

Project: Modelling the Spread of Covid-19

Leontii Shtokolenko ID 14699966

Michał Sanocki ID 14623366

October 2022

Contents

1	Introduction	1
2	Problem 1: Gillespie's Direct Algorithm	2
2.1	Five hallmarks of stochastic SIR dynamics	3
3	Problem 2: Spatial Models	6
3.1	Meta-population model	6
3.2	Theoretical overview of used network models	9
3.2.1	Barabasi-Albert model	9
3.2.2	Watts-Strogatz model	10
3.2.3	Erdos-Renyi model	10
3.2.4	Models comparison	10
3.3	Epidemic models on networks	12
3.3.1	Basic SIR model on the network	12
3.3.2	Practical comparison	12
3.3.3	Influence of selective infection	14
3.3.4	Selective vaccination	14
4	Appendix	17
4.1	Network generation	17
4.2	Basic SIR model simulation	17
5	Conclusion	17

1 Introduction

In the follow-up to the previous paper on the possibilities of modelling infectious diseases, we will explore the behaviour of the stochastic SIR models and spatial models. In the first part dedicated to stochastic models, we will focus on the event-driven

approach towards the modelling of infectious diseases. We will discuss the different properties of those models and how they differ from the deterministic SIR ODEs including disease extinction. In order to achieve this goal we will utilize Gillespie's Direct Algorithm [1], which is an algorithm that uses the weighted chance to decide what event happens based on the previous state of the model. This weighted chance is based on the population in the model, in combination with the probability of that event happening.

The second part is going to be dedicated to the spatial models of infectious diseases, first, we are going to discuss the meta-population model after which we are going to explore the behaviour of network models and how they can be applied to evaluate vaccination strategies.

2 Problem 1: Gillespie's Direct Algorithm

It is worth highlighting that in the event-driven approach, all values must be integers, and in addition, the stochastic models' behaviour is dependent on the population size. Therefore, we cannot rely on the proportions of the population as in our previous study. Events still occur randomly and sequentially, but their rates depend on the given parameters and equations. Each event causes an integer change in the population variables, and the process is repeated [2]. In his original paper, Gillespie proposed two different et mathematically equivalent algorithms [1]: Direct Algorithm and First Reaction Method. In this paper, we will utilize the direct algorithm due to its increased computational efficiency. The Gillespie's Direct algorithm works as follows:

1. Initialize the time $t=t_0$ and the system's state $\mathbf{x}=\mathbf{x}_0$
2. Evaluate the time τ until the next event occurs

$$\tau = \frac{-\log(\text{RAND})}{\sum_j a_j} \quad (1)$$

RAND - random number from the uniform distribution and range(0,1).

3. With the system in state \mathbf{x} at time t , evaluate all the $a_j(\mathbf{x})$, and their sum

$$\sum_j a_j(\mathbf{x}) \quad (2)$$

4. Choose which event will be performed, based on the RAND_2
5. Effect the next reaction by replacing $t \leftarrow t + \tau$, and $\mathbf{x} \leftarrow \mathbf{x} + \mathbf{v}_j$
6. Record (\mathbf{x}, t) and return to step 1, or else end the simulation.

The events which we will consider in this chapter are (in the version without demography, two last events will be omitted):

- Infection $S \rightarrow S-1, I \rightarrow I+1$

- Recovery $I \rightarrow I-1, R \rightarrow R+1$
- Death (three events with different corresponding rates) $S \rightarrow S-1, I \rightarrow I-1, R \rightarrow R-1$
- Birth $S \rightarrow S+1$

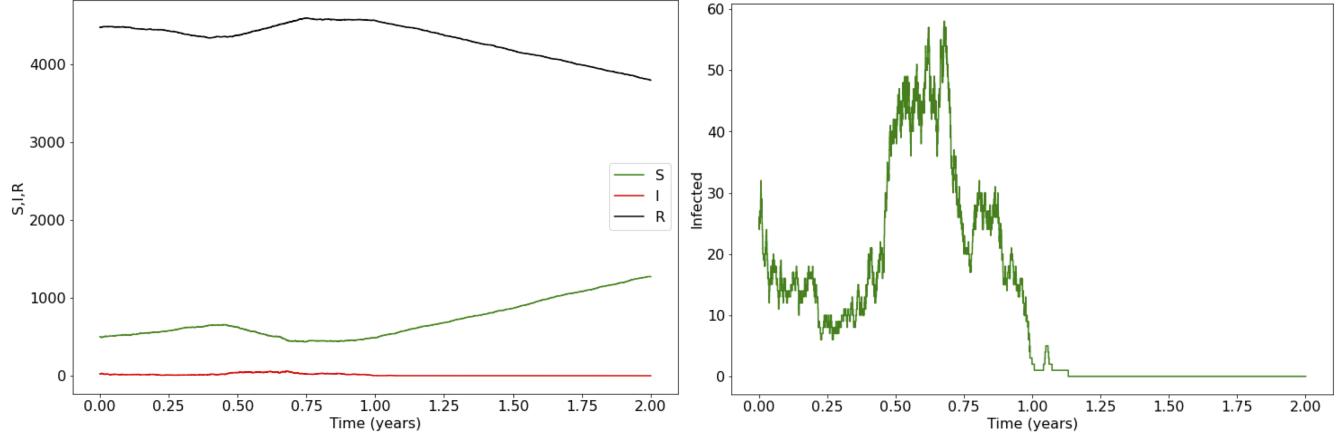
2.1 Five hallmarks of stochastic SIR dynamics

The SIR discrete event approach based on this algorithm will allow us to explore five properties of the stochastic SIR models:

1. Variability between simulations Due to the random nature of the stochastic SIR models, we will observe different outcomes each time. This is quite problematic because it disables us to predict the exact disease prevalence in the future, even though the statistical properties might be accurate (Figure1)
2. Variances and covariance Random deviations may give rise to variations in the prevalence of the diseases, which in combination with the deterministic dynamics creates the negative covariance between I and S. Additionally, this can cause the mean populations to deviate from their deterministic equilibria Figure 2.
3. Increased transients When the stochastic process manages to pull the system out of its deterministic equilibria, the underlying forces will try to push the system back to the equilibrium. This phenomenon often results in oscillation around the single deterministic value. It is also important to highlight that the strength of those forces depends on the distance from equilibrium, so states with higher deviations will be less stable.
4. Stochastic Resonance Perturbations can excite the oscillations of the system close to the natural frequency leading to cyclic-like behaviour.
5. Extinctions As a result of introducing the stochasticity to our model, it is now possible for the disease to go extinct, even when the R_0 is higher than 1, due to the random fluctuations in our model. The probability of these events in the given timeframe is dependent on the population size, which means that a larger population can on average support the endemic state for longer periods of time. Crucially all closed populations will eventually experience extinction in the long run. Thus the endemic state can only be sustained by the imports of the pathogen from the outside.

In order to better understand the extinctions we will examine how often those events occur in different variations of the SIR event-driven model. First, we will perform simulations for models without demography, later we will add demography (this version was used in previous examples) and finally, we will examine the influence of imports. In the first case, our model is going to be modified simply by deleting the rates for deaths and births. However, for imports we are going to add the import rate [2]: $I \rightarrow I - 1$, Rate = $0.0625\mu(R_0 - 1)\sqrt{N}$

SIR discrete event model



Variability in Simulations

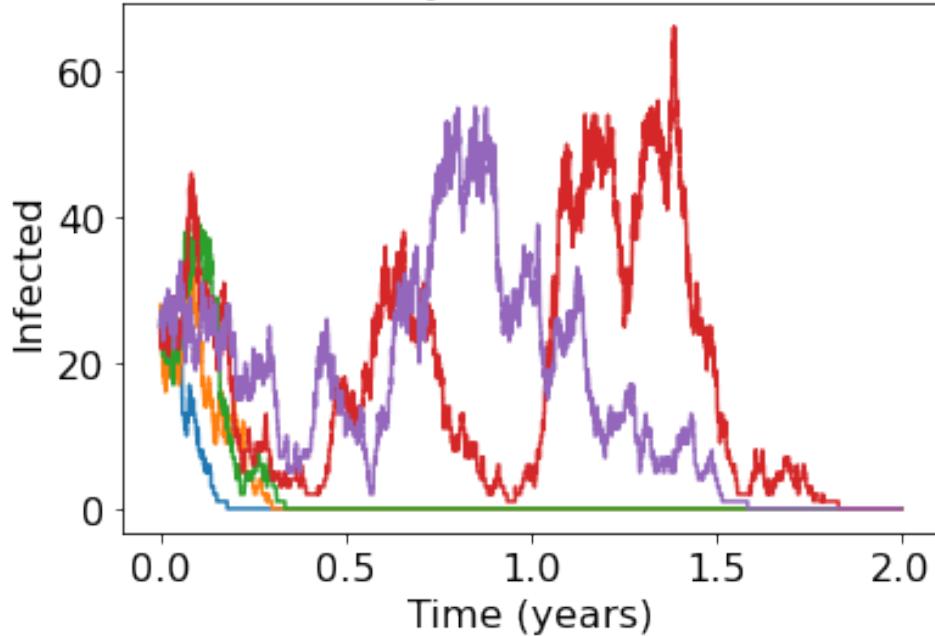
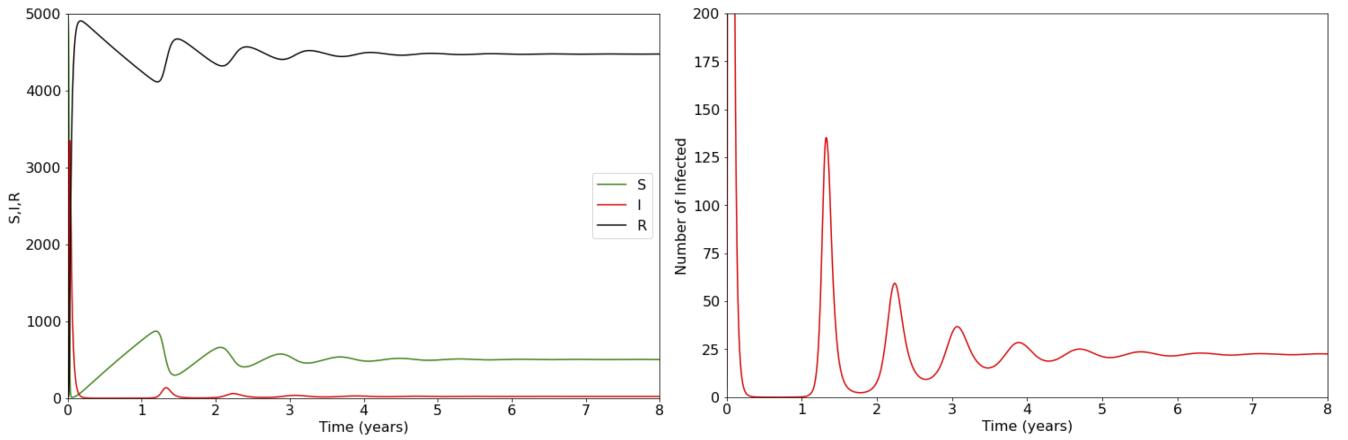


Figure 1: The plots above show the behaviour of the SIR discrete event model based on Gillespie's direct algorithm. As we can observe, once the disease is extinct, the number of susceptibles decreases due to death caused by natural causes. The covariance between S and I is -2251.9, which is a result of interactions between stochasticity and deterministic nonlinear dynamics. Not only does the number of infected deviate significantly from its deterministic equilibrium, but also after approximately one year the disease goes extinct, even though R_0 was set to 10. Due to the stochastic nature of the discrete event models, each run gives us different results, this makes predictions on the future state of the system impossible based on just one run. We can, however, make some observations based on multiple runs. For example, in the runs presented in the bottom plot extinction happened in all of the runs and in three out of five cases, it happened in less than a year. We can approximate the distribution of the variables in the future but not the exact prevalence of the diseases.

