Information Travel and Epidemics

Group project as part of the seminar:
“Epidemics, Infodemics and Mobility”, SS 2022.

Introduction

Information spreading changes the behavior of individuals and thus a population during a pandemic significantly. In this analysis, we want to study the effect of information spreading on containing an epidemic. We based our work on Funk et al. 2009. They conclude that high-quality information can change human behavior. We extended a basic SIR-Model with an awareness component to replicate their results. The model we created is based on our research questions:

1. "Which conditions can lead to outbreaks being contained by the information spread?"

2. "What is the impact of self-isolation of knowingly infected individuals on the spread of a disease, and under which conditions is it most impactful?"

Based on the study of Funk et al., we assume that information quality decreases after a certain time and the number of individuals passing information on (grapevine effect). We extended the original model with further aspects like the isolation rate. We would like to investigate how the infection changes after an individual are infected and isolate under what circumstances the isolation would be particularly successful.

How our Program is Structured

The core files are **main.ipynb**, **awareness.ipynb**, **simulation.ipynb** and **plot.ipynb**. The file **rk4.ipynb** includes the numerical approach to solving partial differential equations. The folder data stores the results of the executed computations. They provide the results used in the model's diagrams. The **"figures" folder** holds the created figures, while the "**model" folder** contains custom-made figures describing the workings of the model.

You can download the final code from the Github Repository at the following link: **https://github.com/Mela/Information-Travel-and-Pandemics**

The file requirements.txt lists the requirements for the setup. Those will be installed automatically if the program runs in the deepnote environment. If the program runs in another environment, you need to run the "pip install" command.

Reading Order

Our suggested reading order for the different jupyter notebooks is as follows:

1. **documentation.ipynb** for the introduction, explanation of the model and the parameters
2. **main.ipynb** for the main functions
3. **rk4.ipynb** for the algorithm to solve partial differential equations (PDEs) and a comparison between Runge-Kutta of the 4th order (rk4) and the Euler forward step method
4. optional: **example_heat.ipynb** for a simple example using rk4 and a model for the heat equation
5. **awareness.ipynb** is the file where SIR and awareness are modeled
6. **analysis.ipynb** is the file for running and evaluating all of the simulations.

Our Team

- **Lisa Beckmann**, Georg-August-University of Göttingen, Physics MSc.
- **Melanie Eckenfels**, Karlsruhe Institute of Technology, European Culture and History of Ideas MA.
- **Nathanael Gutknecht**, Karlsruhe Institute of Technology, Physics MSc.
- **Zuzanna Mielczarek**, University of Lübeck, Media Informatics MA.

The Model

Note: Depending on the situation, the term "information" or "awareness" is used. However, in our model, they are used interchangeably.

We chose a basic SIR model to model the spread of the epidemic (**see Fig. 1**). The rate for susceptibility to infection is modified by the decay constant ρ and the parameter i, which describes the awareness, with $(1-\rho^i)$. It is to note that $i = 0$ corresponds to first-hand information and, therefore, the highest awareness.

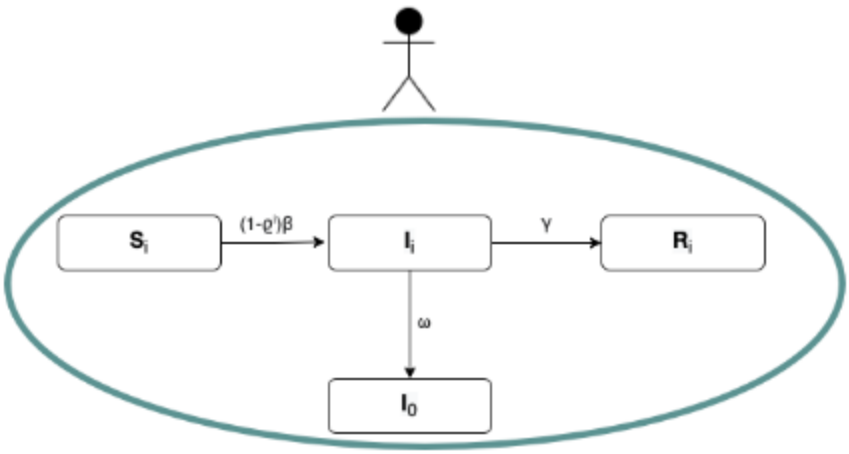


Fig. 1: Flowchart of the SIR model. The boxes describe individuals having one of three possible conditions: susceptible, infected, or recovered. The arrows show the transition rates.

Information quality is characterized by the number of times it was passed on until arriving at individual X (**see Fig. 2**). Thus, the lower i is, the better the information quality and the less susceptible an individual is. Information is generated by infected individuals who also can notice their infection with a rate of ω. Information transmission is achieved by interactions with the neighbors with a rate of α, but the information quality wanes by $i+1$ for the receiving individual. The waning of information quality by time is modeled with a rate of λ.

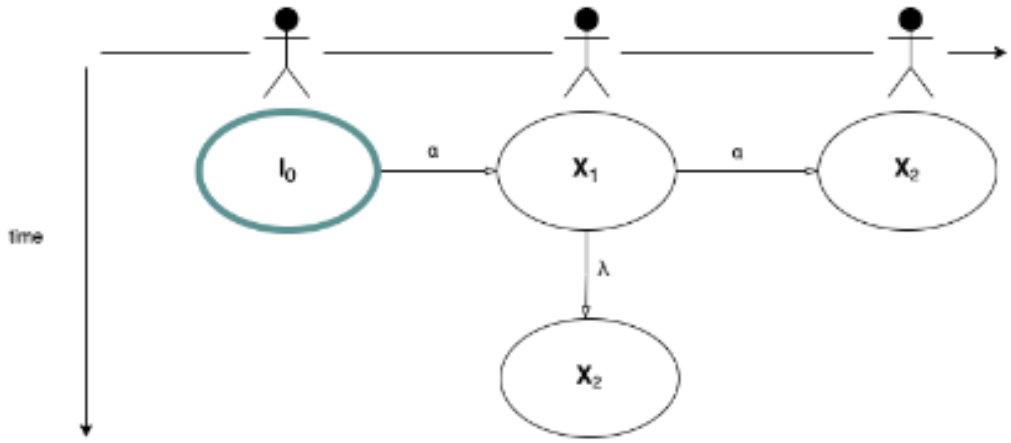


Fig. 2: Flowchart of the awareness generation. The boxes describe the individuals, and the arrows represent the information transmission/fading with time. The index describes the grade of awareness.

An overview of our model's parameters:

Parameter	Description
β	Normal infection rate
γ	Recovery rate
ρ	Decay constant
ω	Information generation rate
α	Information transmission rate
λ	Information fading rate
i	Awareness
κ	Self-isolation rate

To simplify the model and to match the individual-based analysis described by Funk et al. 2009, we chose to model the population on a 2D-grid with each point having four neighbors.

Extension of the Model

In the original model described above, only susceptible individuals acted on the information they received. We extended the model with a self-isolation rate of knowingly infected individuals of **κ**. This way, the infected individuals also acted on information (although generated by themselves).

Results

Unless otherwise noted, these default parameters are assumed for all simulations:

Parameter	Default Value	Description
β:	0.03	normal infection rate
γ:	0.03	recovery rate
ω:	0.06	information generation rate
α:	0.06	information transmission rate
λ:	0.05	information fading rate
ρ:	0.6	decay rate for information quality
κ:	0.7	self-isolation probability
gs:	30x30	grid size
Δt:	1	time step

Awareness with no new Infections

First, we had a look at how the awareness spreads and decays in our network when there are no new infections present, which could add more information/awareness to the system.

