1. Introduction

Real systems of interactions between humans are commonly studied as interaction networks. In this setting, an edge in a network represents a possible interaction between the connected nodes. The mechanism of disease transmission upon contact provides a strong case for the use of these networks in modelling epidemics. Spreading processes of infectious diseases take place over these networks, thus their structures can be essential for understanding disease transmission. In particular, knowledge on the interplay between network structure and transmission parameters can improve predictions as well as prevention and control strategies.

A key factor in an epidemic spread is the population network structure where a disease may spread. However, a complete picture, in principle, may require the knowledge of every individual in a population and its relationships, for example, as in contact tracing (Ferretti et al., 2020). Networks used in different fields have some common geometric characteristics regarding the distribution of the nodes and edges, for example, a few nodes that may act as hubs and the vast majority of nodes with a few neighbours (Pastor-Satorras et al., 2015). A classification in this context is given by network generation models, used as generators of synthetic networks, with controlled topological properties.

Several types of networks have been proposed over time (De Arruda et al., 2018; M. M. Keeling & Eames, 2005). Erdős and Rényi (1959) proposed a simple model where nodes are connected according to a uniform probability without any preference. However, some structural properties observed in real-world networks cannot be reproduced by this model as empirically verified by M. E. J. Newman and Park (2003). It has been observed in real networks that the degree distribution of individuals is far from homogeneous; on the contrary, only a few individuals have several connections and the majority have a few (Albert et al., 1999, 2000). Barabási and Albert proposed a model to generate scale-free networks (Barabási & Albert, 1999) with a connection mechanism that mimics the natural formation of social contacts. Recent results highlight criticalities in the use of their approach to model the realistic spread of an epidemic. In Bertotti and Modanese (2019), the authors remark about an unexpected, recently discovered assortative behaviour of Barabási-Albert networks. In Chang et al. (2020), the authors show a correlation between assortativity and epidemic spread, which might result in misleading simulations. Moreover, the algorithm proposed by Barabási and Albert (1999) cannot control the value of the exponent of the power-law distribution (M. M. Keeling & Eames, 2005). In Thedchanamoorthy et al. (2014), assortativity is taken into account and discussed in greater detail in the context of epidemics and vaccinations.

However, Caldarelli and Vespignani (2007) state that the configuration model can overcome these issues, generating a network with a given degree
sequence. The configuration model is used, for example, in social dynamics, as it captures connectivity features of this class of networks (Boccaletti et al., 2006). The study of how epidemics evolve on networks has been addressed using various approaches, theoretical and computational. Theoretical results on this topic can be found, for example, in the monographs Estrada (2011); Kiss et al. (2017) or in Boguná et al. (2013). In López-García (2016), the author derives results on a small heterogeneous population before validating them via numerical simulations. In Strona et al. (2018), the authors explore the interplay between network properties and disease characteristics on the spread. Many other articles are devoted to the study of the impact of the network structure on the evolution of an epidemic (e.g., household structure (Ball et al., 2010), the influence of network topology on epidemic spread (Draief et al., 2006; Ganesh et al., 2005; M. Keeling, 2005), or community structure (Zhang et al., 2013)). In particular, researchers have been interested in identifying the nodes which, if infected, would cause the largest epidemic, the so-called influential spreaders (Kitsak et al., 2010; Mín, 2018; Radicchi & Castellano, 2016). Another interesting feature, studied in Rocha et al. (2011), is the implementation of dynamic contacts, meaning a network in which nodes may delete and create edges in time. Other studies have been devoted to the extinction time for epidemics on networks (Hindes & Schwartz, 2016; Holme, 2013; Holme & Túpikina, 2018) and epidemics thresholds (Cator & Van Mieghem, 2012; Chakrabarti et al., 2008; Van Mieghem et al., 2008). Bounds and estimates on the final size of the epidemics have been investigated assuming both homogeneity and heterogeneity of viral transmission (Miller, 2008, 2012). These estimates have played a major role in quantifying the consequences of different restrictions on the Covid-19 pandemic (Aleta et al., 2020; Firth et al., 2020; Sun et al., 2021).

In Cliff et al. (2018) and Zachreson et al. (2018) the authors study the effect of seeding on the evolution of the epidemic to consider various realistic scenarios for its beginning. Different scenarios, in this regard, can be integrated into the model. Examples to these scenarios include multiple infected individuals arriving in a country almost simultaneously in different airports, or a localised initial cluster which is then spread by the individuals from the same starting area. In particular, the former scenario is included in our analysis, in Section 4.1, in the “random case” with multiple initially infected individuals.

In this paper, we investigate how the interplay between the connectivity of different configuration model networks and the contagiousness of the disease affect the magnitude of the epidemic. Observations on empirical data show that standard epidemiological metrics often fail to predict the evolution of an epidemic due to their lack of integrating various aspects of the population that influence the spread dynamics. In this respect, the capability to perform stochastic simulations on graphs indicates a more empirical approach in studying the spread and behaviour of epidemics while maintaining a theoretical grip. To this end, the present work lays the foundation for such refinements to study SIR (Susceptible-Infected-Removed) compartmental models stochastically on graphs to assess the factors that are otherwise difficult or even impossible. To achieve this, we verify our approach for theoretical rigour by showing an agreement with theoretical results. With the integration of stochastic elements, in the generation of the configuration model and the simulations, we introduce an empirical component that captures the variability in real-world epidemics. This step, in turn, provides new capabilities to perform in silico experiments by simulations, possibly by introducing additional components that model various epidemic influences.

In the following, after a summary of the relevant mathematical background, we derive a lower bound and a closed formula for the probability of having a minor epidemic of the disease in a network. In other words, we model the contacts in a community as a scale-free network and study how an epidemic spreads over these contacts. For this, we use the derivation of the probability of extinction for a branching process as a function of all the parameters involved and the degree of the initially infected node. A similar analysis of the extinction probability was carried out in Rogers (2015) where the author considers a more general case for the distribution of the infectious period and proceeds with the so-called cavity method and in Karrer and Newman (2010) where the authors obtain a result similar to ours through the message passing method. In this work, we provide a different exact result on the probability of a minor epidemic, however, in a form that only allows for numerical exploration. We consequently deduce from it a lower bound for the probability of a minor epidemic, given the parameters of the epidemics and the degree of the initially infected node.

