The University of
Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Ball, Frank and House, Thomas (2017) Heterogeneous
network epidemics: real-time growth, variance and
extinction of infection. Journal of Mathematical Biology .
pp. 1-43. ISSN 1432-1416

Mathematical Biology

CrossMark

Heterogeneous network epidemics: real-time growth, variance and extinction of infection

Frank Ball[1] · **Thomas House**[2]

Abstract Recent years have seen a large amount of interest in epidemics on networks as a way of representing the complex structure of contacts capable of spreading infections through the modern human population. The configuration model is a popular choice in theoretical studies since it combines the ability to specify the distribution of the number of contacts (degree) with analytical tractability. Here we consider the early real-time behaviour of the Markovian SIR epidemic model on a configuration model network using a multitype branching process. We find closed-form analytic expressions for the mean and variance of the number of infectious individuals as a function of time and the degree of the initially infected individual(s), and write down a system of differential equations for the probability of extinction by time t that are numerically fast compared to Monte Carlo simulation. We show that these quantities are all sensitive to the degree distribution—in particular we confirm that the mean prevalence of infection depends on the first two moments of the degree distribution and the variance in prevalence depends on the first three moments of the degree distribution. In contrast to most existing analytic approaches, the accuracy of these results does not depend on having a large number of infectious individuals, meaning that in the large population limit they would be asymptotically exact even for one initial infectious individual.

✉ Thomas House
thomas.house@manchester.ac.uk

Frank Ball
frank.ball@nottingham.ac.uk

1 School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK

2 School of Mathematics, University of Manchester, Oxford Road, Manchester M13 9PL, UK

Springer

1 Introduction

1.1 Background

Models of infectious disease transmission have, from relatively modest beginnings (e.g. Bailey 1957), developed a rich domain of applicability covering the whole spectrum of human, animal and plant pathogens, and informing the study of questions from viral evolution, through epidemiology of infectious diseases, to public health policy (see Heesterbeek et al. 2015). Increasingly, networks have been seen as a way of modelling the complex, heterogeneous patterns of contacts between individuals (Danon et al. 2011).

In theoretical studies, the configuration model has been a popular choice due to the ability to specify the number of contacts each individual has that are capable of spreading disease, while allowing for analytic results to be obtained (e.g. Molloy and Reed 1995; Newman 2002). Ball and Neal (2008) used an effective degree approach (which we describe in Sect. 2.2 below—cf. Lindquist et al. 2010) to derive a system of ordinary differential equations that describes the deterministic limit of the epidemic model as the population size $N \to \infty$. A much simpler (equivalent) system of only 4 ordinary differential equations was obtained by Volz (2008) and subsequently shown by Miller (2011), Miller et al. (2012) to be essentially one-dimensional (the 4 ODEs were also shown by House and Keeling (2010) to be a special case of the much higher dimensional pair approximation model of Eames and Keeling (2002), in which the degree structure is explicit). Fully rigorous proofs of convergence in probability of the scaled stochastic model to the deterministic limit are given by Decreusefond et al. (2012), Bohman and Picollelli (2012), Barbour and Reinert (2013) and Janson et al. (2014). These works are primarily concerned with the temporal behaviour of *proportions* of the population in different epidemiological compartments (susceptible, infectious and removed) over the main body of a large epidemic. Here, we are also concerned with temporal behaviour, but focus on *numbers* infected early in the epidemic, including the possibility of early stochastic extinction.

In a recent paper, Graham and House (2014) use a pairwise approximation in conjunction with the central limit theorem for density dependent population processes (Ethier and Kurtz 1986, Chap. 11) to obtain a closed-form approximation to the mean and variance of prevalence in the linearised model which approximates the early asymptotic exponential growth phase of a Markovian SIR epidemic on a configuration network. In particular, they find that, under these approximations, the variance in disease prevalence is determined by the first three moments of the network degree distribution. In this paper, we use the effective degree approach of Ball and Neal (2008) to approximate the early stages of the epidemic by a continuous-time, multitype Markovian branching process, which is then analysed in detail. For $t \geq 0$, let $Z(t)$ denote the total number of individuals alive in this branching process at time t, so $Z(t)$ approximates disease prevalence in the epidemic model during its early asymptotic exponential growth phase. Explicit closed-form expressions are derived for the mean and variance of $Z(t)$, the covariance of $Z(t)$ and $Z(s)$ to give the behaviour over time, and also for the probability of extinction $\pi(t) = \mathbb{P}(Z(t) = 0)$. As in Graham and House (2014), the mean and variance in disease prevalence

depends on the degree distribution only through its first two and three moments, respectively.