For this, we placed all the individuals in a state of no-awareness and then moved one individual to the compartment with the highest awareness (awareness level index 0). The average amount of awareness in the population is calculated using

$$\text{awareness} = \langle \sum_{i=0}^{N_c-2} P(S_i) \cdot \rho^i \rangle,$$

where $P(Si)$ is the probability of an individual being in the i-th compartment and ρ is the decay constant for transmissions.

The last compartment has zero awareness and therefore is not included in the sum. The average is taken over the grid. We chose the grid large enough so that the awareness cannot reach the edge and is therefore not cut off. The resulting plot is shown in **Figure 3**. **Figure 4** shows the awareness and susceptibility associated with each awareness compartment.

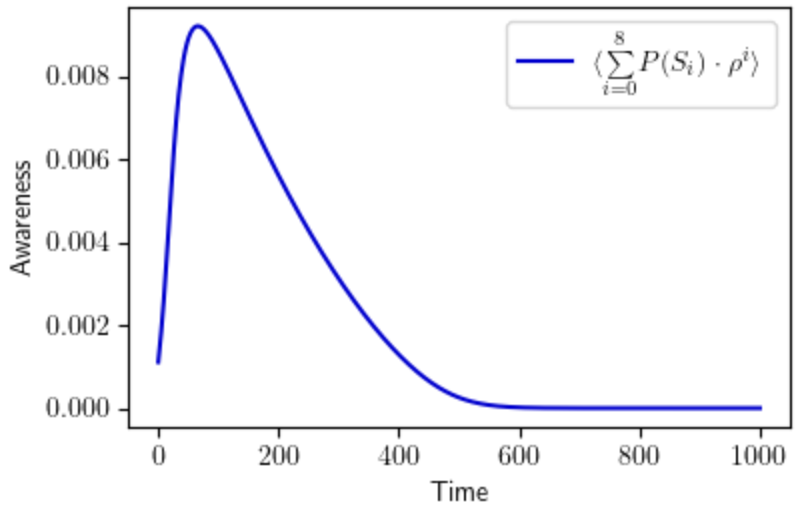


Fig. 3: Average awareness on a 30x30 grid over time. In the beginning, the awareness quickly spreads throughout the network. Then the awareness slowly fades over time until there is basically none left. Also, compare **Figure 1** from Funk et al. 2009.

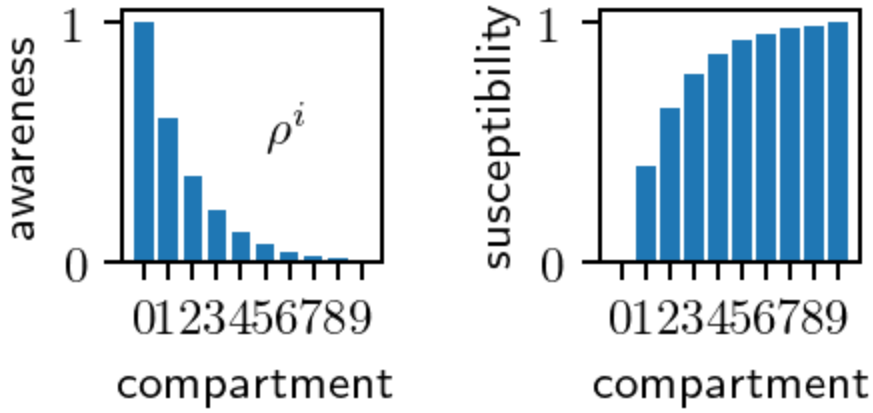


Fig. 4: Awareness and susceptibility for each awareness compartment.

The Impact of Awareness on an Outbreak

An Outbreak without Awareness

As a baseline, we simulated an outbreak without awareness in the population. The infections moved outward in a circular shape until nearly everyone was in the recovered compartment in the end. **Figure 5** shows the results.

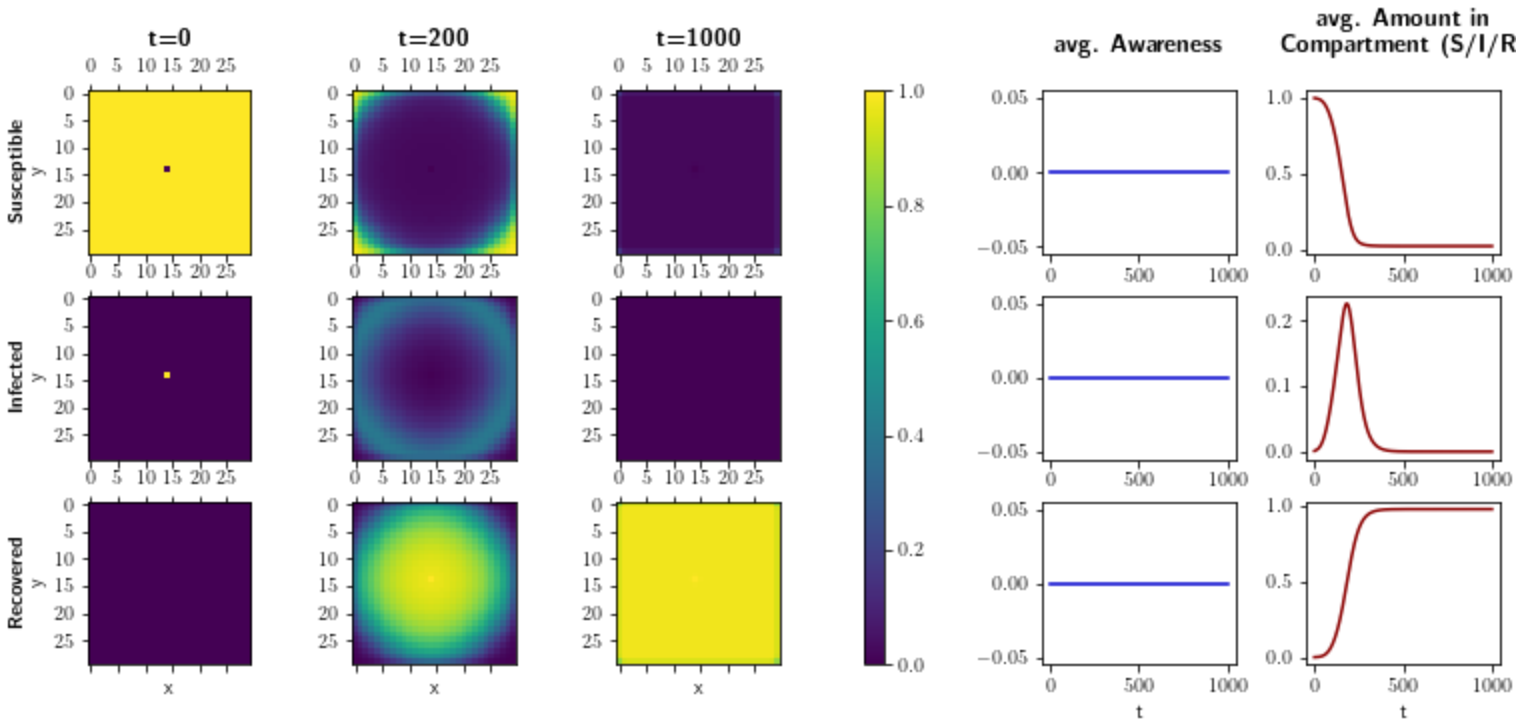


Fig. 5: Outbreak with no awareness. The outbreak is not contained and in the end, the probability of having been infected at some point in time is close to one.

