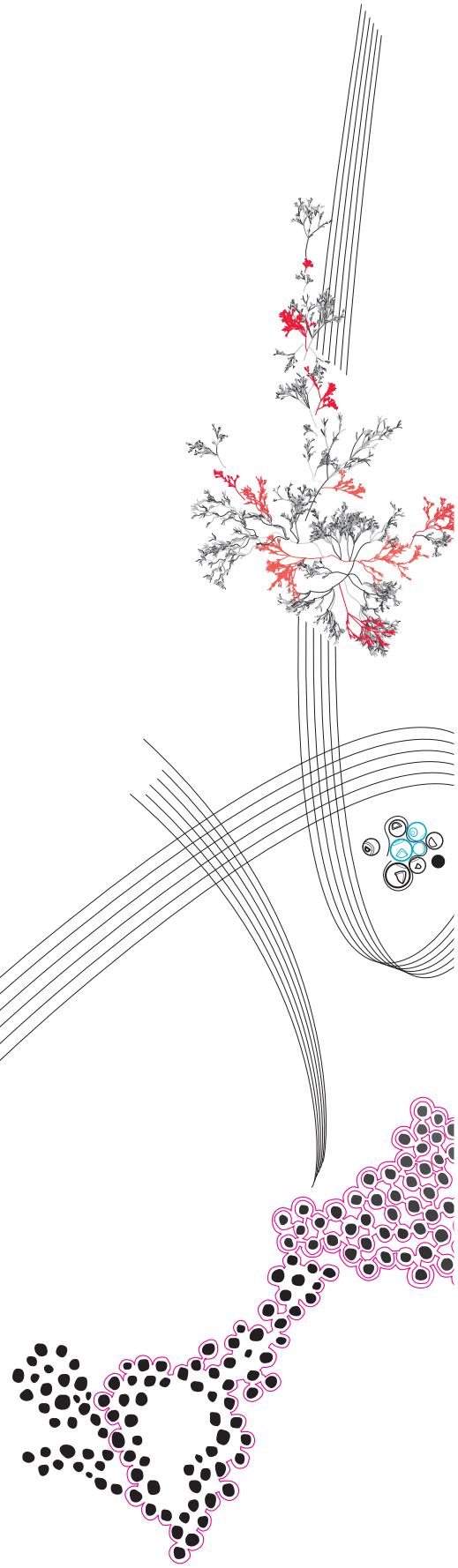

MSc Thesis Applied Mathematics

Effects of Behaviour Adaptation on the Spread of Infectious Disease on Networks with Community Structures

Ioannis Linardos

Supervisors: Nelly Litvak & Petter Holme

October 18, 2021

Department of Applied Mathematics
Faculty of Electrical Engineering,
Mathematics and Computer Science

UNIVERSITY OF TWENTE.

Preface

The present work was executed as a Master's thesis, the culmination of a two-year Master's degree programme in Applied Mathematics with a specialization in Mathematics of Data Science at the University of Twente. As the project's theme suggests, it was primarily inspired by the global scientific effort to mitigate the effects of the COVID-19 pandemic; a crisis that is still unfolding at the time of this writing, hopefully being in its last stages.

This pandemic provoked significant disruptions in the way we were used to working and presented us with new obstacles to overcome. In a time of physical isolation, I would like to express my gratitude to my friends and family for the much-needed moral and psychological support that they offered me. Furthermore, I would like to sincerely thank the academic and administrative staff of the University of Twente, who managed to keep the institution up and running without discounts on the quality of work despite the crisis.

Moreover, I would like to thank my supervisors, Nelly Litvak and Petter Holme, the latter joining from far-away Japan, as well as Clara Stegehuis, who was my stand-by supervisor in a time of need. All of them helped form the project and offered valuable insight drawn from their respective expertise, without which I would not be able to carry it to completion.

Last but not least, I would like to thank the other members of the graduation committee, Maria Vlasiou and Maurits de Graaf, who took time away from their busy schedules to read and evaluate my thesis.

Abstract

The study investigates the influence of awareness in the spread of infectious disease on networks with community structures. Awareness is defined as the reaction of individuals to the presence of infections in their environment. As the number of infections rises, they adjust their behaviour so that the probability of becoming infected decreases. We distinguish between local, community and global awareness, that is, awareness of the number of infected among one's direct neighbours (local awareness), one's community (community awareness) and the entire network (global awareness). The impact of awareness is studied on an SIS epidemic model using stochastic simulations and the mean-field approach. The results are reported for two major characteristics of an epidemic: the epidemic prevalence and the epidemic threshold. As expected, each of these three types of awareness reduces the epidemic prevalence. Interestingly, the epidemic threshold is lowered only by the local awareness and possibly by the community awareness when the communities are small.

Keywords: network epidemiology, infection awareness, community structures, Hierarchical Configuration Model, SIS model

Contents

1 List of Abbreviations

Abbreviation	Meaning
CM	Configuration Model
COVID-19	Coronavirus Disease 2019
DBMF	Degree-Based Mean-Field
HCM	Hierarchical Configuration Model
IBMF	Individual-Based Mean-Field
MF	Mean-Field
MSE	Mean Squared Error
ODE	Ordinary Differential Equation
QS	Quasi-Stationary
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus-2
SEIR	Susceptible-Exposed-Infected-Recovered
SEIS	Susceptible-Exposed-Infected-Susceptible
SIR	Susceptible-Infected-Recovered
SIRS	Susceptible-Infected-Recovered-Susceptible
SIS	Susceptible-Infected-Susceptible
SS	Steady-State
UAU	Unaware-Aware-Unaware

2 Introduction

First identified in late 2019 in the Wuhan province of China, the virus SARS-CoV-2, which causes the disease COVID-19, has spread worldwide, infecting and killing millions and causing significant and unprecedented disruption on the daily lives of people around the globe. The reaction of the scientific community to the virus was equally unprecedented on multiple fronts; treatment protocols and vaccines were developed on record time, and mathematical models were brought to the spotlight, informing political decision-making on non-pharmaceutical interventions [4] [31]. The present master's assignment comes to contribute to this growing literature on epidemiological modelling.

Until the development and full deployment of vaccination programmes, our primary weapon against the virus was behaviour modification. Similar to other respiratory infections, SARS-CoV-2 primarily spreads through social contacts. Consequently, behaviours that decrease the number or the form of social contact can effectively lower the transmission rate. According to the World Health Organization, such behaviours include social distancing, increased attention to personal hygiene and mask-wearing, as well as quarantining in case of observing related symptoms [27]. The importance of behaviour modification is highlighted by the evolution of the pandemic which seems to indicate that behaviour, either self-adjusted or imposed through government lockdowns, is one of the primary driving factors of upward and downward trends in the infection rates. Moreover, it has been observed that past epidemic outbreaks caused behaviour adjustments depending on the risk perception of the population [7]. Thus, examining the interaction between perception-based behaviour modification and infection spread is paramount to understand the state of the unfolding pandemic and gain valuable insights on curbing the development of similar infections that have plagued humanity since the dawn of time and will probably continue reappearing in the future.

The present investigation is not specific to the ongoing (as of 2021) pandemic. Instead, it is an abstract investigation of the spread of infections in contact networks. Various infections spread in this manner, such as sexual contact in the case of venereal disease or social contact in the case of respiratory diseases. The contact networks on which infections spread can be modelled as a graph on which vertices represent individuals and edges represent the contacts through which the infection can be transmitted. It has been observed that the topology of the network plays a significant role in the outbreak's evolution. The study of the spread of epidemics on networks has given rise to the field of network epidemiology [19].

Real-life social networks tend to exhibit community structures in which certain individuals interact with other members of their community with a higher frequency than with members of other communities [38]. Furthermore, individuals tend to be aware and react to the presence of the infection on the community level, and individuals belonging to the same community tend to have similar risk perception [21]. Lastly, to the best of the author's knowledge, epidemic spreading in networks with community structures has not been examined adequately, especially regarding risk perception. Thus, in the present study, we focused our investigations on such networks.

The main research question that we set to investigate was *"What is the influence of awareness on the spread of infectious disease on networks with community structures?"*

To answer this question, we needed to model the network's community structure, the spread of the disease, and the awareness-based behaviour adjustment. For the first, we used the Hierarchical Configuration Model (HCM) [38], a model that allows us to mathematically describe as well as generate instances of random networks with community

structures. For the second, we applied the network SIS model [29], which is the simplest model of epidemic spread. By applying a simple model, we were able to focus on the effects of the variable of interest, which was awareness. The modelling of awareness is a more complicated matter because it is essentially an attempt to quantify a psychological and sociological phenomenon. The awareness was divided into local, community and global awareness, calculated proportionally to the prevalence of infection in the corresponding regions. The effects of awareness were limited on a decrease of the infection rate.

In the field of network epidemiology, the study of the development of an outbreak on a network can be conducted in two ways, both of which were deployed in the duration of this thesis: compartmental models and agent-based models [12]. Compartmental models rely on separating the population in compartments and modelling the process using systems of differential equations that describe the evolution of the disease. Agent-based models employ simulations that follow the interactions of agents on the individual level. In simulations, specific instances of the outbreak were realized in specific networks, while the stochastic nature of the process causes the events in each simulation to be unique. However, stochastic analysis allows us to derive a set of differential equations, also called master equations, that describe the approximate time evolution of the system on a large scale. In this way, an agent-based model is approximated using a compartmental model. The method used to derive the master equations is the mean-field approach which consists of averaging over different attributes of the system [19] over time.

In Chapter 3, we expand on the mathematical preliminaries required to follow the subsequent study and report on the relevant literature on the effects of awareness on the spread of disease on networks. Afterwards, in Chapters 4 and 5, we formulate the mathematical model and define the relevant notation, respectively. Subsequently, in Chapter 6, we derive the master equations using the mean-field approach. We follow with Chapter 7, in which we explain the simulation methodologies that were deployed. Then, in Chapter 8, we present the analysis results on specific types of networks and compare the mean-field predictions with the results of the simulations. Finally, in Chapter 9, we discuss the results and limitations of the study, propose possible future research directions and close with a conclusion in Chapter 10.

3 Literature Review

In this chapter, we present some preliminary knowledge that is necessary to follow the proceeding analysis. We commence with a presentation of the random graph models used to model the networks, first the Configuration Model and then its extension to the Hierarchical Configuration Model, which can also express the presence of community structures. Afterwards, we proceed in analyzing the epidemiological SIS model and its applications on networks. Thereupon, we explain the algorithm used to perform stochastic agent-based simulations. Eventually, we conduct a literature review on the effects of awareness on the spread of infectious diseases.

3.1 Configuration Model

Considering the numerous different types of networks available in our modelling repertoire, we shall begin by restricting our scope to a specific type of network. Namely, we may investigate simple undirected graphs, that is, undirected graphs without self-loops and multiple edges.

Researchers are often required to generate a random graph with a certain number of nodes and a given degree sequence. The Configuration Model (CM) was formulated to answer that specific need [37].

Let N denote the number of nodes in a random graph and $\mathbf{d} = (d_i)_{i\in[N]}$ a sequence of degrees. We assume that $d_i \geq 1,$ for all $i \in [N]$, since isolated nodes with zero degree cannot participate on disease spread. An inviolable constraint for such a graph to exist is that the sum of the degrees $\ell_N = \sum_{i\in[N]} d_i$ should be an even number. Then, it is always possible to construct a multigraph with the exact degree sequence $(d_i)_{i\in[N]}$. Nonetheless, this multigraph may contain multiple edges between nodes and self-loops.

The construction of a configuration model is conducted as follows:

1. A graph is initialized with N nodes; each one with an associated degree given by the degree sequence $\mathbf{d} = (d_i)_{i\in[N]}$.

2. Node i is assigned d_i half-edges. Every half-edge needs to be connected to another half-edge to form an edge. The total number of half-edges is ℓ_N, which equals the sum of all degrees. Half-edges are numbered arbitrarily from 1 to ℓ_N.

3. We start by connecting the first half-edge randomly with one of the $\ell_N - 1$ remaining half-edges. Next, we remove these two half-edges from the list of half-edges that need to be connected.

4. We continue until we have connected all half-edges.

Definition 3.1 (Configuration Model)**.** The network constructed with the process described above is called the Configuration Model with degree sequence $\mathbf{d}$, abbreviated as $CM_N(\mathbf{d})$.

As mentioned above, we are interested in producing simple undirected networks. However, this may not be possible for a given degree sequence $\mathbf{d}$, even when ℓ_N is even. In the literature, two approaches to deal with the problem of generating a CM without multiple edges and self-loops are described [3]:

Erased Configuration Model: A CM is first generated using the procedure described above. Then, all self-loops are removed, and multiple edges are merged into one single edge.

Repeated Configuration Model: The process of producing a CM is repeated until a simple graph is generated.

Each method has advantages and disadvantages. On the one hand, the Erased CM is faster and computationally cheaper, but it generates graphs on which the degree sequence differs from the one prescribed. On the other hand, the Repeated CM is computationally costly, but it produces a graph with the given sequence when possible.

In the present work, the importance of the exactness of degree distributions is secondary to the primary goal, the study of the effects of awareness in the spread of infectious disease. Thus, the Erased CM was deemed preferable.

3.2 Community structures

The notion of community originates in the social sciences and has been extended to mathematical network theory because of the prevalence of community structures in real-life social networks [28]. The prevalence of community structures was demonstrated in [14] where the community detection algorithm of "edge betweenness" was introduced and used to establish that many real-life social networks exhibit clustering in groups, showing that investigating the community structures was a meaningful endeavour.

In mathematical network science, communities are seen as groups of nodes in a graph that are tightly connected or cohesive. The notion of cohesion can be defined in many different ways, which means that there are multiple mathematical definitions of the term "community". The strongest definition is that of a clique that consists of nodes adjacent to each other. A more general class of definitions is based on the relative frequency of connections; communities are defined as sets of nodes within which connections are dense and between which they are sparse [2]. In [32], two definitions of the community are given:

- **Strong definition of community:** A subgraph V is a community in the strong sense if $k_i^{in} > k_i^{out}$, for all $i \in V$.
- **Weak definition of community:** A subgraph V is a community in the weak sense if $\sum_{i \in V} k_i^{in} > \sum_{i \in V} k_i^{out}$.

Here, k_i^{in} and k_i^{out} are the intra-community degrees and inter-community degrees of node $i \in V$ respectively. In the strong case, each node in a community has more connections with other nodes in the community than with nodes outside the community, while in the weak sense, there are more intra-community edges than inter-community edges. In the present work, the weak definition of community may be assumed unless otherwise stated. However, it should be noted that the model used to generate random graphs, the Hierarchical Configuration Model, allows for the definition of communities that are not even communities in the weak sense. This is useful when a community is defined as a group of nodes that share the same risk perception or disease awareness, even though they do not form a cohesive community.

The model used in this study is the Hierarchical Configuration Model (HCM), a model expressive enough to describe arbitrary community structures with given degree sequences. This network model has a hierarchical structure that consists of two levels, connections between communities and within communities [38]. The model builds upon the well-studied Configuration Model for which many efficient random network generator algorithms exist, and transforming these algorithms to generate an HCM is relatively effortless, as we shall see in Chapter 4.

Definition 3.2 (Hierarchical Configuration Model)**.** Let random graph G with n communities. A community H is represented by $H = ((V_H, E_H), (d_u^{out})_{u \in V_H})$ where (V_H, E_H) is a simple, connected graph and d_u^{out} is the number of edges from vertex $u \in V_H$ to other communities.

Each vertex v has inter-community degree (hereafter called out-degree) d_v^{out} and intra-community degree (hereafter called in-degree) d_v^{in}. The degree of the vertex is then $d_v = d_v^{out} + d_v^{in}$. We also define the total number of edges out of community H as $d_H = \sum_{v \in V_H} d_v^{out}$. We should stress that the terms in-degree and out-degree as used in this work should not be confused with their meaning in directed graphs, where the prefixes in- and out- signify the direction of the edge.

Since we examine the spread of diseases through edges, we shall exclude from our research vertices with zero degrees because they are inconsequential. Moreover, by definition, we examine non-overlapping communities; each node belongs to exactly one community and belonging to a community requires the vertex to be connected to other vertices inside the community. Therefore, for the in-degrees we shall set $d_u^{in} \geq 1$, for all $u \in H$ and for all $H \in G$, while for the out-degrees we have $d_u^{out} \geq 0$, for all $u \in H$ and for all $H \in G$.

We shall also define the denseness of the network δ_{netw} as the number of edges in the network divided by the number of edges in a complete graph of the same size. Similarly, we define the denseness of community H, δ_{com}^H, as the number of edges in the community divided by the number of possible edges in the community [36].

3.3 SIS model

Epidemiological models assume that the population can be separated into different categories (often referred to as compartments) depending on the status of the infection. In the simplest case, the model considers a fixed population of N individuals and ignores demographic changes such as births and deaths.

Compartmental models operate under two fundamental assumptions. The first assumption is that individuals in a particular compartment behave in the same manner, which is not always realistic, but it is a necessary simplification to allow for aggregation. The second assumption is the law of mass action, an idea borrowed from chemistry. According to this law, the rate of change of the number of individuals in a compartment is proportional to the number of individuals in this compartment [12].

The SIS model is the epidemiological model in which each individual can be in one of two possible states: S for susceptible individuals who can contract the disease and I for infected/infectious individuals who have contracted the disease and can spread it. Note that infected individuals are immediately infectious and can transmit the disease to other susceptible individuals. More complicated models differentiate between the states "exposed infected" E and "infectious" I. However, there is no such discrimination in the SIS model, and the terms "infected" and "infectious" may be used interchangeably to signify the I state. Additional compartments can be appended in different models to signify other possible states of infection, i.e. R for recovered. [29].

In the SIS model, two transitions are possible (see Figure 1):

- $S \to I$ or from susceptible to infected: This transition occurs when a susceptible individual becomes infected/infectious through contact with another infected individual.

- $I \to S$ or from infected to susceptible: This transition happens when an infected individual recovers from the infection and becomes susceptible again.

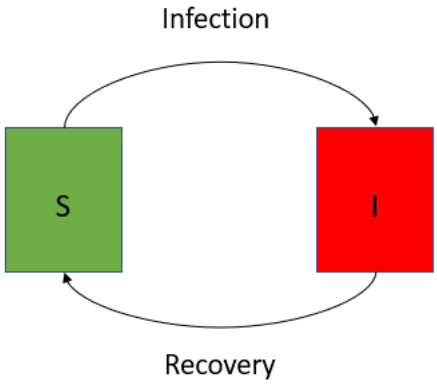

FIGURE 1: Flow diagram of the SIS model.

Since a recovered individual can be immediately reinfected, immunity effects are not considered in the SIS model. Therefore, individuals can undergo a cycle of the form $S \to I \to S$, which reveals the origin of the model's name. Consequently, the long-term regime of the model can exhibit an endemic state characterized by a constant number of infections on average [29], as we can see in Figure 2.

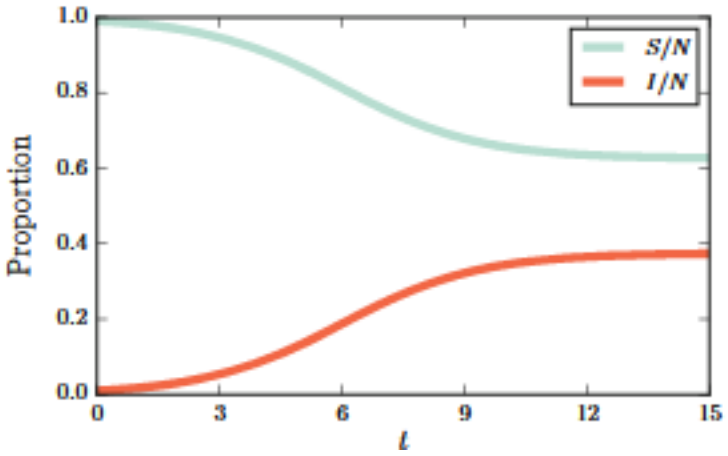

FIGURE 2: Fractions of the susceptible (green) and infected (red) with respect to time in the SIS model [19].

In contrast, we shall mention the SIR model in which recovering individuals enter the R (recovered) state and become hereafter immune to the infection or die; in both cases, the epidemiological effect is indistinguishable since they cannot be reinfected. In the SIR system, the long-term number of infected individuals always tends to zero, and the infection dies out in finite populations [29].

3.3.1 Deterministic Compartmental SIS

The time evolution of the SIS model can be described by a system of Ordinary Differential Equations (ODE), also called "master equations". This method of analysis, that reduces the stochastic variation of the system to a deterministic average, is called the mean-field approach.