Following the theoretical groundwork, we introduce a stochastic model on scale-free random networks, which we use to run simulations on different instantiations of the model. The simulations, carried out via a specialised modification of the Gillespie algorithm, validate our theoretical results regarding the probability of a minor epidemic of the disease. Consequently, we provide a thorough analysis via simulations that highlights the influence of network connectivity (which decreases in α, the exponent of the power law) and infectiousness (the parameter β of our SIR model) on three key epidemic indices: the peak of infected individuals, the total number of eventually
infected individuals, and the duration of the epidemic. In particular, we focus on the role of the position of the nodes from which the disease starts spreading as well as the number of initially infected nodes, exploring simulations for different instantiations of the model parameters. We compare the evolution of the epidemic, measured by the aforementioned indices, in four different cases for the initially infected nodes, categorised by their position in the respective networks: hub (degree in the tail of the distribution), mean degree, peripheral (low degree), and randomly chosen. The stochastic simulations with our model are in good agreement with the analytical lower bound for the probability of extinction we derived. The comparison between our analytical and numerical results provides a quantification of the role of model parameters in the spread of the epidemic on scale-free networks.

Overall, our results illustrate how theoretical methods in epidemiology can be coupled with discrete stochastic simulations in a rigorous manner. In this regard, our theoretical results set a baseline for discrete simulations for comparison and validation. Our implementation, available through a Github repository, optimises Gillespie’s stochastic simulation algorithm Gillespie (1977) by exploiting the network structure for increased efficiency, which is otherwise a bottleneck for simulating networks. Our implementation may thus be used, by us or other researchers, to study broader questions while maintaining the baseline given by our theoretical results.

2. Power-law networks

In this section, we describe scale-free networks, which can be used to mathematically evaluate a given model or describe a real network structure, mimicking social contacts between individuals. We then describe the classic algorithm used to generate networks with a certain degree distribution, that is, configuration model (CM).

2.1. Scale-free networks

Many real-world graphs follow power-law degree distributions (Artico et al., 2020; Eguiluz et al., 2005; Fox Keller, 2005; however their actual frequency is debated (Broido & Clauset, 2019; Holme, 2019)), that is, degree distributions with the probability of a node to have k direct neighbours given by \( p_k \sim k^{-\alpha} \). These kinds of networks are also called scale-free random graphs. The preferential attachment model of Barabási and Albert (1999) produces a degree distribution with \( \alpha = 3 \). However, there are many examples in which \( \alpha \in (2, 3] \) (see e.g., (Durrett, 2006, Sec. 1.4)). To explain the scale-free property, Barabási defined the power-law distribution both with a discrete and a continuum formalism; we refer to Barabási (2015) for a more in-depth introduction to this topic and recall here the definitions and results we need for our analysis.

In a scale-free network, the probability of having k neighbours is \( p_k = Ck^{-\alpha} \) where C is a normalisation constant. The main difference between an E-R graph and a scale-free network comes in the tail of the degree distribution, representing the high-k region of \( p_k \): high-degree nodes, called hubs, are naturally present in scale-free networks, contrary to random networks. Since all real networks are finite, we may expect that nodes assume a maximum degree, \( k_{\text{max}} \), called the natural cut-off of the degree distribution \( p_k \) (see, Boguná et al. (2004)). This quantity represents the expected size of the largest hub in a network and it is determined by

\[
K_{\text{max}} = \left\lfloor K_{\text{min}} N^{\frac{1}{\alpha}} \right\rfloor, 
\]

where \( K_{\text{min}} \) is the smallest degree we allow a node to have. Hence, the highest attainable degree is directly proportional to a power of \( N \) between \( \frac{1}{\alpha} \) (corresponding to \( \alpha = 3 \)) and 1 (\( \alpha = 2 \)).

2.2. Configuration model

In the case of large networks, the adjacency matrix may not be immediately available (Kiss et al., 2017). However, in the case of real networks of which we know the degree sequence, we can generate a graph with precisely the same degrees. Given the number of nodes \( N \) and the sequence of degrees \( \{k_i\}_{1 \leq i \leq N} \) of length \( N \) (we omit the subscript from now on, for ease of notation), the aim is to construct an undirected graph with \( N \) nodes, in which the \( i \)-th node has precisely degree \( k_i \). We denote such graphs with \( G(N, \{k_i\}) \). Given a degree distribution obtained from observing a stochastic network, the algorithm used to fit this distribution is called the configuration model. To construct the network, we start by assigning to each node \( i \) in the set of nodes a random degree \( k_i \), drawn from the chosen probability distribution \( p_k \). Clearly, \( k_{\text{max}} \leq N - 1 \) since no node can have a degree larger than \( N - 1 \). The degrees of the nodes are represented as half-links or stubs, thus we impose \( \sum_{i=1}^N k_i = 2m \), where \( m \) is the total number of edges.

First, two stubs are connected to form an edge. After that, another pair of stubs are chosen from the remaining \( 2m - 2 \) stubs and connected, respecting the preassigned degrees. The network is completed by repeating this procedure until the stubs run out. The result of this construction is a random network whose degrees are distributed according to \( p_k \) (M. Newman, 2010). If \( L \) denotes the numbers of degrees assumed in
the network and $N_1, N_2, \ldots, N_l$ the number of nodes of each degree, the average degree in the network is given by Kiss et al. (2017)

$$
\langle k \rangle = \frac{1}{N} \sum_{i=1}^{L} N_i k_i.
$$

This formula is equivalent to $\langle k \rangle = \sum_i k_i p_i$ for this specific realisation. Note that this linkage procedure does not exclude self-loops or multiple edges, but their expected number is bounded (see e.g., (Hofstad, 2016, Prop. 7.1)). When the size of the graph $N \to +\infty$ with a fixed degree distribution, self-loops and multiple edges become less and less apparent in the global dynamics (see e.g (Durrett, 2006, Th. 3.1.2)).