An Outbreak with Awareness

As the next step, we ran the simulation with the same initial conditions. But this time, awareness was generated from infections, which spread throughout the network and reduced susceptibility. **Figure 6** shows the results.

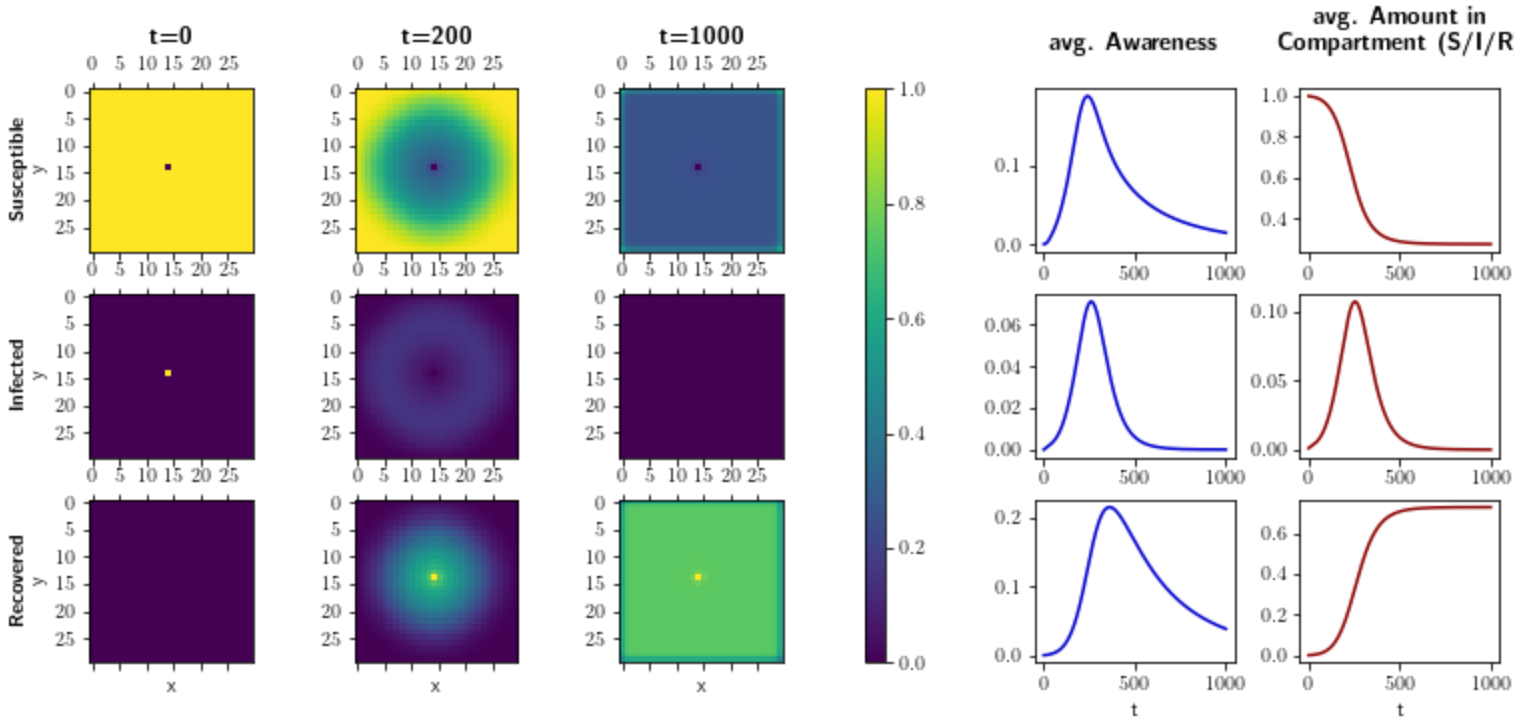


Fig. 6: Outbreak with awareness. The outbreak is reduced both in speed and intensity.

Comparison

When comparing the two scenarios, one can notice the awareness has reduced the total number of infections and the maximum number of simultaneous infections, as shown in **Figure 7**.

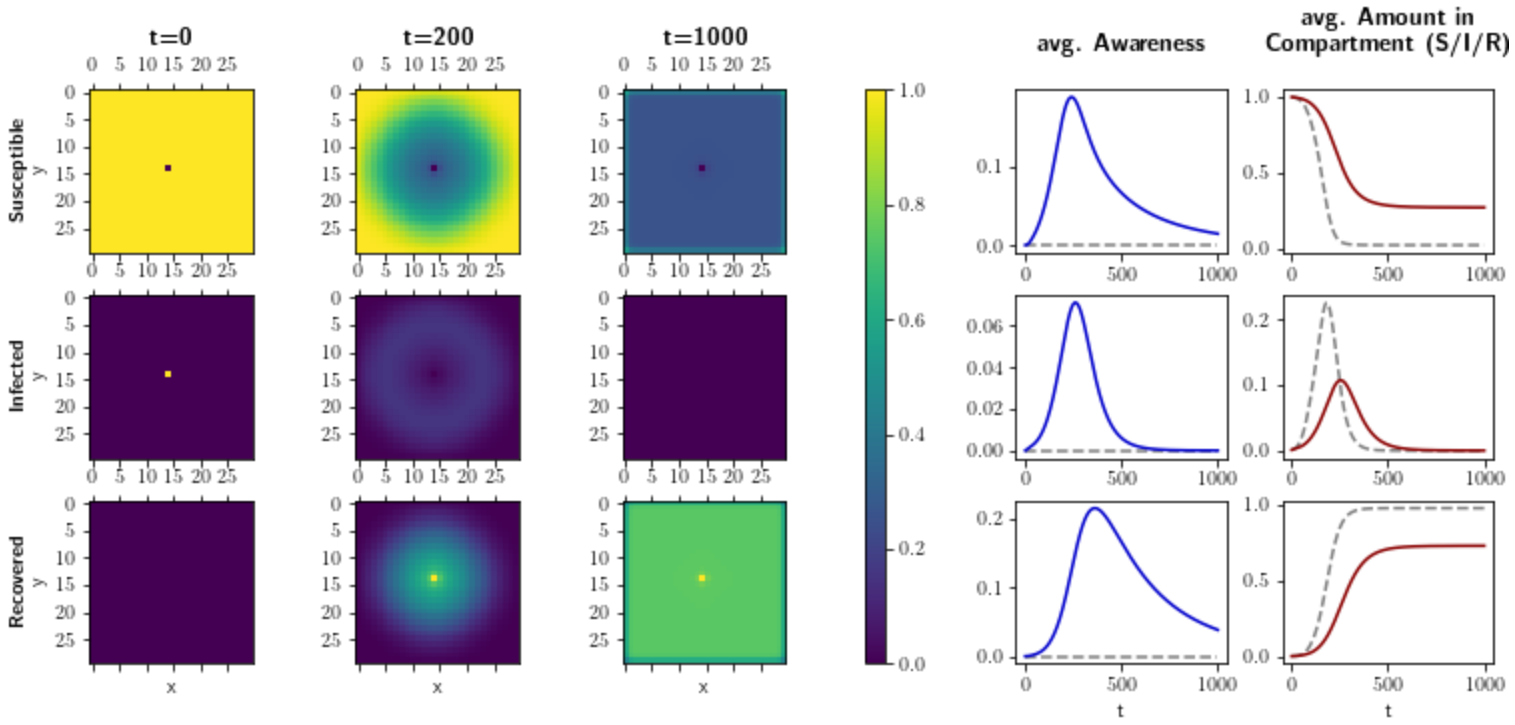


Fig. 7: Comparison of an outbreak with awareness (solid line) and without awareness (dashed line, gray). Here, the awareness reduces both the speed and the intensity.