Let i and s represent the fractions of infected and susceptible individuals in the population respectively. By definition, we have that $s+i=1$ and consequently, by differentiating with respect to time, $\frac{di}{dt} = -\frac{ds}{dt}$. Hence, in order to fully describe the behaviour of the system, it is sufficient to deduct an expression for either i or s.

In the SIS system, we assume that the transition rate $I \to S$ is constant. In discrete time formulation, we define the average number of recoveries per time unit, or recovery rate, γ. Then, the mean time an individual remains in the I state is γ^{-1} time steps. In continuous time formulation, if we assume that recovery is a Poisson process, the recovery time follows an exponential distribution with mean γ^{-1} [19].

Moreover, we assume that the transmission rate $S \to I$ is constant. This is the rate that an infectious individual will transmit the infection to a susceptible individual with whom they come into contact in one time step. In discrete time, we define the average number transmissions, or infection rate, τ. Similarly, in continuous time formulation, assuming that infection is a Poisson process, the time until a susceptible individual is infected by an infectious individual is an exponential distribution with mean τ^{-1} [19].

Homogeneous mixing is the assumption that every infected individual can infect every susceptible individual. Under this assumption, we have that:

$$\begin{aligned} \frac{di}{dt} &= \tau i s - \gamma i \\ &= \tau i(1-i) - \gamma i. \end{aligned} \tag{1}$$

Intuitively speaking, τ is the rate at which infected individuals have infection-transmitting contacts. Then, the total rate of infection-transmitting contacts is τi, but only a fraction s of them is with susceptible individuals and lead to new infections. Hence, the rate of change of the fraction of infections equals the fraction of new infections $\tau i s$ minus the number of existing infections that recovered γi [19]. This is an application of the law of mass action that was discussed above.

It is interesting to investigate the steady-state solutions of equation (1). We see that:

$$\begin{aligned} &\frac{di}{dt} = 0 \implies \\ &\tau i(1-i) - \gamma i = 0. \end{aligned}$$

The last equation has two solutions indicating two steady-states or equilibria: the infection-free steady-state $i=0$ and the endemic state $i = 1 - \frac{\gamma}{\tau}$ which only exists when $\frac{\gamma}{\tau} < 1$ or inversely $\frac{\tau}{\gamma} > 1$. The last fraction is called basic reproductive ratio $R_0 = \frac{\tau}{\gamma}$ and is a significant value in epidemiological models. Specifically, in the case under examination, when $R_0 \leq 1$ the infection-free steady-state is stable and the endemic state unstable and when $R_0 > 1$ the endemic state is stable and the infection-free state unstable [29].

We shall call the fraction of infected individuals in the steady-state "epidemic prevalence" i_{ss}. Hence, in the infection-free steady-state we have $i_{ss} = 0$ while in the endemic steady-state we have $i_{ss} = 1 - \frac{\gamma}{\tau}$. Moreover, we notice that if we fix the recovery rate γ, there is a critical value of the infection rate τ_c above which the system has an endemic

equilibrium and below which it has an infection-free equilibrium. We shall call this critical value "epidemic threshold". In this case, the epidemic threshold is $\tau_c = \gamma$. These two values shall serve as quantitative metrics of the spread of disease.

3.3.2 SIS in Networks

In the previous section, we assumed that every individual can infect everyone else, an assumption called homogeneous mixing. In reality, this assumption is rarely fulfilled because people tend to have potentially disease-transmitting interactions only with specific individuals. Thus, the network point of view is more realistic: individuals are seen as nodes in a network and their interactions as edges. In this model, the infection can spread from infectious nodes only to their susceptible neighbours.

As in the case of homogeneous mixing, in the SIS network model, each node can be in one of two states: S (susceptible) and I (infected). Infected nodes recover with constant rate γ but, contrary to homogeneous mixing, susceptible nodes can become infected only by their infected neighbours with a constant rate τ per infected neighbour.

The epidemic process can be described as a continuous-time Markov chain. In a network with N nodes, there are 2^N states because each node can be in two possible states. The Markov chain approach is exact, but it has limited use because the size of the system grows exponentially with the size of the graph. Therefore, different mean-field approaches have been developed to decrease the system's degrees of freedom and make it more tractable. Particularly for the SIS system, there are two derivations of a mean-field approximation: the individual-based mean-field (IBMF) and the degree-based mean-field (DBMF) approach [29].

On the one hand, the IBMF approach simplifies the presentation by assuming that the probability that a node belongs to a compartment is statistically independent of the state of its immediate neighbours. Then, we can derive the mean-field equations using the governing equations of the 2^N-state Markov chain, under the assumption that the expected values of variable pairs factorize. The IBMF method takes into consideration the topological structure of the network, encoded in the adjacency matrix, and its solutions depend on the spectral properties of said matrix. On the other hand, the DBMF assumes that all nodes with the same degree are statistically equivalent. Consequently, it does not preserve the totality of the topological structure; it only retains the degree distribution [29].

In the subsequent analysis, the aim is to investigate the effects of the awareness of the infection in the environment of each node. Thus, it is convenient to use a modified version of the DBMF such that nodes that are statistically equivalent under this method also have the same expected awareness. Therefore, we shall build our analysis using the DBMF approach, which we shall hereafter simply refer to as the mean-field (MF) approach. This method was the first one to be used to investigate the SIS model in complex networks [30].

Let $i_k(t)$ be the fraction of infected nodes with degree k at time t and $s_k(t)$ the fraction of susceptible nodes with degree k at time t[1]. That is:

$$i_k(t) = \frac{\#\text{ infected nodes with degree } k \text{ at time } t}{\#\text{ nodes with degree } k},$$
$$s_k(t) = \frac{\#\text{ susceptible nodes with degree } k \text{ at time } t}{\#\text{ nodes with degree } k}.$$

[1] In infinite networks, $i_k(t)$ and $s_k(t)$ are interpreted as the probability that a node of degree k at time t is infected or susceptible respectively.

Also, let $P(k'|k)$ be the probability that a node of degree k is connected to a node of degree k'. Then, the dynamic equation that describes the evolution of the system is:

$$\frac{di_k}{dt} = -\gamma i_k(t) + \tau k(1 - i_k(t)) \sum_{k'} P(k'|k) i_{k'}(t). \tag{2}$$

The first term of the right-hand side accounts for the infected nodes that recovered and became susceptible anew. The second term accounts for the susceptible nodes that were infected. According to the law of mass action, this is proportional to the fraction of susceptible nodes $1 - i_k(t)$, the number of edges k through which the infection can travel to a node of degree k and the probability that their edges point to an infected node. The probability that they are connected to an infected node is calculated by summing the probabilities that they are connected to a node of degree k' times the fraction of these nodes that are infected $i_{k'}$ for all degree classes k' in the network [29].

Since there are only two compartments, we have that $i_k(t) + s_k(t) = 1$ and $\frac{di_k}{dt} = -\frac{ds_k}{dt}$. Thus, it is sufficient to describe the time evolution of i_k for all degree classes k in the network.

Uncorrelated networks are networks in which the degrees of nodes are uncorrelated, such as the networks generated with the Configuration Model method. In this case, we have that $P(k'|k) = \frac{k'}{\langle k \rangle} P(k')$ where $\langle k \rangle = \sum_{k'} k' p(k')$ the mean degree of the network. Then, equation (2) can be written as:

$$\frac{di_k}{dt} = -\gamma i_k(t) + \tau k(1 - i_k(t)) \sum_{k'} \frac{k'}{\langle k \rangle} P(k') i_{k'}(t), \tag{3}$$

where the term

$$\Theta = \sum_{k'} \frac{k'}{\langle k \rangle} P(k') i_{k'}(t) \tag{4}$$

is the probability to find an infected node following a randomly selected edge.

The system of ODE (3) cannot be solved in a closed form for general degree distributions. However, it is possible to investigate its steady-state behaviour by solving the linear system of equations for the stationarity condition $\frac{di_k}{dt} = 0$ [30]. In the steady-state, we obtain:

$$i_k = \frac{k\Theta\frac{\tau}{\gamma}}{1 + k\Theta\frac{\tau}{\gamma}}. \tag{5}$$

We notice that i_k is increasing with k, meaning that more interconnected nodes are more likely to become infected.

Then, the epidemic prevalence is:

$$i_{ss} = \sum_k P(k) i_k = \sum_k P(k) \frac{k\Theta\frac{\tau}{\gamma}}{1 + k\Theta\frac{\tau}{\gamma}}.$$

The parameter Θ is computed by substituting (5) to (4) to obtain the self-consistent equation

$$\Theta = \frac{1}{\langle k \rangle} \sum_k kP(k) \frac{k\Theta\frac{\tau}{\gamma}}{1 + k\Theta\frac{\tau}{\gamma}}. \tag{6}$$

The equation (6) has a non-zero solution for Θ, that leads to the endemic state, only when:

$$\tau > \tau_c = \gamma \frac{\langle k \rangle}{\langle k^2 \rangle},$$

which is the epidemic threshold of the model. The epidemic threshold is proportional to the mean degree and inversely proportional to the mean squared degree. The mean squared in the denominator shows the importance of degree heterogeneity in the spread of the epidemic.

For a fully homogeneous network where all degrees are equal, we have that $\langle k^2 \rangle = \langle k \rangle^2$ and $\tau_c = \gamma \frac{1}{\langle k \rangle}$, an epidemic threshold inversely proportional to the average connectivity. However, in networks with high degree heterogeneity, the second moment of the degree $\langle k^2 \rangle$ grows relative to the mean $\langle k \rangle$, lowering the epidemic threshold. For example, this is observed in networks with power-law degree distribution $P(k) \sim k^{-\alpha}$. In the special case of $2 < \alpha \leq 3$, we have $\langle k^2 \rangle \to \infty$ as the size of the network grows to infinity. In such cases of so-called scale-free networks, the epidemic threshold vanishes $\tau_c \to 0$ as the network size grows to infinity.

3.4 Stochastic Simulations

The mathematical models described in the previous sections attempt to aggregate the degrees of freedom of the systems and predict the time evolution of stochastic processes. They were examples of compartmental models. Another way to study stochastic processes is to perform agent-based stochastic simulations. In this method, there is no aggregation, and each individual's behaviour is taken into consideration.

In this work, stochastic simulations were run on specific networks for multiple repetitions so that the results gain statistical significance. Then, the results of the simulations were compared with the expected results of the mathematical models to verify the models. The simulations shall be treated as the stochastic "reality" that the mathematical models attempt to describe deterministically.

Stochastic simulations can be performed either in discrete or in continuous time. In discrete time simulations, we calculate the probability that a node will transition in status ($S \to I$ or $I \to S$) in each time step, and then we determine which transition happens using random number generators. The main disadvantage of this method is that it cannot take into consideration that transitions that happen within one time step can influence each other since they are seen as happening simultaneously [19].

Since epidemic outbreaks, in reality, happen in continuous time, we opt out for continuous time simulations. In this work, we used the Gillespie algorithm, which was first introduced to simulate chemical reactions [13]. The idea of the algorithm is as follows:

1. In the present time t_{now}, each possible event e has a rate r_e associated with it. In the SIS model, the possible events are of two kinds: each susceptible individual may become infected by one of its infected neighbours and each infected node may recover and become susceptible again.

2. We calculate the total rate $r_{total} = \sum_e r_e$ and generate the time until the next event t_{next} from an exponential distribution with that rate.

3. Next, we need to determine which event occurred. For this purpose, we assign a probability to each possible event $p_e = \frac{r_e}{r_{total}}$ and select randomly which event occurred.

4. We set $t_{now} \rightarrow t_{now} + t_{next}$ and update the state of the system. Then, we calculate the new set of possible events and update the associated rates. The process is repeated until $r_{total} = 0$ or t_{now} becomes larger than a set value t_{max}.

The Gillespie algorithm provides an exact stochastic simulation in the case of a Markovian system [19].

An alternative continuous time simulation method is the event-driven simulation. In this method, a priority queue of upcoming events is kept in memory, and the events are processed sequentially. Each event causes a series of upcoming events that are then added to the queue at the appropriate locations. For example, when individuals become infected, we can calculate when they will recover and which other individuals they will infect in the meantime [19]. In general, the event-driven algorithm is more efficient than the Gillespie algorithm. However, our analysis cannot exploit its efficiency because each new infection affects all subsequent infections until recovery through the awareness mechanism, which causes an adjustment to the infection rates of all individuals in the network, as we shall see below.

3.5 Approximation of the Epidemic Threshold Using Stochastic Simulations

Recall that the epidemic threshold is the critical infection rate τ_c above which we reach the endemic steady-state in the SIS model. When we solve the system using a mean-field approximation, calculating the epidemic threshold is a matter of studying the stability of a dynamic system. However, when we examine the results of stochastic simulations, a specialized methodology is required.

In [35] and [41], the researchers ran multiple realizations of the simulations for an adequate time horizon and averaged the results to reduce stochastic fluctuations. Then, they estimated the epidemic threshold as the minimum infection rate that caused an average infection density over a threshold. However, it is admitted that this methodology may lead to overestimating the epidemic threshold because of finite-size effects, which we will explain next.

There are two types of finite size effects that may affect the estimation of the epidemic threshold. Firstly, the SIS system contains an absorbing state, the infection-free equilibrium, which cannot be exited once visited. Stochastic simulations of finite systems are particularly sensitive to this effect because the absorbing states may be reached even in the supercritical regime because of random stochastic fluctuations. Then, even infection rates above the epidemic threshold may lead to a vanishing infection. This may lead to an overestimation of the epidemic threshold.

The standard solution to this problem is to restrict the simulations to runs that do not visit the infection-free equilibrium, which is called the surviving runs method [9]. This is a rather costly process because of the large number of runs that will not be taken into consideration. A second approach is the so-called quasi-stationary (QS) procedure in which the absorbing configuration is excluded from the dynamics. This is achieved by keeping

track of a set of active states from the history of the simulation and regularly replacing the present configuration with one from the past with a certain probability [6] [9].

A second related issue is that of the degree-distribution related finite-size effects. As we discussed previously, in a scale-free network, the epidemic threshold vanishes. This effect is attributed to the infinite variance of the degree distributions in such networks. However, real-life networks, as well as specific network configurations used in simulations, are not infinite. This means that there is an inevitable cut-off to the network's power-law degree distribution, which leads to an epidemic threshold even in networks with infinite variance degree sequences. In these cases, the epidemic threshold tends to be inversely proportional to the size of the network, meaning that smaller networks tend to overestimate the epidemic threshold [26] [24].

To attend to this problem, the size dependent susceptibility is introduced:

$$\chi = N\frac{\langle i^2\rangle - \langle i\rangle^2}{\langle i\rangle},$$

where the average infection density $\langle i\rangle$ and its second moment $\langle i^2\rangle$ are computed using surviving runs or quasi-stationary simulations [8]. The methodology entails running multiple realizations of networks with the same degree distribution and different sizes N. The size-independent epidemic threshold is the infection rate τ_c for which the susceptibility χ peaks.

3.6 Spread of Infections on Networks with Community Structures

The existence of mesoscopic community structures in networks has been investigated for its effects on the spread of infections. It has often been seen in the literature that such structures have profound effects on the evolution of an outbreak.

In [36], the researchers investigated several real-world networks and extracted several statistics as well as the community structures using a community detection algorithm. Then, they generated networks using three methods. First, they applied the CM algorithm to create networks with identical degree sequences as the real ones while at the same time destroying the community structures. Then, they used the HCM method to rewire only the inter-community edges; in this way, they preserved the community structures themselves and the intra-community microscopic structures. Lastly, they used the HCM algorithm to rewire both the inter-community and intra-community edges with the restriction that intra-community edges shall remain within their respective communities and inter-community edges shall remain between communities. The last method preserves both the mesoscopic and microscopic structures without overfitting to specific cases. The last algorithm was denoted with HCM*.

In these three models, the spread of a SIR epidemic was studied using the bond percolation method, and the results were compared with the original network. In the context of a SIR epidemic, the bond percolation method entails removing edges from the network with probability $1-p$ and retaining them with probability p, where p is the probability that an infected node will transmit the infection to a susceptible neighbour. In the resulting network, the size of the largest component provides information on the final size of the epidemic. It was shown that the HCM and HCM* models follow the development of the epidemic with greater fidelity to the original model than the CM model. This led to the conclusion that the community structures play a vital role in the epidemic. Furthermore, the fact that HCM and HCM* did not differ significantly in how closely they modelled the epidemic shows that the internal structure of the communities is insignificant for the

most part. This is explained by the relative denseness of the edges inside the communities compared to between them; an epidemic reaching a community would eventually infect most of the members while the spread was inhibited by the relative sparseness of the inter-community degrees.

In [23], the researchers generated random networks by first partitioning the nodes to communities and then generating edges between them with different probabilities: q for inter-community edges and p for intra-community edges. They also defined the degree of community $\sigma = \frac{p}{q} >> 1$; with a larger degree of community, the edges within communities become denser in comparison to the inter-community edges. This is a special case of the stochastic block model [17] which generates Erdös-Rényi-like networks with a given number of Erdös-Rényi-like communities. With $\sigma >> 1$ and considering the stochastic element in network generation, we should expect the communities to comply with the weak definition of community.

Afterwards, the spread of an SIS infection was examined using both simulations and deterministic compartmental modelling. Both methods were in agreement that the epidemic threshold was inversely proportional to the degree of community $\tau_c \sim \frac{1}{\sigma}$. Therefore, the denser the communities compared to the inter-community connectivity, the less likely it was for a global epidemic outbreak to occur. The explanation was that outbreaks tend to spread locally and die out before they reach other communities. Nonetheless, community structures were shown once more to play a crucial role in the evolution of the epidemic.

In [21], a variation of the stochastic block model was used, and extensive mean-field analysis, as well as stochastic simulations, were conducted in both the SIS and SIR models. The analysis was performed based on a distinction between infections within the community and infections introduced from outside a community. On this level of analysis, the focus is on the boundary nodes of each community, defined as the nodes that have external connections. The results indicate that only a small fraction of infections happen between different communities. In fact, when the communities were analyzed in the absence of external connections, no considerable change in the infection levels was noted. Hence, the conclusion was that boundary nodes play an essential role in introducing the infection to a community or reintroduce it after dying out in a specific community while still circulating in the network. However, other than that, the role of external connectivity is mostly insignificant.

3.7 Infection Awareness

As we saw in the preceding sections, the rate of infection is proportional to the frequency of contact with other infected individuals. Nevertheless, it has been observed that individuals and communities tend to adjust their behaviour based on the perceived risk related to the levels of infection in the environment [15] [31] [11]. Moreover, risk perception tends to differ between individuals or groups based on various factors such as their location, their demographics, their susceptibility to infection and media exposure. Notably, risk perception does not depend on the real danger but only on the perceived risk, as the name suggests [1]. The behaviour adjustment, which is directed by risk perception, takes the form of precautions that limit the probability of infection. Depending on the nature of the infection, these include decreased frequency of contacts, use of protective apparatuses, social distancing and personal hygiene [27]. Such precautions can be undertaken individually or imposed in some form by authorities or social pressure. In effect, these behaviours decrease the probability of infection.

In [11], the researchers classified the behaviour-disease models based on three aspects:

- **The source of information:** The source could be either global, such as public announcements, or local, coming from their neighbourhood or community.
- **The type of information:** The type was either prevalence-related (directly related to the level of infection) or not prevalence-related, which can be set independently of the progression of the outbreak.
- **The effects of the behavioural change:** The effects can be on three levels: the disease state (e.g. vaccinations making individuals immune), the infection or recovery rate and the contact network on which the disease spreads.

In [1], the authors analyzed the effects of risk perception on an SIS infection spreading on a scale-free network generated using the preferential attachment algorithm [37]. The risk perception differs per susceptible node, and it was modelled by a function $A(k_{inf}, k) \leq 1$, which depends on the degree k and the number of infected neighbours k_{inf}. This function is multiplied with the infection rate τ so that the effective infection rate is decreased by a certain factor. The risk perception was defined as

$$A(k_{inf}, k) = e^{-(K+J(\frac{k_{inf}}{k})^{\alpha_1})}, \tag{7}$$

where J represents the level of precautions adopted in response to new infections in the neighbourhood, $\alpha_1 \leq 1$ models the use of special prophylaxis (additional safety measures), and K quantifies the global influence over the whole population. It should be noted that K remains constant, meaning that the global risk perception does not depend on disease prevalence. J and α_1 can be thought of as quantifying the linear and non-linear responses on the local level. It was demonstrated that for fixed or bounded connectivity and $\alpha_1 = 1$, there was always a finite value J_c that would lead the epidemic into extinction. However, when the variance of the degree distribution diverged to infinity, the disease could only become extinct by setting $\alpha_1 < 1$, namely by applying non-linear risk perception.