The results in Graham and House (2014) assume implicitly that the initial number of infectives is sufficiently large for the density dependent population process central limit theorem to yield a good approximation. In contrast, our results assume any arbitrary, but specified, initial number of infectives. The asymptotic distribution of types in the branching process, when it does not go extinct, is also available in closed-form and enables us to obtain a Gaussian process approximation, with explicit mean and covariance function, for the prevalence in the early asymptotic exponential growth phase of an SIR epidemic, with few initial infectives, which takes off and becomes established. We show that this approximation can be applied together with the methods of Ross et al. (2006) to estimate epidemiological parameters from early prevalence data of a simulated epidemic provided the first three moments of the degree distribution are known.

1.2 Outline of the paper

The paper is organised as follows. The configuration network model and a Markov SIR epidemic on that network are described in Sect. 2.1. The effective degree construction of this epidemic is outlined in Sect. 2.2. Approximation of the early stages of this epidemic by a branching process is outlined in Sect. 2.3, where conditions are given for the mean, variance and covariance functions of the number of infectives in the epidemic process to converge to the corresponding quantities of the approximating branching process as the population size tend to infinity. The representation of the approximating branching process as a continuous-time, multitype Markov branching process is outlined in Sect. 2.3 and described more explicitly in Sect. 2.4. The mean, variance and covariance functions of the total number of individuals alive in the branching process are considered in Sects. 3, 4 and 5, respectively. Explicit closed-form expressions are obtained for each of these quantities and for their limits as time $t \to \infty$. The arguments in Sects. 3, 4 and 5 assume that underlying degree distribution has a maximum degree. In Sect. 6, we show that these expressions continue to hold in the unbounded degree setting, subject to the degree distribution satisfying suitable moment conditions. The probability that the branching process is extinct at time t is studied in Sect. 7. Closed-form expressions for this probability, given the initial state of the branching process, are not available so asymptotic results as $t \to 0$ and $t \to \infty$ are considered.

The mean, variance and covariance functions derived in Sects. 3, 4 and 5 are unconditional, so they include realisations of the branching process which result in extinction. However, in the epidemic setting, we are often interested in analysing the behaviour of epidemics that take off and become established, which correspond to non-extinction of the branching process. In Sect. 8, we first derive the mean and variance of the total number of individuals alive in the branching process at time t, conditional upon the process having survived to time t; fully closed-form results are not available owing to the absence of a closed-form expression for the survival probability. We then consider realisations of the branching process which reach some specified size, K say, with time being set to zero the first time the total number of individuals alive is K. The results in Sect. 3 yield an explicit expression for the asymptotic distribution of types, given that the branching process does not go extinct, which, provided K is sufficiently

large, enables the above branching process starting from K individuals to be approximated by a Gaussian process whose mean and covariance functions are determined explicitly. The theory is illustrated by numerical examples of both forward simulation and inference in Sect. 9 and some concluding comments are given in Sect. 10.

In general, we define notation as it is introduced; we also collect notation that is used in multiple sections in Table 1.

2 Model and approximating branching process

2.1 Model

We consider the spread of an SIR epidemic on a network of N individuals, labelled $1, 2, \ldots, N$, constructed using the configuration model as follows (see e.g. Newman 2002). Let D be a random variable which describes the degree of a typical individual and let $p_k = \mathbb{P}(D = k)$ $(k = 0, 1, \ldots)$. Let $D_1, D_2, \ldots, D_N$ be independent realisations of D and, for $i = 1, 2, \ldots, N$, attach D_i stubs (half-edges) to individual i. Pair up these stubs uniformly at random to form the edges in the network. If $D_1 + D_2 + \cdots + D_N$ is odd, there will be a left-over stub, which is ignored; the resulting network may have other 'defects' such as self-loops and multiple edges between pairs of individuals but, provided that D has finite variance, such imperfections become sparse in the network as $N \to \infty$ (see e.g. Durrett 2007, Theorem 3.1.2). An alternative to the degrees $D_1, D_2, \ldots, D_N$ being random is, for each $N = 1, 2, \cdots$, to replace $\mathbf{D} = (D_1, D_2, \ldots, D_N)$ by $\mathbf{D}^{(N)} = \left(D_1^{(N)}, D_2^{(N)}, \ldots, D_N^{(N)} \right)$, where the degree sequences $\mathbf{D}^{(N)}$ $(N = 1, 2, \ldots)$ are prescribed and satisfy $p_k^{(N)} = N^{-1} \sum_{i=1}^{N} \delta_{k, D_i^{(N)}} \to p_k$ as $N \to \infty$ $(k = 0, 1, \ldots)$, where the Kronecker delta $\delta_{k,j}$ is 1 if $k = j$ and 0 otherwise (see e.g. Molloy and Reed 1995).