Deterministic SIR



Stochastic SIR

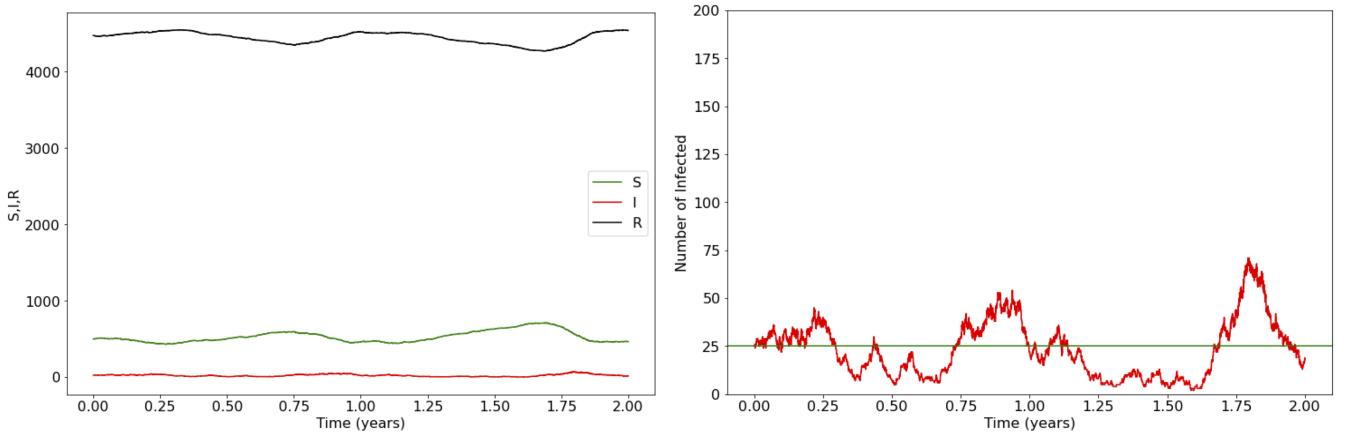


Figure 2: The figure above presents the differences between stochastic and deterministic SIR models. Both simulations were conducted with the same set of parameters. Once the deterministic SIR model reached equilibrium, we employed the event-driven SIR model, which better illustrates the behaviour of endemic diseases in a smaller population, due to its more dynamic behaviour, in comparison with the deterministic version. The oscillations of the number of infected individuals around the deterministic equilibrium can portray the 3rd and 4th hallmarks of the SIR.

For each model, we performed 100 simulations, which enabled us to perform a statistical analysis of our system due to a large sample size. Each run was started from the same starting point, and at the end, we measured when the extinction happened to approximate the distribution of extinction time for different models. Based on the histogram (Figure 3.), we can say that model with demography extinct significantly quicker than other versions. However, we do not have a reason to think that imports significantly change the extinction time.

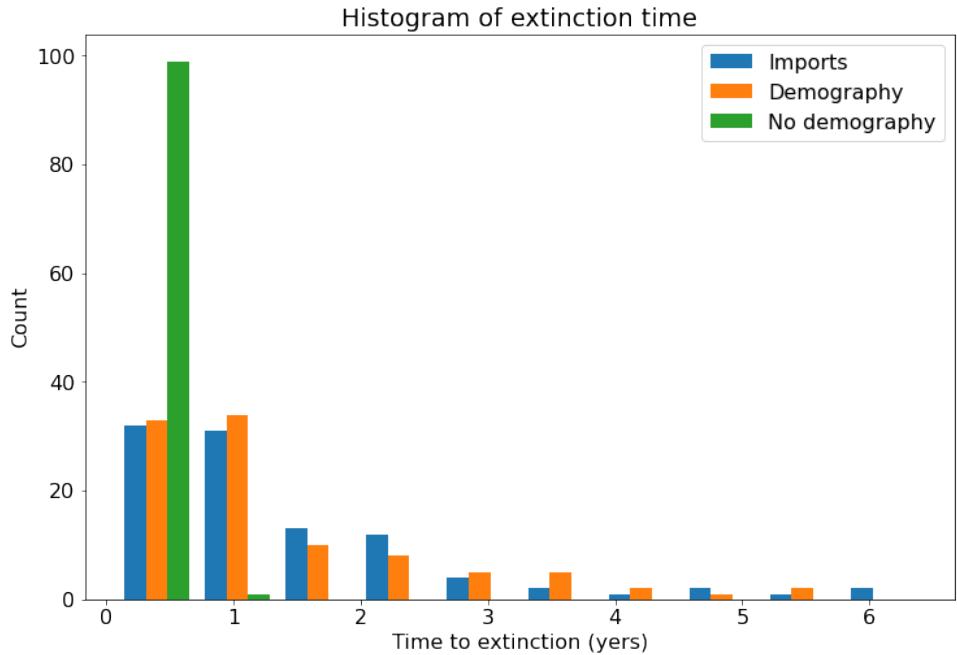
Differences between average extinctions time are not statistically significant between those two models, and the distribution shown on the histogram is not high enough to justify the conclusion that imports can extend the expected time for extinction. The main difference between SIR with demography and SIR with imports doesn't come from the extended expected extinction time but from the ability of the disease to come back even after it went extinct in a given population, which is why they are considered necessary for pathogens to persist in the population for longer periods.