In [20], the same model (7) for the risk perception was used on scale-free networks with power-law degree distribution. The main focus of this paper is to study the effect of risk perception on nodes with different connectivity. Power-law degree distributions result in a few nodes with disproportionately high degrees, and it seems that these highly connected nodes are responsible for the bulk of infections. When the local risk perception J is introduced, the infection levels among highly connected nodes do not change significantly. However, the infection fails to reach nodes in the low end of the degree distribution in the periphery of the network. The effect of K is predictably the lowering of the infectivity across the network.

In [21], we encounter an investigation on the effects of risk perception on networks with community structures. The authors adapted the equation (7) by setting different global awareness K for each community, now called community awareness. As in [1] and [20], K remained constant throughout the epidemic outbreak, independently of the levels of infection. Nonetheless, it allowed the researchers to consider the effects of different levels of community awareness. In high awareness communities, the infection tended to die out and had to be reintroduced by the boundary nodes. Therefore, as the community awareness increases, the effects of external connectivity become more pronounced in sustaining the infection, while their impact was minor in the absence of awareness. Moreover, this effect highlights the importance of low awareness communities in preserving the infection in the network and repeatedly reintroducing it to high awareness communities.

In [35] and [41], the researchers investigate the effect of prevalence-related awareness both on the local and the global level. Furthermore, they introduce the contact awareness $\psi(k)$, which is not dependent on the presence of the infection but on the degree of a node. The rationale of the contact awareness is that highly connected nodes are at greater risk of infection, and therefore they would take precautionary measures of proportional magnitude. In this case, a linear model of awareness is used following the formula:

$$A(k_{inf}, k, i) = \psi(k)(1 - J\frac{k_{inf}}{k})(1 - Ki), \tag{8}$$

where i is the fraction of infected nodes in the network, or infection density.

It is shown, using both stochastic simulations and mean-field analysis, that all forms of awareness affect the epidemic prevalence. However, local and contact awareness were also able to influence the epidemic threshold.

In [35], non-linear effects were also studied in the form of the exponents α_1, α_2 in the formula:

$$A(k_{inf}, k, i) = \psi(k)(1 - J(\frac{k_{inf}}{k})^{\alpha_1})(1 - Ki^{\alpha_2}), \tag{9}$$

and it was shown once more that only the local parameter α_1 affects the epidemic threshold.

In the investigations described thus far, the awareness of the infection was produced either in reaction to the presence of infection or imposed globally. An alternative research direction draws from the literature on the spread of ideas on social networks. As a matter of fact, the study of the spread of ideas has been influenced by epidemiological modelling in networks where opinions take the form of contagious agents spreading through social contact or imitation [22]. In this case, awareness becomes a discrete state in which nodes can enter under the influence of their neighbours, in a similar manner as in the epidemiological models of diseases.

In [16], the researchers defined a multiplex network with two layers. In the first (physical) layer, an SIS infection process unfolds, while the second (virtual) layer is dedicated to the spread of awareness. The awareness process was modelled using two discrete states, aware (A) and unaware (U), forming a cyclical process UAU in which aware nodes could spread their awareness to neighbouring unaware nodes in the virtual layer. Additionally, nodes could also become aware if they received an infection in the physical layer (self-initiated awareness). The awareness status of a node would affect the transmission rate of the infection on the physical layer creating an interesting dynamical interplay between the two processes where the infection and the awareness spread antagonistically to each other. In particular, the epidemic threshold and the epidemic prevalence in the steady-state depended on the awareness dynamics and the topology of the virtual layer.

In [40], the same UAU awareness model was studied with a SEIS infection, including an exposed state in which nodes are infectious but unaware of the fact, making them asymptomatic carriers[2]. The processes spread in a two-layer multiplex network. The presence of asymptomatic carriers in the model inhibited the self-initiated awareness, leading to a decrease in the epidemic threshold and an increase in the epidemic prevalence. In [10], it was observed that centrally imposed awareness can decrease the epidemic prevalence but has a negligible effect on the epidemic threshold. Nonetheless, a local self-initiated

[2]This should not be confused with the exposed state in the regular SEIR or SEIS models in which the agent is not yet infectious. Here, E is used instead of A for "asymptomatic" to avoid confusion with A for "aware".

awareness process can affect the epidemic threshold by preventing an outbreak before it spreads globally, mainly when the awareness and the infection spread on the same network rather than on a multiplex.

In the literature, we encounter different terms such as risk perception, infection awareness, information, alertness and behavioural adaptation. All these terms are used to express the same notion that the awareness of an infection in the environment causes behavioural adaptation. Hereafter, we shall refer to the notion as "awareness" to avoid further confusion.

4 Model Formulation

In this chapter, we shall define the mathematical model that was studied and explain the modelling choices. We begin by disclosing the choices related to the network structures used and then justify our choice of epidemiological model. Lastly, we shall expound on how the awareness was modelled.

4.1 Network Modelling

In the present analysis, we shall use graphs that consist of nodes and edges that connect them. Each node models an individual with a status related to the infection, while edges model the interactions between individuals that can transmit the infection.

The topologies of the networks studied are static, meaning that the connections between nodes do not change over time. This is a convenient approximation since real-life networks are rarely static; edges are being cut, created or rewired at some rate. A static network provides a good approximation for processes that evolve at a much faster scale than the network topology [29].

Moreover, we shall examine networks that are simple, unweighted and undirected graphs. Simple graphs are graphs without multiple edges and self-loops. The fact that edges are unweighted means that each interaction between individuals is equally likely to transmit the infection. In real life, this may not be the case, as some interactions are considered riskier. Similarly, multiple edges could be used to model the fact that nodes may interact more often, increasing the probability of infection, which is reducible to a weighted edge. Undirected edges model the fact that the infection can travel in both directions, a choice that is quite realistic for most infections. However, it should be noted that there are cases in which there is an intrinsic directionality, such as blood transfusions [29]. Finally, self-loops do not have a physical meaning in the study of epidemics.

Furthermore, in this work, we focus on networks with community structures. As we saw in the literature review, community structures have been shown to play a critical role in the spread of disease. It must be stressed that, in this project, we shall not focus on the effect of community structures but on the effects of awareness, which may be shared among individuals belonging to the same community. To generate networks with community structures, we shall use the Hierarchical Configuration Model (HCM) that was described in the literature review. The main advantage of this method is that it makes it possible to generate networks with prescribed degree distributions.

Now, we shall explain how we generated HCM networks. Let N be the number of nodes in the network, n the number of communities, and N_H the number of nodes in community H. Furthermore, let an out-degree sequence $\mathbf{d}^{out}$ of length N and n in-degree sequences $\mathbf{d}_H^{in}$, one for each community. The communities are ordered arbitrarily, $H_1, \cdots, H_n$. Then, the nodes are assigned a number from 1 to N such that the nodes $1, \cdots, N_{H_1}$ belong to H_1, the nodes $N_{H_1} + 1, \cdots, N_{H_2}$ belong to H_2 and so on. Thus, each node is assigned a number, a community, an in-degree and an out-degree. We can construct a Hierarchical Configuration Model with the following process:

S.1 Generate n different graphs, one for each community H using the in-degree distributions $\mathbf{d}_H^{in}$ with the Configuration Model process.

S.2 Generate one graph using the out-degree distribution $\mathbf{d}^{out}$ with the Configuration Model process.

S.3 Superimpose the graph generated in step 2 on the community graphs generated on step 1 using the ordering of the nodes.

In this work, we used the Erased Configuration Model process to create graphs in steps **S.1** and **S.2**. Moreover, there is no guarantee in step **S.2** that an edge will not point to the same community, although the probability of this happening decreases as the fraction of inter-community half-edges of each community by the inter-community half-edges of all other communities decreases. Consequently, the in-degrees and out-degrees in the final network may differ slightly from the prescribed ones.

4.2 Epidemiological Model

The epidemiological model used in the following analysis is the SIS model on networks which was described in Chapter 3. As we mentioned, individuals can be in either of two states I for infected/infectious and S for susceptible.

Real-life diseases rarely comply with this model. In most cases, individuals gain some level of immunity after infection, something more accurately portrayed in the SIR model. Furthermore, immunity may not last forever. Thus individuals may end up in the susceptible state again, as in the SIRS model. Alternatively, it is common for patients not to become infectious immediately after exposure, which is conveyed by the E state in the SEIS, SEIR and SEIRS models.

In any case, the SIS model expresses an idealized case that epidemiologists regularly study because of its conceptual simplicity and mathematical tractability. The results of SIS modelling can still offer essential understandings of the infection dynamics. In our case, we opted for the SIS model because by simplifying the model, we could add complexity and study the effects of other aspects of the system, such as the presence of awareness. After understanding the effects of awareness on networks with community structures on the simple SIS model, we may invite researchers to add realism by researching more complicated epidemiological models.

A susceptible node $j \in S$ becomes infected upon contact with infected node $m \in I$ at some rate. When a susceptible node is infected, it immediately moves to the set of infected nodes I. The transmission rate along edge jm is $A_j T_m$ where A_j is the admission rate of $j \in S$ and T_m the infection rate of $m \in I$. The admission rate A_j depends on the behaviour of the susceptible node and will be used to model the awareness of the disease; nodes with high awareness will have lower A_j that will make it harder to become infected, with $0 \leq A_j \leq 1$. In our model, we set the infection rate constant $T_m = \tau$, where τ is the awareness-free per edge infection rate. Note that in the absence of awareness, we have $A_j = 1$, and the transmission rate is equal to the infection rate. We define τ based on the nonlinear contagion scheme as in [25], where $0 \leq \tau \leq 1$ is the probability that a susceptible node will be infected by one infected neighbour in one time step in the absence of awareness. Thus, if a node $j \in S$ has k_{inf} infected neighbours, the probability that it will become infected in one time step is $1 - (1 - A_j\tau)^{k_{inf}}$. Lastly, let γ be the recovery rate, the rate with which infected nodes recover, which is constant. When a node recovers, it becomes susceptible immediately. Note that γ refers to an infected node while τ refers to an edge that connects a susceptible node with an infected one.

4.3 Infection Awareness

The effects of infection awareness form the core of the present research. Hence, it is essential that we locate our research in the infection awareness literature and highlight our novel contributions.

In this thesis, we shall model the awareness of susceptible node j as a continuous variable:

$$A_j = \phi_j^L \phi_H^C \phi^G,$$

which depends on three types of awareness:

- ϕ_j^L is the local awareness that depends only on the immediate neighbours of node j.
- ϕ_H^C is the community awareness of community H to which the node j belongs.
- ϕ^G is the global awareness that is shared along with the whole network G.

We define the awareness such that $0 \leq \phi_j^L, \phi_H^C, \phi^G \leq 1$ which implies that $0 \leq A_j \leq 1$.

We favoured a multiplicative model in which the effect of each type of awareness is multiplied with the rest. The rationale is that each type of awareness decreases the rate of transmission by a certain factor and that the three types are independent of each other [1]. Therefore, without awareness, node j becomes infected by one infected neighbour with rate τ. After receiving the local awareness, the rate becomes $\tau\phi_j^L$. Subsequently, after receiving the community and global awareness, the rate becomes $\tau\phi_j^L\phi_H^C\phi^G = \tau A_j$.

According to the classification given in [11], the source of awareness A_j is both local and global. Specifically, the global source is captured by the global awareness ϕ^G, while the local and community awareness ϕ_j^L and ϕ_H^C are forms of awareness from local sources with different horizons. Furthermore, the effects of the behaviour change are only captured in the transmission rate; neither the disease state nor the network topology is affected. Lastly, as we shall see, the type of information in our awareness model is prevalence-related, meaning that the awareness is proportional to the prevalence of the infection, either locally or globally. The mechanism with which the awareness depends on the prevalence depends on the specific awareness model.

In this work, we used a linear awareness model. In this model, the awareness has a linear relationship with the prevalence of the infection. Specifically, the three types of awareness are defined below.

Local awareness: $\phi_j^L = 1 - c_L\frac{k_{inf}}{k}$, where c_L is the local awareness coefficient, k_{inf} is the number of infected neighbours and k is the degree of the node j.

Community awareness: $\phi_H^C = 1 - c_C i^H$, where c_C is the community awareness coefficient and i^H the fraction of infected nodes in the community H.

Global awareness: $\phi^G = 1 - c_G i$, where c_G is the global awareness coefficient and i the fraction of infected nodes in the network.

The linear relationship is characterized by the local, community and global awareness coefficients c_L, c_C and c_G, respectively. With $0 \leq c_L, c_C, c_G \leq 1$ we have $0 \leq \phi_j^L, \phi_H^C, \phi^G \leq 1$. Hence, we have:

$$\begin{aligned} A_j &= \phi_j^L \phi_H^C \phi^G \\ &= (1 - c_L\frac{k_{inf}}{k})(1 - c_C i^H)(1 - c_G i) \end{aligned}$$

with $0 \leq A_j \leq 1$. We observe that A_j decreases as the awareness of the infection increases.

With the above formulation, we have that the transmission rate along the edge jm with $j \in S$ and $m \in I$ is:

$$A_j\tau = \phi_j^L \phi_H^C \phi^G \tau = \tau(1 - c_L \frac{k_{inf}}{k})(1 - c_C i^H)(1 - c_G i).$$

5 Notations

5.1 Compartments and Classes of Nodes

Before proceeding, we need to clarify the notation that will be used in the subsequent steps. The symbols I and S have multiple meanings according to the context. Firstly, they are used to denote the state that a node can be in as well as the sets of all nodes at that state, susceptible and infected, respectively (i.e. $v \in S$ means that node v is susceptible). A set of all nodes in the same state is also called a compartment. Moreover, in the mean-field analysis, they will be used to denote the number of infected or susceptible nodes at a specific time t (i.e. $I(t)$ and $S(t)$).

When $S(t)$ and $I(t)$ are in capital letter notation, they denote the absolute number of nodes in a specific state, while in the lower case, they denote the fraction of the nodes in that state, otherwise called infection density. Therefore, we have that $s(t) = \frac{S(t)}{N}$ and $i(t) = \frac{I(t)}{N}$, where N is the number of nodes in the network.

In the SIS model, we have $s + i = 1$ and $S + I = N$. By differentiating with respect to time, we have that $\frac{ds}{dt} + \frac{di}{dt} = 0$ or $\frac{ds}{dt} = -\frac{di}{dt}$.

Let G be a network with N nodes and n communities, as defined above. The communities will be represented as $H_1, \cdots, H_n$, and N_{H_i} represent the number of nodes in community H_i.

We may also need to denote the number or fraction of nodes in a specific state in the community H_i. This will be denoted by S^{H_i}, I^{H_i} and s^{H_i}, i^{H_i} respectively. Thus, we have $i^{H_i} = \frac{I^{H_i}}{N_{H_i}}$ and $s^{H_i} = \frac{S^{H_i}}{N_{H_i}}$.

Each node has an in-degree k_{in}, the number of neighbours that belong to the same community, and an out-degree k_{out}, the number of neighbours that belong to a different community. We shall denote the number (fraction) of infected and susceptible nodes in community H_i with in-degree k_{in} as $I^{H_i}_{k_{in}}$ and $S^{H_i}_{k_{in}}$ $(i^{H_i}_{k_{in}}, s^{H_i}_{k_{in}})$ respectively. These fractions are defined with respect to the number of nodes in the community. Thus, we have $i^{H_i}_{k_{in}} = \frac{I^{H_i}_{k_{in}}}{N_{H_i}}$ and $s^{H_i}_{k_{in}} = \frac{S^{H_i}_{k_{in}}}{N_{H_i}}$.

Furthermore, we will denote the number (fraction) of infected and susceptible nodes with out-degree k_{out} by $I^{out}_{k_{out}}$ and $S^{out}_{k_{out}}$ $(i^{out}_{k_{out}}, s^{out}_{k_{out}})$ respectively (note the absence of the community in the superscript). Hence, we have $i^{out}_{k_{out}} = \frac{I^{out}_{k_{out}}}{N}$ and $s^{out}_{k_{out}} = \frac{S^{out}_{k_{out}}}{N}$. Note that in this case, the fraction is with respect to the total population N.

Finally, we denote the number (fraction) of infected and susceptible nodes in community H_i with in-degree k_{in} and out-degree k_{out} with $I^{H_i}_{k_{in},k_{out}}$ and $S^{H_i}_{k_{in},k_{out}}$ $(i^{H_i}_{k_{in},k_{out}}, s^{H_i}_{k_{in},k_{out}})$ respectively. In this case, we will define the fractions with respect to the total population N, hence we have $i^{H_i}_{k_{in},k_{out}} = \frac{I^{H_i}_{k_{in},k_{out}}}{N}$ and $s^{H_i}_{k_{in},k_{out}} = \frac{S^{H_i}_{k_{in},k_{out}}}{N}$.

5.2 Relations Between Classes of Nodes

Having defined the different classes of nodes, we need to define the relations between them. As we mentioned earlier, the mean-field analysis culminates in the derivation of the master equations. As we shall see, the level of analysis will be on classes of nodes in the same community with the same in-degree and out-degree. This means that we shall assume that nodes in each class defined by their community, in-degree and out-degree are statistically equivalent. Thus, in the master equations, we will examine the time evolution of the fraction of infected or susceptible nodes in each class.

However, on some occasions, we need the fraction of all the infected nodes in a specific community or a specific community with a specific in-degree or a specific out-degree. In this section, we show that these classes are related given the in-degree distribution for each community $p_{k_{in}}^{H_i}$ and the out-degree distribution $p_{k_{out}}^{out}$.

First, we also need to define the maximum out-degree in the network k_{out}^{max}, the maximum in-degree in community H_i, k_{in}^{max,H_i}, and the maximum in-degree in the network $k_{in}^{max} = \max_{H_i} k_{in}^{max,H_i}$. Lastly, by $\sum_{H_i}$ we shall denote the operation of summing over all the communities H_i of the network.

- The in-degree distribution in community H_i is defined as the fraction of nodes in community H_i with in-degree k_{in}:

$$p_{k_{in}}^{H_i} = \frac{\#\text{ nodes in } H_i \text{ with in-degree } k_{in}}{N_{H_i}}.$$

- The out-degree distribution for all $v \in G$ is defined as the fraction of nodes with out-degree k_{out} among all the nodes of the network:

$$p_{k_{out}}^{out} = \frac{\#\text{ nodes in with out-degree } k_{out}}{N}.$$

- The distribution of the nodes in each class defined by the community H_i and the in- and out-degrees k_{in} and k_{out} is:

$$p_{k_{in},k_{out}}^{H_i} = \frac{\#\text{ nodes in community } H_i \text{ with in-degree } k_{in} \text{ and out-degree } k_{out}}{N}.$$

- The number of infected nodes in community H_i with in-degree k_{in} is:

$$I_{k_{in}}^{H_i} = \sum_{k_{out}=0}^{k_{out}^{max}} I_{k_{in},k_{out}}^{H_i} = N \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}.$$

  Thus, the fraction of infected nodes in community H_i with in-degree k_{in} is:

$$i_{k_{in}}^{H_i} = \frac{I_{k_{in}}^{H_i}}{N_{H_i}} = \frac{N}{N_{H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}.$$

- The number of infected nodes in community H_i is:

$$I^{H_i} = \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} I_{k_{in},k_{out}}^{H_i} = \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} N i_{k_{in},k_{out}}^{H_i} = N \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}.$$

  Therefore, the fraction of infected nodes in community H_i is:

$$i^{H_i} = \frac{I^{H_i}}{N_H} = \frac{N}{N_{H_i}} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}.$$

- The number of infected nodes with out-degree k_{out} is:

$$I_{k_{out}}^{out} = \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} I_{k_{in},k_{out}}^{H_i} = N \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} i_{k_{in},k_{out}}^{H_i}.$$

  Therefore, the fraction of infected nodes with out-degree k_{out} in the network is:

$$i_{k_{out}}^{out} = \frac{I_{k_{out}}}{N} = \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} i_{k_{in},k_{out}}^{H_i}.$$

- The number of infected nodes in the whole graph G is:

$$I = \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} I_{k_{in},k_{out}}^{H_i} = N \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}.$$

  Thus, the fraction of infected nodes in the whole graph is:

$$i = \frac{I}{N} = \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}.$$

6 Mean-field Analysis

6.1 Preliminaries

In the mean-field analysis, we will derive a system of nonlinear Ordinary Differential Equations, called the master equations, that describe the evolution of the system over time. Specifically, we derive two differential equations per class of nodes where each class consists of all the nodes that belong to the same community H_i and have the same in-degree k_{in} and out-degree k_{out}. Thus, we shall assume that all nodes of each class behave epidemiologically in the same manner.