### 3. Lower bound for the probability of a minor epidemic

We work on an SIR epidemiological dynamics model on a network built with the CM algorithm, described in Section 2.2. In this network, we introduce infected individuals in an otherwise fully susceptible population; the exact number of infected individuals is specified each time it changes. Later, in Section 4, we study how the introduction of a different number of initially infected individuals affects the evolution of the epidemic. In this section, we consider an epidemic that starts with only one infected individual. We then derive an analytical formula for the probability of a minor epidemic, that is, the probability that one infected individual in the network does not cause a major epidemic. In other words, an epidemic in which a large number of individuals are infected simultaneously, a large number of infections occur in total and the disease remains in the population for an extended period (Thompson et al., 2020). We use this probability as a benchmark to compare with our simulations. As usual, we assume that an individual remains infected for a duration drawn from an exponential distribution with rate parameter $\gamma$. During its infectious period, an individual infects each of its neighbours (independently of the others) according to a Poisson process with rate parameter $\beta$. Note that modellers often assume that $\beta$ does not depend on the number of contacts; however, in most epidemic models (and data that are collected), the infection probability decreases with the number of contacts (M. J. Keeling & Rohani, 2011, Chapter 2). Under the assumptions above, the basic reproduction number $R_0$ is given in Iannelli and Pugliese (2014) as

$$
R_0 = \frac{\beta}{\gamma} \langle k^2 \rangle (1 + CV^2),
$$

where $CV$ is the coefficient of variation, defined as $CV = \sqrt{\frac{\langle k^2 \rangle}{\langle k \rangle}}$. Using this definition, the expression of $R_0$ can be written also as

$$
R_0 = \frac{\beta \langle k^2 \rangle}{\gamma \langle k \rangle}.
$$

We focus on this slightly unrealistic but analytically tractable assumption; we refer to (M. J. Keeling & Grenfell, 2000, App. 1) for a similar construction, applied however to a power law distribution without cut-off. We approximate the initial phase of an epidemic by a branching process where all contacted individuals are susceptible. This is a property of the configuration model as $N$ goes to infinity; thus, cliques and triangles are neglected, and the population is assumed to be large enough. First, we compute the probability that an infected node $i$ with $k_i$ neighbours infects $j$ of them; for ease of reading, we denote the degree of a node with $k$.

We start by considering one of them, conditioning on the length of the infectious period:

$$\Pr(\text{a contact is not infected}) = \gamma \int_0^\infty e^{-\beta t} e^{-\gamma t} dt = \frac{\gamma}{\beta + \gamma} = \frac{1}{R + 1},$$

where $R := \beta/\gamma$.

We can not use the binomial distribution to obtain the probability of having $n$ infected individuals, because infections of different contacts are not independent, but correlated by the length of the infectious period, whereas an important assumption which would allow us to use the binomial distribution is independence between infections (Flanders & Kleinbaum, 1995). This is due to the fact that if the infectious period is short, it is likely that no neighbour will be infected, while if it is long, most of them will. Indeed, if $Q$ is the number of infected neighbours,

$$\Pr(Q = 0) = \gamma \int_0^\infty (e^{-\beta t})^k e^{-\gamma t} dt = \frac{\gamma}{\beta k + \gamma} = \frac{1}{kR + 1}.$$

The other expressions are more complicated:

$$\Pr(Q = j) = \gamma \int_0^\infty \binom{k}{j} (1 - e^{-\beta t})^j (e^{-\beta t})^{k-j} e^{-\gamma t} dt. \quad (3)$$

We can verify that $\sum_{j=0}^k \Pr(Q = j) = 1$ as follows

$$\sum_{j=0}^k \Pr(Q = j) = \gamma \int_0^\infty \left( \sum_{j=0}^k \binom{k}{j} (1 - e^{-\beta t})^j (e^{-\beta t})^{k-j} e^{-\gamma t} \right) dt.$$

Note that $\sum_{j=0}^k \binom{k}{j} (1 - e^{-\beta t})^j (e^{-\beta t})^{k-j} = (1 - e^{-\beta t} + e^{-\beta t})^k = 1$, thus we obtain

$$\sum_{j=0}^k \Pr(Q = j) = \gamma \int_0^\infty e^{-\gamma t} dt = 1.$$

We can clarify the expression in (3) as...
The probability of extinction of this branching process is given by the smallest positive solution of
$$s = f(s),$$
where \(s = (s_1, \ldots, s_L)\) and the components of \(f\) are given by
$$f_k(s) = \sum_{k = (k_1, \ldots, k_L)} p_k(k) s_1^{k_1} \cdots s_L^{k_L},$$
with
$$p_k(k) = \mathbb{P}(\text{an infected node of type } k \text{ infects } k_1 \text{ of type } 1, \ldots, k_L \text{ of type } L).$$

Once we found the smallest solution \(s^*, s_i^* = f_k(s^*)\) represents the extinction probability starting with
one individual of type \(l\). Note that \(f_k(1, \ldots, 1) = 1\); hence \(1 = (1, \ldots, 1)\) is always a solution of \(f(s) = s\).
If we find a smaller solution, then the probability of extinction is smaller than 1; otherwise, the
probability of extinction is 1. We then need to compute \(p_k(k)\). An individual with \(k\) neighbours will have \(k_1\)
neighbours of type 1, \(\ldots, k_L\) of type \(L\) with probability
$$q_i = \frac{f_p}{\sum_{i=1}^L f_p}$$
(size-biased probabilities).

We could then compute the probability of infecting \(m_1\) out of \(k_1\), \(m_2\) out of \(k_2\), \(\ldots, m_L\) out of \(k_L\), and sum
over all possible combinations. We do not need to go through all the components because the probability of
infecting one neighbour is independent of its properties, i.e., its degree. Hence, if \(j = m_1 + \ldots + m_L\),
$$p_k(m) = \mathbb{P}(Q = j) \prod_{i=1}^L q_i^{m_i} d_i^{m_i}.$$  

We then compute the probability of infecting \(j \leq k\)
neighbours; in particular the probability that, among
the \(j\) infected neighbours, \(m_1\) are of type 1, \(m_2\) of type 2, \(\ldots, m_L\) of type \(L\). We can go a further step in the
computation of \(f_k(s)\); indeed
$$f_k(s) = \sum_{m_1=0}^{m_1} \cdots \sum_{m_L=0}^{m_L} \mathbb{P}(Q = j) m_1^{j_1} \cdots m_L^{j_L} q_1^{m_1} \cdots q_L^{m_L}.$$  

One could go beyond this, but the formulae would become increasingly cumbersome. In order to compute
the relevant solution numerically, one can start with a vector \(s^0 \leq s^*\) (for instance,
s^0 = 0 = (0, \ldots, 0)) and then compute \( s^n = f(s^{n-1}) \).