Long Term Stability

In our model, which uses a superposition of SIR and information states for each individual, an outbreak won't completely die out just by chance. Instead, there will always be a small portion of infection remaining in the system. Thus, a second outbreak might occur if the awareness level dropped to zero.

A stable condition without awareness is reached when the rate of new infections is less or equal to the rate of recovery with no awareness being present. This happens when a certain number of people have been removed from the susceptible compartment. For our model, we used $\beta = \gamma = 0.03$ this means that ¾ of the population has to be removed from the susceptible pool, as each individual has four neighbors that they can infect with a rate of β. This corresponds to a mean-field SIR with $\beta = 0.12$, $\gamma = 0.03 \Rightarrow R\_0 = \beta/\gamma = 4$.

However, both in the mean-field approach and in our grid-based model, the outbreak will lead to a much higher amount of infected people than the stability threshold requires. Because when the threshold for stability is reached ($R\_eff = 1$), there are still many infections present in the population, which still generate new cases, even when the effective reproduction number is below one. Therefore, more than 75 % of people might become infected over time.

The awareness in our model acts against overshooting the stability threshold. For our default parameters, the awareness reduces the number of recovered ones at $t = 1000$ to about 73 %, with some awareness remaining in the susceptible population. That will slow down any further spread while being below the threshold. (Edge effects might slightly reduce the threshold.)

When the stability threshold is reached, no further outbreaks will occur even when the awareness has faded (**Figure 8**).

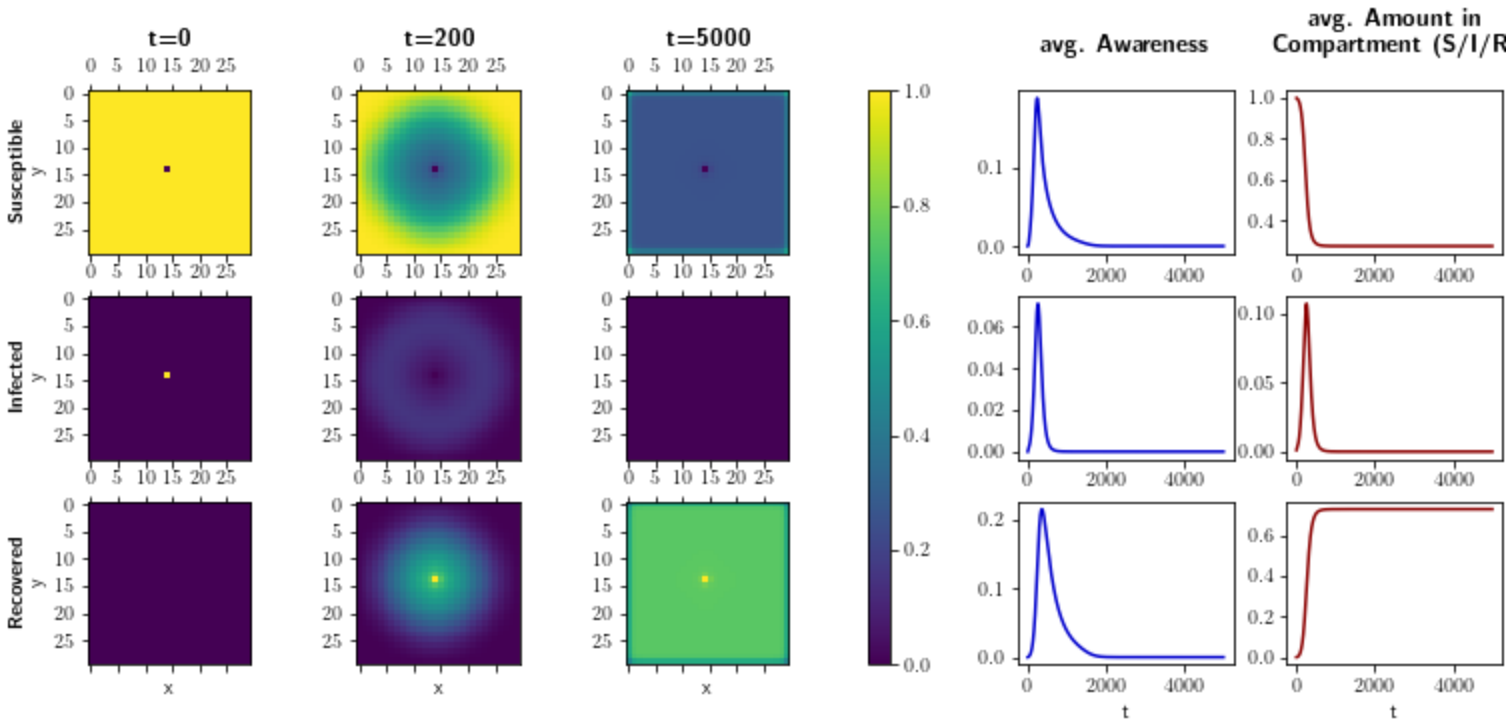


Fig. 8: Long-term stability when 1 - 1/R_0 have been infected.

When using more aggressive parameters for awareness, the outbreak looks contained at an even lower value than the threshold. In that case, the awareness in the susceptible population will stay high so that the further spread of infections is kept at a minimal rate (as depicted in **Figure 9**).