Since we examine the SIS model, infected nodes that recover become susceptible again. Therefore, the rate of change of the fraction of susceptible nodes of a target class equals minus the rate of change of the fraction of the infected nodes in the same set. Thus, it suffices to derive the differential equation for the fraction of infected in each class:

$$\frac{ds^{H_i}_{k_{in},k_{out}}}{dt} = -\frac{di^{H_i}_{k_{in},k_{out}}}{dt}.$$

For these functions, we have the constraints:

$$0 \leq s^{H_i}_{k_{in},k_{out}}, i^{H_i}_{k_{in},k_{out}} \leq p^{H_i}_{k_{in},k_{out}} \text{ and } s^{H_i}_{k_{in},k_{out}} + i^{H_i}_{k_{in},k_{out}} = p^{H_i}_{k_{in},k_{out}},\ \forall H_i, k_{in}, k_{out}$$

We first need to define an ordering of the node classes so that the differential equation can be ordered and written in a vector notation. In other words, we need to define a totally ordered set [5]. Thus, we define the array $\boldsymbol{i}$ which contains all the functions $i^{H_i}_{k_{in},k_{out}}$ ordered using the algorithm described below.

1. The communities are defined with an arbitrary ordering $H_1, H_2, \dots H_n$.
2. The functions are first ordered per community.
3. Then, they are ordered by their in-degree in their community.
4. Lastly, when the community and in-degree are the same, they are ordered by their out-degree.

The number of node classes, and hence the number of differential equations is $L = (k^{max}_{out} + 1)\sum_{H_i} k^{max,H_i}_{in}$. Hence, we have that $\boldsymbol{i} \in \mathbb{R}^L$.

Now, we can define the vector $\boldsymbol{i}$:

$$\boldsymbol{i} = \left[i^{H_1}_{1,0}, i^{H_1}_{1,1}, \dots i^{H_1}_{1,k^{max}_{out}}, i^{H_1}_{2,0} \dots i^{H_1}_{k^{max,H_1}_{in},k^{max}_{out}}, i^{H_2}_{1,0}, \dots, i^{H_n}_{k^{max,H_n}_{in},k^{max}_{out}}\right]^T.$$

Then, the system of differential equations can be written as:

$$\frac{d}{dt}\boldsymbol{i} = F(\boldsymbol{i}),$$

where F is a nonlinear function $\mathbb{R}^L \to \mathbb{R}^L$.

Each susceptible node can become infected either from a neighbour within the community or from a neighbour from a different community. We will examine each case separately and then we will combine the results.

6.2 Infections within a community

Let the in-degrees of the nodes within the community H_i be distributed according to $p_{k_{in}}^{H_i}$. Then, let $\theta_{H_i}^{in}(t)$ be the probability that a randomly chosen edge between nodes of the community H_i points to an infected individual. We have:

$$\begin{aligned}
\theta_{H_i}^{in}(t) &= \frac{\#\text{ edges pointing to an infected node in } H_i}{\#\text{ of edges in } H_i} \\
&= \frac{\sum_{k_{in}=1}^{k_{in}^{max,H_i}} k_{in} p_{k_{in}}^{H_i} i_{k_{in}}^{H_i}(t)}{\sum_{k_{in}=1}^{k_{in}^{max,H_i}} k_{in} p_{k_{in}}^{H_i}} = \frac{\sum_{k_{in}=1}^{k_{in}^{max,H_i}} k_{in} p_{k_{in}}^{H_i} i_{k_{in}}^{H_i}(t)}{\langle k \rangle_{H_i}} \\
&= \frac{\sum_{k_{in}=1}^{k_{in}^{max,H_i}} k_{in} p_{k_{in}}^{H_i} \frac{N}{N_{H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} i_{k_{in},k_{out}}^{H_i}(t)}{\langle k \rangle_{H_i}} \\
&= \frac{N}{N_{H_i}} \frac{\sum_{k_{in}=1}^{k_{in}^{max,H_i}} \sum_{k_{out}=0}^{k_{out}^{max}} k_{in} p_{k_{in}}^{H_i} i_{k_{in},k_{out}}^{H_i}(t)}{\langle k \rangle_{H_i}},
\end{aligned} \tag{10}$$

where $\langle k \rangle_{H_i}$ is the average in-degree in the community H_i.

Then, let $X_{k_{in}}^{H_i}$ be a random variable representing the number of infected neighbours within the community H_i of a node in the community H_i with in-degree k_{in}. This is a binomial random variable with success probability $\theta_{H_i}^{in}$ and number of trials the in-degree k_{in}. Hence:

$$P(X_{k_{in}}^{H_i} = s_{in}) = \binom{k_{in}}{s_{in}} (\theta_{H_i}^{in})^{s_{in}} (1 - \theta_{H_i}^{in})^{k_{in} - s_{in}}, \text{ for } 0 \leq s_{in} \leq k_{in}.$$

Similarly, we define the probability $\theta^{out}(t)$ that a randomly selected edge between nodes of different communities points to an infected node and $p_{k_{out}}^{out}$ the out-degree distribution in the network.

Then we have:

$$\begin{aligned}
\theta^{out}(t) &= \frac{\#\text{ edges between communities pointing to an infected node}}{\#\text{ of edges between communities}} \\
&= \frac{\sum_{k_{out}=0}^{k_{out}^{max}} k_{out} p_{k_{out}}^{out} i_{k_{out}}^{out}(t)}{\sum_{k_{out}=0}^{k_{out}^{max}} k_{out} p_{k}^{out}} = \frac{\sum_{k_{out}=0}^{k_{out}^{max}} k_{out} p_{k_{out}}^{out} i_{k_{out}}^{out}(t)}{\langle k \rangle_{out}} \\
&= \frac{\sum_{k_{out}=0}^{k_{out}^{max}} k_{out} p_{k_{out}}^{out} \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} i_{k_{in},k_{out}}^{H_i}(t)}{\langle k \rangle_{out}} \\
&= \frac{\sum_{k_{out}=0}^{k_{out}^{max}} \sum_{H_i} \sum_{k_{in}=1}^{k_{in}^{max,H_i}} k_{out} p_{k_{out}}^{out} i_{k_{in},k_{out}}^{H_i}(t)}{\langle k \rangle_{out}},
\end{aligned} \tag{11}$$

where $\langle k \rangle_{out}$ is the average out-degree in the network.

We also define $X_{k_{out}}^{out}$ to be a random variable denoting the number of infected neighbours outside the community of a node with out-degree k_{out}. This random variable is binomially distributed with success probability $\theta^{out}(t)$ and number of trials the out-degree of the node:

$$P(X_{k_{out}}^{out} = s_{out}) = \binom{k_{out}}{s_{out}} (\theta^{out}(t))^{s_{out}} (1 - \theta^{out}(t))^{k_{out} - s_{out}}, \text{ for } 0 \leq s_{out} \leq k_{out}.$$

A susceptible node j in the community H_i with degree $k = k_{in} + k_{out}$ and one infected neighbour is infected with rate $A_j\tau$. Since we defined the infection rate τ based on the nonlinear contagion scheme [25], this means that the node is infected with probability $A_j\tau$ per time unit. Therefore, the probability of not becoming infected is $1 - A_j\tau$ per infected neighbour per time unit. Since infections along each edge are assumed independent, the probability of not becoming infected when it has s_{in} infected neighbours in its community and s_{out} infected neighbours outside its community is:

$$P(\text{not infected from inside}|s_{in}, s_{out}) = (1 - A_j\tau)^{s_{in}} = (1 - \tau\phi_j^L\phi_{H_i}^C\phi^G)^{s_{in}}.$$

Note that the probability is dependent on both s_{in} and s_{out} because they both affect the local awareness ϕ_j^L.

6.3 Infections from outside the community

We have already defined $p_{k_{out}}^{out}, \theta^{out}(t)$ and $X_{k_{out}}^{out}$ in the previous section.

With a similar reasoning as in the previous section, the probability that a node j from community H_i with degree $k = k_{in} + k_{out}$, s_{out} infected neighbours outside the community and s_{in} infected neighbours within its community will not become infected from outside the community is:

$$P(\text{not infected from outside}|s_{in}, s_{out}) = (1 - A_j\tau)^{s_{out}} = (1 - \tau\phi_j^L\phi_{H_i}^C\phi^G)^{s_{out}}.$$

6.4 Combination of infections from inside and outside of the community

Consider a node j belonging to the community H_i with in-degree k_{in}, out-degree k_{out} infection density in the community i^{H_i} and global infection density i. The probability of not becoming infected depends on the number of infected neighbours from the same community $X_{k_{in}}^{H_i}$, the number of infected neighbours from outside the community $X_{k_{out}}^{out}$ as well as i^{H_i} and i. With i^{H_i} and i given, the probability that the node will not become infected in a time unit is:

$$\begin{aligned}
P_{\text{not infected}} &= \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [P(\text{not infected}|X_{k_{in}}^{H_i}, X_{k_{out}}^{out})|i^{H_i}, i] \\
&= \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [P(\text{not infected from inside}|X_{k_{in}}^{H_i}, X_{k_{out}}^{out}) P(\text{not infected from outside}|X_{k_{in}}^{H_i}, X_{k_{out}}^{out})|i^{H_i}, i] \\
&= \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [(1 - \tau\phi_j^L\phi_{H_i}^C\phi^G)^{X_{k_{in}}^{H_i}} (1 - \tau\phi_j^L\phi_{H_i}^C\phi^G)^{X_{k_{out}}^{out}} |i^{H_i}, i] \\
&= \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [(1 - \tau\phi_j^L\phi_{H_i}^C\phi^G)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}} |i^{H_i}, i] \\
&= \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [\big(1 - \tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}]
\end{aligned}$$

$$= \sum_{s_{out}=0}^{k_{out}} \sum_{s_{in}=1}^{k_{in}} P(X_{k_{out}}^{out} = s_{out}) P(X_{k_{in}}^{H_i} = s_{in}) \big(1 - \tau(1 - c_L \frac{s_{in} + s_{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)\big)^{s_{in}+s_{out}}.$$

In the above derivation we used the assumption that the probabilities of (not) getting infected from within and from outside the community are independent:

$$P(\text{not infected}|X_{k_{in}}^{H_i}, X_{k_{out}}^{out}) = P(\text{not infected from inside}|X_{k_{in}}^{H_i}, X_{k_{out}}^{out}) P(\text{not infected from outside}|X_{k_{in}}^{H_i}, X_{k_{out}}^{out}).$$

Since it is enough for a node to be infected from one infected neighbour in order to become infected, the probability of infection is:

$$\begin{aligned} P_{\text{infected}} &= 1 - P_{\text{not infected}} \\ &= 1 - \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [\big(1 - \tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}] \\ &= \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [1 - \big(1 - \tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}]. \end{aligned} \tag{12}$$

6.5 Derivation of the master equation

To facilitate the understanding of the reasoning, we shall start by examining the discrete time dynamics of the system, that is, what the state of the system is after one unit of time $t + 1$ given the status at the present time t. From the law of mass action, we have that the number of infected nodes in a specific class of nodes at time $t + 1$ equals the number of those who were infected at time t plus those susceptible at time t who became infected minus those infected at time t that recovered. Hence:

$$\begin{aligned} i_{k_{in},k_{out}}^{H_i}(t+1) &= i_{k_{in},k_{out}}^{H_i}(t) - \gamma i_{k_{in},k_{out}}^{H_i}(t) + s_{k_{in},k_{out}}^{H_i}(t) P_{\text{infected}} \overset{(12)}{\Longrightarrow} \\ i_{k_{in},k_{out}}^{H_i}(t+1) - i_{k_{in},k_{out}}^{H_i}(t) &= -\gamma i_{k_{in},k_{out}}^{H_i}(t) + (p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i}(t)) \\ &\quad \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [1 - \big(1 - \tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}]. \end{aligned}$$

Let us now move on to the continuous time dynamics and examine the infinitesimal time interval $(t, t + h]$ as in [33]. Since the amount of time until a transition (infection or recovery) is assumed to follow an exponential distribution, the probability of two or more transitions in time h is $o(h)$. Recall that a function $g(h) \in o(h)$ if $\lim_{h\to 0} \frac{g(h)}{h} = 0$. Thus, the probability that an infected node recovers at time h equals the probability that a transition occurs within time h plus a small value compared to h or $\gamma h + g_r(h)$ with $g_r(h) \in o(h)$, and $g_r(0) = 0$ since at time $h = 0$ no events can happen. Similarly, the probability that a susceptible node gets infected within time h is $h\tau A_i + g_i(h)$ per infected neighbour with $g_i(h) \in o(h)$ and $g_i(0) = 0$ [34].

Therefore, in this infinitesimal (as $h \to 0$) interval we have:

$$\begin{aligned} i_{k_{in},k_{out}}^{H_i}(t+h) - i_{k_{in},k_{out}}^{H_i}(t) &= -\gamma h i_{k_{in},k_{out}}^{H_i}(t) + g_r(h) + (p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i}(t)) \\ &\quad \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [1 - \big(1 - h\tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i) + g_i(h)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}]. \end{aligned} \tag{13}$$

We should note that, generally, $g_i(h)$ is going to be different for each value that $X_{k_{in}}^{H_i}$ and $X_{k_{out}}^{out}$ can take. Then, we divide both sides of the equation with h and take the limit for $h \to 0$:

$$\frac{di_{k_{in},k_{out}}^{H_i}}{dt} = -\gamma i_{k_{in},k_{out}}^{H_i} + (p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i})$$
$$\lim_{h\to 0} \frac{1}{h} \mathop{\mathbb{E}}_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}} [1 - \big(1 - h\tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i) + g_i(h)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}] \quad (14)$$

since $\lim_{h\to 0} \frac{g_r(h)}{h} = 0$ from $g_r(h) \in o(h)$.

Now, let us define the function:

$$f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(h) = 1 - \big(1 - h\tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i) + g_i(h)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}},$$

which is the expression inside the expectation as a function of h.

Note that the random variables $X_{k_{in}}^{H_i}, X_{k_{out}}^{out}$ are bounded with maximum values k_{in}, k_{out} respectively. Therefore, $f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(h)$ is analytic in the neighbourhood of $h = 0$. So, we can take the Taylor expansion of the function around 0:

$$f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(h) = f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(0) + h f'_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(0) + hR(h)$$

with $\lim_{h\to 0} R(h) = 0$.

At $h = 0$, we have:

$$f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(0) = 0.$$

The derivative is:

$$f'_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(h) = \big(\tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i) - g_i'(h)\big)(X_{k_{in}}^{H_i} + X_{k_{out}}^{out})$$
$$\big(1 - h\tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i) + g_i(h)\big)^{X_{k_{in}}^{H_i} + X_{k_{out}}^{out} - 1}.$$

Note that $\lim_{h\to 0} \frac{g_i(h)}{h} = 0 \implies g_i'(0) = 0$, so:

$$f'_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(0) = \tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)(X_{k_{in}}^{H_i} + X_{k_{out}}^{out}).$$

Thus, the Taylor expansion of $f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}$ around 0 is:

$$f_{X_{k_{in}}^{H_i}, X_{k_{out}}^{out}}(h) = h\tau(1 - c_L \frac{X_{k_{in}}^{H_i} + X_{k_{out}}^{out}}{k_{in} + k_{out}})(1 - c_C i^{H_i})(1 - c_G i)(X_{k_{in}}^{H_i} + X_{k_{out}}^{out}) + hR(h).$$

We shall now compute the remaining limit in equation (14) by substituting the Taylor expansion for $f_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}$. Since h is deterministic, $\frac{1}{h}$ can get inside the expectation. Moreover, as we mentioned, $X^{H_i}_{k_{in}}, X^{out}_{k_{out}}$ are bounded, hence $f_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}$ is bounded, and from the Dominated Convergence Theorem [37], we can interchange the limit and the expectation. So we have:

$$
\begin{aligned}
&\lim_{h\to 0}\frac{1}{h}\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[1-\big(1-\tau(1-c_L\frac{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}{k_{in}+k_{out}})(1-c_Ci^{H_i})(1-c_Gi)+g_i(h)\big)^{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}]\\
&=\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[\lim_{h\to 0}\frac{1-\big(1-\tau(1-c_L\frac{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}{k_{in}+k_{out}})(1-c_Ci^{H_i})(1-c_Gi)+g_i(h)\big)^{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}}{h}]\\
&=\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[\lim_{h\to 0}\frac{h\tau(1-c_L\frac{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}{k_{in}+k_{out}})(1-c_Ci^{H_i})(1-c_Gi)(X^{H_i}_{k_{in}}+X^{out}_{k_{out}})+hR(h)}{h}]\\
&=\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[\tau(1-c_L\frac{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}{k_{in}+k_{out}})(1-c_Ci^{H_i})(1-c_Gi)(X^{H_i}_{k_{in}}+X^{out}_{k_{out}})]\\
&=\tau(1-c_Ci^{H_i})(1-c_Gi)\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[(1-c_L\frac{X^{H_i}_{k_{in}}+X^{out}_{k_{out}}}{k_{in}+k_{out}})(X^{H_i}_{k_{in}}+X^{out}_{k_{out}})]\\
&=\tau(1-c_Ci^{H_i})(1-c_Gi)\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[(X^{H_i}_{k_{in}}+X^{out}_{k_{out}})-(X^{H_i}_{k_{in}}+X^{out}_{k_{out}})^2\frac{c_L}{k_{in}+k_{out}}]\\
&=\tau(1-c_Ci^{H_i})(1-c_Gi)\big(\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[X^{out}_{k_{out}}+X^{H_i}_{k_{in}}]-\frac{c_L}{k_{in}+k_{out}}\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[(X^{out}_{k_{out}}+X^{H_i}_{k_{in}})^2]\big).
\end{aligned}
$$

We can now return in (14) and substitute the limit we just calculated:

$$
\begin{aligned}
\frac{di^{H_i}_{k_{in},k_{out}}}{dt}=&-\gamma i^{H_i}_{k_{in},k_{out}}+(p^{H_i}_{k_{in},k_{out}}-i^{H_i}_{k_{in},k_{out}})\tau(1-c_Ci^{H_i})(1-c_Gi)\\
&\big(\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[X^{out}_{k_{out}}+X^{H_i}_{k_{in}}]-\frac{c_L}{k_{in}+k_{out}}\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[(X^{out}_{k_{out}}+X^{H_i}_{k_{in}})^2]\big). \quad (15)
\end{aligned}
$$

Moreover, we know that $X^{out}_{k_{out}}\sim Binomial(k_{out},\theta^{out})$ and $X^{H_i}_{k_{in}}\sim Binomial(k_{in},\theta^{H_i}_{in})$. Hence:

$$
\begin{aligned}
\mathbb{E}[X^{out}_{k_{out}}]&=k_{out}\theta^{out},\\
\mathbb{E}[X^{H_i}_{k_{in}}]&=k_{in}\theta^{H_i}_{in},\\
Var(X^{out}_{k_{out}})&=k_{out}\theta^{out}(1-\theta^{out}),\\
Var(X^{H_i}_{k_{in}})&=k_{in}\theta^{H_i}_{in}(1-\theta^{H_i}_{in}).
\end{aligned}
$$

So:

$$
\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}},X^{out}_{k_{out}}}[X^{out}_{k_{out}}+X^{H_i}_{k_{in}}]=\mathbb{E}[X^{out}_{k_{out}}]+\mathbb{E}[X^{H_i}_{k_{in}}]=k_{in}\theta^{H_i}_{in}+k_{out}\theta^{out}. \quad (16)
$$

Assuming $X^{out}_{k_{out}}$ and $X^{H_i}_{k_{in}}$ are independent, we also have:

$$\mathop{\mathbb{E}}_{X^{H_i}_{k_{in}}, X^{out}_{k_{out}}}[(X^{out}_{k_{out}} + X^{H_i}_{k_{in}})^2] = \mathop{\mathbb{E}}_{X^{H_i}_{k_{in}}, X^{out}_{k_{out}}}[(X^{out}_{k_{out}})^2 + (X^{H_i}_{k_{in}})^2 + 2X^{out}_{k_{out}} X^{H_i}_{k_{in}}]$$
$$= \mathbb{E}[(X^{out}_{k_{out}})^2] + \mathbb{E}[(X^{H_i}_{k_{in}})^2] + 2\,\mathbb{E}[X^{out}_{k_{out}}]\,\mathbb{E}[X^{H_i}_{k_{in}}]$$
$$= Var(X^{out}_{k_{out}}) + (\mathbb{E}[X^{out}_{k_{out}}])^2 + Var(X^{H_i}_{k_{in}}) + (\mathbb{E}[X^{H_i}_{k_{in}}])^2 + 2\,\mathbb{E}[X^{out}_{k_{out}}]\,\mathbb{E}[X^{H_i}_{k_{in}}]$$
$$= k_{out}\theta^{out}(1-\theta^{out}) + (k_{out}\theta^{out})^2 + k_{in}\theta^{H_i}_{in}(1-\theta^{H_i}_{in}) + (k_{in}\theta^{H_i}_{in})^2 + 2k_{out}\theta^{out}k_{in}\theta^{H_i}_{in}. \quad (17)$$