The epidemic is defined as follows. Initially some individuals are infective and the remaining individuals are susceptible. Infective individuals have independent infectious periods, each having an exponential distribution with rate γ (and hence mean γ^{-1}), after which they become recovered and play no further role in the epidemic. Throughout its infectious period, an infective contacts each of its susceptible neighbours in the network at the points of independent Poisson processes having rate τ, so the probability that a given infective contacts a given neighbour before the infective recovers is $\tau/(\gamma + \tau)$. Any contacted susceptible immediately becomes infective and may transmit the infection to any of its neighbouring susceptibles; i.e. there is no latent period. All the infectious periods and Poisson processes governing transmission of infection are mutually independent. The epidemic ends as soon as there is no infective present in the network.

2.2 The effective degree model

Ball and Neal (2008) introduced an 'effective-degree' construction of the above epidemic, in which the network is constructed as the epidemic progresses. The process starts with some individuals infective and the remaining individuals susceptible, but

Table 1 Notation used in multiple sections of this paper

Primary notation	Meaning	Equivalent notation
Network properties		
N	Size of the population	
D	A random variable for an individual's degree	$D_1, D_2, \ldots$
p_k	Probability mass function for D evaluated at k	
$k_{\max}$	The maximum degree	
$\mathcal{K}$	The set of possible degrees	$\{0, 1, \ldots k_{\max}\}$
$\tilde{D}$	A random variable for an individual's size-biased degree	$\tilde{D}_1, \tilde{D}_2, \ldots$
$\tilde{p}_k$	Probability mass function for $\tilde{D}$ evaluated at k	$\mu_D^{-1} k p_k$
Vectors and matrices		
$\mathbf{v}$	A column vector whose kth entry is v_k	(v_k)
$\mathbf{v}^\top$	A row vector (transpose of a column vector)	
$\boldsymbol{M}$	A matrix with (k, l)th entry M_{kl} or m_{kl}	$[M_{kl}], [m_{kl}]$
$\|\boldsymbol{M}\|$	Determinant of matrix $\boldsymbol{M}$	
$\mathbf{1}$	A column vector whose entries are all equal to 1	
$\mathbf{n}$	A column vector whose ith entry is i	
$\mathbf{n_2}$	A column vector whose ith entry is i^2	
$\boldsymbol{I}$	The identity matrix	$[\delta_{k,l}]$
Probability		
$\mathbb{P}(e)$	Probability of event e	
$\mu_{f(X)}$	Expected value of a function f of a random variable X	$\mathbb{E}[f(X)]$
$M_X(\theta)$	Moment generating function for random variable X	$\mathbb{E}\left[\exp(\theta X)\right]$
$\text{var}(X)$	Variance of random variable X	$\mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$
$\text{cov}(X, Y)$	Covariance of random variables X and Y	$\mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$
Epidemic and branching process dynamics		
τ	Rate of transmission across a network link	
γ	Rate of recovery from infection	
ω_k	Death rate for individual of type k	
t	Real time	s
$\mathcal{B}$	The limiting branching process	
E_N	The epidemic process in a population of size N	
K	A large value of infectious population size	
$Z_i^{(k)}(t)$	Random number of individuals of type i in the branching process at time t given initial type k	
$\pi_k(t)$	Probability that the branching process is extinct at time t given initial type k	$1 - q_k(t)$