Iterating this, we converge to the required fixed point as a consequence of the Dominated Convergence Theorem.

**Remark 1.** One may notice the property \( s_k^* = f_k(s^*) > f_k(0) = \frac{1}{1 + Rk} \). Hence, a lower bound for the probability of extinction is given by

\[
l := \frac{1}{1 + Rk}.
\]

We recall that \( R = \frac{\beta}{\gamma} \) and \( k \) is the degree of the initially infected node. For example, if \( R = 0.002 \) we get \( s_k^* > \frac{1}{1 + Rk} \approx 0.91 \); thus, we immediately see that the probability of a minor epidemic is very high.

Note that the probability of extinction of a branching process approximates the probability of having a minor epidemic in the SIR model. Indeed, since the population is fixed and acquired immunity is permanent, the epidemics always end in finite time, so the epidemics end with extinction of the disease with probability 1.

**Remark 2.** In Rogers (2015), the author analyses a more general setting, keeping the recovery rate generic in the form of a function \( \gamma(\cdot) \) before specialising it for the numerical simulations. Our construction from (3) onwards could be generalised as well to consider infectious periods that follow laws more realistic than the geometric distribution, which we use throughout our analysis in this paper.

In order to see if the probability of extinction is 1 or lower, we can resort to \( R_0 \), the spectral radius of the \( N \times N \) matrix \( M \), whose elements \( m_{jk} \) are the expected number of infected of type \( j \) generated by an infected of type \( k \). This is easy to compute, conditioning on the number of neighbours:

\[
m_{jk} = \mathbb{E}(\text{infected of type } j \mid k_j \text{ neighbours of type } j)
\]

\[
= \mathbb{E}(k_j R + 1) = \frac{R}{R + 1} \sum_{l=1}^L k_{jl} \lambda_l.
\]

Notice that \( M \) is a matrix of rank 1; hence, its spectral radius is easy to compute. From \( Mv = \rho v \), we obtain

\[
\sum_{k=1}^L m_{jk} v_k = \frac{R}{R + 1} \sum_{l=1}^L \lambda_l \sum_{k=1}^L k_{jl} v_k = \rho v_j.
\]

Since exists \( C > 0 \) such that \( v_j = C \lambda_j \), thus

\[
C = \frac{R}{R + 1} \frac{\lambda_j}{\sum_{l=1}^L \lambda_l \sum_{k=1}^L k_{jl} v_k} = C \lambda_j \rho v_j,
\]

which implies

\[
\rho = \frac{R}{R + 1} \frac{\sum_{k=1}^L k^2 p_k}{\sum_{k=1}^L k p_k}.
\]

This is exactly the expression in (2).

### 4. The stochastic model

We build a network following the CM algorithm, choosing \( k_{\text{min}} = 1 \), and fixing \( N = 10000 \); \( k_{\text{max}} \) then varies with \( \alpha \) as described by

\[
k_{\text{max}} = \left[ k_{\text{min}} N^{\alpha} \right].
\]

We focus on the interval \( 2 \leq R_0 \leq 3 \). The value of \( R_0 \) varies with the parameters \( \alpha \in [2, 3] \), which is the most commonly used interval for the power-law and \( \beta \in [0.002, 0.619] \), following (2). First, we consider five couples of values for the parameters \( \alpha \) and \( \beta \), and we compare the behaviour of the epidemics with different numbers of initially infected nodes. Later, we expand the spectrum of parameters taking into account ten values for each. It should be noted that the recovery rate \( \gamma \) has been fixed to \( \gamma = 1 \), without loss of generality, since this amounts to a rescaling of the chosen time unit by \( \gamma \). In particular, notice that this implies that all the times in our simulations have been rescaled by that same factor \( \gamma \). Depending on the disease, it would be necessary to multiply those times by the average recovery period to obtain a result in days.

The simulation algorithm implements a specialised version of the Gillespie algorithm (Gillespie, 1977), which drastically reduces the simulation times compared to the standard implementation.

In the case of the SIR model, we have two kinds of events. In the standard chemical reaction network notation, these are the following: \( S + I \rightarrow I + I \), which is a second-order reaction since it requires two connected individuals, one in the susceptible state and one in the infected state to happen: this results in the infection of the susceptible individual with probability \( \beta \); and \( I \rightarrow R \), which is a first-order reaction since it happens at a node level. The standard implementation of the algorithm would require \( 2m + N \) reactions; recall that \( m \) is the number of edges in the network and \( N \) is the number of nodes. This computational overhead can be prohibitive for a large scale analysis.

The main difference in our implementation is in the computation of the second-order reaction propensities: on a network, the number of possible infection events is given by the number of edges that connect the susceptible and infected nodes in the network. By exploiting this fact, our algorithm computes the propensity of the second-order reactions as the product of}

\[
\rho = \frac{R}{R + 1} \frac{\sum_{k=1}^L k^2 p_k}{\sum_{k=1}^L k p_k}.
\]
the number of these edges and the reaction rate constant \( \beta \). This simplification drastically reduces the simulation times in comparison to a direct encoding.

In each simulation, we recorded the peak of infected nodes and the number of eventually infected, that is, the total number of recovered at the end of the epidemic with \( I = 0 \) in the SIR model. For each combination of values \( \alpha \) and \( \beta \), we simulated the epidemic 100 times, and we computed the mean of the aforementioned epidemic indices as depicted in Figure 1. With 100 simulations for each pair, the largest coefficient of variation which can be found (namely, the one computed for the eventually infected in the random case) is \( CV_{\text{max}} = 8.71 \); the mean of the \( CV \) s in the random case is 2.273, while in all the other settings it is close to 1.