3 Problem 2: Spatial Models

3.1 Meta-population model

Meta-population models are often used in the modelling of infectious diseases because they allow us to take into account the following concepts: heterogeneity, interactions, isolations, local extinctions and scale(populations and simulation). The basic framework for this method works as follows we divide the total population into groups (cities, villages, or based on administrative boundaries) with their own independent dynamics and a limited form of interaction. This approach allows us to better model most human diseases, especially for populations, where there are a lot of semi-separated communities [2]. Although it is possible to use the metapopulation model both in the deterministic and stochastic form in our study, we are going to apply the event-driven approach from the previous chapter. As shown in Figure 4 stronger interactions will speed up the epidemic due to the quicker transport of the disease. This leaves us with a very important conclusion connected to the development of the epidemic, it is crucial in the early stages of the epidemic two minimize the number of interactions as much as possible in order to 1) maximize the chances for extinction, 2) minimize transport and thus slow down its progression which will give the healthcare system more time to react.

In the second model, we used the meat population model comprised of six equal-sized subpopulations. Populations were numbered from one to six. If they were neighbours (i.e. 1-2, 2-3, 5-6 etc.) then ρ_{ij} was set to 0.4. If they were second-degree neighbours (I.e. 1-3, 2-4), then ρ_{ij} was set to 0.3, and so on. However, even very distant populations had ρ_{ij} set value to 0.1 to better represent the smaller interactions between populations. This model can be interpreted as six equal cities placed in a line since interactions between one and six were considered small. In this example, we will also compare the results of introducing demography to our model. After the addition of



Event driven SIR with imports

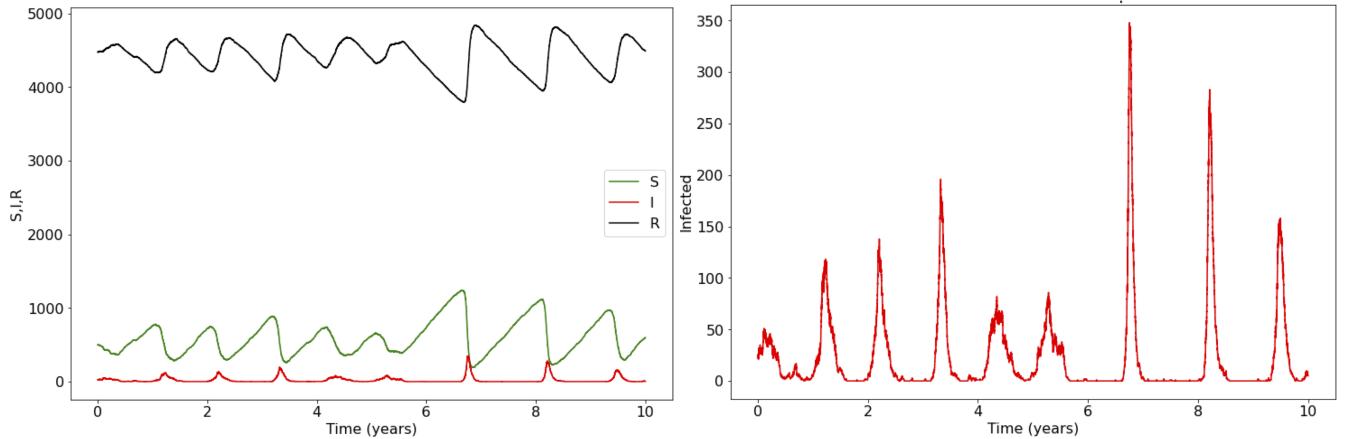


Figure 3: The histogram presented above shows the distribution of extinction time in different models based on 100 runs for each version. Even though the import model has the highest extinction time out of all the modifications, the differences are not high enough to justify the hypothesis that imports extend the time needed for extinction. They, however, enable previously extinct diseases to come back even if $I = 0$. This phenomenon displays why imports are necessary for diseases to persist in the population for extended periods of time.

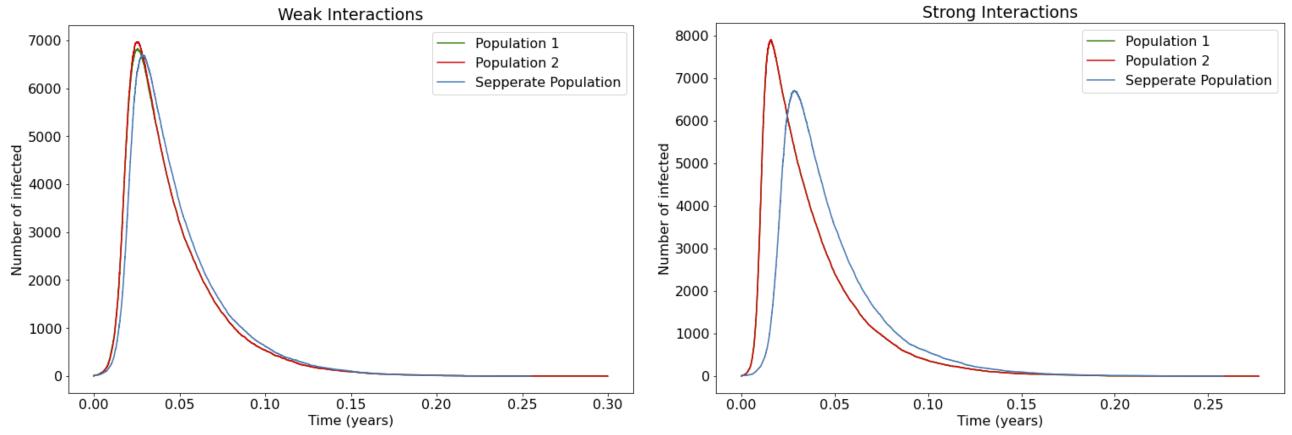


Figure 4: Comparison of two meta-population models fully susceptible populations with the same size but with different interactions between subpopulations. In the first example ρ_{ij} was set to 0.1, and in the second to 0.8. In both cases, the results were compared to the event-driven SIR model without demography. For the model with stronger coupling between subpopulations, the disease burns through the population significantly quicker than in the other example. The differences between metapopulations and separated populations in the first example aren't high enough to be considered meaningful.