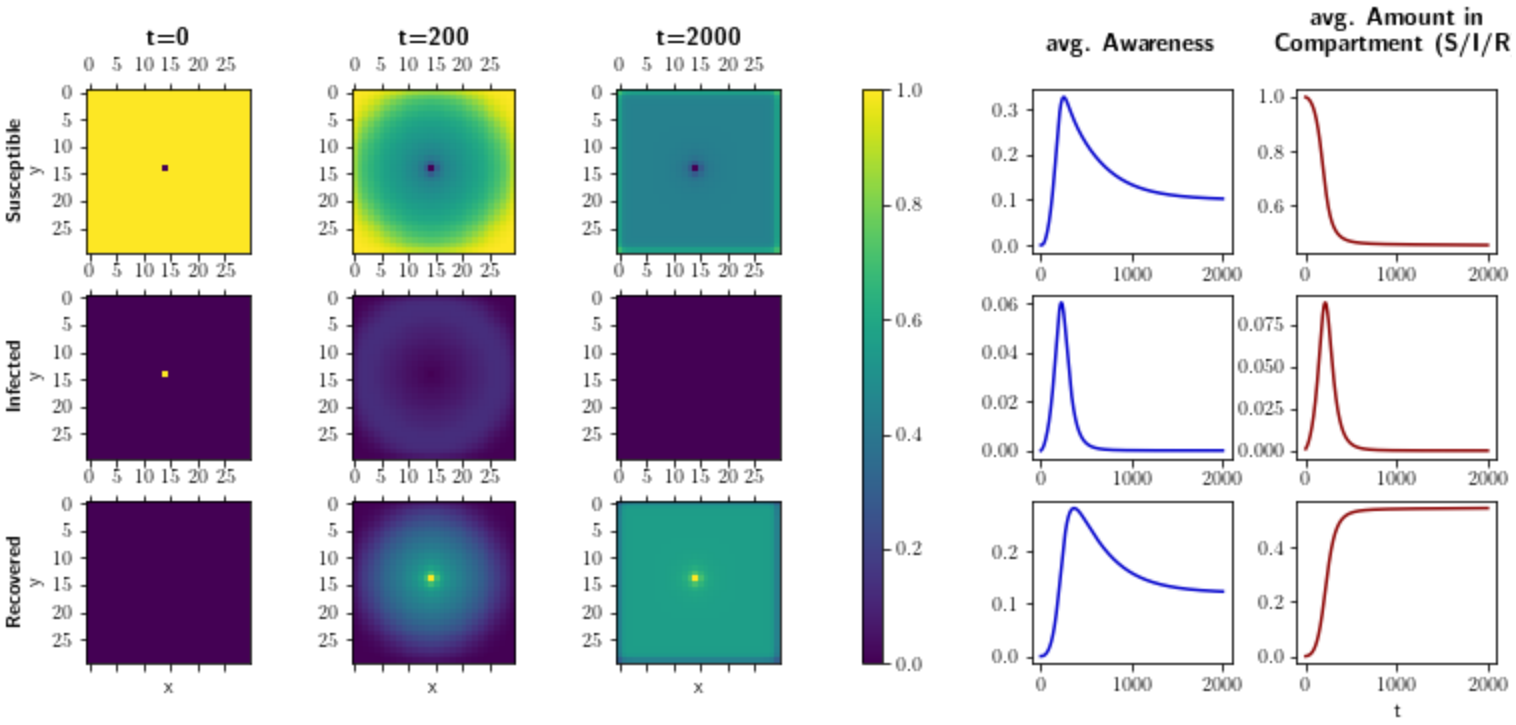


Fig. 9: Long-term stability when 1 - 1/R_0 have been infected.

However, if the awareness suddenly drops during that time, while the threshold has not been reached, a second wave is possible, as **Figure 10** shows.

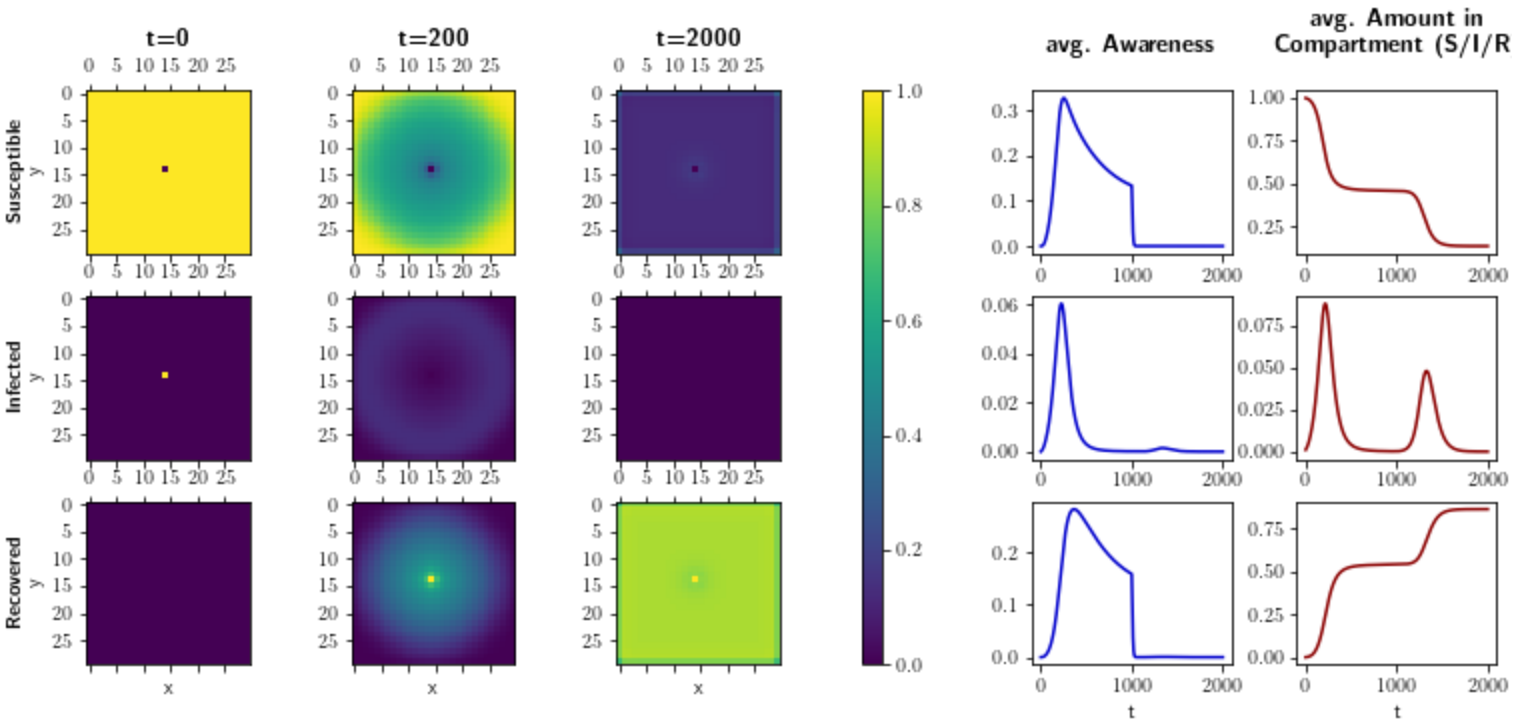


Fig. 10: Long term stability when a total of 1 - 1/R_0 have been infected.

Sensitivity Analysis

To understand better how much the different parameters influence the outcome of our system, we ran a sensitivity analysis on our model. Using our default parameters as a point of reference, we took one parameter and adjusted it to see how much the outcome would change. We repeated this for all parameters. Since the model is always assumed to reach at least approximately 75 % recovered individuals (in the (very) long run) and because the time of the end of the initial wave is influenced by our parameters, we chose the maximum number of simultaneously infected people to be the measure for the effectiveness of the awareness. **Figure 11** shows the results from the sensitivity analysis.

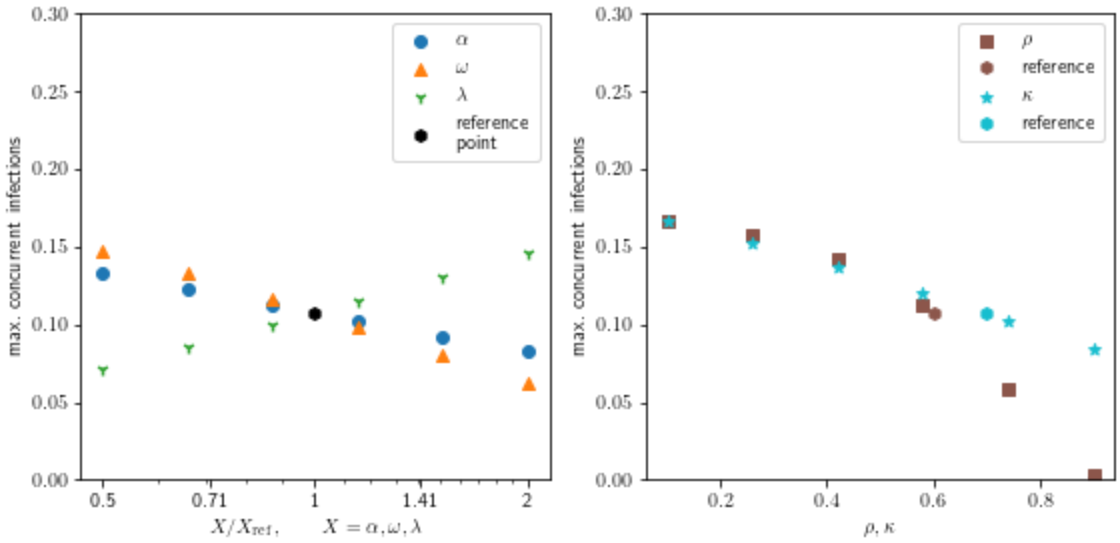


Fig. 11: Sensitivity analysis around the default values.