4.1. Comparison between different numbers of initially infected nodes

We first considered as separate cases epidemics with three different values of initially infected nodes, namely \( I(0) = 1, I(0) = 5 \) and \( I(0) = 10 \), with four possible initial positions in the network: hub, meaning a node whose degree is in the tail of the distribution; mean degree; peripheral, meaning a node with a low degree; and randomly chosen. In this section, we illustrate how changing the position of the initially infected node influences the dynamics of the spread of the disease on a network. Indeed, when the first infected nodes are on the periphery, the epidemic is diffusing slowly compared to the case in which the epidemic starts in a node with more contacts, as shown in Figure 1 for the greatest value of the parameter \( \beta \). The random case is qualitatively intermediate between the peripheral case and the mean-degree case: this follows from the distribution of the degree of nodes in the network since we have few nodes in the tail of the distribution (hub). The resulting numbers of infected nodes do not change significantly between the \( I(0) = 5 \) and \( I(0) = 10 \) cases, depicted in Figure 1 by green and blue lines, respectively, suggesting a saturation effect even for a small number of initially infected nodes (compared to the total population). Moreover, in the random, mean-degree and

![Log-plot of the average of the peak of infected nodes, comparing different numbers of initially infected nodes and their positions.](image1)

![Logplot of the average of the eventually infected nodes, comparing different numbers of initially infected nodes and their positions.](image2)

**Figure 1.** Comparison of the effect of taking different numbers of initially infected nodes in different positions on the average peak or the average number of eventually infected nodes. The hub case is analysed in greater detail below. Solid line: random case. Dashed lines: peripheral case. Dotted lines: mean case. Blue: 10 initially infected individuals. Green: 5. Black: 1.
peripheral cases, the numbers of infected nodes (both simultaneously and eventually) increase as the initial number of infected nodes grows.

The hub case, shown in Figure 2, deserves particular attention: as expected, the epidemic spreads more with an increase in $\beta$ and less with an increase in $\alpha$; recall that a greater $\alpha$ indicates an overall less connected network. Moreover, the more central the initially infected node is, the more the disease can spread through the network and more so with higher numbers of initially infected nodes. This is because the higher number of initially infected nodes, the smaller the probability of all of them being in a hub, disconnected from the giant cluster, except for the greatest value of $\beta$. This behaviour can be explained in two distinct ways: firstly, the positions of the initially infected nodes are assigned in a “hierarchical manner”. That is, since the hubs live in the tail of the power-law distribution, the second infected node inserted in the network has a degree lower than the first one. Secondly, if two neighbouring nodes are surrounded by several infected ones, they will be infected and they will recover approximately simultaneously. This, in return, will create a saturation in the number of available susceptible nodes.

![Figure 2](image.png)

- **(a)** Peak of infected nodes.
- **(b)** Number of eventually infected nodes.
- **(c)** End time for the epidemic.

**Figure 2.** Comparison between different numbers of initially infected nodes in the hub-degree.
The hub case is a “worst case scenario”: the epidemic spreads even if it starts from only one infected node, if it has enough strength in terms of its parameters to do so. Moreover, it is very unlikely that more than one infected node enters a susceptible population simultaneously. Hence, in the next section, we analyse the hub scenario with \( I(0) = 1 \) in greater detail.

Furthermore, we observe a pattern in the final time in Figure 2c, which can be explained as follows: for lower values of \( \beta \) and greater values of \( \alpha \) the epidemic does not take off. On the other hand, for greater values of \( \beta \) and lower values of \( \alpha \) the epidemic quickly goes to extinction because it starts from a node or nodes with more connections. The noise in the final time, compared to the other quantities, is to be expected since the final part of the epidemic is a subcritical branching process, characterised by variability in its duration (Windridge, 2018).

4.2. Comparison between different positions of one initially infected node

In this section, we expand the analysis above on (Figure 3) to a broader setting. Recalling Equation (2), the higher values of \( \beta \) correspond to higher values of \( R_0 \), thus causing a greater spread of the epidemic for fixed values of \( \alpha \). The opposite is true for \( \alpha \): greater values of \( \alpha \) mean a less connected network, which impairs the spread of the disease as it can be noticed in Figure 3. However, it is less evident from Equation (2) that greater values of \( \alpha \) imply smaller values of \( R_0 \). In this regard, our simulations confirm the intuition and what was known in the literature (Barabási, 2015).

The more central and connected the initially infected node is, the greater the magnitude of the spread, as the heatmaps labelled “hub”, “mean degree” and “peripheral” illustrate; the “random” heatmaps give, qualitatively, an average of the other three. The only quantity which remains noisy is the final time, visualised in Figure 3c. For the “hub node” initial position, the tendency is the same as in Figure 2c. However, in the other cases, the behaviour in the measurements follow a similar trend, i.e., values decreasing with \( \alpha \) and increasing with \( \beta \) due to the weakness of the disease together with the effect of the initially infected node positions.

4.3. Analytical vs. stochastic “hub case”

In this section, we verify our stochastic model with respect to the analytical results in Section 3, for the case where the infection starts from a hub of the network. The first infected node is chosen as the node with the maximum degree in the network and the degree is inversely proportional to \( \alpha \); this follows from expression (1).

We considered 100 stochastic simulations for each pair of parameters \( \alpha \) and \( \beta \). For each choice of these parameters, we evaluated the probability of a minor epidemic using the analytical derivation in Section 3 for a branching process. We denote this with \( P^{\alpha,\beta}_{\text{ext}} \). Note that for the stochastic simulations there is no formal definition of a minor epidemic. We thus considered different threshold values to have a small number of eventually infected nodes. The thresholds considered in this study are 1, 2, 3 and 4 nodes, which represent epidemics that do not take off.

After collecting the simulation data, we obtained the probability of a minor epidemic for each pair of parameters \( (\alpha, \beta) \) for the four different thresholds. Since we ran 100 simulations, we normalised the results to have a probability measure such that, in the notation \( P^{\alpha,\beta}_{\text{stoc}} \), \( T \) is the value of the chosen threshold. We then computed the difference between the simulation data (for each threshold) and the analytical probability of a minor epidemic. With \( T \in \{1, 2, 3, 4\} \), we compute

\[
\text{diff}(\alpha, \beta, T) = P^{\alpha,\beta}_{\text{stoc}} - P^{\alpha,\beta}_{\text{ext}}. 
\]

Since we consider 10 different values for each of the parameters \( \alpha \) and \( \beta \), we obtained 100 values of \( \text{diff}(\alpha, \beta, T) \) (for each threshold \( T \)).