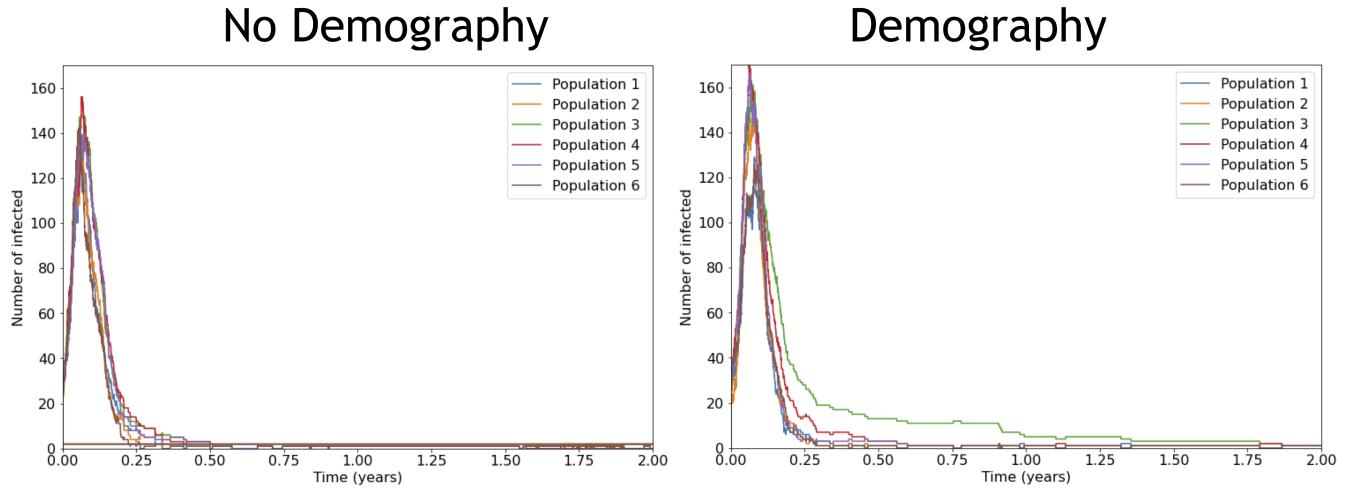


Figure 5: Comparison between metapopulation models with and without demography. The addition of demography extends extinction time due to the introduction of the new stream of susceptibles.

demography and new subpopulations from the starting point based on deterministic equilibrium. The progression of the disease is significantly more chaotic than in the previous case. Although the simulations start from the deterministic equilibrium, at the beginning we can observe a spike in the number of infected. Additionally, the last population to go extinct is population 3 in version with demography, which is located in the middle and thus has stronger interactions with neighbours.

3.2 Theoretical overview of used network models

Although spatial model described in the Section 3.1 includes possible geographical effect on the infection dynamics, it still relies on quantitative qualities inside each sub-population with a certain level of interaction between them. Following the same logic as with transition from a single population to meta-population models, we will assume that each sub-population is small enough or, possibly, consisting even of only one individual, so the internal dynamics of infection in this population is negligible on the scale of whole simulation period. Let also set the transition rates between these populations as equal to each other. In other words if A and B - two different sub-populations and $A \rightarrow B$, $B \rightarrow A$ - transition rates between them, then $A \rightarrow B = B \rightarrow A$. With these assumptions the spread of an infection can be modeled using *networks* and corresponding to them graphs.

In this work three types of randomly generated networks was used: Barabasi-Albert, Watts-Strogatz and Erdos-Renyi models [3]. This paper also provides us with sufficient for our purposes metrics for model comparison. We will analyse and use only final (static) state of these models, ignoring the dynamics of network generation. This corresponds to a highly established population structure in reality with negligible structural changes in a certain time period.

3.2.1 Barabasi-Albert model

The Barabasi-Albert (BA) model is an algorithm of generation of complex scale-free networks originally described in [4]. The algorithm can be defined with the following statements:

1. Some number of connected nodes m_0 is set as an initial state of the model.
2. For each time step one node is added to the network with $m < m_0$ connections to other nodes.
3. The probability of the new node connected to some i-node is $p_i = \frac{k_i}{\sum_j k_j}$, where j is indexing over all nodes that have been existed before the current time-step.

As a scale-free network, BA's main structural characteristics are growth and preferential attachment. First one is representing the algorithm's dynamics and hardly influence the final state of the model. Preferential attachment, on the other hand, has a considerable influence on the final state. It results in diversity of nodes' degrees since, by definition, in BA model nodes with a higher degree tend to increase it during the generation process.

3.2.2 Watts-Strogatz model

Another random graph generation model that we are considering is Watts-Strogatz (WS) model. As an input, model takes the total number of nodes N , mean degree of a node K and the probability of creation of an edge β . The generation algorithm can be defined with the following statements:

1. An initial ring lattice with N nodes each connected to K neighbours, $K/2$ on each side. The condition of creation of an edge (i, j) is $0 < |i - j| \text{mod}(N - 1 - \frac{K}{2}) \leq \frac{K}{2}$.
2. Every node i one edge that satisfies the condition $(i, j \text{mod} N)$ with $i < j \leq i + K/2$) is rewired with probability β .
3. Rewiring is done by replacing $(i, j \text{mod} N)$ node with (i, k) is randomly chosen from all possible nodes in such way, that all self-loops and duplicates are avoided.

Due to the underlying lattice structure, WS graphs often have locally clustered networks with small number of connection between them. The structure of model can significantly distinct for a different parameters, therefore, WS models must be generated with high attention to interaction between the parameters.