Before drawing any conclusions about real-world epidemics, one should remember that the sensitivity analysis depends on the chosen reference point (compare **Figures 12 and 13**). (The gradient varies depending on where you are in the parameter space.) Therefore, the default values chosen for our model do not necessarily reflect a realistic view of the world. Instead, we selected them such that the different effects of the model become visible.

Furthermore, one should remember that the effectiveness of the chosen parameters heavily depends on the assumptions of the model, which might not represent reality accurately (square lattice, etc.).

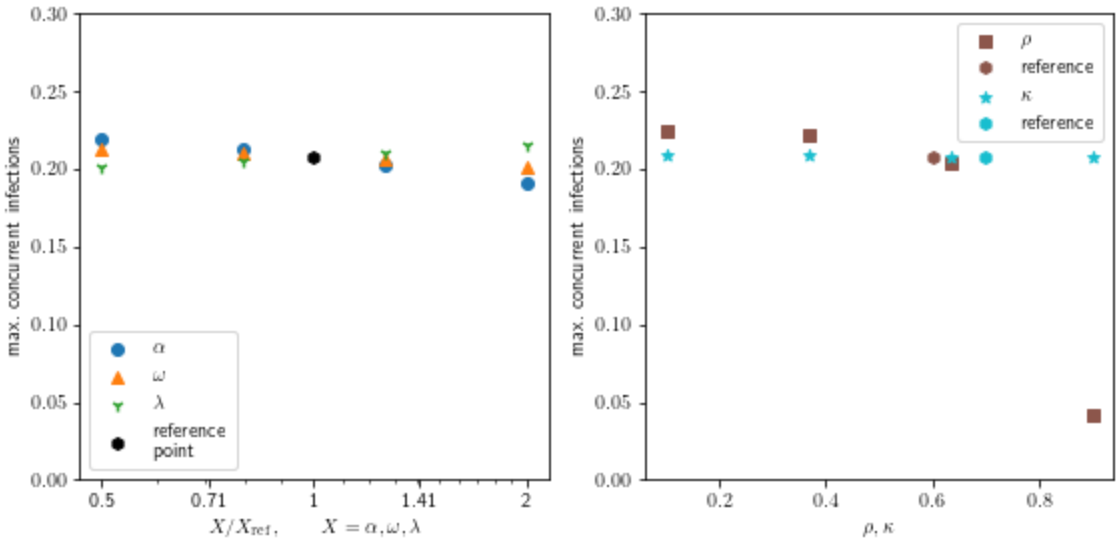


Fig. 12: Sensitivity analysis in a scenario where the spread is dominated by a fast transmission of information (α = 0.1) while the information generates at a slow rate (ω = 0.001). Therefore self-isolation only has little impact as the spreading of the virus is much faster than the virus is being detected.

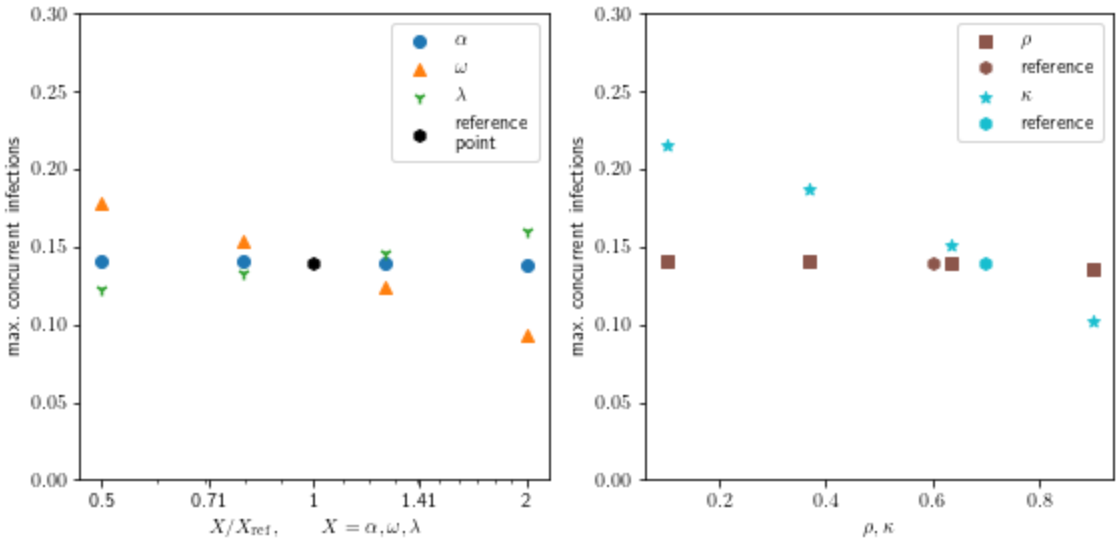


Fig. 13: Sensitivity analysis in a scenario where the transmission of information is very slow (α = 0.001), but the detection rate for generating new information is large (ω = 0.1), which leads to self-isolation having a much more significant impact.

Particular Features of Our Model

Note: This section has not been part of our regular group work but came about during debugging and trying to understand why our model behaves in that specific way. It led to some backup slides of our presentation in case of more detailed questions about our model. As the findings give some interesting insights into the workings of our model, they are included here. Furthermore, it shows how important it is to carefully consider which aspects are to put into a model.

We used a square lattice to reflect a straightforward, local contact structure when building our model. This distinguishes our model from a mean-field approach, where everybody is connected to everybody. As another simplification, we assumed the contact structure for spreading infections and awareness to be the same.

Instead of using a Monte Carlo simulation, we decided to use the expectation values for the interactions and a compartmentalized model, where each individual on the grid is in a superposition of different states.

We decided to have $3 \cdot 10 = 30$ compartments for each individual (3 for SIR, 10 for awareness) to reflect the correlations between awareness and the SIR state. For example, individual A might have a high awareness probability because individual A is partly infected. But also, awareness should not reduce individual A's susceptibility for the part of individual A that is still susceptible.

To keep things simple and to reduce computation time, we do not respect correlations between the SIR states of neighbors. This, however, means that individual A could infect their neighbor (individual B). Individual B, in turn, can increase individual A's probability of infection.

In that way, the local behavior of our model is similar to a mean-field approach.

To better understand the differences between a mean-field approach, a Monte Carlo simulation on a grid (which automatically respects correlations between neighbors), and our superposition model, we wrote a quick Monte Carlo simulation and a mean-field simulation for a SIR model (without awareness; the models can be found in model_inspection.ipynb). For better comparison, the Runge-Kutta method is dropped, and the Euler forward method is used.

Running the simulation with our default parameters, we saw a huge difference in the results from the Monte Carlo method compared to the mean-field analysis and our model. The outbreak from the Monte Carlo simulation happens considerably slower and has a real chance of dying out (see **Figure 14**). Because the infections there either happen (with a set probability) or they don't happen (instead of using rates and averages), each individual effectively loses one neighbor to pass on the virus (the one that passed on the infection). Also, people in the same area often share their neighbors, thus further reducing the number of effective contacts.