In order to choose the “best threshold value”, we computed the mean and the standard deviations for all the datasets \( \{\text{diff}(\alpha, \beta, T)\}_{(\alpha, \beta)} \). From Table 1, we can observe that the best agreement is with the threshold \( T = 3 \).

After choosing the threshold to define a minor epidemic, we compared the analytical and simulation results to verify our model. This comparison is depicted in Figure 4. The simulation results are in good agreement with the analytical results, especially for the smallest values of the transmission rate \( \beta \); in the other cases, the results differ no more than the 3%. This behaviour can be explained by the fact that theoretically we consider a network with an infinite number of nodes, but practically we use a finite network with a large number of nodes. In Figure 4, we compare, for the probability of a minor epidemic, the analytical value (blue) obtained in Section 3 with the stochastic realisations (black) by simulations. For each \( (\alpha, \beta) \) pair, we performed 100 simulations. We observe that for the smallest value of \( \beta \), probability of a minor epidemic seems to be almost independent of the connectivity of the network. Moreover, when \( \beta = 0.002 \), even in the cases where the epidemic does not immediately die out, the number of eventually infected individuals in each of the 100 iterations does not exceed 15. However, when the transmission rate increases, the probability of having
a minor epidemic increases for greater values of $\alpha$. This behaviour can be noted also in Figure 3b (“Hub node”).

**Remark 3.** The probability of a minor epidemic, even in the analytical case, depends on the degree of the initially infected node chosen as the “hub node”. However, even though the value of the maximum degree of each network follows (1) (and thus it should decrease with $\alpha$) when the networks are simulated, the degree $k_{\text{max}}$ is rarely reached. Therefore, the degree of the hub node is not monotonic with $\alpha$, and the blue bars are accordingly sometimes not monotonic.
5. Discussion and outlook

We have shown how an epidemic affects a synthetic network, following a cut-off power-law distribution for the degree of the nodes through the use of the configuration model. We provided an analytical result on the probability of minor epidemic of the infectious disease, which is based on the initial condition of the network and the degree of the initially infected node. Our analytical results and stochastic simulation results are in good agreement; the gap in the values stems from the fact that we have considered an infinite network in the analytical framework, and a finite network, yet with a large number of nodes, in the stochastic one.

We have analysed various possible initial conditions for the networks we simulated and compared them with the available analytical insight. In particular, we explored how the positions of the initially infected individuals influence the whole epidemic; measured through three indices: eventually infected individuals, peak of infected individuals and overall duration of the epidemic. Our analysis and numerical exploration confirm that the infectiousness of the disease is directly proportional to the spread of the epidemic; moreover, we conclude that the same disease (i.e., characterised by the same parameters $\beta$ and $\gamma$) infects more individuals in networks generated with a lower exponent power-law, as one would expect intuitively.
We have simulated the epidemic as an SIR model on a network with stochastic dynamics using a specialised and computationally less expensive version of the Gillespie algorithm. Our algorithm is versatile; many different additional features can be implemented, for example, dynamics edges, contact tracing, quarantine. Other generalisations of this work may regard the use of a different model for the epidemic process, e.g., considering a disease in which recovered individuals lose their immunity after a certain period of time (i.e., a SIRS model) or a disease in which there are different types of infected individuals. Moreover, it could be interesting to analyse different structures of the networks, studying how the geometry of the graph may affect the dynamics which occurs on it.

Other topics which we leave for further investigation include choosing the initially infected nodes based on different centrality measures such as betweenness, percolation centrality (Piraveenan et al., 2013), eigenvector centrality, random walk centrality, and compare the outcomes in the these settings. A thorough comparison of different centrality measures would highlight the prevalent one in the transmission of infectious diseases in populations modelled by a scale-free network of contacts.

Notes
1. A positive solution $s^*$ of $s = f(s)$ is called the smallest if $s^*_i \leq r_i$ for all $i = 1, \ldots, L$, for any other solution $r$.
2. All the codes and additional data are available on https://github.com/SaraSottile/StochasticSIRnetwork. Animations of sample simulations with different rates are available at https://www.youtube.com/playlist?list=PLdDHYeVsbaLUY7-9gt9F01JEgIFm8D09m

Acknowledgments
The authors thank Prof. Andrea Pugliese for his helpful comments and expert help in various stages of the manuscript. MS would like to thank the University of Trento for supporting his research during the final months of his PhD.

Disclosure statement
No potential conflict of interest was reported by the author(s).

ORCID
Sara Sottile https://orcid.org/0000-0001-7810-2099
Ozan Kahramanoğulları https://orcid.org/0000-0001-5178-7470
Mattia Sensi https://orcid.org/0000-0002-6631-0242