3.2.3 Erdos-Renyi model

The last randomly generated graph model that we will consider is Erdos-Renyi (ER) model. It takes two parameter on the input: total number of nodes N and probability of creating an edge p . The generation of the graph is defined with the following steps:

1. Some initial set of N nodes is generated.
2. For every pair of vertices i and j edge (i, j) could be created with given probability p .

Despite relative simplicity of the algorithm, it can generate fairly complex models, some of which are shown in the Section 3.2.4.

3.2.4 Models comparison

It is important to outline some distinctive properties between networks generated with BA, WS and ER algorithms. Some examples of these networks are provided in the Figure 5.

Some special cases of generated graphs are represented in the Figure 6. These output was mostly avoided for analysis of infection dynamics in order to maintain frame of reference for model comparison.

As the frame of reference, three main properties has been chosen: degree distribution l , clustering coefficient C and Pathlength $P(k)$. Mentioned properties for all three types of graphs found in [3] are represented in the Table 1.

Although all types of graphs mostly have distinct properties, one of the most significant difference between them is in the clustering coefficient C . In particular, from the Table 1 it is seen that in WS model clustering coefficient $C = \text{const}$ in the limit for

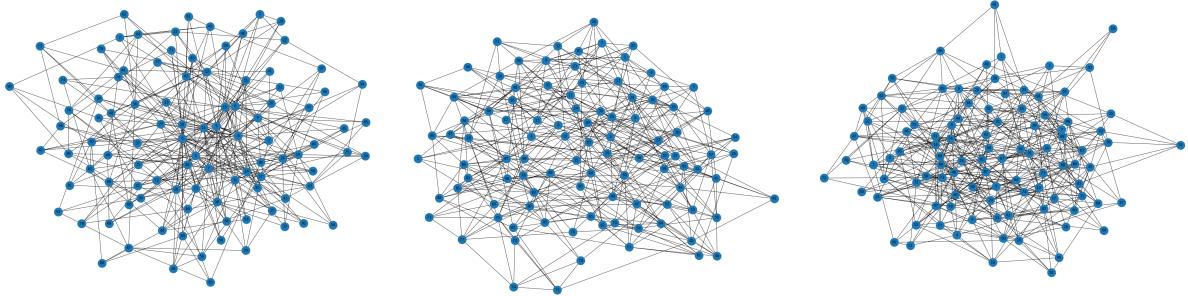


Figure 6: Examples of generated networks. From left to right: Barabasi-Albert, Watts-Strogatz, Erdos-Renyi models. Initial parameters was chosen for graph to have the same number of nodes $N = 100$ and same order of numbers of edges.

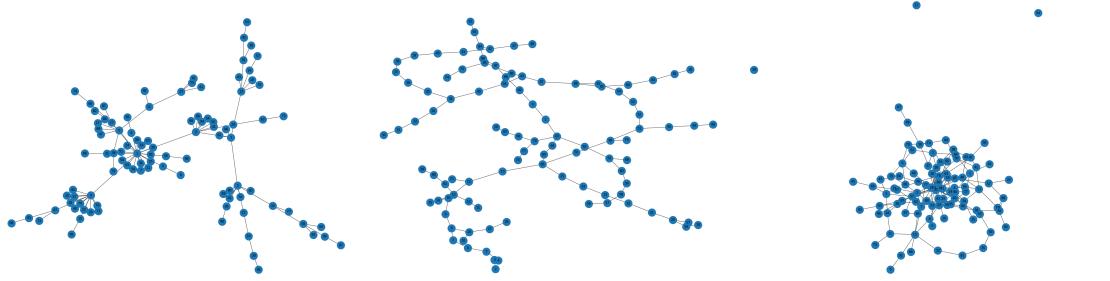


Figure 7: Special cases of used graph generation algorithms. From left to right: Barabasi-Albert, Watts-Strogatz, Erdos-Renyi models. For some initial parameters generated graphs can have absolutely different structures

Table 1: Table with the main properties of the network models. All considered models have distinct properties with exception of equal l in WS and ER networks

Model	Degree distribution l	Clustering coefficient C	Pathlength $P(k)$
Barabasi-Albert	$= \frac{\ln N}{l \ln N}$	$\sim \frac{(\ln N)^2}{N}$	$\sim k^{-\gamma}$
Watts-Strogatz	$= \frac{\ln N}{\ln \langle k \rangle}$	$\sim \text{const}$	$\sim \exp$
Erdos-Renyi	$= \frac{\ln N}{\ln \langle k \rangle}$	$= \frac{\langle k \rangle}{N}$	$= e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$

$N \rightarrow \infty$ is smaller compared to other models. On the other hand, equal degree distribution of WS and ER models does not necessary lead to a considerable similarities between them for arbitrary parameters.

3.3 Epidemic models on networks

3.3.1 Basic SIR model on the network

Analysis of basic SIR model on the randomly generated ER network was performed in order to described the spread of an infection over the population. Figure 7 provides with a representation of the specific network with initial parameters $N = 80, p = 0.1$ used in the model. Output of the SIR model with parameters with parameters $\beta = 0.01, \gamma = 0.005$ and proportion of initially infected nodes $I_0 = 0.02$ and corresponding network are provided in the Figure 7.

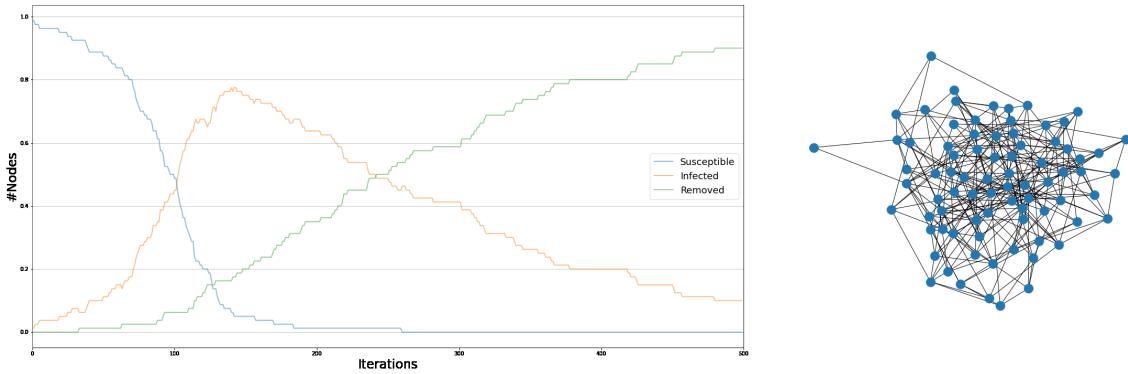


Figure 8: On the left: Dynamics of the SIR model on the network is represented by proportions of Susceptible, Infected and Recovered individuals. On the right: representation of network on which SIR model was performed.