We can now substitute the expressions (16) and (17) as well as the relations (10) and (11) and the expressions for $i, i^{H_i}, i^{out}_{k_{out}}, i^{H_i}, i^{H_i}_{k_{in}}$ from Chapter 5 in the equation (15):

$$\frac{di^{H_i}_{k_{in},k_{out}}}{dt} = -\gamma i^{H_i}_{k_{in},k_{out}} + (p^{H_i}_{k_{in},k_{out}} - i^{H_i}_{k_{in},k_{out}})\tau(1 - c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}})(1 - c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}})$$
$$\big(k_{in}\theta^{H_i}_{in} + k_{out}\theta^{out} - \frac{c_L}{k_{in}+k_{out}}(k_{out}\theta^{out}(1-\theta^{out}) + (k_{out}\theta^{out})^2 + k_{in}\theta^{H_i}_{in}(1-\theta^{H_i}_{in}) + (k_{in}\theta^{H_i}_{in})^2 + 2k_{out}\theta^{out}k_{in}\theta^{H_i}_{in})\big)$$

or

$$\frac{di^{H_i}_{k_{in},k_{out}}}{dt} = -\gamma i^{H_i}_{k_{in},k_{out}} + (p^{H_i}_{k_{in},k_{out}} - i^{H_i}_{k_{in},k_{out}})\tau(1 - c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}})(1 - c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}})$$
$$\Big(k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}i^{H_i}_{s_{in}}}{\langle k\rangle_{H_i}} + k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}i^{out}_{s_{out}}}{\langle k\rangle_{out}}$$
$$- \frac{c_L}{k_{in}+k_{out}}\Big(k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}i^{out}_{s_{out}}}{\langle k\rangle_{out}}(1 - \frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}i^{out}_{s_{out}}}{\langle k\rangle_{out}})$$
$$+ (k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}i^{out}_{s_{out}}}{\langle k\rangle_{out}})^2 + (k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}i^{H_i}_{s_{in}}}{\langle k\rangle_{H_i}})^2$$
$$+ k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}i^{H_i}_{s_{in}}}{\langle k\rangle_{H_i}}(1 - \frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}i^{H_i}_{s_{in}}}{\langle k\rangle_{H_i}}) +$$
$$+ 2k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}i^{out}_{s_{out}}}{\langle k\rangle_{out}}k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}i^{H_i}_{s_{in}}}{\langle k\rangle_{H_i}}\Big)\Big)$$

or

$$\frac{di^{H_i}_{k_{in},k_{out}}}{dt} = -\gamma i^{H_i}_{k_{in},k_{out}} + (p^{H_i}_{k_{in},k_{out}} - i^{H_i}_{k_{in},k_{out}})\tau(1 - c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}})(1 - c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}})$$
$$\Big(k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}\frac{N}{N_{H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}}}{\langle k\rangle_{H_i}} + k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}} i^{H_j}_{s_{in},s_{out}}}{\langle k\rangle_{out}}$$
$$- \frac{c_L}{k_{in}+k_{out}}\Big(k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}} i^{H_j}_{s_{in},s_{out}}}{\langle k\rangle_{out}}(1 - \frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}} i^{H_j}_{s_{in},s_{out}}}{\langle k\rangle_{out}})$$
$$+ (k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p^{out}_{s_{out}}\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}} i^{H_j}_{s_{in},s_{out}}}{\langle k\rangle_{out}})^2 + (k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p^{H_i}_{s_{in}}\frac{N}{N_{H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}}}{\langle k\rangle_{H_i}})^2$$

$$+ k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p_{s_{in}}^{H_i}\frac{N}{N_{H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i}}{\langle k\rangle_{H_i}}(1-\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p_{s_{in}}^{H_i}\frac{N}{N_{H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i}}{\langle k\rangle_{H_i}})+$$
$$+ 2k_{out}\frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out}p_{s_{out}}^{out}\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}} i_{s_{in},s_{out}}^{H_j}}{\langle k\rangle_{out}}k_{in}\frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in}p_{s_{in}}^{H_i}\frac{N}{N_{H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i}}{\langle k\rangle_{H_i}})\Big). \quad (18)$$

6.6 Linearization

The equation (18) defines a system of nonlinear Ordinary Differential Equations that describes the evolution of the system. To proceed with our analytical reasoning, we shall establish a linearization of the system by omitting higher orders of $i_{k_{in},k_{out}}^{H_i}$.

For reasons of clarity, we shall proceed step by step. We first note that θ^{out} and $\theta_{in}^{H_i}$ contain a sum of $i_{k_{in},k_{out}}^{H_i}$. Therefore, the terms $(k_{out}\theta^{out})^2$ and $(k_{in}\theta_{in}^{H_i})^2$ should be omitted. Moreover, we have that $(k_{out}\theta^{out}(1-\theta^{out}) \approx k_{out}\theta^{out}$ and $k_{in}\theta_{in}^{H_i}(1-\theta_{in}^{H_i}) \approx k_{in}\theta_{in}^{H_i}$. Similarly, we omit the term $k_{out}\theta^{out}k_{in}\theta_{in}^{H_i}$. This leaves us with:

$$\frac{di_{k_{in},k_{out}}^{H_i}}{dt} = -\gamma i_{k_{in},k_{out}}^{H_i} + (p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i})\tau(1-c_C\frac{N}{N_{H_i}}\sum_{k_{in}}\sum_{k_{out}} i_{k_{in},k_{out}}^{H_i})(1-c_G\sum_{H_j}\sum_{s_{in}}\sum_{s_{out}} i_{s_{in},s_{out}}^{H_j})$$
$$(k_{in}\theta_{in}^{H_i} + k_{out}\theta^{out} - \frac{c_L}{k_{in}+k_{out}}k_{in}\theta_{in}^{H_i} - \frac{c_L}{k_{in}+k_{out}}k_{out}\theta^{out})$$
$$= -\gamma i_{k_{in},k_{out}}^{H_i} + (p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i})\tau(1-c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i})(1-c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_j})$$
$$(1-\frac{c_L}{k_{in}+k_{out}})(k_{in}\theta_{in}^{H_i} + k_{out}\theta^{out}).$$

Then, we shall expand the parentheses:

$$\frac{di_{k_{in},k_{out}}^{H_i}}{dt} = -\gamma i_{k_{in},k_{out}}^{H_i} + (p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i})\tau(1-c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i})(1-c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_j})$$
$$(1-\frac{c_L}{k_{in}+k_{out}})(k_{in}\theta_{in}^{H_i} + k_{out}\theta^{out})$$
$$= -\gamma i_{k_{in},k_{out}}^{H_i} + \tau(1-\frac{c_L}{k_{in}+k_{out}})$$
$$(p_{k_{in},k_{out}}^{H_i} - i_{k_{in},k_{out}}^{H_i} - p_{k_{in},k_{out}}^{H_i}c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i} + i_{k_{in},k_{out}}^{H_i}c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i}$$
$$- p_{k_{in},k_{out}}^{H_i}c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_j} + i_{k_{in},k_{out}}^{H_i}c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_j}$$
$$+ p_{k_{in},k_{out}}^{H_i}c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i}c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_j}$$
$$- i_{k_{in},k_{out}}^{H_i}c_C\frac{N}{N_{H_i}}\sum_{s_{in}=1}^{k_{in}^{max,H_i}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_i}c_G\sum_{H_j}\sum_{s_{in}=1}^{k_{in}^{max,H_j}}\sum_{s_{out}=0}^{k_{out}^{max}} i_{s_{in},s_{out}}^{H_j})(k_{in}\theta_{in}^{H_i} + k_{out}\theta^{out})$$
$$= -\gamma i_{k_{in},k_{out}}^{H_i} + \tau(1-\frac{c_L}{k_{in}+k_{out}})$$

$$
\begin{aligned}
&(p^{H_i}_{k_{in},k_{out}} k_{in}\theta^{H_i}_{in} - k_{in}\theta^{H_i}_{in} i^{H_i}_{k_{in},k_{out}} - p^{H_i}_{k_{in},k_{out}} k_{in}\theta^{H_i}_{in} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} \\
&+ k_{in}\theta^{H_i}_{in} i^{H_i}_{k_{in},k_{out}} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} \\
&- p^{H_i}_{k_{in},k_{out}} k_{in}\theta^{H_i}_{in} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} + k_{in}\theta^{H_i}_{in} i^{H_i}_{k_{in},k_{out}} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} \\
&+ p^{H_i}_{k_{in},k_{out}} k_{in}\theta^{H_i}_{in} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} \\
&- k_{in}\theta^{H_i}_{in} i^{H_i}_{k_{in},k_{out}} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} \\
&+ p^{H_i}_{k_{in},k_{out}} k_{out}\theta^{out} - k_{out}\theta^{out} i^{H_i}_{k_{in},k_{out}} - p^{H_i}_{k_{in},k_{out}} k_{out}\theta^{out} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} \\
&+ k_{out}\theta^{out} i^{H_i}_{k_{in},k_{out}} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} \\
&- p^{H_i}_{k_{in},k_{out}} k_{out}\theta^{out} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} + k_{out}\theta^{out} i^{H_i}_{k_{in},k_{out}} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} \\
&+ p^{H_i}_{k_{in},k_{out}} k_{out}\theta^{out} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}} \\
&- k_{out}\theta^{out} i^{H_i}_{k_{in},k_{out}} c_C \frac{N}{N_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}} c_G \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_j}_{s_{in},s_{out}}).
\end{aligned}
$$

Now, we shall omit any factor that contains a higher order of $i^{H_i}_{k_{in},k_{out}}$, considering the fact that $\theta^{H_i}_{in}$ and θ^{out} contain multiple factors $i^{H_i}_{k_{in},k_{out}}$ in their formula. Then, the final linearization is:

$$
\begin{aligned}
&\frac{di^{H_i}_{k_{in},k_{out}}}{dt} = -\gamma i^{H_i}_{k_{in},k_{out}} + \tau p^{H_i}_{k_{in},k_{out}} (1 - \frac{c_L}{k_{in}+k_{out}})(k_{in}\theta^{H_i}_{in} + k_{out}\theta^{out}) \\
&= -\gamma i^{H_i}_{k_{in},k_{out}} + \tau p^{H_i}_{k_{in},k_{out}} (1 - \frac{c_L}{k_{in}+k_{out}})(k_{in} \frac{\sum_{s_{in}=1}^{k_{in}^{max,H_i}} s_{in} p^{H_i}_{k_{in}} \frac{N}{N_{H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} i^{H_i}_{s_{in},s_{out}}}{< k >_{H_i}} \\
&+ k_{out} \frac{\sum_{s_{out}=0}^{k_{out}^{max}} s_{out} p^{out}_{k_{out}} \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} i^{H_j}_{s_{in},s_{out}}}{< k >_{out}})
\end{aligned}
$$

or

$$
\begin{aligned}
&\frac{di^{H_i}_{k_{in},k_{out}}}{dt} = -\gamma i^{H_i}_{k_{in},k_{out}} + \tau p^{H_i}_{k_{in},k_{out}} (1 - \frac{c_L}{k_{in}+k_{out}}) \\
&(\frac{N k_{in}}{N_{H_i} < k >_{H_i}} \sum_{s_{in}=1}^{k_{in}^{max,H_i}} \sum_{s_{out}=0}^{k_{out}^{max}} s_{in} p^{H_i}_{s_{in}} i^{H_i}_{s_{in},s_{out}} + \frac{k_{out}}{< k >_{out}} \sum_{H_j} \sum_{s_{in}=1}^{k_{in}^{max,H_j}} \sum_{s_{out}=0}^{k_{out}^{max}} s_{out} p^{out}_{s_{out}} i^{H_j}_{s_{in},s_{out}}). \qquad (19)
\end{aligned}
$$

We notice that the only awareness coefficient that remains in the equation after linearization is the local coefficient c_L. Note that this result follows because we defined the awareness as a linear multiplicative function; other awareness models may lead to linearized master equations that preserve the community and global awareness coefficients.

7 Simulation Study

In this chapter, we shall discuss the methodology of the simulations conducted in the course of the research. We commence by explaining how the Gillespie algorithm was adjusted so that it includes infection awareness. Afterwards, we shall present the methodology that was used to approximate the epidemic threshold using simulations.

7.1 Adjusted Gillespie Algorithm

The version of the Gillespie algorithm that was used in the current project was adapted from the SIR simulation algorithm presented in [19]. We transformed it into the SIS simulation algorithm by replacing the final R state with the S state. Furthermore, as the name of the project suggests, the most important modification was the addition of infection awareness which was calculated every time a new infection or recovery was recorded and affected the infection rates of all the nodes in the network. The pseudo-code that was used is presented in Algorithm 1.

7.2 Calculation of the Epidemic Threshold

Much of the work performed in this project centres around the approximation of the epidemic threshold, which was defined above for the SIS model as the critical infection rate τ_c such that for $\tau > \tau_c$ we reach the endemic equilibrium and for $\tau \leq \tau_c$ we reach the infection-free steady-state.

As we mentioned in the literature review, there are two types of finite-size effects that may lead to overestimating the epidemic threshold using stochastic simulations. One of the causes is the degree sequence cut-off in scale-free networks, a type of network that was not examined during this work. The second cause affects all finite networks with an absorbing state, meaning that our case studies are susceptible to it. The cause of the problem is that once a realization reaches the infection-free absorbing state, the outbreak cannot be revived. The stochastic nature of the simulations causes fluctuations that may drive the system to the absorbing state, even for infection rates $\tau > \tau_c$. Therefore, this may lead to an overestimation of the threshold.

In the literature, we saw that the most common method for addressing this issue is to only take into consideration the surviving runs [8] [9]. In our case studies, we shall follow this method. Specifically, we shall run each configuration with a specific infection rate τ until we have r_1 realizations that are not in the infection-free steady-state. However, when the infection rate τ is below the epidemic threshold, all the realizations will reach the infection-free equilibrium in the long run. Thus, we need an upper bound $r_2 > r_1$ for the number of realizations that we will run to avoid simulating ad infinitum. After running r_2 realizations, the simulations will stop, and the τ under consideration will be considered lower than the epidemic threshold. To reduce the probability of a supercritical system reaching the infection-free equilibrium, we initialized the simulations with all the nodes in the system infected.

After executing r_1 surviving runs, we calculate the average infection density. To further reduce the fluctuations within each realization, we averaged over the last t_{last} time steps. To make sure that we have reached the equilibrium of each realization, we let it unfold for adequate time steps t_{max}. Finally, we compare the average infection density $\langle i \rangle$ with a given threshold as in the literature. In [35], the threshold was set to $\theta = 0.0025$ while in [41] it was set to $\theta = 0.0005$. In our case, we shall set it to the higher one $\theta = 0.0025$. It should be noted that in some of the specific network cases that were examined, the threshold is surpassed even with one infected node. Finally, if the τ under consideration is not found to be the τ_c, we repeat the process by incrementing τ by a predefined step.

Data: Network G, per-edge infection rate τ, recovery rate γ, $initial\_infections$, maximum time t_{max}, awareness function A

Result: array of timestamps $times$, S and I containing the number of nodes in each state at each time in $times$ and $I^H_{k_{in},k_{out}}$, $S^H_{k_{in},k_{out}}$ containing the number of nodes in each node class at each time in $times$.

$times$, S, $I \leftarrow$ [0], [$|G|$ - len($initial\_infections$)], [len($initial\_infections$)];
$infected\_nodes \leftarrow initial\_infections$;
$at\_risk\_nodes \leftarrow$ uninfected nodes with infected neighbour;
for each node u in at_risk_nodes **do**
  $infection\_rate[u] \leftarrow \tau \times$ # infected neighbours of u;
end
calculate $awareness[v]$ for each node v in at_risk_nodes depending on the number of infected nodes in its neighbourhood, community and network and the awareness function A;
$effective\_infection\_rate \leftarrow awareness \times infection\_rate$;
$total\_infection\_rate \leftarrow \sum_{u \in at\_risk\_nodes} effective\_infection\_rate[u]$;
$total\_recovery\_rate \leftarrow \gamma \times$ len($infected\_nodes$);
$total\_rate \leftarrow total\_recovery\_rate + total\_infection\_rate$;
$time \leftarrow exponential\_distribution(total\_rate)$
while $time < t_{max}$ and $total\_rate > 0$ **do**
  $r \leftarrow uniform\_random(0, total\_rate)$;
  if $r < total\_recovery\_rate$ **then**
    $u \leftarrow random\_choice(infected\_nodes)$;
    remove u from $infected\_nodes$;
    update $infection\_rate[v]$ for all susceptible neighbours v of u;
  end
  else
    choose u from at_risk_nodes with probability $\frac{effective\_infection\_rate[u]}{total\_infection\_rate}$;
    remove u from at_risk_nodes;
    add u to $infected\_nodes$;
    for susceptible neighbours v of u **do**
      if v not in at_risk_nodes **then**
        add v to at_risk_nodes;
      end
      update $infection\_rate[v]$;
    end
  end
  add $time$ to $times$;
  update S, I and $I^H_{k_{in},k_{out}}$, $S^H_{k_{in},k_{out}}$ for each node class;
  update $awareness[v]$ for each node v in at_risk_nodes depending on the number of infected nodes in its neighbourhood, community and network and the awareness function A;
  $effective\_infection\_rate \leftarrow awareness \times infection\_rate$;
  $total\_infection\_rate \leftarrow \sum_{u \in at\_risk\_nodes} effective\_infection\_rate[u]$;
  $total\_recovery\_rate \leftarrow \gamma \times$ len($infected\_nodes$);
  $total\_rate \leftarrow total\_recovery\_rate + total\_infection\_rate$;
  $time \leftarrow time + exponential\_distribution(total\_rate)$
end

Algorithm 1: Gillespie Algorithm with Awareness

8 Results

In this chapter, we shall perform numerical experiments in the form of simulations and test whether the results agree with the predictions of the mean-field analysis. The mean-field master equations will be solved numerically because of the non-linearity of the system. The comparison will be performed on multiple levels. First, we will plot a simulated outbreak and compare the findings with the results of the mean-field approximation. Then, we will examine the relationship between the awareness coefficients and the epidemic prevalence. Lastly, we shall compare the epidemic threshold as approximated using simulations and mean-field analysis.

Without loss of generality, in the following analysis, we shall set the recovery rate $\gamma = 1$. This is equivalent to rescaling the time such that one time unit is equal to the recovery time.

Because of the complicated nonlinear nature of the master equations (18), we shall restrict ourselves to computing the steady-state solution of the system. As we saw, the SIS system has two equilibria: a disease-free steady-state and an endemic steady-state. The condition for the equilibrium is:

$$\frac{di^{H_i}_{k_{in},k_{out}}}{dt} = 0, \text{ with } 0 \leq i^{H}_{k_{in},k_{out}} \leq p^{H}_{k_{in},k_{out}} \; \forall H_i, k_{in}, k_{out} \text{ in the network.}$$

This leads to a nonlinear system of equations with restrictions which can be solved using the usual equation solvers. In the following analysis, we used the nonlinear least-square equation solver [18] from the SciPy library with the default options [39]. The initial guess was set by default to the maximum value, that is, $i^{H}_{k_{in},k_{out}} = p^{H}_{k_{in},k_{out}}$. When the least-squares failed to converge to a solution with this initial guess, we used different values by trial and error and chose the one that produced the lowest value for the cost function. Then, the epidemic prevalence was calculated by:

$$i_{ss} = \sum_{\forall H, k_{in}, k_{out}} i^{H}_{k_{in},k_{out}}.$$

To estimate the epidemic prevalence using stochastic simulations, we simulated the development of the outbreak for $t_{max} = 100$ time steps because it was empirically noticed that this is enough time for the system to reach the steady-state. The system was initialized with 10% of the population infected, selected randomly. To reduce stochastic fluctuation, we ran the system for $r = 10$ realizations and averaged the final epidemic prevalence. For the same reason, for each realization, we averaged over the $t_{last} = 10$ last time steps.

In order to estimate the epidemic threshold using simulations, we ran each configuration until we had $r_1 = 10$ realizations that did not reach the infection-free equilibrium (surviving runs method) or $r_2 = 20$ realizations in total. The parameters r_1 and r_2 were selected based on practical considerations regarding the execution time of the algorithm. Then, we averaged over the $t_{last} = 10$ last time steps of each run to calculate the average infection density. Each realization was allowed to reach a steady-state by being executed for $t_{max} = 100$ time steps, a temporal horizon that was empirically verified to be adequate for the system to reach an equilibrium. However, in many cases, the system has reached the infection-free equilibrium earlier. Each realization was initialized with all the population being infected to reduce the finite-size effects, as explained in Chapter 7. This process was repeated for gradually increasing infection rates τ from 0 to 1 with a step of 0.05. The

epidemic threshold was the smallest τ for which the final average infection density over the surviving runs was above the threshold $\theta = 0.0025$, similar to the threshold used in the literature [35].

