(k_h^t, |h|)$, where k_h^t is the number of the infectious in h, f is an $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ spreading function with an image set in the $[0, 1]$ interval, and w_h is the spreading rate of the hyperedge. The latent period and the infectious period are both governed by Poisson processes with parameters λ and γ, i.e. in every timestep an exposed individual transitions to infectious with probability λ or stays in the exposed state else. Similarly, an individual ends its infectious period with probability γ or stays infectious else. Here λ and γ could be chosen according to the latent period and the infectious period of the disease. After an individual stops being infectious, then she or he either dies with probability p or recovers with probability $1 - p$.

The process ends at the first τ, when there is not any exposed or infectious individual $(E(\tau) + I(\tau) = 0)$, therefore infections can not occur furthermore. We call this τ as end time of the epidemic. Let Z denote the number of individuals who have been infected during the epidemic. We will call this number as the total infected or final size. Then $Z = N - S(\tau) = R(\tau) + D(\tau) - I(0)$.

In this model, we could adjust the spreading function f to the disease that we are aiming to investigate. This level of freedom guarantees that the model is more flexible than models which use graphs as underlying structures.

Here, we assumed that the disease has identical dynamics within every unit belonging to the same type. Thus, for example in every unit of size k and spreading rate w_h if there is exactly one infectious individual between the members, then each of the susceptible member may contract the disease with the same probability from the

unit. What we also assume that in every unit the members create a homogeneously mixing community. Thus, every person has the same contact rate with one another inside the community unit. If there are more than two infectious individuals in a unit then it is not clear who is responsible for an infection inside the unit unlike to the models using graphs as underlying structures. This is the consequence of using hypergraphs for the representation of society. Because if there is more than one infectious in one community, then we can not pick a person who is responsible for the spreading. The cause is more like the infectious people together. This idea makes sense in the case of airborne diseases because the people in a community share their common ground and spend time together in one closed place. We might say they share their common "air" too, so this is more like polyadic than many separated dyadic relations.

3.2 The clique graph model

There are numerous studies on epidemic spreading on graphs. But the models that use a hypergraph as the underlying structure are not so widely used. Our aim is to investigate the differences of disease spreading dynamics on hypergraphs and on their clique graphs for some special spreading functions f.

Now the question is, how do we define the spreading process on the clique graph? While creating the edges, we have to fix the spreading rates of each edge, which determine the probabilities of the disease transfer through the edges. If e edge was created due to hyperedge h (i.e. e is in the clique which represents h), then $\frac{w_e}{|h|-1}$ is a reasonable choice for the probability of the disease transfer. And we stick to this idea throughout the paper.

After this one could say that the clique graph model operates the usual way as we expect from an epidemic model on graph writing $\frac{w_e}{|h|-1}$ weights on the edges. However, if someone unfamiliar with the epidemic spreading on graphs we could define the clique graph model using the hypergraph model's definition. Then we just have to replace words hypergraph with clique graph and hyperedge with edge. The edge spreading rates inherit from the hypergraph model as follows: If e edge was created due to hyperedge h, then $w_e = w_h \frac{1}{|h|-1}$.

3.3 Comparison between the models

We focus on three main topics, where we might find differences between the two model's dynamics.

Early stages: Which is the lowest initial number of infectious for given parameters that could lead to an epidemic outbreak? What is the basic reproductive number? Is there any difference in these between the two models?

Peak height: Which model has a higher peak in the number of infectious?

Final stages: Is there any asymptotic difference between the models? Which leads to a larger total number of infection? Is there any difference in the decrease of the infectious number?

Let A_{vh}^t note the event that a susceptible individual v gets the disease through an incident hyperedge h in timestep t. The probability of A_{vh}^t depends on two factors, the spreading rate of h, w_h and the function of the edge size and the number of infectious individuals at timestep t in h, $f(k_h^t, |h|)$. Let's denote this event with A_{vh}^t, then

$$\mathbb{P}(A_{vh}^t) = w_h f(k_h^t, |h|).$$

Let us consider a similar process in the clique graph model. Let's note the event with B_{vh}^t that in timestep t a susceptible individual v gets the disease from a particular clique which is originated from hyperedge h in its hypergraph. The probability of B_{vh}^t calculated as follows:

$$\mathbb{P}(B_{vh}^t) = 1 - \mathbb{P}(\overline{B_{vh}^t}) = 1 - \left(1 - w_h \frac{1}{|h| - 1}\right)^{k_h^t},$$

where k_h^t is the number of infectious individuals in the clique a timestep t. Thus, the choice of the function f may determine the differences between the hypergraph and its clique graph model.

The basic reproductive number in the hypergraph model also depends on the choice of f, while in the clique graph it is known that

$$R_0 \sim \frac{\overline{dw}}{\gamma},$$

where $\overline{d}$ is the average degree and $\overline{w}$ is the average spreading rate of the hyperedges. We picked some reasonable choices for the spreading function f. First, we use the $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$ and then we use $f(k_h^t, |h|) = \frac{\sqrt{k_h^t}}{|h|-1}$ for our investigations. Before presenting numerical results and analysis, let us take a brief overlook of our implementation.

3.3.1 Implementation

Our Python implementation of the model described in Section 3.1. operates the following way:

1. It creates the nodes of the hypergraph and sets the initial viral states of the individuals according to input rates or numbers.

2. It generates the hyperedges of the hypergraph using arbitrary combinations of the models described in Chapter 1. We can declare the preferred models and their parameters as inputs to the algorithm.

3. Then it runs a discrete time simulation of the SEIR model described in Section 3.1.

The codes and the data used and generated can be found here [23]. We have to note, that if we present numerical simulation results throughout the document what concludes computing averages, then we calculate the average using 80% of the results cutting down the upper and lower 10% of them.

3.3.2 Epidemic in a closed community

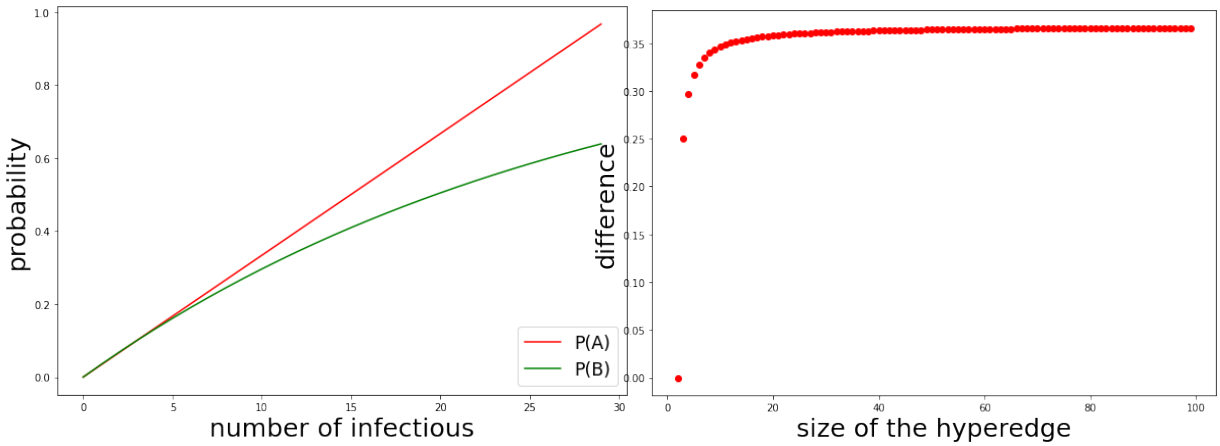
Let us consider the case when there is only one community in the population. Equally say, the hypergraph in the model consists of only one hyperedge. These examinations may bring us closer to understanding an epidemic spreading on more complicated hypergraphs, which we present in the next section. Note that if the hypergraph stands from one hyperedge then its clique graph is a complete graph with the size of the hyperedge.

Firstly, let us choose the spreading function f to be the function $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$. Note that now it is easy to see that

$$R_0 \sim \frac{\overline{dw_h}}{\gamma},$$

for both models. So we might not experience large difference in the early stage of the spreading.

Let us suppose that there is a timestep t, when both in the hypergraph model and the clique graph model the number of infectious is equal to k_h^t. Recall the same notations for events A_{vh}^t and B_{vh}^t. A_{vh}^t is the event that a susceptible individual v gets the disease through an incident hyperedge h in timestep t in the hypergraph model. The event B_{vh}^t is a similar event in the clique graph model, so it notes that a susceptible individual v gets the disease from a particular clique which is originated from hyperedge h in its hypergraph in timestep t. If we increase the number of infectious then the difference between $\mathbb{P}(A_{vh}^t)$ and $\mathbb{P}(B_{vh}^t)$ also grows with it (see Figure 3.1a). We can also see that the larger size of the hyperedge leads us to larger difference, but the correlation is not linear.



(a) $\mathbb{P}(A_{vh}^t), \mathbb{P}(B_{vh}^t)$ in the function of k_h^t, where $|h| = 30$

(b) Maximum difference of $\mathbb{P}(A_{vh}^t)$ and $\mathbb{P}(B_{vh}^t)$) in the function of $|h|$

Figure 3.1: The differences of getting the infection from a community in the hypergraph and its clique graph model, if there is k_h^t infectious.

Precisely, if the size of hyperedge is $|h|$ and k_h^t is the number of infectious in the community at timestep t, then

$$\mathbb{P}(A_{vh}^t) - \mathbb{P}(B_{vh}^t) = w_h \frac{k_h^t}{|h|-1} - \left(1 - \left(1 - \frac{w_h}{|h|-1}\right)^{k_h^t}\right).$$

Which has its maximum value when $k_h^t = |h| - 1$ (because here the $k_h^t = |h|$ does not make sense in the aspect of the disease transfer probability). So, the maximum difference between $\mathbb{P}(A_{vh}^t)$ and $\mathbb{P}(B_{vh}^t)$ is

$$\begin{aligned}\mathbb{P}(A_{vh}^t) - \mathbb{P}(B_{vh}^t) &= w_h \frac{|h|-1}{|h|-1} - \left(1 - \left(1 - \frac{w_h}{|h|-1}\right)^{|h|-1}\right) \\ &= w_h - 1 + \left(1 - \frac{w_h}{|h|-1}\right)^{|h|-1}.\end{aligned}$$

By taking $|h|$ to infinity,

$$\lim_{|h|\to\infty} \left(\mathbb{P}(A_{vh}^t) - \mathbb{P}(B_{vh}^t)\right) = w_h - 1 + e^{-w_h}. \tag{3.1}$$

Which is 0 when $w_h = 0$ and monotone increasing in $[0, 1]$, for $w_h = 1$ is equal to e^{-1}. Let us choose a relatively large community size $|h|$ and let us fix $w_h = 1$, so we could expect some differences between the dynamics of the hypergraph and the clique graph model. We have set $|h| = 30$ and run 1000 simulations of the hypergraph and after its clique graph model. We fixed the latent period's parameter $\lambda = 0.2$, the infectious period's parameter $\gamma = 0.1$, death rate $p = 0.1$ and spreading rate $w_h = 1$ for the one hyperedge. We started the simulations with being one infectious and the rest susceptible in the community. A simulation ends at the first timestep τ when there was no exposed or infectious individual. We call this τ as end time.

One can see in Figure 3.2 that there are only little deviations between the average values. We can say that the red curve has a larger maximum value than the blue one. That is the result of the difference between the infection transmission in the two models that we mentioned before (see 3.1).