Overall dynamics of the model on networks has the same properties as the classical except the randomness that comes from network generation and represents different connections between individuals.

3.3.2 Practical comparison

This section is dedicated to comparison between epidemic models on BA, WS and ER networks. For all considered models initial parameters was chosen in a way to keep the same number of nodes N and same order of numbers of edges. Possible behaviour of the number of infected nodes (small groups, individuals) is shown on the Figure 8.

For further analysis statistical analysis is used to define average maximum of infected population based on the models' outputs. From Figure??? we can conclude that ER and BA models are considerably similar in terms maximum proportion of infected population. Means of this property for these models are considerably higher compared to WS. This phenomena can be explained with generally lower clustering of WS networks as it was discussed in the Section???.

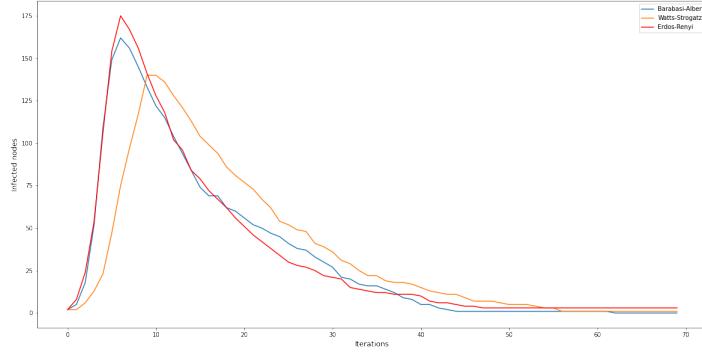


Figure 9: Possible behaviour of SIR model on different networks. As can be seen on the plot, ER and BA models show relatively similar dynamics, while WS model has lower peak and slower epidemic outbreak

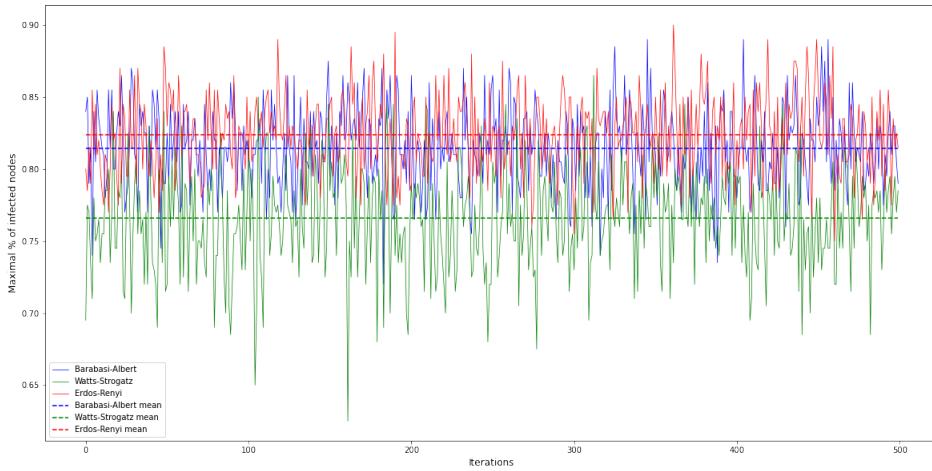


Figure 10: Statistical analysis of SIR models based BA, WS and ER networks. Simulation was iterated 500 times with $N = 500$ nodes and with similar number of edges in each network. Dotted line of the corresponding color shows average maximum of infected population for each model. WS model shows big difference compared to other models with this metric

3.3.3 Influence of selective infection

To access the spread of infection, nodes degrees k were chosen as the main property, influence of which to be determined. In this and further sections we consider only SIR models on BA networks to investigate the effect of selective infection of nodes with highest degrees on the epidemic dynamics. As it is shown in the Figure 10, initial infection of such nodes causes significantly faster growth of number of infected in early behaviour of the system. Lower variance between trajectories with selective infection corresponds to more established and stable initial infection algorithm compared to a random initial infection.

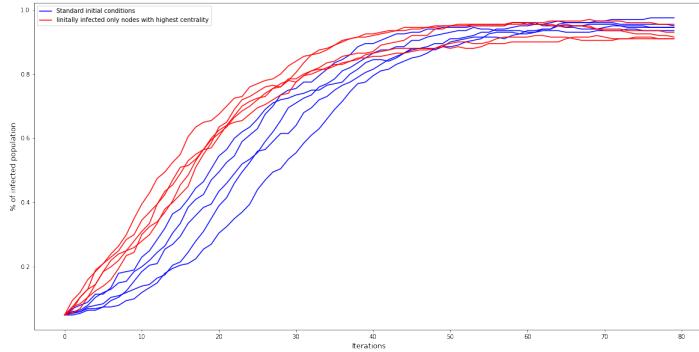


Figure 11: Some trajectories of the SIR model with and without selective infection are provided on the plot. There is a clear separation of trajectories of both cases with significantly higher rate of infection for model with initially infected the most connected nodes. Red curves also show lower variance in the output

3.3.4 Selective vaccination

Based on the idea from the previous Section, our selective vaccination model is designed to remove the most connected nodes from the network to slow down the spread of an epidemic and lower maximum value of infected individuals. Vaccination algorithm can be described in the following steps:

1. Certain number N of nodes m_i with highest degrees are chosen from a previously generated network and represented as a set M .
2. All nodes $m_i \in M$ are removed from the network along with the corresponding edges as such that have acquired lifelong immunity to the disease.