In the case of the mean-field approximation, we found the steady-state infection density for the system with increasing infection rates τ from 0 to 1 with a step of 0.05. The threshold was the smallest infection rate τ_c that led to a steady-state infection density that surpassed the same threshold $\theta = 0.0025$.

Below, we present the results of this analysis in three network case studies. Each network consists of several identical communities, while the degree distributions are simple, all have in-degrees $k_{in} = 2$ and out-degrees $k_{out} = 1$. Note that the real degrees may differ from the prescribed ones because of the HCM generation method that we used, as we explained in Chapter 4. The first network, G_1, consists of a small number of relatively sparse communities, which are large compared to the size of the network. The second network, G_2, consists of $n = \sqrt{N}$ communities with $\sqrt{N}$ nodes each, with N the number of nodes in the network; that is, we have that $n \to \infty$ as $N \to \infty$. The third network, G_3, consists of a large number of small communities, which are denser than the communities of the former two graphs. The defining difference of the three networks is the size of the communities relative to the network, large in G_1, growing with the size of the network in G_2, and small in G_3. The driving factor behind the choice of these three case studies was to investigate whether the effects of community awareness depend on the size of the communities.

8.1 Network G_1 - Large Communities

We start by examining a simple network, which we shall call G_1. The network G_1 consists of $n = 3$ communities, each with $N_{H_1} = N_{H_2} = N_{H_3} = 100$ nodes for a total of $N = 300$ nodes. All the nodes in the network have in-degree $k_{in} = 2$ and out-degree $k_{out} = 1$. However, because of the network generation method, the final network may contain nodes that deviate from the prescribed degrees.

G_1 is a relatively sparse network with denseness:

$$\delta_{G_1} = \frac{\#\text{ edges}}{\#\text{ possible edges}} = \frac{\frac{1}{2}\sum_{v\in G_1} d_v}{\frac{1}{2}N(N-1)} = \frac{900}{89700} \approx 0.0100.$$

Moreover, the denseness coefficients of the three communities are all equal to each other $\delta_{com}^{H_1} = \delta_{com}^{H_2} = \delta_{com}^{H_3} = \delta_{com}^{G_1}$ and they are

$$\delta_{com}^{G_1} = \frac{\frac{1}{2}\sum_{v\in H_1} d_v^{in}}{\frac{1}{2}N_{H_1}(N_{H_1}-1)} = \frac{200}{9900} \approx 0.0202.$$

Note that these denseness coefficients were calculated on the prescribed degree distributions. The real degree distributions, and therefore the denseness coefficients, may vary.

In Figure 3, we present the fractions of infected nodes per community[3] $\frac{I^H}{N}$ as well as the total fraction of infected nodes $i = \frac{I}{N}$ for the network G_1 with $\tau = 1$ and $c_L = c_C = c_G = 0.5$ and $t_{max} = 150$ time steps as observed by a run of a stochastic simulation. We initialized the system with 10% of the nodes infected, randomly selected. As we observe, the infection rate $\tau = 1$ appears to be above the epidemic threshold since we have reached an endemic steady-state. The epidemic prevalence is calculated from the simulations as the average fraction of infected over the last 10 time steps (to reduce stochastic fluctuations). This is

[3]This should not be confused with the infection density per community $i^H = \frac{I^H}{N_H}$.

equal to $i_{ss} = 0.483$, which is not far from the mean-field approximation of 0.469. Naturally, as the communities are defined identically, the plots for each community overlap. The prevalence per community is calculated similarly to 0.163, 0.150 and 0.166 for communities 1, 2 and 3, respectively. This agrees with the mean-field approximation of the steady-state infected fraction, which is also approximately 0.15.

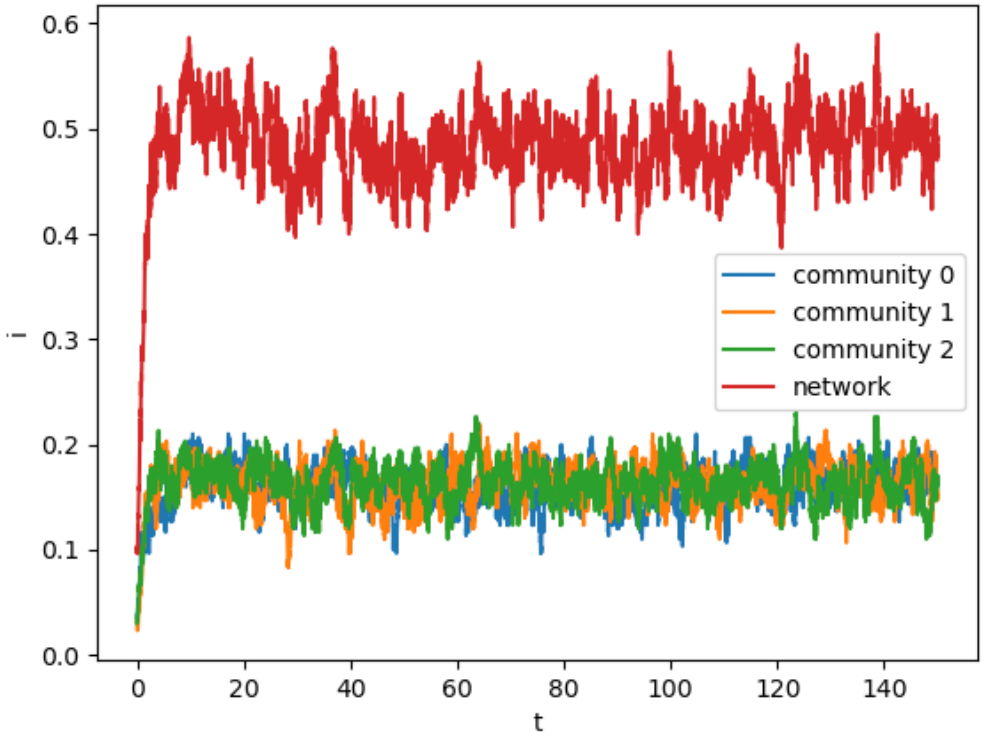

FIGURE 3: Fraction of infected per community for the network G_1 with $\tau = 1$ and $c_L = c_C = c_G = 0.5$

In Figure 4, we present how the results of stochastic simulations compare with the results of the mean-field solution for the network G_1 when it comes to the epidemic prevalence. As we can observe, we approximated the epidemic prevalence for gradually increasing awareness coefficients from 0 to 1 by a step of 0.1. In the top subfigure, we kept the community and global awareness coefficients constant to $c_C = c_G = 0$ while increasing the local awareness coefficient c_L. In the middle subfigure, we performed the same operation by keeping constant $c_L = c_G = 0$ and increasing the community awareness coefficient c_C, while at the bottom subfigure, we kept $c_L = c_C = 0$ and varied the global coefficient c_G. The infection rate was set to $\tau = 1$ and the simulations were initialized with 10% of the population infected.

We notice that in all cases, the numerical solution of the mean-field equations overestimated the epidemic prevalence. However, in all cases, the prevalence follows the same downward trend as the awareness coefficients increase. In Table 1, we present the Mean Squared Errors of the three graphs.

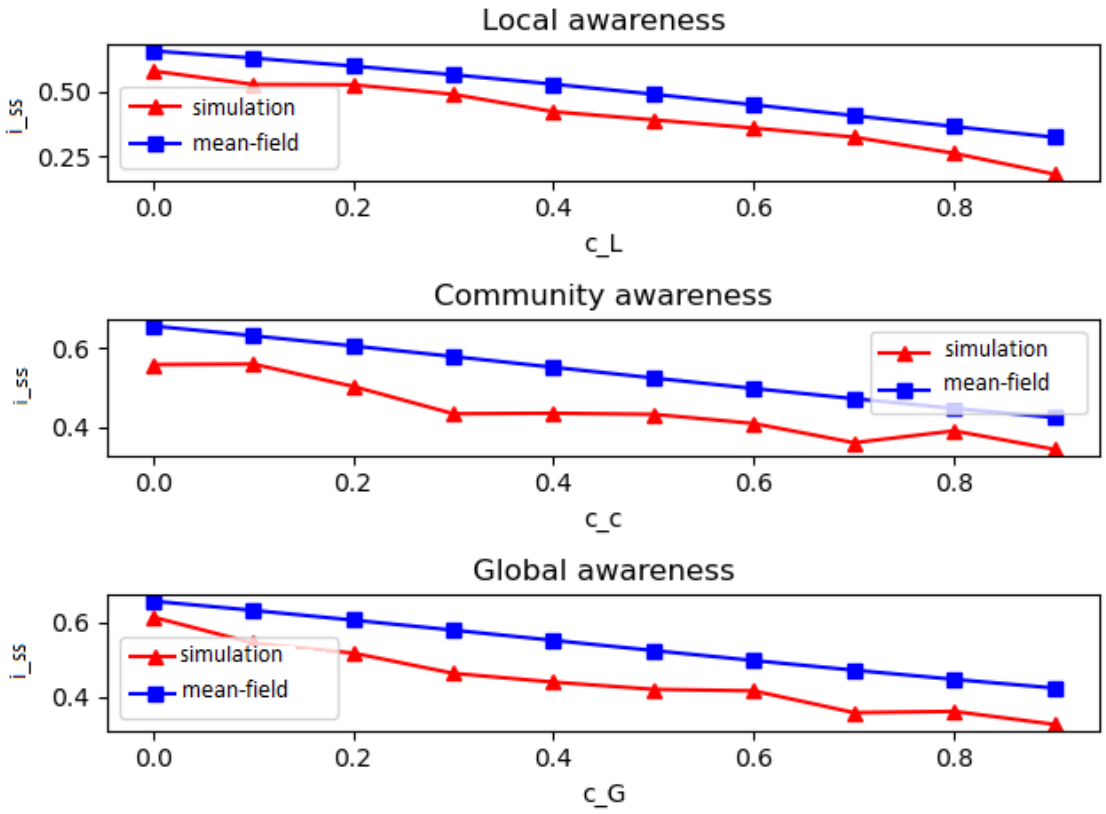

FIGURE 4: Epidemic prevalence i_{ss} as a function of varying awareness coefficients in the network G_1 with $\tau = 1$. In the top subfigure we set $c_C = c_G = 0$ and varied c_L from 0 to 1 by a step of 0.1. In the middle subfigure we kept $c_L = c_G = 0$ and varied c_C and in the bottom one we set $c_L = c_C = 0$ and varied c_G.

TABLE 1: MSE between mean-field approximations and simulations of the epidemic prevalence on the network G_1 while varying c_L, c_C and c_G.

Varying coefficient	c_L	c_C	c_G
Mean Square Error	0.0093	0.0098	0.0089

In Figure 5, we show a comparison of the mean-field approximation for the epidemic threshold and the threshold approximated through the use of simulations for the network G_1. In the top subfigure, we gradually increased c_L from 0 to 1 with a step of 0.1 while keeping $c_C = c_G = 0$. In the middle subfigure, we similarly varied c_C while keeping $c_L = c_G = 0$ and in the bottom one we varied c_G while setting $c_L = c_C = 0$.

We observe that the mean-field approximation consistently underestimates the epidemic threshold. Moreover, we notice that the value of the epidemic threshold increases as the local awareness c_L increases while the community and global awareness coefficients c_C and c_G do not seem to affect it. In Table 2, we present the Mean Squared Errors between the mean-field approximations and the simulations.

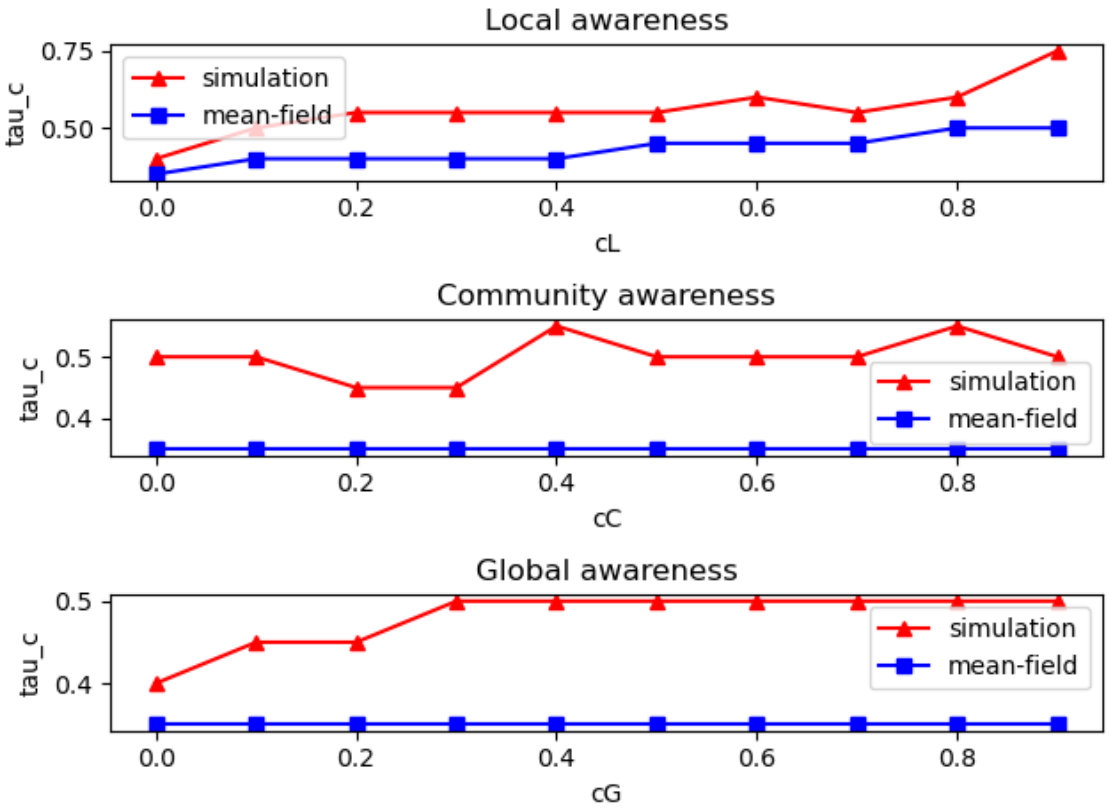

FIGURE 5: Epidemic threshold as a function of the awareness coefficients in the network G_1. In the top subfigure, we set $c_C = c_G = 0$ and varied c_L from 0 to 1 with a step of 0.1. In the middle subfigure, we set $c_L = c_G = 0$ and varied c_C while in the bottom subfigure, we set $c_L = c_C = 0$ and increased c_G.

TABLE 2: MSE of between mean-field approximations and simulations of the epidemic threshold on the network G_1 while varying c_L, c_C and c_G.

Varying coefficient	c_L	c_C	c_G
Mean Square Error	0.0195	0.0235	0.0235

8.2 Network G_2 - $\sqrt{N}$ Communities with $\sqrt{N}$ nodes each

Next, we examine a more complicated network configuration. We shall follow the rule that a network with N nodes consists of $\sqrt{N}$ communities each having $\sqrt{N}$ nodes. In our case, we choose $N = 30^2 = 900$, so the network, which we shall call G_2 contains $n = 30$ communities each with $N_{H_1} = \cdots = N_{H_{30}} = 30$ nodes. Once more, all nodes have in-degree $k_{in} = 2$ and out-degree $k_{out} = 1$, with the note that using the HCM generation algorithm from Chapter 4 may cause some small deviations from the rule.

The denseness of G_2 is

$$\delta_{G_2} = \frac{\frac{1}{2}\sum_{v \in G_2} d_v}{\frac{1}{2}N(N-1)} = \frac{2700}{809100} = 0.0033.$$

Since the communities are identical to each other, the denseness of the communities are all equal to each other $\delta_{com}^{H_1} = \cdots = \delta_{com}^{H_{30}} = \delta_{com}^{G_2}$ and they are equal to

$$\delta_{com}^{G_2} = \frac{\frac{1}{2}\sum_{v\in H_1} d_v^{in}}{\frac{1}{2}N_{H_1}(N_{H_1}-1)} = \frac{60}{870} = 0.0690.$$

We observe that the network G_2 is less dense than G_1, but the communities of G_2 are denser than those of G_1.

In Figure 6, we present the fraction of infected nodes per community $\frac{I^H}{N}$ as well as the total infection density i in the network G_2 for $\tau = 1$ and $c_L = c_C = c_G = 0.5$ and $t_{max} = 150$ time steps as observed by a Gillespie simulation. Once more, we initialized the system with 10% of the nodes infected, selected randomly. We observe that the system reached the endemic equilibrium (hence $\tau = 1 > \tau_c$) and the plots of the infections per community largely overlap. The prevalence calculated using the simulations (averaged over the last 10 time steps) is $i_{ss} = 0.505$ while the mean-field approximation resulted in $i_{ss} = 0.477$. The fraction of infected nodes per community was of the order of 0.016 in both the mean-field and the simulations method, with differences mostly in the third decimal.

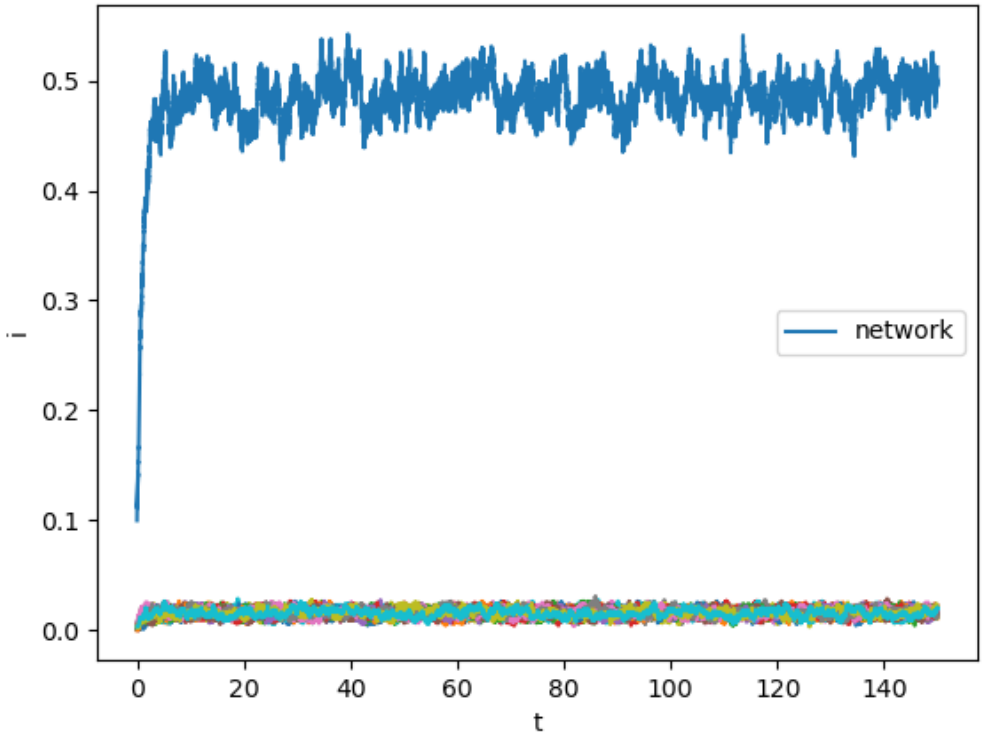

FIGURE 6: Fraction of infected per community for the network G_2 with $\tau = 1$ and $c_L = c_C = c_G = 0.5$

In Figure 7, we show the epidemic prevalence in the network as a function of the awareness coefficients. In each subplot, we set two of the three coefficients equal to zero and vary the next one with a step of 0.1 from 0 to 1. The infection rate was set to $\tau = 1$ and the simulations were initialized with 10% of the population infected. Once more, we notice that the awareness coefficients have a monotone decreasing relationship with the epidemic prevalence, a trend that is followed by both the mean-field and the simulation

approximation. The simulation approximation is consistently lower than the mean-field. In Table 3, we see the Mean Squared Error of the two approximations.