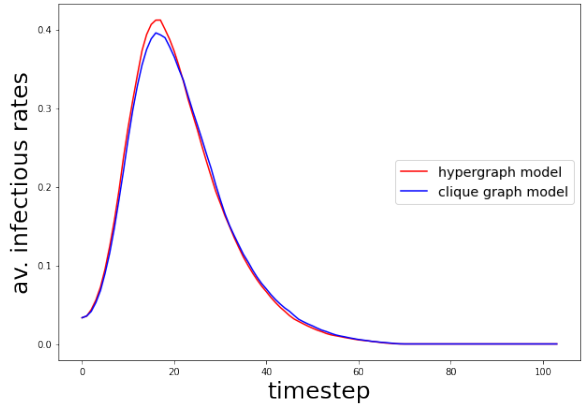


Figure 3.2: Average infectious rates in the function of time running 1000 epidemic simulations in one community with a size of $|h| = 30$. Parameters: $\lambda = 0.2, \gamma = 0.1, p = 0.1, w_h = 1$ and $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$. The blue and the red curves show the average rate of the infectious individuals in the community at timestep t.

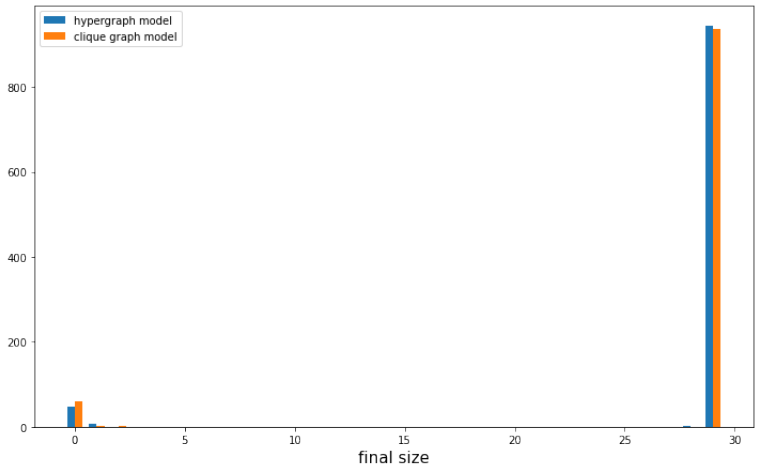


Figure 3.3: The final size distributions of the hypergraph model the clique graph model running 10000 epidemic simulations in one community with a size of $|h| = 30$. Parameters: $\lambda = 0.2, \gamma = 0.1, p = 0.1, w_h = 1$ and $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$.

In Figure 3.3 one can see the distribution of the final size of the 1000 simulations. In most cases, when the first infectious could pass the disease to another member in the community the disease spreading could emerge into an epidemic outbreak infecting almost everyone in the community. But in some cases, the disease could only infect one or two individuals in the community and vanished from the population leaving most of the individuals susceptible.

In Figure 3.4 there is a column diagram from the end time distributions. One may say there are not many differences in this aspect of the models.

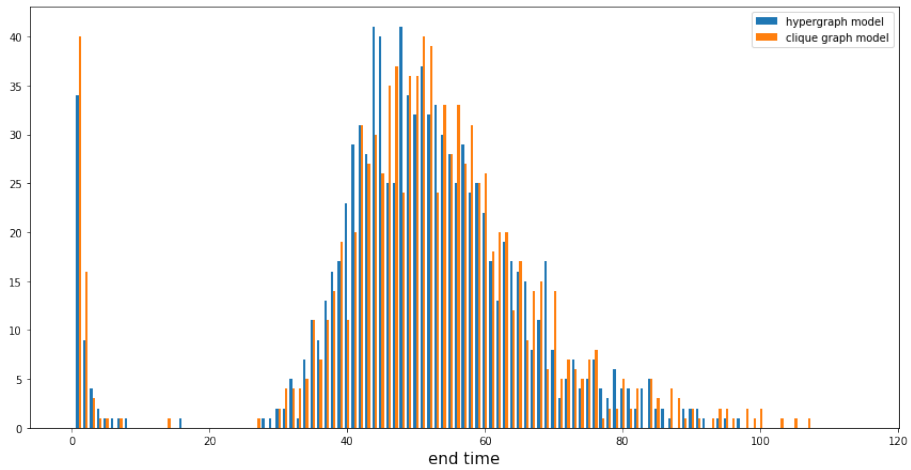


Figure 3.4: The end time distributions of the hypergraph model and the clique graph model running 1000 epidemic simulations in one community with a size of $|h| = 30$. Parameters: $\lambda = 0.2, \gamma = 0.1, p = 0.1, w_h = 1$ and f.

Now, let us choose the $f(k_h^t, |h|) = \frac{\sqrt{k_h^t}}{|h|-1}$ function for f. We ran 1000 simulations with the same parameters as before with the previous function. One can see in Figure 3.5 that using this spreading function the epidemic spreading slows down as it is expected. The average rate of infectious peaks later and smaller than when we used the $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$ function or in the clique graph model.

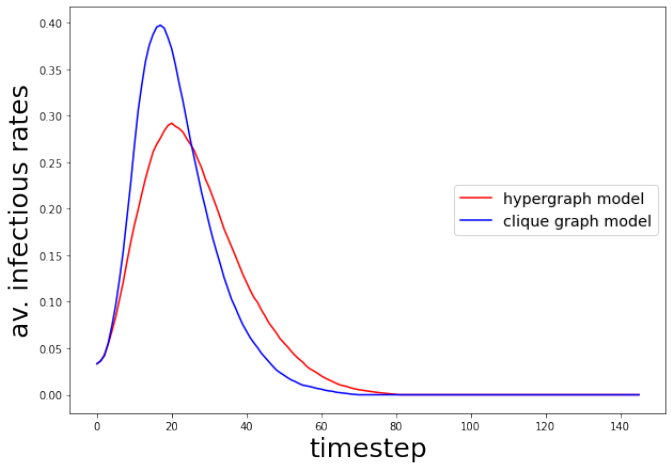
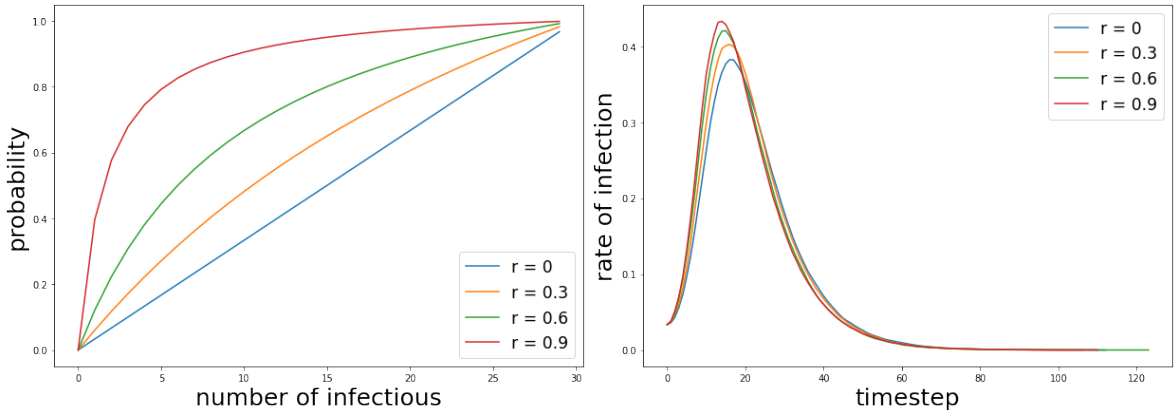


Figure 3.5: Infectious rates in a closed community of size 30 using $\frac{\sqrt{k_h^t}}{|h|-1}$ for spreading function f.

This indicates that the choice of the function f has a significant effect on the dynamics of the spreading. If a spreading function f is above another spreading function g for all possible k_h^t for a fixed $|h|$ then we expect that using f we gain faster epidemic dynamics and a higher peak in the infectious rate. To demonstrate this we used spreading functions $f_r(k_h^t, |h|) = \frac{(1+r)^{\frac{k_h^t}{|h|}}}{2r^{\frac{k_h^t}{|h|}-r+1}}$, for $r = 0, 0.3, 0.6, 0.9$ (see in Figure 3.6).



(a) The spreading functions f_r dependence on k_h^t if $|h| = 30$, and we use different parameters for r.

(b) Average infectious rates using four different parameter for r in spreading function f_r.

Figure 3.6: The dependence of the epidemic dynamics on the spreading function in one community of size 30. In the left Figure, one can see the dependence of the spreading function on k_h^t. On the right side, we present the average infectious rates belonging to them in the function of time. Here we ran 1000 simulations for every value to be tested to r. Parameters: $\lambda = 0.2, \gamma = 0.1, p = 0.1, w_h = 1$ during all simulations.

From now in the document, we will use our first choice the function $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$ for spreading function.

3.3.3 Epidemic on d-regular hypergraph

In this subsection, we investigate epidemic processes on random d-regular hypergraphs, which we have described in Subsection 1.2.2. We make comparisons between running epidemic simulations on d-regular hypergraphs and on their clique graphs. The special type of d-regular hypergraphs what we have described in

Subsection 1.2.2 can be seen as d partitions of the nodes, i. e. of the population in our case. Thus, everyone in the population has d communities which he or she is a part of.

Let us investigate how the epidemic dynamics depend on the size of the hyperedges on a random k-uniform 3-regular hypergraph. Recall that we have fixed the function f to be $f(k_h^t, |h|) = \frac{k_h^t}{|h|-1}$. We generated 50, 3-regular hypergraphs on 10000 nodes from each 3-uniform, 5-uniform, 8-uniform and 10-uniform hypergraphs. We ran an epidemic simulation process on each hypergraph and on their clique graph model. We started every simulation with one infectious individual and the rest susceptible. In all simulations we chose the same values $\lambda = 0.2, \gamma = 0.1, p = 0.1$ and spreading rate $w_h = 0.1$ for all hyperedge in the hypergraph. We ended the simulations in timestep 200.

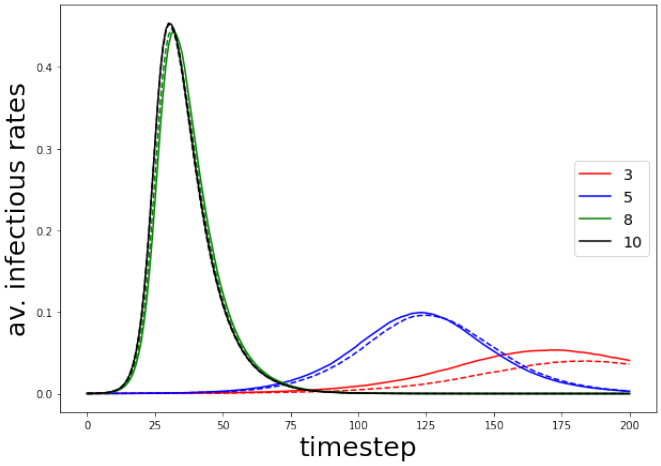


Figure 3.7: The average rate of the infectious in the function of time running 50 epidemic simulations on random k-uniform 3-regular hypergraphs (continuous lines) and on their clique graph (dashed lines).

If we look at the shared plot in Figure 3.7, as one could expect that the maximum of the average rate of infectious individuals is increasing as we increase k. One can also observe that the difference between the runs on the hypergraph and its clique graph model are more visible with smaller edge sizes.

3.4 Epidemic spreading on different hypergraph structures

In this subsection, our aim is to investigate what are the effects of the underlying hypergraph structures on the dynamics of the epidemic spreading. We make a comparison between epidemics on Erdős-Rényi(ER), Barabási-Albert(BA) hypergraph and d-regular hypergraphs. We have described how to generate these structures in Chapter 1.

First of all, let us investigate how the epidemic dynamics change as we increase the degree of a regular hypergraph. We ran 50 simulations on 4-uniform d-regular hypergraphs for $d = 2, 4, 6, 8$. We generated new hypergraphs with size 1000 for every simulation. One can see the way of the generations in Subsection 1.2.2. We started all simulations from the initial vector $(999, 0, 1, 0, 0)$, so with one individual being infected and the rest being susceptible. We used the same parameters for $\lambda = 0.2, \gamma = 0.1, p = 0.1$ and spreading rates $w_h = 0.2$ in every run.