References
Albert, R., Jeong, H., & Barabási, A. (1999). Diameter of the world-wide web. Nature, 401(6749), 130–131. https://doi.org/10.1038/43601
Albert, R., Jeong, H., & Barabási, A. (2000). Error and attack tolerance of complex networks. Nature, 406 (6794), 378.381. https://doi.org/10.1038/35019019
Aleta, A., Martin-Corral, D., Y Piontelli, A. P., Ajelli, M., Litvinova, M., Chinazzi, M., Dean, N. E., Halloran, M. E., Longini Jr, I. M., Jr, Merler, S., Pentland, A., Vespignani, A., Moro, E., & Moreno, Y. (2020). Modelling the impact of testing, contact tracing and household quarantine on second waves of COVID-19. Nature Human Behaviour, 4(9), 964–971. https://doi.org/10.1038/s41562-020-0931-9
Artico, I., Smolyarenko, I., Vinciotti, V., & Wit, E. C. (2020). How rare are power-law networks really? Proceedings of the Royal Society A, 476(2241), 20190742. https://doi.org/10.1098/rspa.2019.0742
Ball, F., Sirl, D., & Trapman, P. (2010). Analysis of a stochastic SIR epidemic on a random network incorporating household structure. Mathematical Biosciences, 224(2), 53–73. https://doi.org/10.1016/j.mbs.2009.12.003
Barabási, A. L., & Albert, R. (1999). Emergence of scaling in random networks. Science, 286(5439), 509–512. https://doi.org/10.1126/science.286.5439.509
Barabási, A. L. (2015). Network science. Cambridge University Press.
Bertotti, M. L., & Modanese, G. (2019). The configuration model for barabasi-albert networks. Applied Network Science, 4(1), 1–13. https://doi.org/10.1007/s11109-019-0152-1
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., & Hwang, D. (2006). Complex networks: Structure and dynamics. Physics Reports, 424(4–5), 175–308. https://doi.org/10.1016/j.physrep.2005.10.009
Boguná, M., Castellano, C., & Pastor-Satorras, R. (2013). Nature of the epidemic threshold for the susceptible-infected-susceptible dynamics in networks. Physical Review Letters, 111(6), 068701. https://doi.org/10.1103/PhysRevLett.111.068701
Boguñá, M., Pastor-Satorras, R., & Vespignani, A. (2004). Cut-offs and finite size effects in scale-free networks. European Physical Journal B, 38(2), 205–209. https://doi.org/10.1140/epjb/e2004-00038-8
Broido, A. D., & Clauset, A. (2019). Scale-free networks are rare. Nature Communications, 10(1), 1–10. https://doi.org/10.1038/s41467-019-08746-5
Caldarelli, G., & Vespignani, A. (2007). Large scale structure and dynamics of complex networks: From information technology to finance and natural science. World Scientific Publishing Company.
Cator, E., & Van Mieghem, P. (2012). Second-order mean-field susceptible-infected-susceptible epidemic threshold. Physical Review E, 85(5), 056111. https://doi.org/10.1103/PhysRevE.85.056111
Chakrabarti, D., Wang, Y., Wang, C., Leskovec, J., & Faloutsos, C. (2008). Epidemic thresholds in real networks. ACM Transactions on Information and System Security (TISSEC), 10(4), 1–26. https://doi.org/10.1145/1284680.1284681
Chang, S. L., Piraveenan, M., & Prokopenko, M. (2020). Impact of network assortativity on epidemic and vaccination behaviour. Chaos, Solitons, and Fractals, 140, 110143. https://doi.org/10.1016/j.chaos.2020.110143
Cliff, O. M., Harding, N., Piraveenan, M., Erten, E. Y., Gambhir, M., & Prokopenko, M. (2018). Investigating spatiotemporal dynamics and synchrony of influenza epidemics in Australia: An agent-based modelling approach. *Simulation Modelling Practice and Theory, 87*, 412–431. https://doi.org/10.1016/j.simpat.2018.07.005
de Arruda, G. F., Rodrigues, F. A., & Moreno, Y. (2018). Fundamentals of spreading processes in single and multilayer complex networks. *Physics Reports, 756*, 1–59. https://doi.org/10.1016/j.physrep.2018.06.007
Draief, M., Ganesh, A., & Massoulié, L. (2006). Thresholds for virus spread on networks. Proceedings of the 1st international conference on performance evaluation methodologies and tools (pp. 51–es).
Durrett, R. (2006). *Random graph dynamics*. Cambridge University Press.
Eguiluz, V. M., Chialvo, D. R., Cecchi, G. A., Baliki, M., & Apkarian, A. V. (2005). Scale-free brain functional networks. *Physical Review Letters, 94*(1), 018102. https://doi.org/10.1103/PhysRevLett.94.018102
Erdös, P., & Rényi, A. (1959). On random graphs I. *Publicationes Mathematicae Debrecen, 6*, 290. https://doi.org/10.1515/9781400841356.38
Estrada, E. (2011). *The structure of complex networks: Theory and applications*. Oxford University Press.
Ferretti, L., Wyman, C., Kendall, M., Zhao, L., Nurtay, A., Abeler-Dörner, L., Parker, M., Bonsall, D., & Fraser, C. (2020). Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science, 368* (6491), 6491. https://doi.org/10.1126/science.abb6936
Firth, J. A., Hellewell, J., Klepac, P., Kissler, S., Kucharski, A. J., & Spurgeon, L. G. (2020). Using a real-world network to model localized COVID-19 control strategies. *Nature Medicine, 26*(10), 1616–1622. https://doi.org/10.1038/s41591-020-1036-8
Flanders, D., & Kleinbaum, D. (1995). Basic models for disease occurrence in epidemiology. *International Journal of Epidemiology, 24*(1), 1–7. https://doi.org/10.1093/ije/24.1.1
Fox Keller, E. (2005). Revisiting "scale-free" networks. *BioEssays, 27*(10), 1060–1068. https://doi.org/10.1002/bies.20294
Ganesh, A., Massoulié, L., & Towsley, D. (2005). The effect of network topology on the spread of epidemics. Proceedings ieee 24th annual joint conference of the ieee computer and communications societies. (Vol.2, pp. 1455–1466).
Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry, 81*(25), 2340–2361. https://doi.org/10.1021/j100540a008
Harris, T. E. (1963). *The theory of branching processes*. Springer-Verlag.
Hindes, J., & Schwartz, I. B. (2016). Epidemic extinction and control in heterogeneous networks. *Physical Review Letters, 117*(2), 028302. https://doi.org/10.1103/PhysRevLett.117.028302
Hofstad, R. V. D. (2016). *Random graphs and complex networks* (Vol. 1). Cambridge University Press.
Holme, P. (2013). Extinction times of epidemic outbreaks in networks. *PloS one, 8*(12), e84429. https://doi.org/10.1371/journal.pone.0084429
Holme, P., & Tupikina, L. (2018). Epidemic extinction in networks: Insights from the 12 110 smallest graphs. *New Journal of Physics, 20*(11), 113042. https://doi.org/10.1088/1367-2630/aaf016
Holme, P. (2019). Rare and everywhere: Perspectives on scale-free networks. *Nature Communications, 10*(1), 1–3. https://doi.org/10.1038/s41467-019-09038-8
Iannelli, M., & Pugliese, A. (2014). *An introduction to mathematical population dynamics*. Springer.
Karrer, B., & Newman, M. E. J. (2010). Message passing approach for general epidemic models. *Physical Review E, 82*(1), 016101. https://doi.org/10.1103/PhysRevE.82.016101
Keeling, M. J., & Grenfell, B. T. (2000). Individual-based perspectives on R0. *Journal of Theoretical Biology, 203*(1), 51–61. https://doi.org/10.1006/jtbi.1999.1064
Keeling, M. (2005). The implications of network structure for epidemic dynamics. *Theoretical Population Biology, 67*(1), 1–8. https://doi.org/10.1016/j.tpb.2004.08.002
Keeling, M., & Eames, K. (2005). Networks and epidemic models. *Journal of the Royal Society, Interface/the Royal Society, 10* 2(4), 295–307. https://doi.org/10.1098/rsif.2005.0051
Keeling, M. J., & Rohani, P. (2011). *Modeling infectious diseases in humans and animals*. Princeton university press.
Kiss, I., Miller, J., & Simon, P. (2017). *Mathematics of epidemics on networks: From exact to approximate models*. Springer.
Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. *Nature Physics, 6*(11), 888–893. https://doi.org/10.1038/nphys1746
López-García, M. (2016). Stochastic descriptors in an SIR epidemic model for heterogeneous individuals in small networks. *Mathematical Biosciences, 271*, 42–61. https://doi.org/10.1016/j.mbs.2015.10.010
Miller, J. C. (2008). Bounding the size and probability of epidemics on networks. *Journal of Applied Probability, 45*(2), 498–512. https://doi.org/10.1239/jap/1214950363
Miller, J. C. (2012). A note on the derivation of epidemic final sizes. *Bulletin of Mathematical Biology, 74*(9), 2125–2141. https://doi.org/10.1007/s11538-012-9794-6
Min, B. (2018). Identifying an influential spreader from a single seed in complex networks via a message-passing approach. *The European Physical Journal. B, 91*(1), 1–6. https://doi.org/10.1140/epjb/e2017-80597-1
Mode, C. J. (1971). *Multitype branching processes. Theory and applications*. American Elsevier Pub. Co.
Newman, M. E. J., & Park, J. (2003). Why social networks are different from other types of networks. *Physical Review E, 68*(3), 036122. https://doi.org/10.1103/PhysRevE.68.036122
Newman, M. (2010). *Networks: An introduction*. Oxford University Press.
Pastor-Satorras, R., Castellano, C., Van Mieghem, P., & Vespignani, A. (2015, August). Epidemic processes in complex networks. *Review Mod. Phys, 87*(3), 925–979. https://doi.org/10.1103/RevModPhys.87.925
Piraveenan, M., Prokopenko, M., & Hossain, L. (2013). Percolation centrality: Quantifying graph-theoretic impact of nodes during percolation in networks. *PloS one, 8*(1), e53095. https://doi.org/10.1371/journal.pone.0053095
Radicchi, F., & Castellano, C. (2016). Leveraging percolation theory to single out influential spreaders in networks. *Physical Review E, 93*(6), 062314. https://doi.org/10.1103/PhysRevE.93.062314
Rocha, L. E. C., Liljeros, F., & Holme, P. (2011). Simulated epidemics in an empirical spatiotemporal network of 50,185 sexual contacts. *PLoS Computational Biology, 7* (3), e1001109. https://doi.org/10.1371/journal.pcbi.1001109