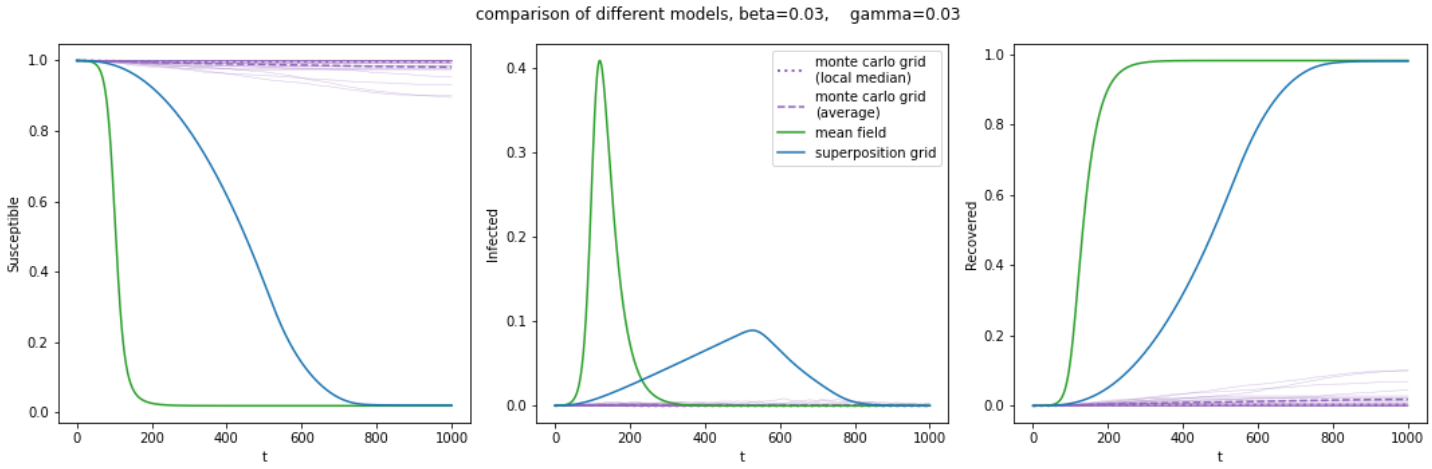


Fig. 14: Comparison between mean-field approach (green), our model (blue), and a Monte Carlo simulation (purple) on a square lattice with size 100x100 for a SIR model (no information/awareness). $\beta = 0.03$, $\gamma = 0.03$. Our model leads to a linear increase in active cases because the circular pattern moves outwards. The Monte Carlo simulation (25 passes) leads to a much slower spread as our model does not respect the correlations between neighbors. Here, the Monte Carlo simulation can freeze out at low numbers and thus significantly reduce the total number of infections.

When increasing the spreading rate β of the infections, the Monte Carlo method results in more infections and is closer to the mean field analysis (**Figure 15**).

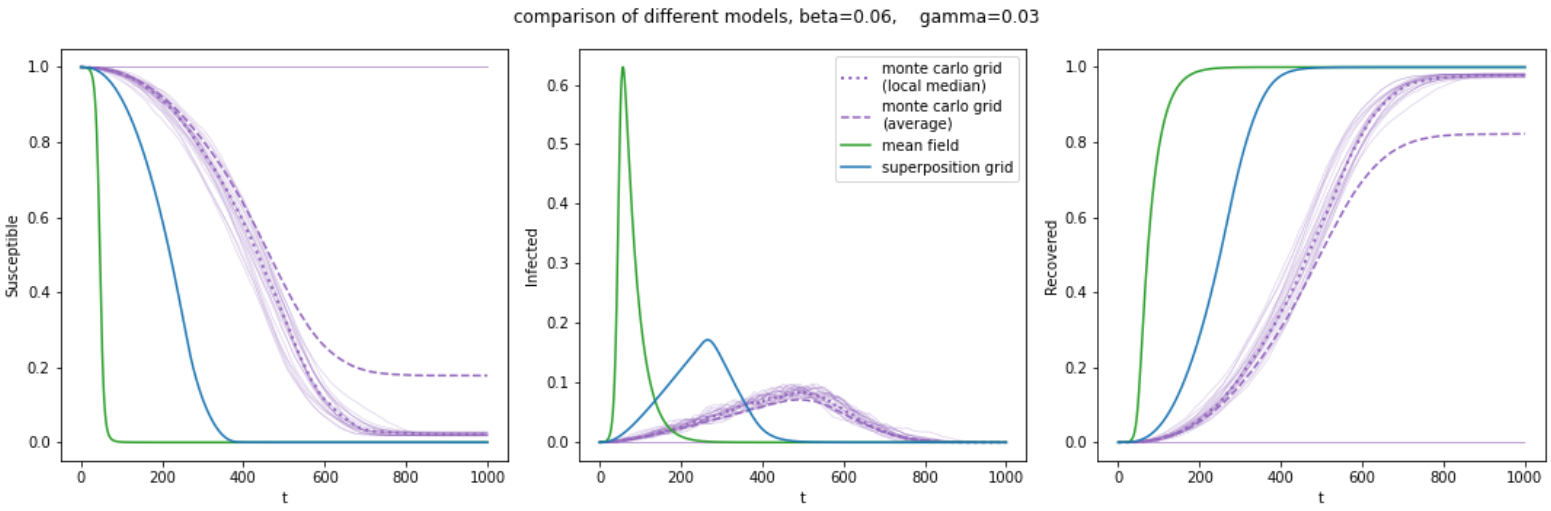


Fig. 15: Comparison between mean-field approach (green), our model (blue), and a Monte Carlo simulation (purple) on a square lattice with size 100x100 for a SIR model (no information/awareness). β = 0.06, γ = 0.03. Our model leads to a linear increase in active cases because the circular pattern moves outwards. The Monte Carlo simulation (25 passes) leads to a much slower spread as our model does not respect the correlations between neighbors.

To prevent infections from bouncing between neighbors, a 1D chain is simulated where the infections can only be passed on in one direction, and the recovery rate is set to zero. Plotting the local spread at different times shows the speed of the infection wave traveling outward. A clear difference can be seen when comparing our superposition approach to a Monte Carlo simulation (**Figure 16**).

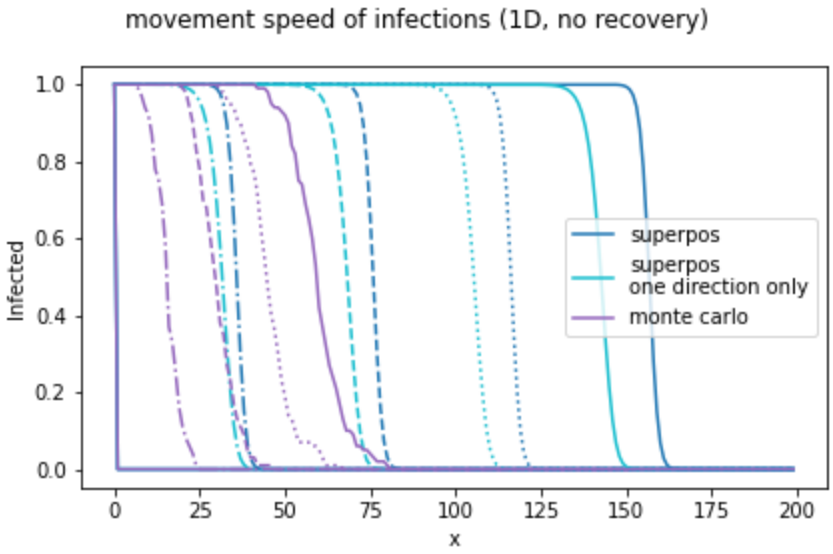


Fig. 16: Movement speed for a 1D chain of people. At t = 0 only the individual at x = 0 is infected. The figure shows the infected people at different time stamps (t = 250, 500, 750, 1000) with line styles: (mixed, dashed, dotted, solid). Blue: Our model. Cyan: A restricted model where infections can only travel from left to right. Purple: A Monte Carlo Simulation with 100 passes. $\beta = 0.06$, $\gamma = 0$. We can see that our model leads to a much faster spread than the Monte Carlo Method.