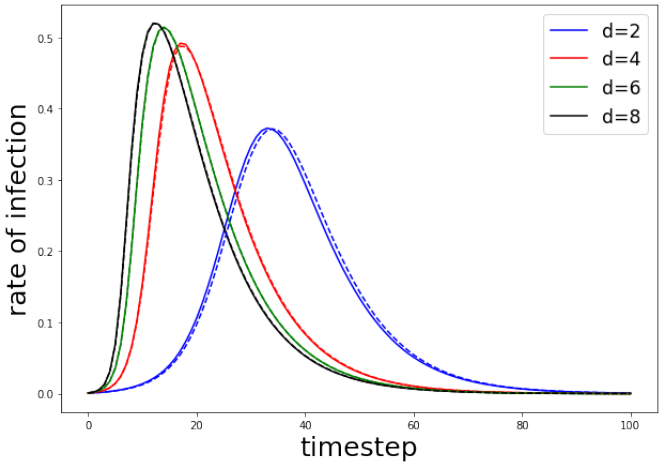


Figure 3.8: The average rates of the infectious in the function of time running 50 epidemic simulations on 5-uniform d-regular hypergraphs with size 1000 (continuous lines) and on their clique graph (dashed lines).

As one could have expected the raise of the degree leads us to faster spreading dynamics. The maximum of the average infectious rate gets larger as the degree grows (see in Figure 3.8). This phenomenon seems to be a trivial consequence of the rise of the number of edges in the hypergraph. In Figure 3.9 one can see the

maximum rate of the average infection in the function of d. So from this and from the simulations in Subsection 3.3.3, we may conclude that the number of the hyperedges has an impact on the spreading dynamics of the disease.

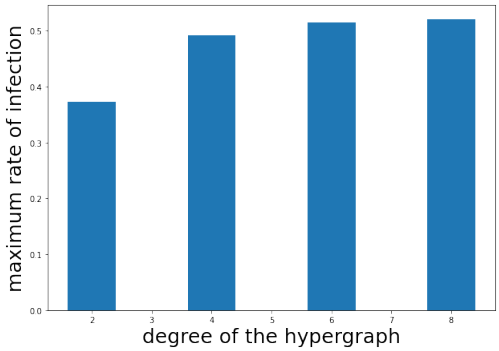
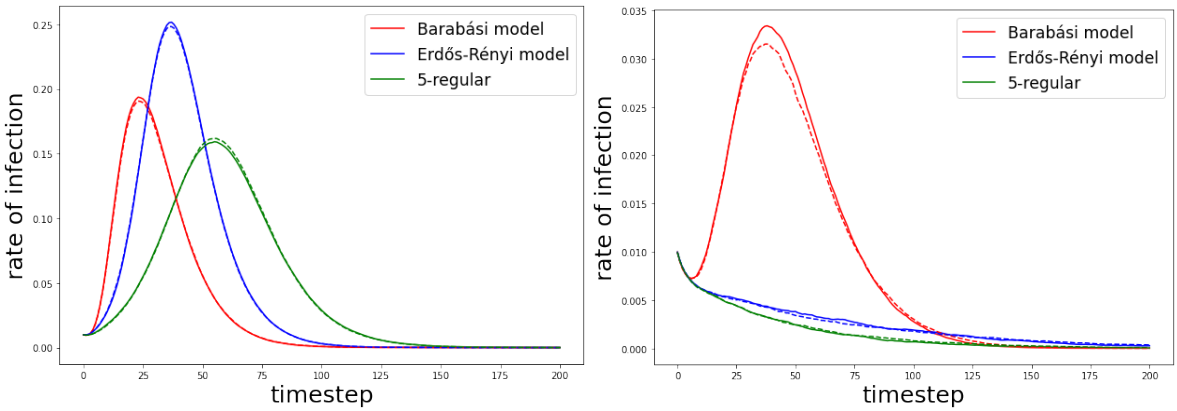


Figure 3.9: The maximums of the average rates of the infections in the function of d the degree of the regular hypergraph

What is the difference if we use different methods for generating the hypergraph but restrict the number of edges to be equal in every model? What kind of graph properties could predict faster or slower epidemic spreading dynamics? In the following subsection, we investigate epidemic simulations when we fix the number of the hyperedges and their sizes to be constant, and the only difference between the underlying hypergraphs is the model how we choose the incident nodes to create the hyperedges.

3.4.1 Hyperedge spreading rate

We investigated how the epidemic dynamics depend on the edge spreading rate w in different hypergraph structures. By different hypergraph structures, we mean the 4-uniform Barabási-Albert(BA), the Erdős-Rényi(ER) and random 4-regular hypergraphs. For one hypergraph we fixed the same spreading rate for all hyperedge. If we take a look at Figure 3.10a then we can see that the BA model leads to fast dynamics, but the size of the epidemic is smaller than in the case of ER hypergraphs as the underlying structure. However, we can see in Figure 3.10b that the Barbási model leads to a small epidemic, but when we use ER hypergraphs or 4-regular hypergraphs for underlying structure and lower spreading rate then we get no large epidemic outbreaks.



(a) $w_h = 0.1$ for every hyperedge in all hypergraphs.

(b) $w_h = 0.02$ for every hyperedge in all hypergraphs.

Figure 3.10: Average infectious rates using three different hypergraph structures. We set different weights for the hyperedges in the left and the right simulations. We gained both plots from performing epidemic processes on 500 different 4-uniform hypergraphs for each type of hypergraph. Every hypergraph had the 1000 hyperedges. We started each simulation with 10 infectious and the rest susceptible in the population. We used the same parameters as we did before in the previous cases except for w_h. For the simulations which aggregate values can be seen on the left, we used $w_h = 0.1$ and for the right side, we used $w_h = 0.02$. It seems that with different underlying structures the change in the spreading rate of the hyperedges has different impacts in the dynamics of the epidemic.

Now let us investigate what happens with the models if we change the hyperedge spreading rates ranging from 0 to 0.2. As it can be seen in Figure 3.11, we chose the hyperedge spreading rates from the interval $[0, 0.2]$ with step size 0.005, so we ran simulations for 40 different spreading rates. We generated 30 hypergraphs on 10000 nodes for each hypergraph model with 10000 hyperedges for each value. All hyperedges had size 4. Then we ran an epidemic simulation on the generated hypergraphs started from being 10 infectious individuals in the population and the rest susceptible. We fixed the parameters $\lambda = 0.2$, $\gamma = 0.1$ and $p = 0.1$. Thus, for example, we ran 30 simulations on BA hypergraphs with 10000 nodes and 10000 hyperedges using the hyperedge spreading rate $w = 0.005$ and the parameters above. We used four numbers to characterise an epidemic simulation, the maximum of the

infectious rate during the epidemic, the peak time of the infectious rate the timestep when the maximum occurs, the total infected and the end time of the epidemic. One can see the dependence of these numbers on the hyperedge spreading rate in Figure 3.11.

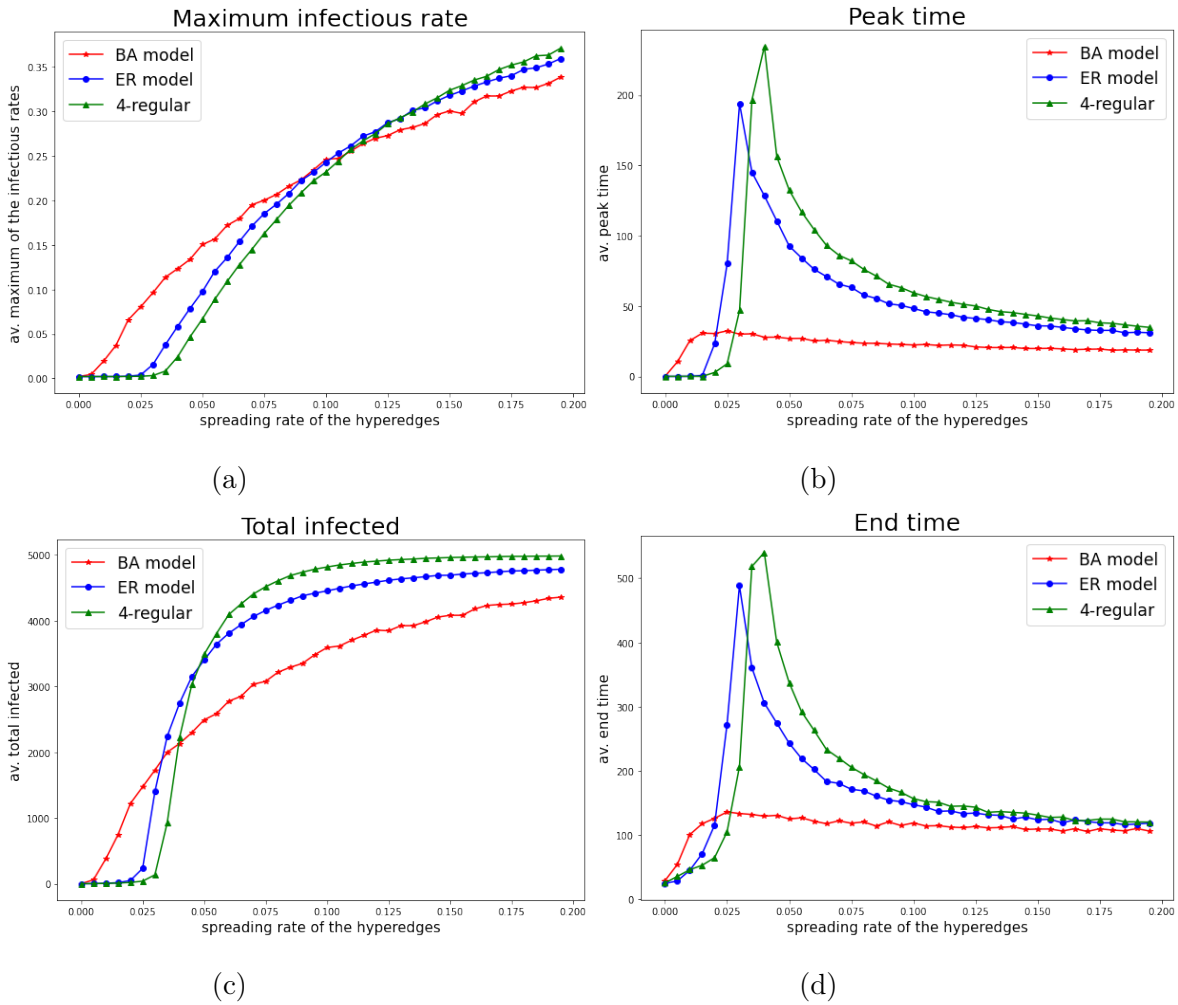


Figure 3.11: The maximum of the infectious rate, the final size and the end time of the epidemic on 4-uniform BA, ER and random 4-regular hypergraphs in the function of the spreading rate of the hyperedges. A point represents an average value from 30 simulations on different hypergraphs from one model. For example, the red curve derived from the 30 simulations on BA hypergraphs for each investigated hyperedge spreading rate.

From the simulations, we could say that the three models depend differently on the spreading rate of the hyperedges. The epidemics on the BA model deviates the most from the other two. If we look at Figure 3.11, then we may recognise that on Plot

3.10a and Plot 3.10b curves of the average epidemic numbers related to the BA model for low hyperedge spreading rates are above the ones which are related to the ER and the 4-regular hypergraphs. One might ask what is the explanation behind that. As in the case of graphs, Barabási says in his book [24] in Chapter 10, that the hubs are responsible for the deviation. By hubs, he means nodes with relatively high degrees. These nodes could function as "centres" of a graph. On scale-free graphs, such as we gain from the BA model, an epidemic spreading can emerge to an outbreak even if the transmission probability between the nodes is relatively low. That is because, if one of the hubs gets infected, then the node could infect a large number of nodes from its neighbours and become a super spreader of the disease. In the BA hypergraph model because these hubs are well presented (see its degree distribution in Figure 1.3), the same reasoning may standstill. The ER and d-regular hypergraphs are lacking of hubs so we can not see the same process on them.