Possible dynamics for different proportions of initially vaccinated population v is shown in the Figure 11. Higher proportion of vaccinated population results in a significant different dynamics of the model.

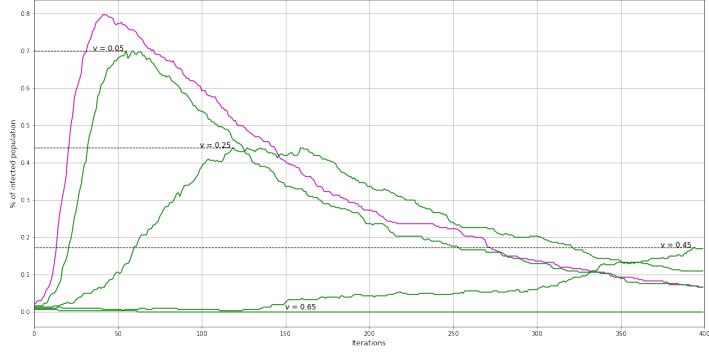


Figure 12: Influence of selective vaccination is represented with proportions of infected population for the BA networks with $N = 300, k = 10$. Blue line represents dynamics without vaccination, while green lines - with vaccination. For this case, vaccination of 45% of population leads to a relatively low maximal number of infected nodes. $v = 0.65$, for this case, leads to practical extinction of the epidemic

In opposition to the previous approach where only several cases of model dynamics was considered, statistical analysis of the vaccination plan was performed, results of which are shown on the Figure 12. This approach makes it possible to examine the effectiveness of the model in more realistic case and propose sufficient proportion of vaccinated population for some specific targeted threshold proportion of infected individuals per time-step. With different purposes and policies, different threshold can be defined. For the simplicity we will assume that if at each time-step not more than half of the population is infected, vaccination plan is regarded as sufficient. With this assumption, analysis of the Figure??? leads to the conclusion that approximately with $v \approx 0.4$ and higher vaccination is sufficient.

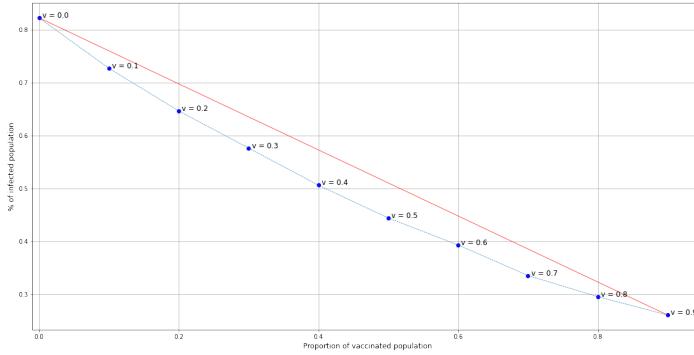


Figure 13: Effectiveness of infection is described with the relation between maximal proportions of infected population and proportions of selectively vaccinated nodes. Simple linearization of this relation was added as the flat red line to express its non-linear nature. As can be seen on the plot, at $v \approx 0.4$ we reach half of the infected population on a certain time-step

4 Appendix

4.1 Network generation

Significant part of this work is dedicated to infection modeling on networks. For creation of these networks based on the randomly generated graphs with programming language Python 3, library NetworkX was used. It provides us with a number of different graph models and methods to work with them. This library was also used for visual representation of networks.

4.2 Basic SIR model simulation

In several sections basic SIR model was performed with Python library NDLib that includes tools for generation of different models with specific parameters and initial condition together with tools for their analysis.

5 Conclusion

All hallmarks of the stochastic SIR models were presented. We managed to identify all the differences between deterministic and event-driven versions of SIR. In particular, we focused on the description of the extinction of endemic diseases. In particular, we managed to identify the necessary factors for the prevention of the disease's extinction. We also compared the behaviour of event-driven models to that of the deterministic models.

Afterwards we investigated, the metapopulation models allowed us to better understand the spatial dynamics of epidemics, especially connected to the speed of spread of the pathogen of the population, and the influence of interactions between multiple subpopulations with different interaction strengths. The disease managed to persist for the longest time in the populations located in the centre of our model. This conclusion, in combination with other results, may suggest that it is necessary for the prevention of epidemics to minimise interaction between subpopulations.

Several scenarios for spread of an infection on networks was studied. In Section 3.3.1 it is shown that SIR model on a network shows similar dynamics compared to standard SIR. Comparison between different network generation models in Section 3.3.2 leads to the conclusion that SIR on BA and ER networks are similar in terms of maximal proportion of infected population and model on WS graphs on average show significantly lower value of this property. This result can be used in further analysis of applications of these models to "real-world" networks. In Section 3.3.3 it was shown that initial infection of nodes with the highest degrees results in a faster epidemic outbreak, higher predictability in trajectories, but eventually leads to a similar late-time dynamics. Particular vaccination plan, provided in Section 3.3.4 can be used to create or verify vaccination policies parameterized on a maximal number of infected individuals.

References

- [1] Exact stochastic simulation of coupled chemical reactions Daniel T. Gillespie, The Journal of Physical Chemistry 1977 81 (25), 2340-2361
- [2] Keeling MJ, Rohani P. Modeling Infectious Diseases in Humans and Animals. Princeton, NJ; Oxford: Princeton University Press (2008).
- [3] Epidemic Processes in complex networks, Reviews of Modern Physics, 87, 925 (2015)
- [4] Réka Albert and Albert-László Barabási Rev. Mod. Phys. 74, 47 – Published 30 January 2002