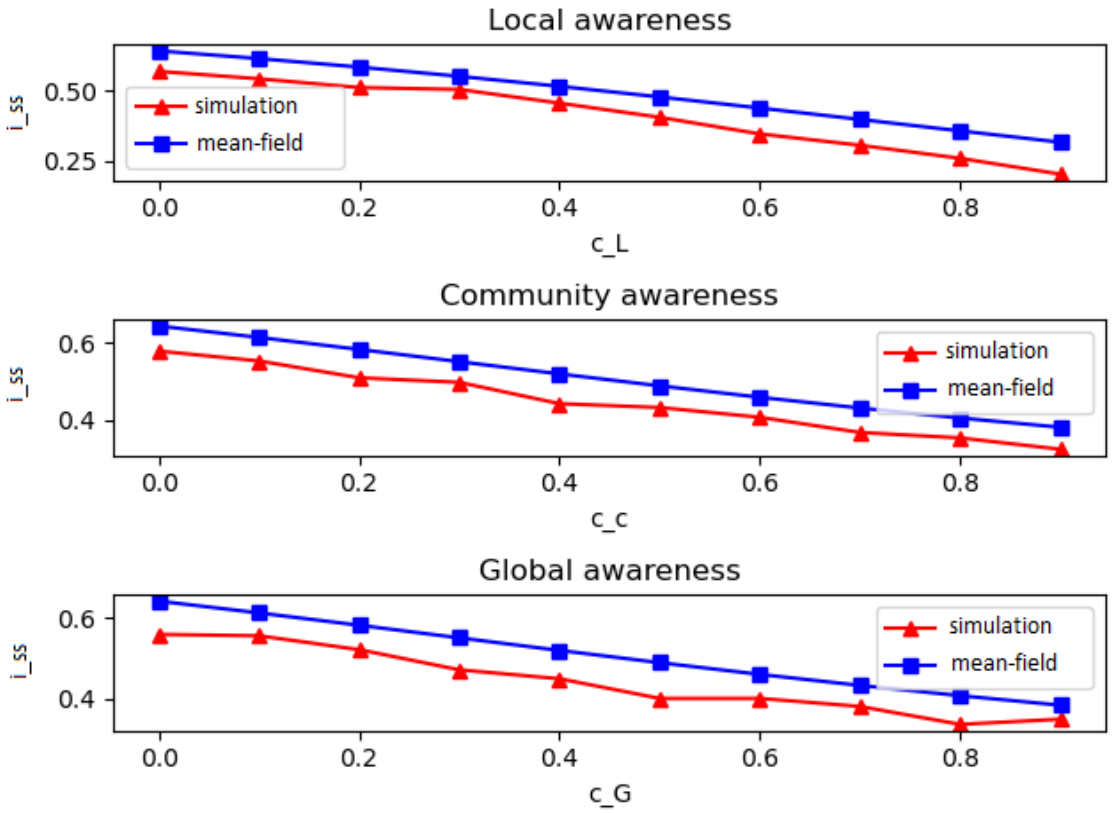

FIGURE 7: Epidemic prevalence i_{ss} as a function of the awareness coefficients on the network G_2 with $\tau = 1$. In the top subplot, we set $c_C = c_G = 0$ and vary c_L from 0 to 1 with a step of 0.1. In the middle subplot we repeat the process with $c_L = c_G = 0$ and varying c_C and on the bottom with $c_L = c_C = 0$ and varying c_G.

TABLE 3: MSE between mean-field approximations and simulations of the epidemic prevalence on the network G_2 while varying c_L, c_C and c_G.

Varying coefficient	c_L	c_C	c_G
Mean Square Error	0.0067	0.0037	0.0046

In Figure 8, we compare the mean-field approximation for the epidemic threshold with the threshold that was approximated by simulations for the network G_2. In each subfigure, we set two out of the three awareness coefficients constant and varied the third from 0 to 1 with a step of 0.1. In Table 4, we see the Mean Squared Error of the two approximations. Once more, we remark that the epidemic threshold seems to increase as c_L increases, a trend that is not apparent in the cases of c_C and c_G. In addition, we notice that the mean-field approximation of the threshold is consistently smaller than the approximation by simulations.

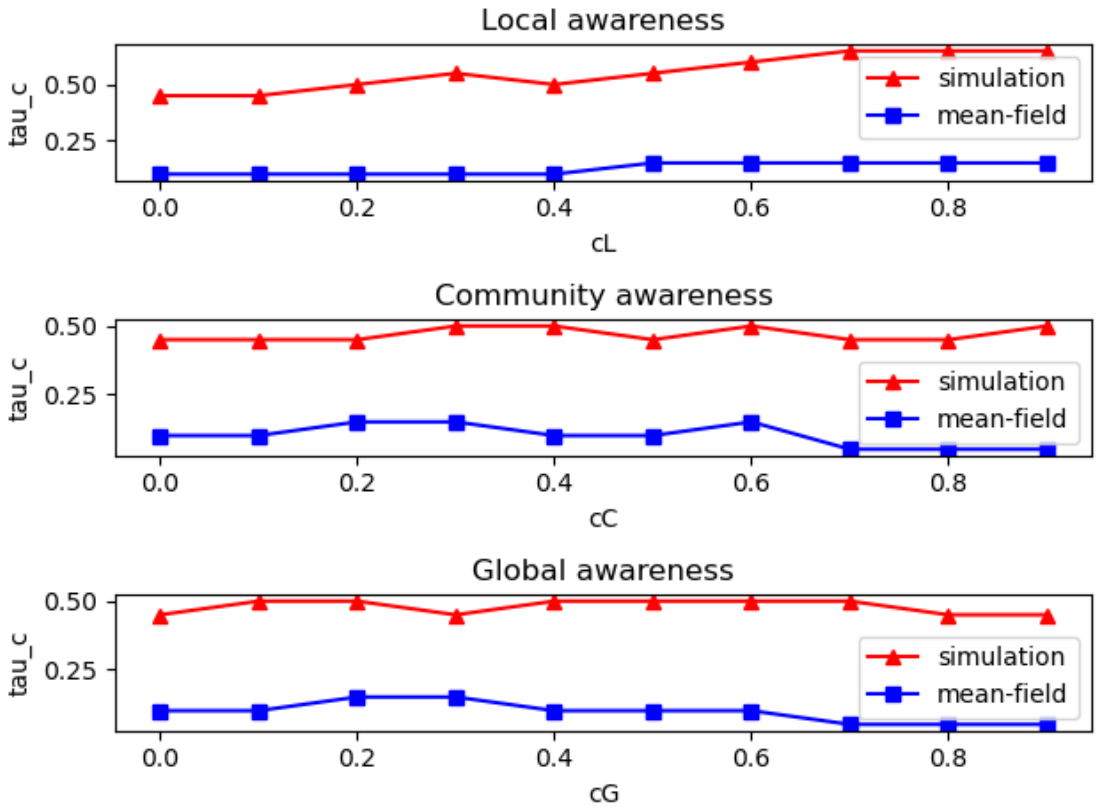

FIGURE 8: Epidemic threshold as a function of the awareness coefficients in the network G_2. In the top subfigure, we set $c_C = c_G = 0$ and varied c_L from 0 to 1 with a step of 0.1. In the middle subfigure, we set $c_L = c_G = 0$ and varied c_C while in the bottom subfigure, we set $c_L = c_C = 0$ and increased c_G.

TABLE 4: MSE between mean-field approximations and simulations of the epidemic threshold on the network G_2 while varying c_L, c_C and c_G.

Varying coefficient	c_L	c_C	c_G
Mean Square Error	0.1880	0.1385	0.1385

8.3 Network G_3 - Small Communities

The last network we examined was characterized by smaller communities. That is, we define network G_3 with $N = 300$ nodes and $n = 30$ communities with $N_{H_1} = \cdots = N_{H_{30}} = 10$ nodes each. All the nodes have in-degree $k_{in} = 1$ and out-degree $k_{out} = 2$. As in the previous examples, we used the HCM generation algorithm descried in Chapter 4 to build it.

The denseness of G_3 is

$$\delta_{G_3} = \frac{\frac{1}{2}\sum_{v \in G_3} d_v}{\frac{1}{2}N(N-1)} = \frac{900}{89700} = 0.0100,$$

which is equal to the denseness of G_1. The denseness coefficients of the communities are equal to each other $\delta_{com}^{H_1} = \cdots = \delta_{com}^{H_{30}} = \delta_{com}^{G_3}$ and they are

$$\delta_{com}^{G_3} = \frac{\frac{1}{2}\sum_{v\in H_1} d_v^{in}}{\frac{1}{2}N_{H_1}(N_{H_1}-1)} = \frac{20}{90} = 0.2222.$$

We observe that $\delta_{G_3} = \delta_{G_1} > \delta_{G_2}$ and $\delta_{com}^{G_3} > \delta_{com}^{G_2} > \delta_{com}^{G_1}$

In Figure 9, we show the fraction of infected nodes per community $\frac{I^H}{N}$ as well as the infection density in the network G_3 with $\tau = 1$ and $c_L = c_C = c_G = 0.5$. As in the previous cases, we ran the simulations for $t_{max} = 150$ time steps and initialized the system with a random selection of 10% of the nodes infected. The average infection density over the last 10 time steps of the simulation is $i_{ss} = 0.526$ while the mean-field equations gave $i_{ss} = 0.584$. The prevalence per community was of the order of 0.017 for the simulations and 0.019 for the mean-field approximation.

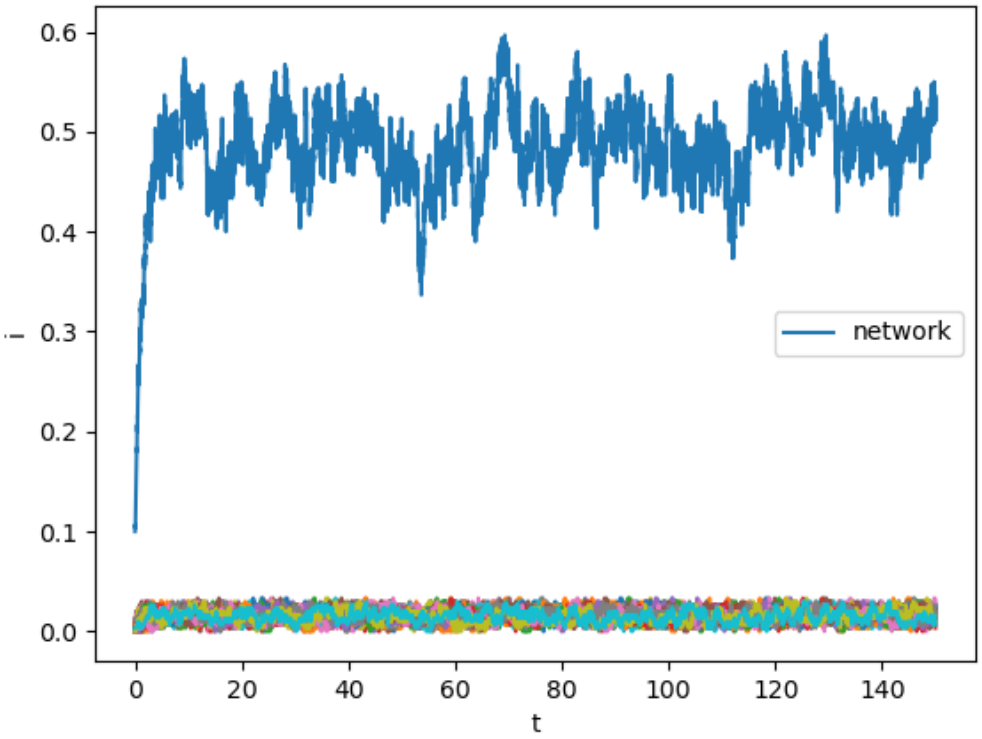

FIGURE 9: Fraction of infected per community for the network G_3 with $\tau = 1$ and $c_L = c_C = c_G = 0.5$

In Figure 10, we see a comparison between the epidemic prevalence computed using stochastic simulations and mean-field approximations while in Table 5, we present the MSE. In each of the three subplots, we varied one awareness coefficient from 0 to 1 with a step of 0.1 while keeping the other two constant and equal to zero. The infection rate was $\tau = 1$ and the simulations were initialized with 10% of the population infected. In all cases, i_{ss} follows a downward trend as the awareness coefficients increase. It it worth noting that the numerical method failed to converge when the initial guess was set to $i_{k_{in},k_{out}}^{H_i} = p_{k_{in},k_{out}}^{H_i}$ as usual and it only converged when set to $i_{k_{in},k_{out}}^{H_i} = \frac{p_{k_{in},k_{out}}^{H_i}}{2}$, a value which was decided upon trial and error.

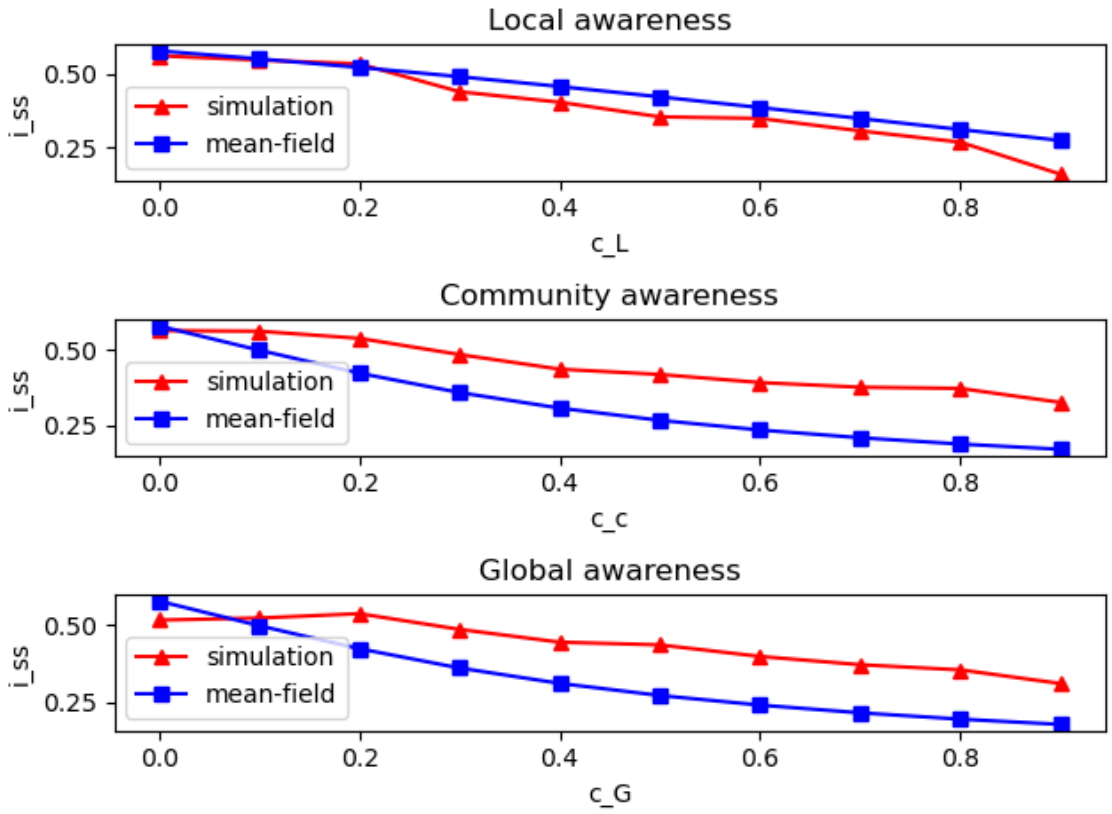

FIGURE 10: Epidemic prevalence i_{ss} as a function of varying awareness coefficients in the network G_3 with $\tau = 1$. In the top subfigure, we set $c_C = c_G = 0$ and varied c_L from 0 to 1 by a step of 0.1. In the middle subfigure, we kept $c_L = c_G = 0$ and varied c_C and in the bottom one, we set $c_L = c_C = 0$ and varied c_G.

TABLE 5: MSE between mean-field approximations and simulations of the epidemic prevalence on the network G_3 while varying c_L, c_C and c_G.

Varying coefficient	c_L	c_C	c_G
Mean Square Error	0.0028	0.0181	0.0169

In Figure 11 we contrast the mean-field approximation of the epidemic threshold with the one calculated using the simulation method. In each subfigure, we set two of the three awareness coefficients equal to zero while increasing the third one from 0 to 1 with a step of 0.1. In Table 6, we see the MSE of the two approximations. As seen in graphs G_1 and G_2, the mean-field method consistently underestimates τ_c. Moreover, we observe a relationship between τ_c and c_L, where τ_c increases as c_L increases. Furthermore, in this case, there seems to be a relation between τ_c and c_C; the epidemic threshold increases slightly with an increasing community awareness in the approximation by simulation. However, this trend does not manifest itself in the mean-field approximation.

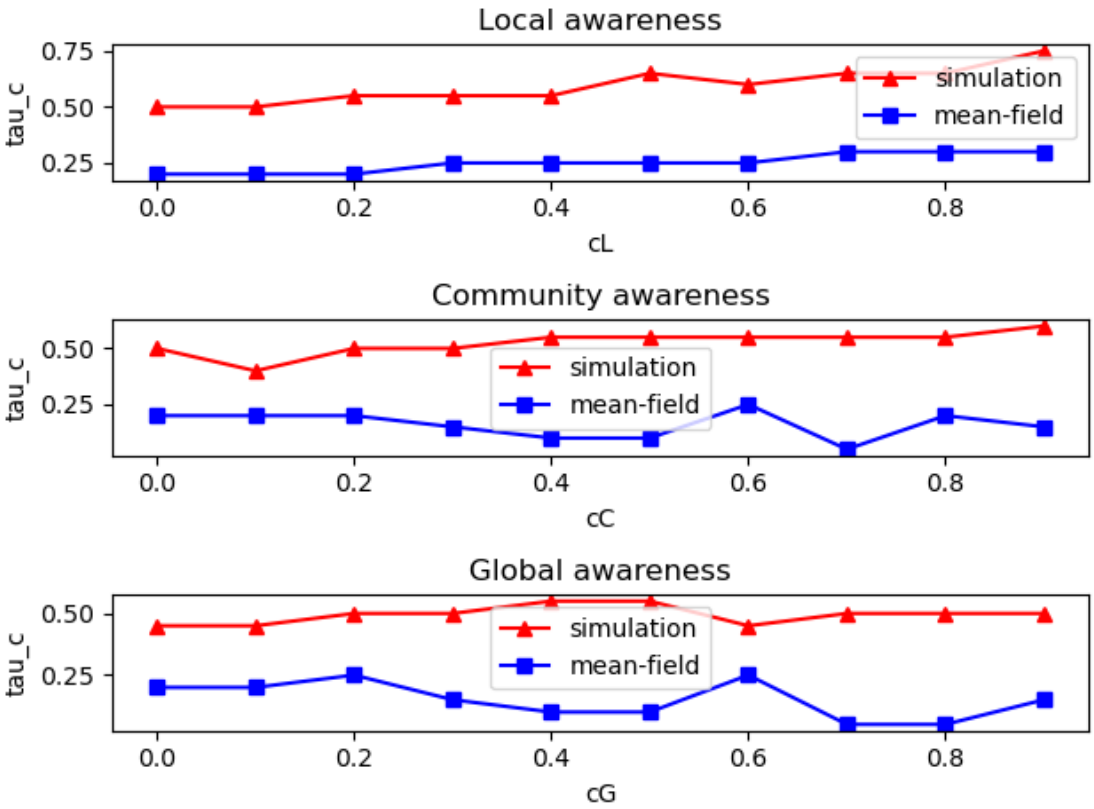

FIGURE 11: Epidemic threshold as a function of the awareness coefficients in the network G_3. In the top subfigure, we set $c_C = c_G = 0$ and varied c_L from 0 to 1 with a step of 0.1. In the middle subfigure, we set $c_L = c_G = 0$ and varied c_C while in the bottom subfigure, we set $c_L = c_C = 0$ and increased c_G.

TABLE 6: MSE between mean-field approximations and simulations of the epidemic threshold on the network G_3 while varying c_L, c_C and c_G.

Varying coefficient	c_L	c_C	c_G
Mean Square Error	0.1213	0.1413	0.1413

9 Discussion, Limitations and Future Research

In this chapter, we shall discuss the analysis and results presented in the previous chapters, explore some of the limitations of our study and suggest possible future research directions.

9.1 Discussion

We start the discussion by highlighting the differences between the present work and related studies in terms of methodology. Then, we proceed by explaining the key findings of the research regarding the two metrics for which the case studies were examined: the epidemic prevalence and the epidemic threshold.

Comparison with the Literature

In the relevant scientific literature, there have been a few attempts to study the effects of awareness on infectious disease. In [35] and [20], the researchers implemented a similar prevalence-related model in the absence of community structures, while in [21], they included community awareness, which, however, remained constant during the outbreak, while they excluded global awareness. In our work, we implemented a model which captures the effects of awareness on all three levels, micro (neighbourhood), meso (community) and macro (global), with an awareness function that was prevalence-related on each level. Furthermore, during the course of this work, we used an alternative community structure model compared to [21], namely the Hierarchical Configuration Model (HCM) [38], which has the advantage of generating graphs with communities using a predefined degree sequence.