Chapter 4

Modelling COVID-19 epidemic spreading on hypergraph

In December of 2019, there were some reports that a novel virus, later known as SARS-CoV-2, started rapidly spreading amongst the people of Wuhan. The most alarming property of this airborne virus was its high reproductive number [25]. Since then, all of the world effected by the virus, it became a pandemic and the human world's most important issue to battle with. Because of the extremely rapid spread of the virus, a huge number of people became ill and needed hospitalization. The wards got overcrowded and the hospitals themselves became the hubs of the spreading. It is one of the most researched questions how can we tackle the virus. Precautions like quarantines, wearing masks, and social distancing was the primary solution when there were not any vaccines available. Now when the vaccination is reachable on large scale, mass vaccination campaigns have been launched in most of the countries all over the world.

In this chapter in Section 4.1, we aim to build a stochastic model for the spread of SARS-CoV-2. Here, we stratify the population into five age groups. For now, it is clear, that the virus is more dangerous to older people [26], so we may consider this into our model. If they contract the virus they have a higher chance to develop symptoms and have more serious effects like heavy illness, need of medical support or even death.

Another aspect of the age groups the individual's number of connections in them. We will assume that the number of communities that a people part of depends on

his or her age. Young adults may part of more and old people may part of fewer communities in the society.

Another important characteristic of SARS-CoV-2, that most people contracts the virus without feeling any sign of the disease. So, without knowing it they carry and spread the virus. In light of these, we split the infectious state into symptomatic and asymptomatic infectious states besides the basic known states from Chapter 3.

After that in Section 4.2, we investigate the effectiveness of precautions, quarantines and testing. Then in Section 4.3, we make observations of different types of vaccination campaigns. Our goal to find the best order for a vaccination campaign if several types of precautions are active in the society.

4.1 The model

This model will be similar to what we used in Chapter 3, but we make further refinements with the aim of a more realistic model. We see the population as a hypergraph. The nodes are the individuals and the communities and gatherings are the hyperedges.

4.1.1 Underlying hypergraph structure

The airborne SARS-CoV-2 spreads through face to face social interactions. With the underlying hypergraph structure, we try to make a good model of them. The hypergraph $H = (V, E)$ what we use from now has $n = |V|$ nodes. There is an 'age' function on the nodes $g : V \longrightarrow [1, \ldots, l]$ which categorises the individuals to age groups. For the sake of simplicity, we will note $g(v)$ with g_v for a $v \in V$. We usually stratify the population into 5 age groups, so if we do not say otherwise l will be 5.

We generate three types of hyperedges which consist E. We call them households E_{hh} , workplaces E_{wp} and events E_{ev}. We require from E_{hh} and E_{wp} hyperedges to be two partiotions on the nodes. Thus, the subhypergraph $(V, E_h \cup E_w)$ creates a 2-regular hypergraph. So we genrate these hyperedges as we have described in Subsection 1.2.2. We sample the size of the edges in E_{hh} and E_{wp} from distributions $Poisson(2.57)$ and $Bin(30, 0.3)$ respectively. In Figure 4.1 one can see the probability mass functions of these distributions.

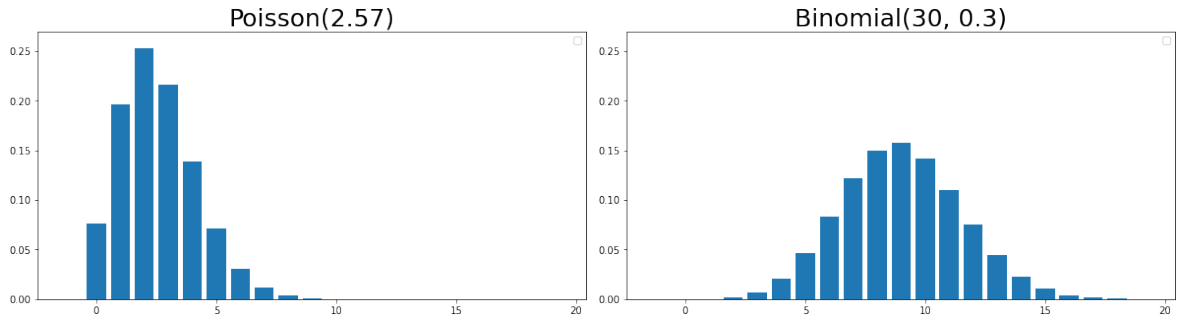


Figure 4.1: The probability mass function of $Poisson(2.57)$ for the hyperedge sizes in E_{hh} on the left and $Binomial(30, 0.3)$ for the hyperedge sizes in E_{wp} on the right side.

We generate the event hyperedges E_{ev} following the Bianconi-Barabási model (see Subsection 1.3.3). We assign the same fitness for every node in the same age group. This way we can adjust the number of social connections per individual for an average person in the groups. We assumed in every model from now that mid-age people in the age group 2 has the most connections from all of the age groups. Every node v with age g_v has fitness η_{g_v} in the model. The proportions of the age groups presented in Table 4.1 is based on real data from [27]. The fitness numbers belonging to the groups will be $\eta_1 = 0.3, \eta_2 = 0.4, \eta_3 = 0.2, \eta_4 = 0.15, \eta_5 = 0.05$.

age group	age	% of the pop	η_i
1	0-24	25.2	0.3
2	25-49	35.8	0.4
3	50-64	19.6	0.2
4	65-79	14.9	0.15
5	80+	4.4	0.05

Table 4.1: The proportions of the age groups and their fitness parameters.

In figure 4.2, one can see a hypergraph's degree distribution generated according to the model above on 10000 nodes. We can freely decide the distribution what we use during the generation of event hyperedge sizes. Here we use $Pareto(2.3)$, which is a fat tailed distribution, so it allows hyperedges with relatively large sizes. These may represent concerts, festivals or some events with many participants. If we do

not say otherwise we will use these parameters and distributions for the creation of the underlying hypergraph structure of the epidemic model. These event hyperedges

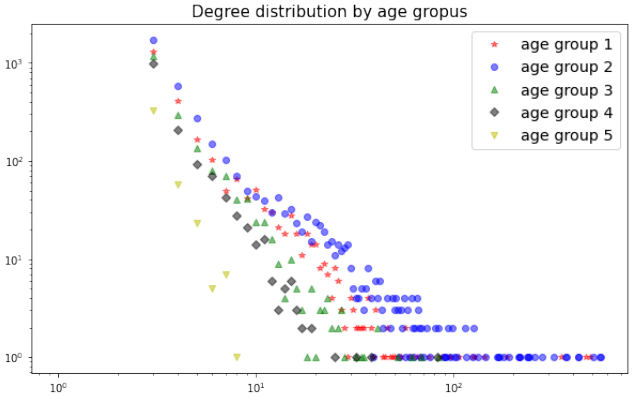


Figure 4.2: The degree distribution in the different age groups

are not always active during the spreading process as we do not meet our friends every single day or go to a concert. We define precisely what we mean by hyperedge activation in the next Subsection.

4.1.2 The discrete model with asymptomatic infectious individuals

In this model an individual can be classified into six different viral state, susceptible (S), exposed (E), asymptomatic infectious (A), symptomatic infectious (I), recovered (R) and dead (D). The S, E, I, R, D have the same definitions in the models before. We note the number of asymptomatic infectious at timestep t with $A(t)$. We call the union of the symptomatic and asymptomatic infectious as infectious. We start our simulation from an initial state distribution $(S(0), E(0), I(0), A(0), R(0), D(0))$ on the nodes. We usually have a small positive number for the initial infectious individuals and the rest is susceptible. In this case, we chose the initial infectious people uniformly random from the population. After we defined the initial states of the nodes, then the epidemic model operates as follows: In timestep t we decide for every hyperedge if it is active or not. Every household hyperedge and workplace hyperedge is active for all t. For each event hyperedge, we flip an unfair coin. It will be active with probability ζ and non-active with $1 - \zeta$ in timestep t. After that, each susceptible individual may contract the

disease from every active incident hyperedge h with probability $w_h \frac{k_h^t}{|h|-1}$, where k_h^t is the sum of infectious in h. We fix the spreading rate w_h differently for the three different types of edges. After contracting the disease the individual becomes exposed, so not immediately infectious, and stays exposed at least for fixed exposed timesteps τ^E. In every timestep t, an exposed individual in age group $g_v = i$, who had been exposed for at least τ^E, transitions to symptomatic infectious with probability λ_i^I or asymptomatic infectious with probability λ_i^A or else stays in the exposed state. An individual ends its symptomatic period with probability γ^I, or else stays symptomatic or asymptomatic infectious states. After an individual in age group $g_v = i$ stops being symptomatic infectious, then she or he either dies with probability p_i or recovers with probability $1 - p_i$. An asymptomatic individual recovers with probability γ_i^A or gets symptomatic infectious with β_i or keeps its viral state else. We can also define fix period times τ^A and τ^I for being in the infectious states respectively as we did in the case of exposed.

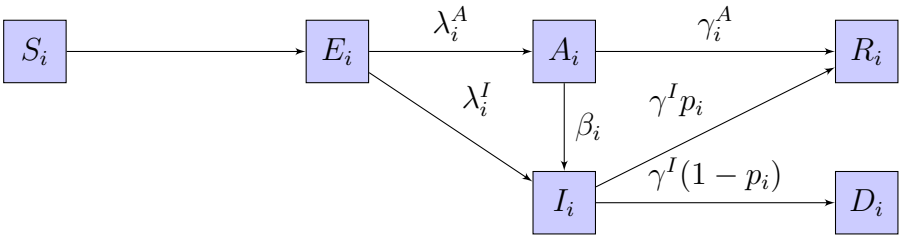


Figure 4.3: Compartmental model for the nodes $g_v = i$

The process ends at the first τ, when there is not any exposed or infectious individual $(E(\tau) + I(\tau) + A(\tau) = 0)$, therefore infections can not occur furthermore. We call this τ as the end time of the epidemic. Let Z denote the number of individuals who have been infected during the epidemic. We will call this number the total infected or final size. Then $Z = N - S(\tau) = R(\tau) + D(\tau) - I(0)$.

Henceforth, we investigate this model for given parameter settings, while different precautions and measurements are active in the population. If we don't fix a parameter from the model above for a simulation, then we use our basic parameter. We included these basic parameters in Section 4.4 in Table 4.3. We describe the reason why we chose these parameter settings in that section.

Social distancing and wearing masks

One of the essential ways to protect us against the virus is wearing masks in public areas and during work and keep a safe distance from others. Here we investigate how effective is that according to our model. In Figure 4.4 one can see two epidemic simulation on the same hypergraph model with a population size 10000. We stratified the population into 5 age group as we described in Subsection 4.1.1. The event hyperedges were generated according to the Bianconi-Barabási hypergraph model (see Subsection 1.3.3) with its hyperedge size distribution sampled from distribution $Pareto(2.3)$. The fitness of the nodes was distributed according to their age, so the mid-age people had a higher chance to have more connections than the others (see more detailed in Subsection 4.1.1). We set parameters of the spreading process according to our basic settings described in Section 4.4. We started the simulations by being 5 infectious in the population and the rest susceptible.