If recovery is turned on, the infections along a 1D chain will quickly decay along the distance, as shown in Figure 17. However, for a wavefront in 2D, people can also infect each other laterally. Thus, the infection wave will not necessarily decay with distance.

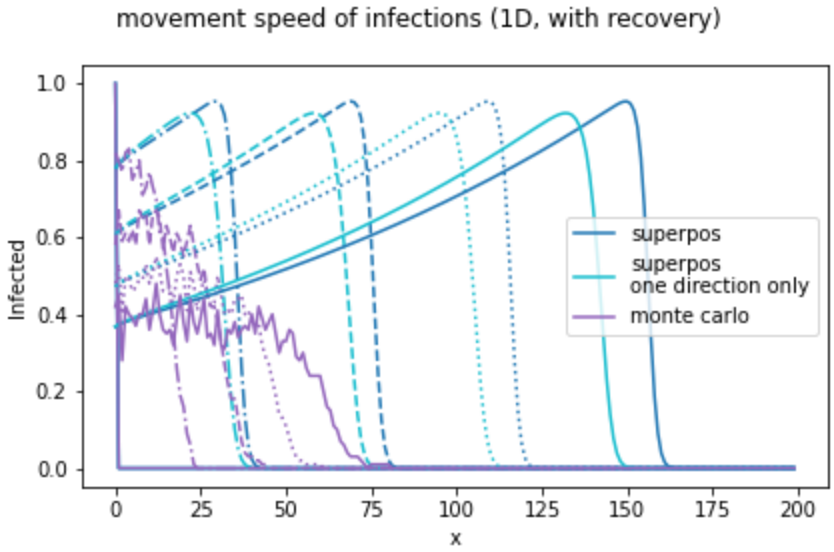
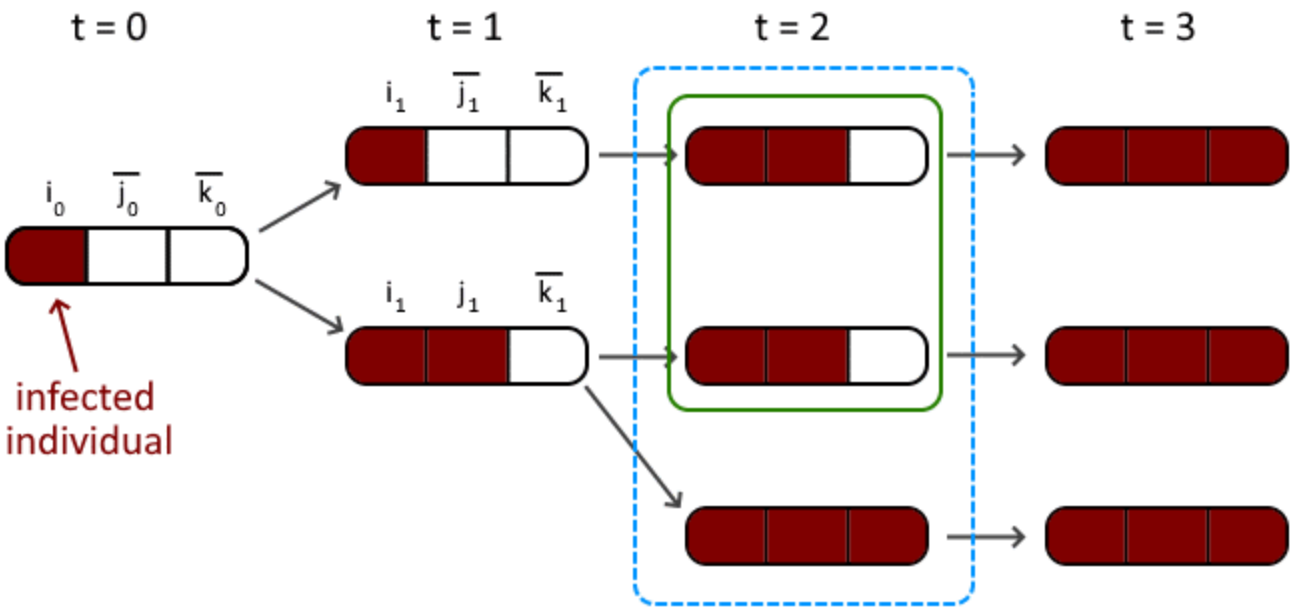


Fig. 17: Movement speed for a 1D chain of people. At t = 0 only the individual at x = 0 is infected. The figure shows the infected people at different time stamps (t = 250, 500, 750, 1000) with line styles: (mixed, dashed, dotted, solid).
Blue: Our model. Cyan: A restricted model where infections can only travel from left to right. Purple: A Monte Carlo Simulation with 100 passes. $\beta = 0.06$, $\gamma = 0.001$. We can see that our model leads to a much faster spread than the Monte Carlo Method.
The recovery does not limit the speed of the spread, but it can heavily reduce the reach as there is always a chance of the infection not being passed on.

The difference between the results from our superposition approach and the Monte Carlo simulation can be explained by the negligence of correlations between neighbors in our approach as indicated in **Figure 18**.



$$
\begin{aligned}
P(k_3) &= P(k_2) + \boxed{P(k_3|\overline{k_2})} \cdot P(\overline{k_2}) \\
&= P(k_2) + \boxed{\beta\Delta t \cdot P(j_2 \cap \overline{k_2})} \cdot P(\overline{k_2}) \\
&\leq P(k_2) + \boxed{\beta\Delta t \cdot P(j_2)} \cdot P(\overline{k_2})
\end{aligned}
$$

Fig. 18: Simplifications in the model (correlations between neighbors are not respected) lead to a larger and faster epidemic outbreak than in a Monte Carlo simulation. The example shown is a chain of three people, with the first one being infected. Blue (dashed) line: assumptions used in the model. Green (solid) line: expected behavior corresponding to a Monte Carlo simulation.

Conclusion

1. **Which conditions can help to contain outbreaks through information spreading?**
   A high decay rate (ϱ) can "freeze" the outbreak if awareness is maintained for long.

2. **What is the impact of self-isolation of knowingly infected individuals on the spread of disease, and under which conditions is it most impactful?**
   Self-isolation will help limit the spread of the disease, yet it is harder to discern the ideal conditions.

The combination of parameters we chose for our model can lead to the containment of the disease. Self-isolation will result in a decline in infections.

Reflection

We think our model can be used to deepen the understanding of how information influences infection patterns.

Given more time to expand or enhance the model, it would have been interesting to add — for example — some nodes dispersing higher quality information. Those could represent influencers in the community like sports coaches, teachers, social workers, and cultural or spiritual leaders. Another approach could have included clusters with individuals that will not be reached by information using traditional methods. For example, people highly influenced by anti-vaccine disinformation (anti-vaxxers) or people lacking resources to deal with the anxiety that comes with the uncertainties of a pandemic and who tend to tune out information — called 'blunters'.

As a final worthwhile attempt, we brainstormed ways to expand the model to include the technique called "pre-bunking", meaning the dissemination of high-quality information before the first infection in a community.

Limitations

Due to time constraints, we had to maintain a very basic model, lacking more complex approaches. The grid approach of four contacts per individual won't represent any community's communication and infection pattern in a realistic fashion. We also had to limit the community size to deal with scant computing resources. The percentual approach would have run indefinitely, a non-realistic outcome for a pandemic. Also, the negligence of neighbor correlations leads to deviations compared to a Monte Carlo Simulation. Randomizations would have led to more accurate results.