Rogers, T. (2015). Assessing node risk and vulnerability in epidemics on networks. *EPL (Europhysics Letters)*, 109(2), 28005. https://doi.org/10.1209/0295-5075/109/28005

Strona, G., Carstens, C. J., Beck, P. S. A., & Han, B. A. (2018). The intrinsic vulnerability of networks to epidemics. *Ecological Modelling*, 383, 91–97. https://doi.org/10.1016/j.ecolmodel.2018.05.013

Sun, K., Wang, W., Gao, L., Wang, Y., Luo, K., Ren, L., Zhan, Z., Chen, X., Zhao, S., Huang, Y., Sun, Q., Liu, Z., Litvinova, M., Vespignani, A., Ajelli, M., Vibound, C., & Yu, H. (2021). Transmission heterogeneities, kinetics, and controllability of SARS-CoV-2. *Science*, 371(6526), 6526. https://doi.org/10.1126/science.abe2424

Thedchanamoorthy, G., Piraveenan, M., Uddin, S., & Senanayake, U. (2014). Influence of vaccination strategies and topology on the herd immunity of complex networks. *Social Network Analysis and Mining*, 4(1), 213. https://doi.org/10.1007/s13278-014-0213-5

Thompson, R. N., Gilligan, C. A., & Cunniffe, N. J. (2020). Will an outbreak exceed available resources for control? Estimating the risk from invading pathogens using practical definitions of a severe epidemic. *Journal of the Royal Society, Interface*, 17(17), 20200690. https://doi.org/10.1098/rsif.2020.0690

Van Mieghem, P., Omic, J., & Kooij, R. (2008). Virus spread in networks. *IEEE/ACM Transactions On Networking*, 17(1), 1–14. https://doi.org/10.1109/TNET.2008.925623

Windridge, P. (2018). The extinction time of a subcritical branching process related to the SIR epidemic on a random graph. *Journal of Applied Probability*, 52(4), 1195–201. https://doi.org/10.1239/jap/1450802763

Zachreson, C., Fair, K. M., Cliff, O. M., Harding, N., Piraveenan, M., & Prokopenko, M. (2018). Urbanization affects peak timing, prevalence, and bimodality of influenza pandemics in Australia: Results of a census-calibrated model. *Science Advances*, 4(12), eaau5294. https://doi.org/10.1126/sciadv.aau5294

Zhang, H., Guan, Z.-H., Li, T., Zhang, X.-H., & Zhang, D.-X. (2013). A stochastic SIR epidemic on scale-free network with community structure. *Physica A: Statistical Mechanics and Its Applications*, 392(4), 974–981. https://doi.org/10.1016/j.physa.2012.10.016