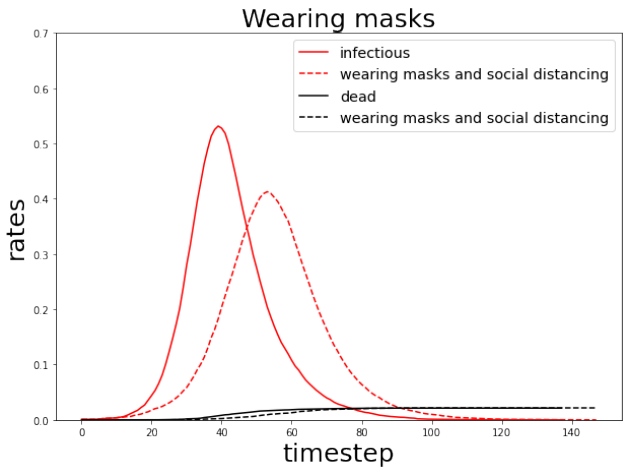


Figure 4.4: The effects of wearing masks and social distancing.

In Figure 4.4 we illustrated with the continuous lines how the infectious and death rates behave if people don't take any precautions. We set the spreading rates of the workplace and event hyperedges rather high to 0.5, this means if a susceptible person meets with 3 of his friends then he gets the virus with probability 1/6 if exactly one of them is infectious. We fixed the event hyperedge appearance probability ζ to 0.05. With the dashed lines we presented the epidemic dynamics on the same hypergraph with different hyperedge spreading rate and appearance probability. Let us say by the right hygienic precautions, social distancing and wearing masks, we can half the

probability of getting the disease from our workplaces and events, so we set w_{wp} and w_{ev} to 0.25. At the same time by making responsible decisions the population avoids the events by half-chance than in the model presented on the right side, so we set $\zeta = 0.025$.

From these two runs, we may say that reducing the spreading probability is an effective method for flatten the infectious curve. Which is crucial from the point of the healthcare capacity. Now let us estimate the rate of symptomatic infectious individuals who need hospitalisation. From the Our World in Data database [28] gained that on average 15.92% of the confirmed cases were needed hospitalisation in the time period from 14.02.2021 till 16.05.2021 in Hungary. In Section 4.4 we used this period for fitting our parameters. Thus, we may at this first time calculate the hospitalised people from the 15.92% of the symptomatic infectious. After that, we can draw the rates of the hospitalised individuals in the population for both epidemic spreads, see in Figure 4.5.

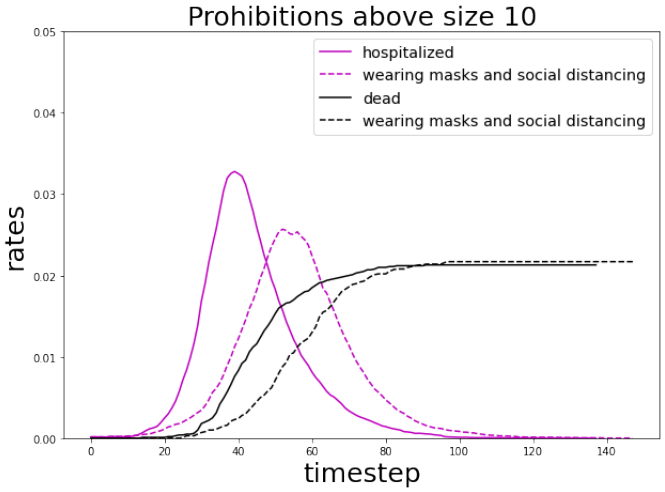


Figure 4.5: The effects of wearing masks and social distancing on the healthcare system. Here we present the same runs as in Figure 4.4 but different rates. The simple lines come from the case when there is no any precaution against the virus and the dashed ones when people wear masks and keep a safe distance.

Surprisingly, the number of deaths did not decrease after lowing the spreading rates and appearance probabilities. This could happen because the virus reached almost everyone in both cases so it ran through the population despite the active precautions. The most welcomed effect is the delay in the peak of infectious and the

lower maximum number of hospitalised people in one timestep. These two outcomes crucial to healthcare efficiency, because the later comes the peak the more time they have for preparations and if there are fewer people in the hospitals they have a better chance for recovery because of the supplies.

Restrictions for the number of people in one place

What if there are not any precautions like wearing masks or distancing, just restrictions for the size of the number of people in one place? In Figure 4.6, we can see the effects of prohibition workplaces and events larger than 10. We generated a hypergraph model with a population size 10000 and simulated epidemic spreading on the same hypergraph twice. First, we allowed hyperedges with arbitrary sizes, but in the second run, we removed the hyperedges with size above 10. The size of the workplace hyperedges sampled from $Bin(0.3, 30)$ and the event hyperedges from $Pareto(2.3)$. The $Pareto(2.3)$ distribution is a fat-tailed distribution so it allows the creation of hyperedges with relatively large sizes.

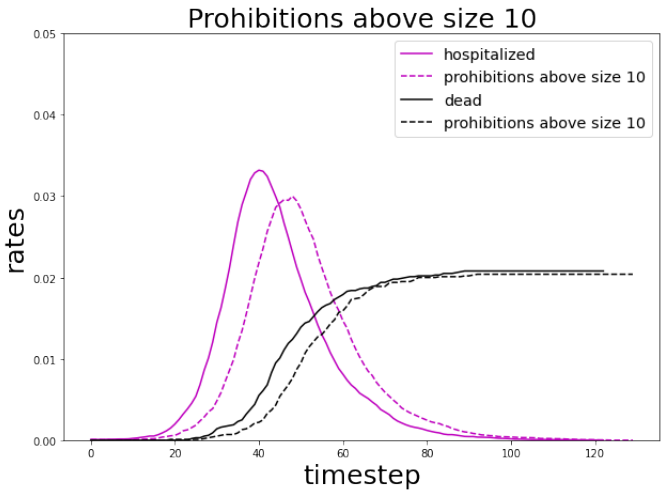


Figure 4.6: The effects of prohibition workplaces and events with sizes above 10: We can see how the epidemic evolves if there are not any restrictions on the size of the events and workplaces. This time we generated the event hyperedges with the BB model using $Pareto(2.3)$. So there were some event hyperedges with large size.
The curves with the dashed lines represent an epidemic run on the same hypergraph, but this time we removed all workplace and event hyperedges larger than 10. This caused a minor reduction in the peak height of the symptomatic infectious.

We could conclude the similar outcomes of the prohibitions as we did in the case of wearing masks and social distances. From both precautions it seems that if we do not take into account the effects of the number of hospitalised people in one timestep, i.e. we assume that the healthcare has a decent capacity to handle the virus, then we can not observe improvements in the death rates. We may need other precautions or other actions to lower the deaths.

4.2 Tests and quarantine

In the early stage of the virus before vaccines have become available the best-known way to fight against the virus was tests, quarantines and contact tracing. Most of the tests were able to detect the virus when its latent period has ended already. So we follow the assumption that an exposed individual can not be tested as positive. From country to country the procedures after a positive coronavirus test can still differ. In our model, if a person performs a positive test (so she or he is either asymptomatic or symptomatic infectious), then we isolate her or him and her or his household from their workplaces and events for q timestep. This means that we remove them from their workplace and event hyperedges, and after q timestep, we place them back. We will set this q equal to 14.

We compare two testing method and the basic case when there are not any testing procedures in progress. The first method if we test randomly, so we test every individual with the same probability p_{test} not paying attention to their viral state. We call the second method targeted testing. We assume in this scenario that the individuals are well aware of the virus. Thus, if a person feels any of the symptoms, he or she goes for a test with a higher probability. So, we test the symptomatic infectious individuals with a higher probability than the others.

We generated 20 hypergraphs according to the model described in Subsection 4.1.1 on 10000 nodes. We set the hyperedge spreading rates $w_{hh} = 0.9$ and $w_{wp} = 0.5, w_{ev} = 0.5$ and $\zeta = 0.05$ as we did when we assumed that there are not any precautions like wearing masks or social distancing. We ran three simulations on each hypergraph one without tests, one with random testing scenario where $p_{test} = 0.01$, and one with targeted testing scenario using the probabilities

$p^S = 0.005, p^E = 0.005, p^I = 0.3, p^A = 0.005, p^R = 0.005, p^D = 0.005$ for the tests. Here p^S means the probability that we test a susceptible individual, we noted the other probabilities analogously. As for the other parameters of the model we used the parameters presented in Section 4.4 in Table 4.3.

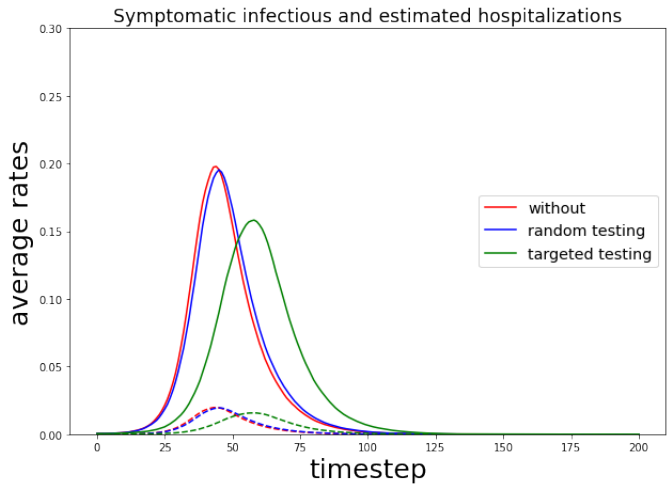


Figure 4.7: The average rates of the symptomatic infectious in the function of time running 20 epidemic simulations, investigating the impact of testing and quarantines.

If we look at Figure 4.7, we see that using targeted testing and quarantines could flat the average infectious curve. What does this mean in the number of death or hospitalised individuals? In Figure 4.8 one can see the estimated hospitalised numbers that we gain by taken 10% of the symptomatic infectious. If we use targeted tests and quarantines without any other actions like wearing masks or distancing then we can not see a reduction in the final death rate of the epidemic. However, if we combine the two methods the targeted tests and wearing masks, then we see that the final death rate has decreased.

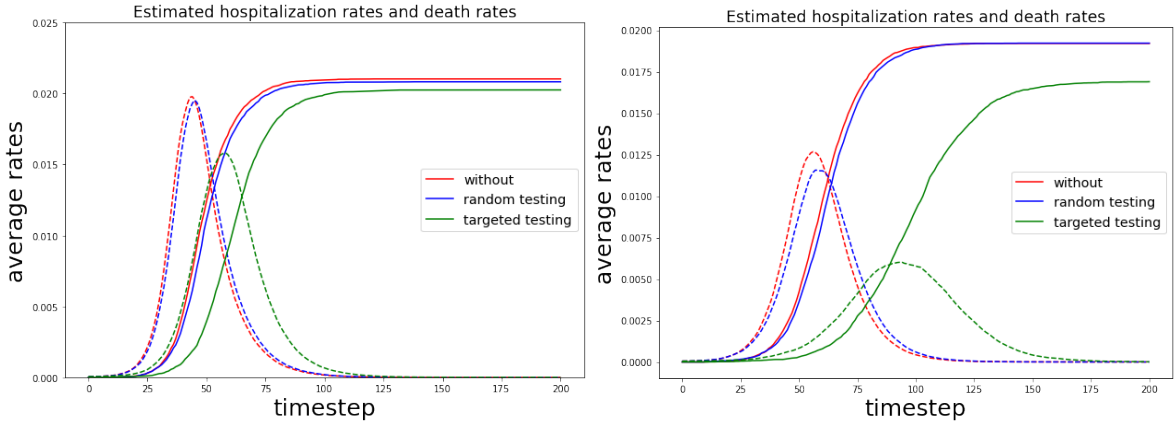


Figure 4.8: On both sides, we see the average death rates and the average rates of the estimated hospitalised individuals using different testing methods. On the left the testing methods launched without wearing masks in the population, so we set $w_{hh} = 0.9, w_{wp} = 0.5, w_{ev} = 0.5$ and $\zeta = 0.05$. On the right we assumed wearing masks and distancing population, so $w_{hh} = 0.9, w_{wp} = 0.25, w_{ev} = 0.25$ and $\zeta = 0.025$.