The epidemiological model used was the network SIS model, a well-studied model in network epidemiology. Based on previous research [35], we proceeded in analyzing the model using a variation of the Degree-Based Mean-Field (DBMF) approach [29]. Typically, DBMF methods separate the vertices of a graph in classes based on their degree and assume that the behaviour of all vertices in each class is equivalent. In contrast, to capture the community structure in our model, we discriminated the nodes in classes based on their in-degree (internal connectivity), out-degree (external connectivity), and on the community to which they belong.

Epidemic Prevalence

The first characteristic that was studied was the epidemic prevalence. The main observation was that the epidemic prevalence is decreasing as the awareness coefficients c_L, c_C and c_G increase, a trend that was derived by both the mean-field analysis and the simulations. In fact, there was significant agreement between the values computed with the two methods, which indicates that the mean-field approach that we presented in Chapter 6 accurately describes the evolution of the epidemic in the given networks. In the networks G_1 and G_2 (see Figures 4 and 7), the mean-field approach slightly overestimated the epidemic prevalence while in G_3 (see Figure 10) this effect was not observed; in some configurations, the mean-field results were larger while in most cases they were smaller than the outcomes of the simulations. A likely reason for that may be the potential sensitivity of the numerical solver used to find the mean-field solutions to the initial guess; as we discussed, the mean-field equations were solved numerically using the nonlinear least-squares method, which requires an initial guess that may affect the convergence to the solution. In networks G_1 and G_2, the initial guess was set to $i^{H_i}_{k_{in},k_{out}} = p^{H_i}_{k_{in},k_{out}}$ while for G_3 it was $i^{H_i}_{k_{in},k_{out}} = \frac{p^{H_i}_{k_{in},k_{out}}}{2}$.

The initial guesses were found through a process of trial and error; if the original guess produced a high value for the cost function of the numerical solver, an alternative guess was tried.

The fact that the epidemic prevalence decreases as the awareness coefficients increase makes intuitive sense; as nodes become more aware of the presence of the infection, their infection rates drop, leading to a lower probability of becoming infected and hence to a lower epidemic prevalence overall. In the networks G_1 and G_2, the effect of the local awareness appears to be more pronounced, while the effects of community and global awareness appear to be of a similar magnitude. A possible explanation for this effect could be the relative scale of the three areas; a neighbourhood is quite smaller than the communities and the entire graph in these two cases. Therefore, it is easier for local awareness to reach its maximum value. In fact, since all nodes have degree $k = 3$, having one infected neighbour leads to the local awareness acquiring $\frac{1}{3}$ of its maximum value. In graph G_3, the size of the communities is closer to the size of the neighbourhood, making the strength of the effects of community awareness comparable to that of local awareness. Furthermore, as we observed in the linearization of the mean-field equations in Chapter 6, the community and global awareness coefficients were multiplied to higher-order infection densities, which means that their effect is expected to be weaker.

Epidemic Threshold

The second characteristic of the epidemic that was investigated was the epidemic threshold, which was computed in the three network case studies for varying awareness coefficients. The first observation is that the outcomes of the mean-field approximation and the results of the surviving runs method for simulations appear to differ; the mean-field approximation consistently estimated a smaller epidemic threshold, something that will be discussed below. However, it was also apparent that in all case studies, the epidemic threshold exhibits a monotone increasing relationship with the local awareness coefficient, a trend seen in both approximation methods. Moreover, we can notice that the community and global awareness coefficients do not seem to affect the epidemic threshold. This lack of correlation is seen in all networks with both methods apart from a curious exception; the epidemic threshold appears to follow an upward trend as the community awareness increases in network G_3, a tendency only observed in the approximation by simulation (see Figure 11).

As we discussed in the case of the epidemic prevalence, the local awareness has a clear effect on the epidemic threshold because of the relatively smaller scale of the neighbourhood compared to the community or the entire network. This was also hinted at by the linearization results, in which the community and global awareness coefficients vanished, and it aligns with the findings of the relevant literature [35] [41]. Additionally, the difference in effect of the different types of awareness may become more pronounced close to the critical value because the infection densities in the community or network have not yet stabilized to a level that would cause these types of awareness to become consequential. This also suggests a reason behind the possible connection between community awareness and epidemic threshold in G_3. The communities of G_3 are of comparable size with the neighbourhoods, bringing the scale of the community closer to the local level. However, this connection is still weak and speculative, while it is not captured by the mean-field approximation. A possible reason the mean-field does not capture these dynamics may be that the community awareness is multiplied with higher order terms in the equations, making its effect weak even with small communities.

Intuitively speaking, it can be understood that an infection outbreak is better stopped at its initial stages, before reaching the endemic state, by increased awareness in the

neighbourhood of the first infections. At these early stages, the infection density in the community and the network are not large enough for the effects of community and global awareness to be substantial. By the same token, when the community and global awareness take effect, the outbreak has already reached the endemic equilibrium and is difficult to be extinguished. This latter point highlights the difference between the two metrics that we explored; all types of awareness affect the epidemic prevalence, albeit at different degrees, because we examined the systems in the endemic equilibrium where the infection densities were large enough for the community and global awareness to become consequential.

Discrepancy in the Epidemic Threshold Approximation

In contrast to the epidemic prevalence, we noticed that the mean-field and the simulation approximation of the threshold differ significantly, although they generally follow the same trends. The MSE was smaller in the simple network G_1 (see Table 2) than in the more complicated cases G_2 and G_3 (see Tables 4 and 6), but still significantly larger than in the epidemic prevalence estimation. In [35], the researchers also observed this discrepancy, although to a smaller scale. The inconsistency could be attributed to multiple reasons.

Firstly, as we saw above, the mean-field approximation tended to overestimate the epidemic prevalence, which would lead to an underestimate of the threshold. However, in the case of G_3, where the mean-field calculated prevalence was sometimes an underestimate, the difference in the threshold remained. Another indication that the discrepancy may originate in the mean-field approximation is that the estimations are closer in the simple network case G_1, where there are fewer mean-field equations, and therefore the system is easier to solve. This would make the mean-field approach potentially more accurate in this case, while there is no apparent reason for the surviving runs method to work better in simpler networks. Nevertheless, the inaccuracies in the mean-field epidemic prevalence estimations were not as significant as the epidemic threshold estimations and cannot sufficiently account for them.

A second possible cause may be the finite-size effects, which is the explanation given in [35]. In [35], these effects could be induced by the cut-off in the power-law degree distribution as well as the by the presence of an absorbing state. In our case, we did not use power-law distributions, but we were still vulnerable to the absorbing state issue. As we discussed in Chapter 7, stochastic fluctuations during the simulations may lead the system to the absorbing infection-free equilibrium even when $\tau > \tau_c$, especially in the neighbourhood of τ_c. Although we attempted to combat this effect by only considering the surviving runs, there was also an upper limit to how many simulations would be run before proceeding to the next τ, a limit that may have been set too low. Another parameter that could play a role in the approximation is the threshold θ which was drawn from literature that dealt with larger networks [35]. However, we deem it unlikely to have had an effect because, in the cases of the endemic equilibrium, the densities vastly surpassed it. Similarly, the infection rate step, which was set to 0.05, could have been set lower to improve the accuracy, but it is unlikely that this caused the discrepancy because the differences between the approximations were multiples of that step.

It should be reminded that contrary to the case of the prevalence where the simulations were exact representations of the physical system, the simulation computed threshold is still an estimate calculated using the surviving runs method. If finite-size effects cause the discrepancy in the approximations, it would mean that it is the surviving runs method that overestimates the value rather than the mean-field underestimating it. It is also very likely that it is a combination of both that causes the divergence.

9.2 Limitations and Future Research

The findings of this study have to be considered in light of some limitations, which in turn point to new potential research directions.

Alternative Epidemiological Models

Firstly, the epidemiological model investigated was the SIS model, a simple model with only two possible states. Although it is a commonly used model because of its simplicity, it should be noted that it is not a particularly realistic one; real-life infections tend to be more complex than the SIS model can capture. Infections may lead to a permanent or temporary immunization of the population, an effect that can be described by the SIR and SIRS models, respectively, which include the recovered state R [29]. Moreover, it is possible that individuals may not become immediately infectious after exposure, leading researchers to propose models which include an exposed E state, such as in the SEIR and SEIRS models [29]. Investigating the effects of awareness on networks with community structures in such systems may lead to more complicated master equations, but the research is recommended such that the studies gain in veracity.

Node Classes Aggregation

Another limitation of our study is that the mean-field analysis presented previously distinguishes vertices based on their in-degree, out-degree and community. In networks with degree distribution with non-zero variance, this classification will lead to a significant increase in the number of equations, something which may render the master equations computationally intractable or complicate the convergence to a solution. Moreover, there may be classes that contain a small number of nodes, something that will obscure the aggregation and make it more difficult for the mean-field equations to capture the dynamics of the system. Therefore, it could be helpful to follow up this study with a mean-field derivation that aggregates the vertices in fewer classes. In [21], the researchers derived one equation per community by counting the nodes in each community that have $k_{out} \neq 0$ and calculating the force of infection each community receives from within and from outside. Alternatively, we could bin the in- and out-degrees in degree classes and derive a different equation for each degree class. These are only two approaches that could lead to a simpler system of ODEs. It should be noted that the accuracy of such an aggregated method may be reduced.

Network case studies

In our case studies, the investigated networks were small, simple, static and had identical degree distributions. This means that we were not exposed to the limitation discussed in the previous paragraph. Nonetheless, the narrow case studies limit the generalization of the results. Thus, we suggest that the findings be verified with larger networks and networks with non-zero variance in the degree distributions. The need for such investigation is highlighted in [20], where the researchers investigated networks with power-law degree distributions, and they discovered that in the case of infinite variance distributions, linear awareness function did not affect the epidemic threshold because it tends to vanish. However, they showed that nonlinear local awareness did affect the threshold. Moreover, real-life networks are rarely static and usually change over time. An investigation on the effects of awareness on so-called temporal networks would shed more light on this topic.

Alternative awareness models

The awareness function that we used includes a set of assumptions that impose further limitations on our study. We assumed that the awareness is linear and prevalence-related, the awareness coefficients were equal all over the network, individuals were immediately aware of infections in their environment and their reaction was also prompt, and the effects were limited to the transmission rate. These limitations point to new research directions. Awareness models do not need to be linear, and as seen in [20], nonlinear effects could potentially change the results. Moreover, the awareness coefficients could be different across the network. Specifically, we could define different community awareness coefficients per community, as in [21]. Furthermore, we could envision models in which nodes would be infectious, yet their environment would not be immediately aware or immediately react to them, and the awareness would propagate in the network with a delay. The delayed awareness may cause some interesting dynamics while providing a more accurate representation of reality. Additionally, the awareness could cause effects that are not limited to the transmission rate, such as leading to rewiring in the network.

A potentially fruitful research direction seems to be the study of discontinuous awareness functions. In the currently unfolding COVID-19 pandemic, individuals did not seem to react linearly to the presence of new infections. On the contrary, private, municipal and national authorities imposed rules and regulations limiting or weakening disease-transmitting interactions when infection densities reached specific levels. This behaviour could be captured better by a multiple level awareness function. In the simplest case, awareness would be activated when infections in an area attained a level and deactivated when they decreased below that level, with multiple levels added in more complex models. In a preliminary investigation conducted during this work, which is not included in the present thesis, there were indications that the infection densities would oscillate around the awareness activation level. An even more realistic model would have different activation and deactivation levels; the awareness would be activated when infections reach a certain height and deactivated when they decline to a level lower than the activation level. This last model would probably lead to wave-like behaviours, with infection densities oscillating between the activation and deactivation levels. Nonetheless, it should be noted that this model is not Markovian and does not have an endemic equilibrium, which would make deriving and solving mean-field expressions a complicated task.

Resolving the Discrepancy in the Epidemic Threshold Approximation

As discussed above, there is a discrepancy between the mean-field and surviving runs approximations of the epidemic threshold, primarily attributed to an overestimation by the surviving runs method. It is certainly possible to improve the method by increasing the number of surviving run iterations r_1 as well as the upper bound on the number of runs r_2. However, this would make the already costly method significantly less practical. A more efficient method that could be explored in future research is the quasi-stationary (QS) procedure in which the absorbing state is artificially excluded from the dynamics [9] [6].

Effects of Community Awareness

Lastly, we should mention the interesting, albeit weak, result that smaller communities may make the community awareness relevant in terms of the epidemic threshold. Our research points to this direction but cannot provide conclusive outcomes. Therefore, we would suggest that this result be further investigated since it provides one more course in preventing an infection from reaching the endemic state.

10 Conclusion

The main research topic that we set out to investigate was the influence of awareness on the spread of infectious disease on networks with community structures. In this direction, we studied an SIS epidemic on networks generated using the Hierarchical Configuration Model. We defined a prevalence-related awareness function for each susceptible node that consisted of three levels of awareness, local, community and global, and would affect the transmission rate of the node. At each level, the awareness depends on an awareness coefficient. Afterwards, we analysed the epidemic using Gillepsie-style stochastic simulations and a Degree-Based Mean-Field approach, adjusted to our model. For the mean-field method, we separated nodes in classes based on their community as well as their intra- and inter-community connectivity and derived differential equations that would predict the infection density in each class of nodes.

The system was examined for two metrics, the epidemic prevalence in the steady-state and the epidemic threshold, which were approximated using the mean-field and the stochastic simulations methods. The two metrics were studied for their relation with the three awareness coefficients on three network case studies. It was revealed that all coefficients affect the epidemic prevalence with higher coefficients leading to lower prevalence, an outcome on which both the mean-field and the simulation results were aligned. The local awareness coefficient seems to have a more substantial effect in this case. Regarding the epidemic threshold, the local awareness coefficient appears to have an impact, with higher coefficients leading to higher thresholds. The mean-field and simulation approximations differed by a margin, but they agreed on the trend. Contrary to the epidemic prevalence, the community and global coefficients did not seem to influence the threshold. However, there are some indications that in small communities, the community awareness becomes consequential, but further research is required because this result only showed in the stochastic simulations and not in the mean-field.

In conclusion, our research verifies that awareness-based behaviour modification can significantly affect the time evolution of an SIS epidemic outbreak in networks with community structures. Local, community and global awareness can decrease the epidemic prevalence, while local awareness can even stop an outbreak from occurring. Therefore, our study suggests that an effective way to stop new infections from reaching the status of an epidemic or becoming endemic is to focus on rapid reactions on the local level in the neighbourhood of the initial infections.

11 References

[1] Franco Bagnoli, Pietro Lio, and Luca Sguanci. Risk perception in epidemic modeling. Physical Review E, 76(6):061904, 2007.

[2] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. Physics reports, 424(4-5):175–308, 2006.

[3] Tom Britton, Maria Deijfen, and Anders Martin-Löf. Generating simple random graphs with prescribed degree distribution. Journal of statistical physics, 124(6):1377–1397, 2006.

[4] Thiago Carvalho, Florian Krammer, and Akiko Iwasaki. The first 12 months of covid-19: a timeline of immunological insights. Nature Reviews Immunology, 21(4):245–256, 2021.

[5] Brian A Davey and Hilary A Priestley. Introduction to lattices and order. Cambridge university press, 2002.

[6] Marcelo Martins de Oliveira and Ronald Dickman. How to simulate the quasistationary state. Physical Review E, 71(1):016129, 2005.

[7] Neil Ferguson. Capturing human behaviour. Nature, 446(7137):733–733, 2007.

[8] Silvio C Ferreira, Claudio Castellano, and Romualdo Pastor-Satorras. Epidemic thresholds of the susceptible-infected-susceptible model on networks: A comparison of numerical and theoretical results. Physical Review E, 86(4):041125, 2012.

[9] Silvio C Ferreira, Ronan S Ferreira, and Romualdo Pastor-Satorras. Quasistationary analysis of the contact process on annealed scale-free networks. Physical Review E, 83(6):066113, 2011.

[10] Sebastian Funk, Erez Gilad, Chris Watkins, and Vincent AA Jansen. The spread of awareness and its impact on epidemic outbreaks. Proceedings of the National Academy of Sciences, 106(16):6872–6877, 2009.

[11] Sebastian Funk, Marcel Salathé, and Vincent AA Jansen. Modelling the influence of human behaviour on the spread of infectious diseases: a review. Journal of the Royal Society Interface, 7(50):1247–1256, 2010.

[12] Shannon Gallagher and J Baltimore. Comparing compartment and agent-based models. In Joint Statistical Meeting, Baltimore, 2017.

[13] Daniel T Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. Journal of computational physics, 22(4):403–434, 1976.

[14] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. Proceedings of the national academy of sciences, 99(12):7821–7826, 2002.

[15] Robin Goodwin, Shamsul Haque, Felix Neto, and Lynn B Myers. Initial psychological responses to influenza a, H1N1 (" swine flu"). BMC Infectious Diseases, 9(1):1–6, 2009.

[16] Clara Granell, Sergio Gómez, and Alex Arenas. Dynamical interplay between awareness and epidemic spreading in multiplex networks. Physical review letters, 111(12):128701, 2013.

[17] Paul W Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. Stochastic blockmodels: First steps. Social networks, 5(2):109–137, 1983.

[18] Carl T Kelley. Iterative methods for optimization. SIAM, 1999.

[19] István Z Kiss, Joel C Miller, Péter L Simon, et al. Mathematics of epidemics on networks. Cham: Springer, 598, 2017.

[20] Stephan Kitchovitch and Pietro Lio. Risk perception and disease spread on social networks. Procedia Computer Science, 1(1):2345–2354, 2010.

[21] Stephan Kitchovitch and Pietro Liò. Community structure in social networks: applications for epidemiological modelling. PloS one, 6(7):e22220, 2011.

[22] Marc Lelarge. Diffusion and cascading behavior in random networks. Games and Economic Behavior, 75(2):752–775, 2012.

[23] Zonghua Liu and Bambi Hu. Epidemic spreading in community networks. EPL (Europhysics Letters), 72(2):315, 2005.

[24] Robert M May and Alun L Lloyd. Infection dynamics on scale-free networks. Physical Review E, 64(6):066112, 2001.

[25] Viktor Nagy. Mean-field theory of a recurrent epidemiological model. Physical Review E, 79(6):066105, 2009.

[26] Ronen Olinky and Lewi Stone. Unexpected epidemic thresholds in heterogeneous networks: The role of disease transmission. Physical Review E, 70(3):030902, 2004.

[27] World Health Organization. Advice for the public on COVID-19.

[28] Gergely Palla, Albert-László Barabási, and Tamás Vicsek. Quantifying social group evolution. Nature, 446(7136):664–667, 2007.

[29] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. Reviews of modern physics, 87(3):925, 2015.

[30] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic spreading in scale-free networks. Physical review letters, 86(14):3200, 2001.

[31] Meirui Qian and Jianli Jiang. Covid-19 and social distancing. Journal of Public Health, pages 1–3, 2020.

[32] Filippo Radicchi, Claudio Castellano, Federico Cecconi, Vittorio Loreto, and Domenico Parisi. Defining and identifying communities in networks. Proceedings of the national academy of sciences, 101(9):2658–2663, 2004.

[33] William J Reed. A stochastic model for the spread of a sexually transmitted disease which results in a scale-free network. Mathematical biosciences, 201(1-2):3–14, 2006.

[34] Sheldon M Ross. Introduction to probability models. Academic press, 2014.

[35] Yilun Shang. Modeling epidemic spread with awareness and heterogeneous transmission rates in networks. Journal of biological physics, 39(3):489–500, 2013.

[36] Clara Stegehuis, Remco Van Der Hofstad, and Johan SH Van Leeuwaarden. Epidemic spreading on complex networks with community structures. Scientific reports, 6(1):1–7, 2016.

[37] Remco Van Der Hofstad. Random graphs and complex networks, volume 1. Cambridge university press, 2016.

[38] Remco van der Hofstad, Johan SH van Leeuwaarden, and Clara Stegehuis. Hierarchical configuration model. arXiv preprint arXiv:1512.08397, 2015.

[39] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261–272, 2020.

[40] Huan Wang, Chuang Ma, Han-Shuang Chen, and Hai-Feng Zhang. Effects of asymptomatic infection and self-initiated awareness on the coupled disease-awareness dynamics in multiplex networks. Applied Mathematics and Computation, 400:126084, 2021.

[41] Qingchu Wu, Xinchu Fu, Michael Small, and Xin-Jian Xu. The impact of awareness on epidemic spreading in networks. Chaos: an interdisciplinary journal of nonlinear science, 22(1):013101, 2012.