4.3 Vaccination methods

In this section, we assume that it is possible to vaccinate individuals. Our goal here is to find the best method for the vaccination process. First, we need to write about what do we mean by "vaccination" and what is the "best". By vaccination we mean, if we vaccinate a susceptible individual at timestep t, then after ν steps he or she becomes immune to the virus. So he or she transitions to the recovered state at timestep $t + \nu$ no matter, what states he or she had been unless he or she has died during the ν steps. In this case, he or she stays dead any other case, so we assume that the vaccination has 100% efficiency.

We have to pay attention to the limits of mass vaccination. We assume that in one timestep we can choose a constant z individuals to be vaccinated. We call this z vaccination capacity. Obviously, we won't vaccinate dead people and also those who have been tested positive before. In these cases, the vaccination does not affect the individuals. We are aiming to find the best vaccination method which means minimising the number of deaths until the end time of the epidemic. So the goal is for the hypergraph model described in Section 4.1, fix a vaccination order for the

individuals that minimise the deaths.

We can follow two main tactics. One when we choose the individuals for vaccination according to their age. Because if we vaccinate old people first who has a lower chance to live after being infectious, that way we might reduce the death tolls. We will call this type of vaccination age based vaccination.

The other scenarios come from slowing the virus by vaccinating the potentially super spreaders or important nodes in the population based on the network's topology. This way we might able to prevent the rapid spreading caused by these nodes. We wrote about the importance of the nodes in hypergraphs in Section 1.4. There we presented some methods for importance calculation. We will use these methods to define the importance of each node. After we picked a method and calculated the importance measures then we sort the individuals for vaccination based on that.

The vaccination campaign efficiency does not depend exclusively on the vaccination capacity per day, but also on the precaution measurements which active in the popuation. Yet we have three measurements that can be active besides the vaccination campaign:

- wearing masks and social distancing: this can be controlled by spreading rates w_{wp}, w_{ev} and appearance probability ζ,
- gathering restrictions: remove every hyperedge with size above 10,
- targeted: tests and quarantines: if this method is active then we use testing probabilities $p^S = 0.005, p^E = 0.005, p^I = 0.3, p^A = 0.005, p^R = 0.005, p^D = 0.005$ and quarantine length $q = 14$.

Firstly, let us assume that there is no measurement active and we can vaccinate from the first timestep. In this case, we investigated the effectiveness of three different vaccination methods. The age based method is when we vaccinate in descending order in the age of the individuals. The degree based method means when we vaccinate in descending order in the degree of the individuals. The weighted degree based method is when we have a weight on the hyperedges and we calculate the degree of the nodes by the weighted sum of the incident hyperedges. We defined weight function for the hyperedges $weight_E(h) = \frac{w_h}{|h|-1}$ for all hyperedges. We investigated these vaccination methods with different vaccination limits z and all for fixed immunization time by vaccines $\nu = 14$.

In Figure 4.9 one can see how the death rates behaved for vaccination capacities 50 and 100 in a population with 10000 size.

We generated 20 hypergraphs using the model described in Section 4.1. For spreading rates on the workplace and event hyperedges we fixed the parameters that we used before when there were not any precautions active in the society. Thus we fixed $w_{wp}, w_{ev} = 0.5$ and the appearance probability of the event hyperedges to 0.05. We used $Pareto(2.3)$ for the size distribution of the event hyperedges. As for the other parameters of the model we used the parameters presented in Section 4.4 in Table 4.3. We illustrated the average death rates of these runs in Figure 4.9.

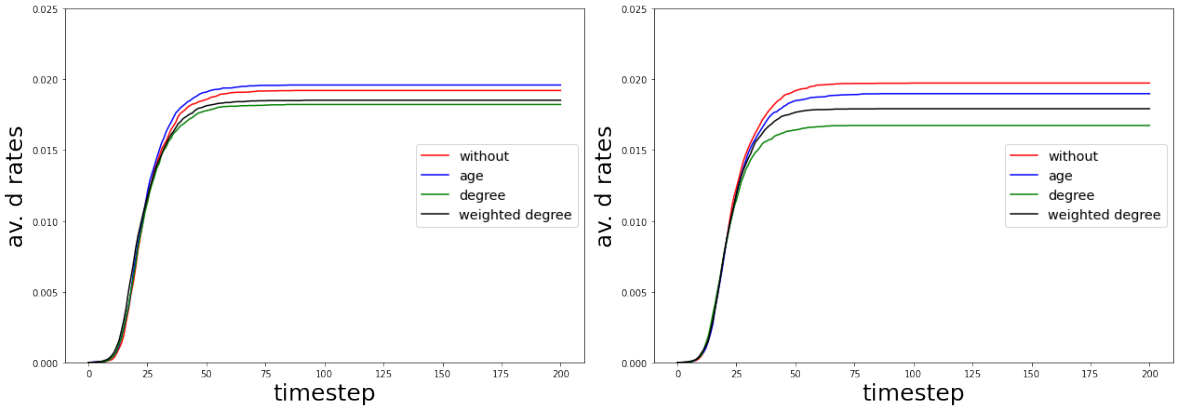


Figure 4.9: The effect of rising the vaccination capacity:
In the left figure belonging to the vaccination capacity 50, we can see that all of the vaccination methods has small effects on the death rates. This could happen because with this speed of vaccination and spreading rates the virus spreads too fast, which the vaccination campaigns can not handle.
In the right figure, we have raised the vaccination capacity to 100. We can see with these parameters the vaccination methods lead to lower final death tolls.

The fact that a person becomes immune 14 timestep later the vaccination results that the vaccination might be more effective on slower epidemic dynamics. What if we can encourage people to wear masks or keep social distance with the aim of lower the spreading rate of the workplace and event hyperedges. In the next figure, we investigated the three vaccination methods as before for fix vaccination capacities in the function of the speed of the epidemic spreading. We controlled the speed of the epidemic spreading by the spreading rate of the event and workplace

precaution level	w_{hh}	w_{wp}	w_{ev}	ζ
j	0.9	$\frac{0.5}{j}$	$\frac{0.5}{j}$	$\frac{0.05}{j}$

Table 4.2: Hyperedge spreadig rate for the levels of precautions.

hyperedges. So we could say there are several levels of precautions that can be active in the society each of them associated with a set of spreading rates on the hyperedge types and the appearance probabilities of the event hyperedges. In Table 4.2, one can see the assignment for level j, what we used for our investigations. We made simulations for levels $j = 1, \ldots, 6$.

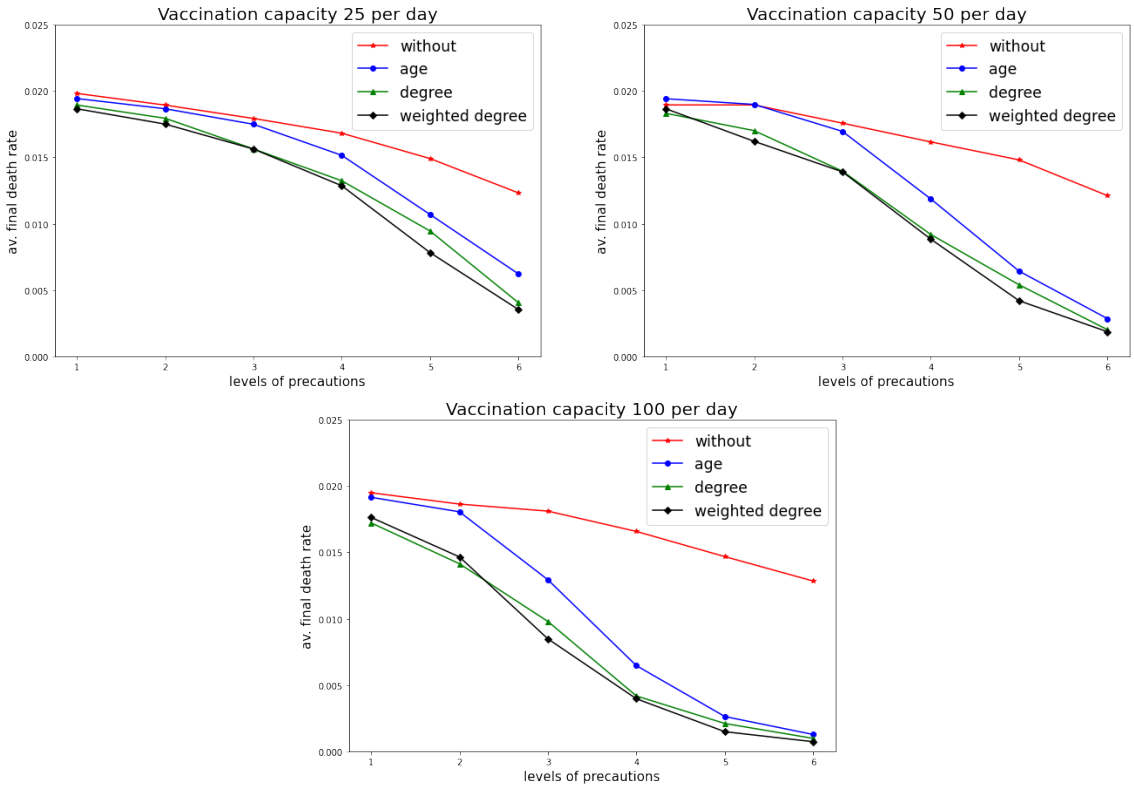


Figure 4.10: The performance of the age based, degree based and weighted degree based vaccination methods for different precaution levels when we have restrictions for gatherings and targeted testing method in the population. The settings of the levels can be seen at Table 4.2.

For every level, we generated 20 hypergraphs and ran the epidemic simulations for without vaccination and for applying the three vaccination methods. In all runs,

we set the parameters except the hyperedge spreading rates to the same values as we did in the previous runs in this section. As we can see from the three cases in Figure 4.10, if we lower the hyperedge spreading rate then the vaccination methods have larger effects on the death rates. We may note the obvious fact that larger vaccination capacity leads to lower death tolls. We could also conclude that the best vaccination method relies on the parameters of the model. As we see for vaccination capacity 100 on precaution level 2 the degree vaccination method came out as the best according to our simulations, but on level 3 the weighted degree vaccination proved to be the best.

Heretofore in this section, we have not activated the restrictions on the gathering sizes and targeted testing. In the following Figure, we see how the death rates behave if we use these measurements in the function of precaution levels.

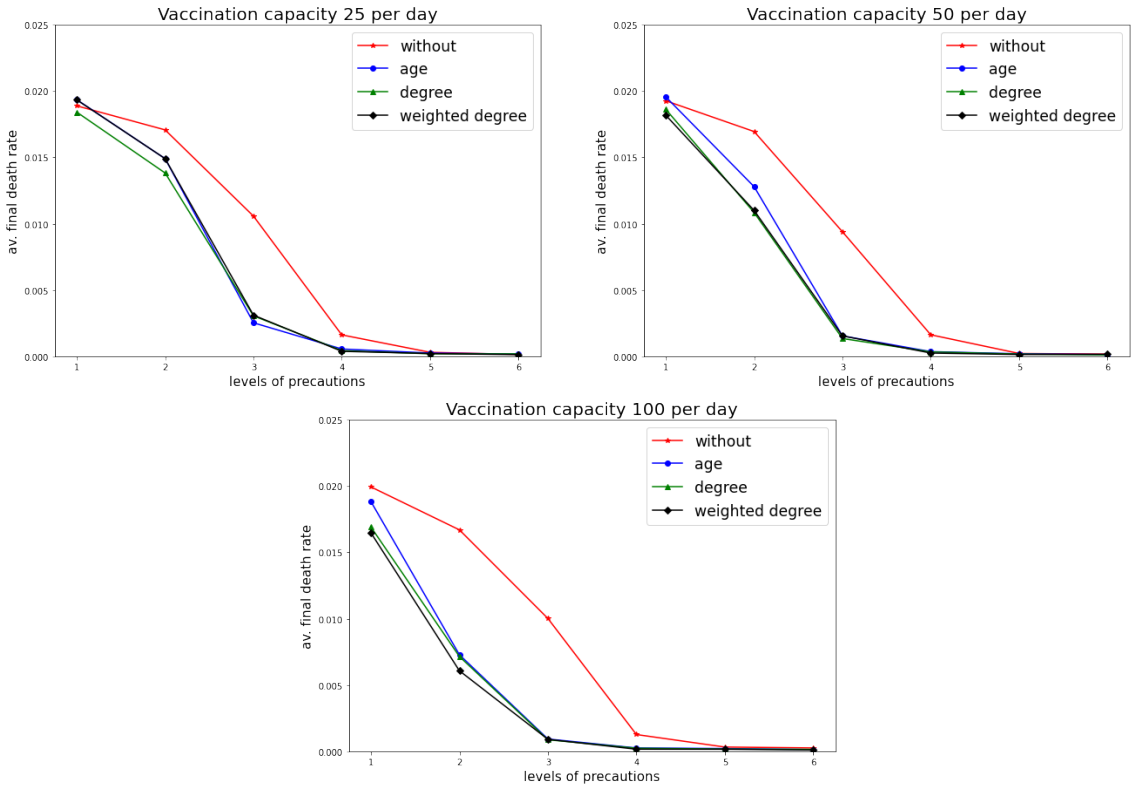


Figure 4.11: The effects of the age based, degree based and weighted degree based vaccination methods for different precaution levels and with targeted testing and gathering restrictions. The settings of the levels can be seen in Table 4.2.

Let us investigate now when we use node centrality measures for defining the importance of the individuals. We run our simulations on a hypergraph having

5000 nodes with the same parameters as before with targeted testing and gathering restrictions. This time we determined the vaccination order by the node's importance measure given by the methods described in Section 1.4.

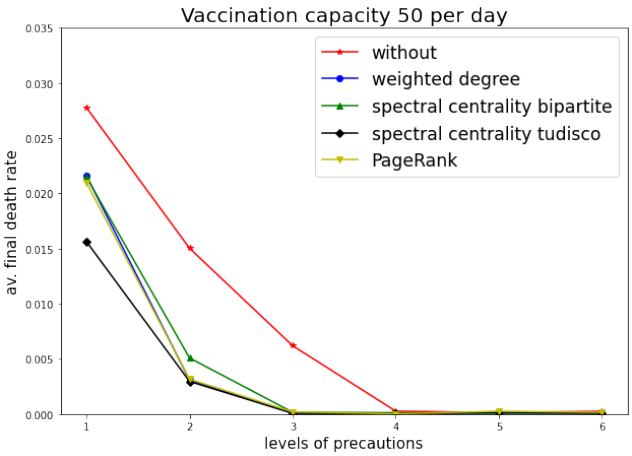


Figure 4.12: The effects of weighted degree based, and node centrality based vaccination methods for different precaution levels and with targeted testing and gathering restrictions. The settings of the levels can be seen in Table 4.2.

As we can see in Figure 4.12, on the method from Tudisco [17] gave the best vaccination order for these parameter settings and precaution levels in our model. One can see from all of the figures in this section, that the difference between the death rates without vaccination and with vaccination is the largest on precaution levels 2 and 3.

Until now, we assumed that we launch our vaccination campaign in the first timestep of the epidemic spreading. But what happens if we have a fixed timestep which defines when we can start sorting out the vaccinations? We run our simulations with a population size 10000 and with the same parameter settings as before in this section. We picked the level 2 for precautions and vaccination capacity 100. Our simulation results can be seen in Figure 4.13.

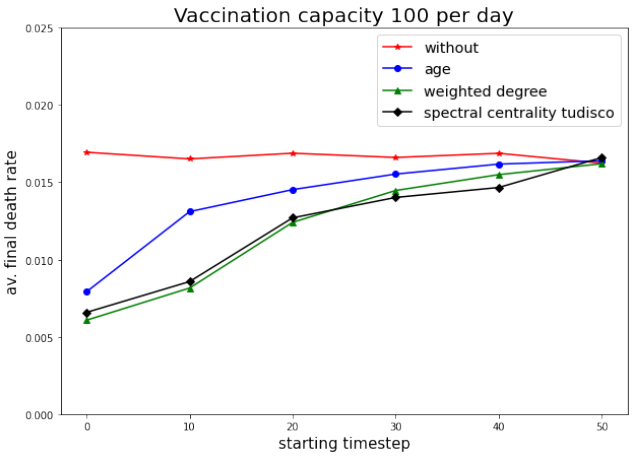


Figure 4.13: The effects of age based, weighted degree based, and node centrality based vaccination methods with targeted testing and gathering restrictions for different vaccination start time. The precaution level is 2.

From these results, we can say that in mass vaccination campaigns, it is crucial that we are able to:

- guarantee sufficient amount of vaccines per day,
- make eligible precautions to slow down the virus,
- start the vaccinations in time.

If one of these conditions is not met, then the vaccination campaign will not have as large effect as it could have on the final death toll. What we can also conclude, that in our model the vaccination methods based on the network's topology could reduce the final death rates more than the method based on age. So prioritising people for vaccination according to their position in the social network could lead us to lower death tolls.

4.4 Model fitting

In this section, our goal is to find a set of parameters for our model described in Section 4.1 that could lead to similar final epidemic rates to the Hungarian real data. According to the Our World in Data database [28] in Hungary by the end of May 2021 the number of deaths caused by coronavirus reached 29581, which roughly 0.003 times the Hungarian population. The database also reports that the total number of confirmed cases was 802510. It is well known, that the reports are under the real the

total number of infected people through the spreading. So we do not want to fit our model's total number of infected to this number. Instead of that we will calculate the total number of confirmed cases in the model by summing up the first positive tests provided by our testing method. If we want to fit our simulation to these final numbers then the total number of positive tests on a 10000 sized population must be around 1000. We may also require from our model that at the end of the epidemic the number of deaths is close to 30.

Firstly, we model the whole outbreak with one wave of the epidemic. We have to consider that in Hungary there have been active precautions and campaigns to slow down the virus spreading. Thus, we must take account into our model:

- wearing masks, social distancing: spreading rates $w_{hh} = 0.9, w_{wp} = 0.2, w_{ev} = 0.2, \zeta = 0.02$,
- restrictions for the number of people in one place: remove hyperedges with size above 10,
- targeted tests, quarantines: $p^S = 0.005, p^E = 0.005, p^I = 0.4, p^A = 0.005, p^R = 0.005, p^D = 0.005$ and quarantine time $q = 14$,
- vaccinations from a fixed timestep: age based vaccination from timestep 40 with 100% efficiency.

We made several attempts to get the correct final numbers from the simulations using brute force. The best settings that we have found is presented below in Table 4.3. Recall the parameters from Section 4.1:

- $\lambda_i^I : E \to I$ transmission probabilities;
- $\lambda_i^A : E \to A$ transmission probabilities;
- $\gamma_i^A : A \to R$ transmission probabilities;
- $\beta_i : A \to I$ transmission probabilities;
- p_i: probability of fatalities from symptomatic infectious state;
- τ^X: fix time period in state X;
- γ^I infectious period parameter

age group	age	λ_i^A	λ_i^I	γ_i^A	β_i	p_i
1	0-24	0.35	0.05	0.1	0	0.001
2	25-49	0.3	0.1	0.09	0.01	0.012
3	50-64	0.2	0.2	0.08	0.02	0.025
4	65-79	0.1	0.3	0.07	0.03	0.08
5	80+	0.05	0.35	0.05	0.05	0.16

τ^E	τ^A	τ^I	γ^I
3	2	2	0.1

Table 4.3: Parameters for model in Subsection 4.1 with final numbers close to real data.

We estimated the probabilities of fatalities from the infectious fatality rates (IFR) which were calculated in [26]. From Table 4.3 we can get the infectious fatality rate, i.e. the probability of death if a person gets the virus. The IFR in our model for age group i is given by:

$$IFR_i = \lambda_i^I p_i + \lambda_i^A \beta_i p_i.$$

IFR_1	IFR_2	IFR_3	IFR_4	IFR_5
0.000125	0.0039	0.015	0.066	0.15

Table 4.4: IFRs by age group in the model.

With these parameter settings, we generated 20 hypergraphs with the model described in Section 4.1.1. We ran epidemic simulations and got 0.0034 for the average number of the final death rates and 910 for the average sum of the first positive tests.

The other way for fitting our model is if we try to reproduce just one wave of the epidemic. Let us chose the last wave of the epidemic in the spring of 2021 in Hungary. From database [28] we collected data from 14.02.2021 till 16.05.2021.

Now, if we want to fit our model to this data, then we have to see that the initial input must be quite special for the simulation. Firstly, we still do not have correct estimations for epidemic numbers like the number of asymptomatic infectious or the recovered people. So we may run into miscalculations if we would just use some guessed initial numbers for these. The only fix points we have are the new reported cases and the number of deaths. On the other hand, if we would have known the

exact numbers for the people in each virus state at 14.02.2021 in Hungary, then we can not just sit back and sample uniformly from the population to set the viral states according to the initial numbers. Why wouldn't this easy method help us? Let us imagine, that we start an epidemic and stop it somewhere in the middle of the process. Now, if we would colour the people according to their viral states, then we wouldn't see a uniformly random colouring. We might see more like colour fronts and areas with one colour.

Thus, with this knowledge in our pocket, we may search for a better idea than random initialisation. Let us run a 'pre-wave' of this wave of the virus to get the initial state distribution on our hypergraph. We ran this first wave simulation with a slower and weaker virus on a hypergraph with 10000 nodes generated according to our model described in Subsection 4.1.1. We set spreading rates $w_{hh} = 0.8$, $w_{wp} = 0.1$,$w_{ev} = 0.1$ and appearance probability $\zeta = 0.015$ and for all age group i the death probabilities $p_i' = p_i/2$. The rest of the parameters were the same as in Table 4.3. We have not used the vaccination method for the first wave as in Hungary before 14.02.2021 there was not any notable vaccination campaign. However, the targeted testing and gathering restrictions were active measurements, so we also used our methods for them with $p^S = 0.005, p^E = 0.005, p^I = 0.3, p^A = 0.005, p^R = 0.005, p^D = 0.005$ and 10 for the limit of event and workplaces sizes.

We ran this simulation until the first timestep when the rate of the total deaths was over the rate in Hungary at 14.02.2021. and the rate of infectious was under 0.004. With this move, we found ourselves where we exactly wanted to, in a middle of an epidemic with the required initial death rates. So, after this we set back the death probabilities according to the values in Table 4.3 and the spreading rates $w_{wp} = 0.2$, $w_{ev} = 0.2$, and appearance probability $\zeta = 0.02$. Then we continued the simulation with the changed parameters considering one day as one timestep in our model. We also started vaccinating the population from the first timestep in this second run. We used our age based method (see Section 4.3) for the order of the mass vaccination. We set the vaccination capacity for each day according to the rate of the real 7-day smoothed new vaccinations in the database [28] from Hungary. So, we vaccinated exactly the same proportion of the population per day as in the smoothed real vaccination data (see Figure 4.14).

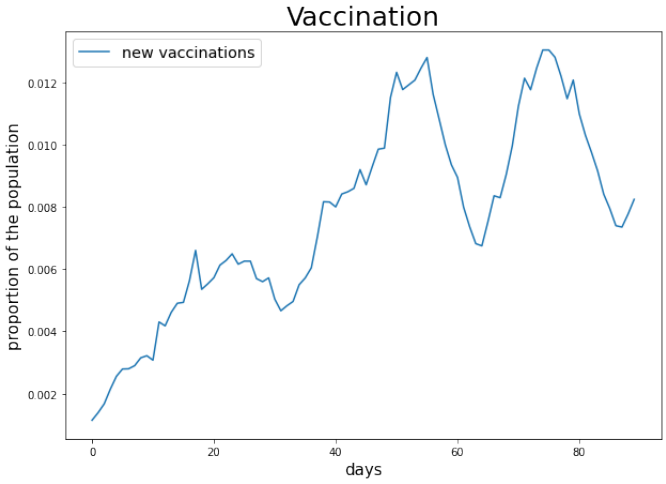


Figure 4.14: Vaccination proportions in the function of time

If we look at Figure 4.15 then we see the result of the simulation with dashed lines and the real data with continuous lines. We tried to fit our simulation to these data gained from the database:

smoothed new cases: This rate comes from a 7-day average of the reported new cases of the coronavirus in Hungary. This means we sum the confirmed cases before and after 3 days and the cases on the day and divide it by 7 and also with the size of the population.

confirmed infectious: This rate comes from summing the smoothed new cases from days no later than 13 days before the day.

total deaths rate: The sum of total deaths caused by the virus divided by the size of the population.

Our highlighted simulation rates in Figure 4.15:

smoothed new cases: This rate comes from a 7-day average of the first positive tests in the simulation. This means we sum the first positive tests before and after 3 days and tests on the day and divide it by 7 and also with the size of the population.

confirmed infectious: This rate comes from summing the smoothed new cases from timesteps no later than 13 timesteps before the timestep.

total deaths rate: The sum of total deaths caused by the virus divided by the size of the population.

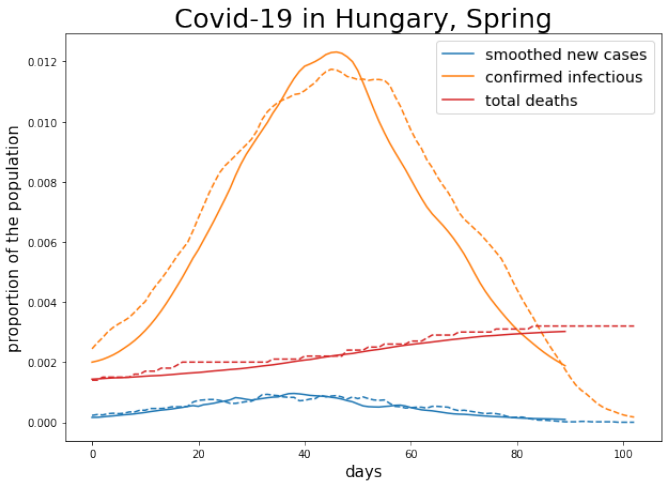


Figure 4.15: The continuous lines are representing real data from the last wave of the COVID-19 in Hungary spring, 2021 from Our World in Data ([28]). The dashed lines are a simulated epidemic spreading with our model.

If we look on the fitting results in Figure 4.15, we might say that our model with the right parameter settings and initialisation can simulate almost correctly one wave of the virus.

Chapter 5

Summary

In this final chapter, we would like to make some conclusions and mention the main results and questions of our paper. We had two main goals were to investigate the differences between epidemics on graph models and hypergraphs, and to create a simulation environment that we can use to examine the effects of the different measurements against the virus spreading and not least fit it to real data of an epidemic spreading.

In Chapter 1, we wrote an introduction about the random hypergraph models. There we extended some well-known graph model like the Hidden Parameter or the Barabási-Albert models to hypergraphs. We also implemented these generative models in Python and used them for our epidemic model later.

In the second chapter, we presented some epidemic models that we used as inspirations for our epidemic model.

In Chapter 3, we developed our discrete stochastic SEIRD model on hypergraphs. However, the larger part of the analytical study of the model is still a job for the future, but we have made empirical experiments from the differences between the hypergraph and its clique graph model. From our simulation results and our calculations in the case of one community, we found that the difference between the epidemic model on hypergraph and on its clique graph depends on the spreading function of the hypergraph model, the hyperedge size distribution of the hypergraph, and spreading rates of the hyperedges.

In Chapter 4, we built up and fit an epidemic model which was the refinement of the model in the previous chapter. With this model, we were able to reproduce the data

from the final wave of the coronavirus in Hungary. We gained from our empirical investigations of the precaution levels and measurements in the population that it is very important to apply these methods against the virus at the same time to get lower death rates.

As for the future, it would be interesting to implement and use hypergraph models for epidemic spreads which are not the extension of graph models. These models may reflect real group networks more accurate. See for example [29] or [30].

In our model for COVID-19 we used appearance probabilities for hyperedges, so in a typical timestep, the virus cloud not spread through every hyperedge of the hypergraph just through the active edges. It may be a task for the future to compare these random hyperedge activations to one with stable hyperedges but the same probability of infection in one timestep.

Before fitting our model to the Hungarian data, we used a weaker virus to set the initial distribution of the states on the hypergraph. We have mentioned that the state distribution on the nodes after or during an epidemic spreading is not uniformly random distributed in the population. We may suggest that it would be very interesting to study these epidemic mid-time state distributions.

It is also a job for the near future to fit our model to data of other countries or different diseases from COVID-19.

Appendix A

Algorithm 1: Nonlinear Power method for hypergraph centrality from [17]

Input: Incidence matrix B of the hypergraph; diagonal weight matrices W_E and W_V for hyperedges and nodes; nonlinear functions f, g, ϕ, ψ; desired vector norm $\|.\|$; stopping tolerance tol

Output: Centrality for nodes $\mathbf{x}$ and hyperedges $\mathbf{y}$ such that $\|\mathbf{x}\| = \|\mathbf{y}\| = 1$

$\mathbf{x}^{(0)}, \mathbf{y}^{(0)} > 0$

while $\|\mathbf{x}^{(r+1)} - \mathbf{x}^{(r)}\|/\|\mathbf{x}^{(r+1)}\| + \|\mathbf{y}^{(r+1)} - \mathbf{y}^{(r)}\|/\|\mathbf{y}^{(r+1)}\| > tol$ **do**

- $\mathbf{u} \leftarrow \sqrt{g\left(BW_E f(\mathbf{y}^{(\mathbf{r})})\right)}$
- $\mathbf{v} \leftarrow \sqrt{\psi\left(B^T W_V \phi(\mathbf{x}^{(r)})\right)}$
- $\mathbf{x}^{(r+1)} \leftarrow \mathbf{u}/\|\mathbf{u}\|$
- $\mathbf{y}^{(r+1)} \leftarrow \mathbf{v}/\|\mathbf{v}\|$

Bibliography

[1] William O Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the Royal Society of London A: mathematical, physical and engineering sciences* 115.772 (1927), pp. 700–721.

[2] Tom Britton. *Stochastic epidemic models: a survey*. 2009. arXiv: 0910.4443 [math.PR].

[3] Håkan Andersson and Tom Britton. "Stochastic Epidemic Models and Their Statistical Analysis, Volume 151 of Lecture Notes in Statistics". In: vol. 151. Jan. 2000. DOI: 10.1007/978-1-4612-1158-7.

[4] István Z Kiss, Joel C Miller, Péter L Simon, et al. "Mathematics of epidemics on networks". In: *Cham: Springer* 598 (2017).

[5] P. Erdős, Chao Ko, and R. Rado. "INTERSECTION THEOREMS FOR SYSTEMS OF FINITE SETS". In: *The Quarterly Journal of Mathematics* 12.1 (1961), 313–320. DOI: 10.1093/qmath/12.1.313.

[6] C. Berge and D. Ray-Chaudhuri. *Hypergraph Seminar: Ohio State University, 1972*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2006. ISBN: 9783540378037. URL: https://books.google.hu/books?id=1Jt7CwAAQBAJ.

[7] P. Erdös and A. Rényi. "On Random Graphs I". In: *Publicationes Mathematicae Debrecen* 6 (1959), p. 290.

[8] M. KARÓŃSKI and T. ŁUCZAK. "Random hypergraphs". In: *Bolyai Soc. Math. Stud.* 2 (1996), pp. 283–293.

[9] E. N. Gilbert. "Random Graphs". In: *Annals of Mathematical Statistics* 30.4 (1959), pp. 1141–1144.

[10] Albert-László Barabási, Réka Albert, and Hawoong Jeong. "Scale-free characteristics of random networks: the topology of the world-wide web". In: *Physica A: statistical mechanics and its applications* 281.1-4 (2000), pp. 69–77.

[11] Pierre Deville et al. "Scaling identity connects human mobility and social interactions". In: *Proceedings of the National Academy of Sciences* 113.26 (2016), pp. 7047–7052.

[12] D.J. Watts and S.H. Strogatz. "Collective dynamics of 'small-world' networks". In: *Nature* 393 (1998), pp. 440–442.

[13] Réka Albert and László Barabási-Albert. "Topology of evolving networks: local events and universality". In: *Physical Review Letters* (1999).

[14] B. Bollobás and O. Riordan. "The Diameter of a Scale-Free Random Graph". In: *Combinatorica* 24 (2004), pp. 5–34.

[15] Frédéric Giroire et al. "Preferential attachment hypergraph with high modularity". In: *arXiv preprint arXiv:2103.01751* (2021).

[16] Chen Avin et al. "Random preferential attachment hypergraph". In: *Proceedings of the 2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. 2019, pp. 398–405.

[17] Francesco Tudisco and Desmond J. Higham. *Node and Edge Eigenvector Centrality for Hypergraphs*. 2021. arXiv: 2101.06215 [cs.SI].

[18] Austin R Benson. "Three hypergraph eigenvector centralities". In: *SIAM Journal on Mathematics of Data Science* 1.2 (2019), pp. 293–312.

[19] Lawrence Page et al. *The PageRank citation ranking: Bringing order to the web*. Tech. rep. Stanford InfoLab, 1999.

[20] Amy N Langville and Carl D Meyer. "A survey of eigenvector methods for web information retrieval". In: *SIAM review* 47.1 (2005), pp. 135–161.

[21] D. J. Daley and J. Gani. *Epidemic Modelling: An Introduction*. Cambridge Studies in Mathematical Biology. Cambridge University Press, 1999. DOI: 10.1017/CBO9780511608834.

[22] Ágnes Bodó, Gyula Katona, and Péter Simon. "SIS Epidemic Propagation on Hypergraphs". In: *Bulletin of Mathematical Biology* 78 (Sept. 2015). DOI: 10.1007/s11538-016-0158-0.

[23] URL: https://github.com/kolokcse/Epidemicmodel.git.

[24] Albert-László Barabási and Márton Pósfai. *Network science*. Cambridge: Cambridge University Press, 2016. ISBN: 9781107076266 1107076269. URL: http://barabasi.com/networksciencebook/.

[25] Ying Liu et al. "The reproductive number of COVID-19 is higher compared to SARS coronavirus". In: *Journal of Travel Medicine* 27.2 (Feb. 2020). taaa021. ISSN: 1708-8305. DOI: 10.1093/jtm/taaa021. eprint: https://academic.oup.com/jtm/article-pdf/27/2/taaa021/32902430/taaa021.pdf. URL: https://doi.org/10.1093/jtm/taaa021.

[26] Andrew T. Levin et al. "Assessing the age specificity of infection fatality rates for COVID-19: systematic review, meta-analysis, and public policy implications". In: *European Journal of Epidemiology* 35.12 (2020), 1123–1138. DOI: 10.1007/s10654-020-00698-1.

[27] John Doe. *Population: Demographic Situation, Languages and Religions*. 2021. URL: https://eacea.ec.europa.eu/national-policies/eurydice/content/population-demographic-situation-languages-and-religions-35_en.

[28] Max Roser et al. "Coronavirus pandemic (COVID-19)". In: *Our world in data* (2020). https://ourworldindata.org/coronavirus.

[29] Gergely Palla et al. "Social group dynamics in networks". In: *Adaptive Networks*. Springer, 2009, pp. 11–38.

[30] Manh Tuan Do et al. "Structural patterns and generative models of real-world hypergraphs". In: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2020, pp. 176–186.