with none of the stubs paired up. For $i = 1, 2, \ldots, N$, the effective degree of individual i is initially D_i. Infected individuals transmit infection by pairing their stubs with stubs attached to susceptible individuals in the following fashion. An infected individual makes infectious contacts down its unpaired stubs independently at rate τ and is removed at rate γ. When an infective, individual i say, transmits infection down a stub that stub is paired with a stub (attached to individual j, say) chosen independently and uniformly at random from all the unpaired stubs, to form an edge. The effective degrees of individuals i and j are both reduced by 1. If $i = j$ then the effective degree of individual i is reduced by 2 but this will not significantly affect the dynamics for large populations since the probability of it happening is $O(N^{-1})$. If individual j is susceptible then it becomes infective and can transmit infection down any of its unattached stubs. As before, the epidemic ends as soon as there is no infective present. The network is then typically only partially constructed but that does not matter if interest is focussed on properties of the epidemic. In the original formulation of Ball and Neal (2008), when an infective recovers its unpaired stubs, if any, were paired with stubs chosen uniformly at random without replacement from the set of unpaired stubs but that is unnecessary; the stubs from such an infective can simply be left in the set of unpaired stubs.

2.3 Approximating multitype branching process

Suppose that the size N of the network is large and the initial number of infectives is small. Then during the early stages of an epidemic it is very likely that each time an infective individual transmits infection down a stub that stub is paired with a stub belonging to a susceptible individual. It follows that the early stages of such an epidemic can be approximated by a branching process in which each newly-infected individual has their "full" effective degree (i.e. their actual degree minus one for the stub that is paired with their infector). This approximation can be made fully rigorous by considering a sequence of epidemics, indexed by N, and using a coupling argument; see e.g. Ball and Neal (2008), which treats a more general model in which infective individuals also make contacts with individuals chosen uniformly at random from the population. Let E_N denote the epidemic on a network of N individuals and let $\mathscr{B}$ denote the approximating branching process. Then following Ball and Neal (2008) (see "Appendix 1") if $\mu_D = \mathbb{E}[D]$ is finite then the epidemics $E_1, E_2 \ldots$ and the branching process $\mathscr{B}$ can be constructed on a common probability space so that, with probability one, over any finite time interval $[0, t]$ the process of infectives in E_N and the branching process $\mathscr{B}$ coincide for all sufficiently large N. The same result holds for the model with prescribed degree sequences provided that $p_k^{(N)} \to p_k$ ($k = 0, 1, \ldots$) and $\mu_D^{(N)} = \sum_{k=0}^{\infty} p_k^{(N)} k \to \mu_D$ as $N \to \infty$, where $\sum_{k=0}^{\infty} p_k = 1$ and $\mu_D < \infty$.

As indicated in "Appendix 1", the branching process $\mathscr{B}$ is not an almost sure upperbound for the process of infectives in E_N, so unlike in Theorem 3.1 of Ball and Donnelly (1995) which considers homogeneously mixing epidemics, one cannot simply use the dominated convergence theorem to deduce convergence of moments

of the number of infectives in the epidemic process to corresponding moments of the branching process as $N \to \infty$. If there is a maximum degree $k_{\max}$ (i.e. $p_k = 0$ for all $k > k_{\max}$, or $p_k^{(N)} = 0$ for all $k > k_{\max}$ and all N in the model with prescribed degrees) then, for all N, the process of infectives in E_N is bounded above by a branching process in which each newly-infected individual has the maximun effective degree $k_{\max} - 1$, so in that case the dominated convergence theorem can be used to prove convergence of moments. In "Appendix 1", we consider the case when there is no maximum degree and use uniform integrability arguments to determine sufficient conditions for the mean and variance of the number of infectives at any given time $t \geq 0$, and the covariance of the number of infectives at any given times $t, s \geq 0$, in the epidemic E_N to converge to the corresponding mean, variance and covariance of the branching process $\mathscr{B}$ as $N \to \infty$. Specifically, we prove that (i) in the model with prescribed degrees these moments converge if, in addition to the conditions given above, there exists $\delta > 0$ such that $\mu_{D^{3+\delta}}^{(N)} = \sum_{k=0}^{\infty} p_k^{(N)} k^{3+\delta} \to \mu_{D^{3+\delta}} = \sum_{k=0}^{\infty} p_k k^{3+\delta}$ as $N \to \infty$, where $\mu_{D^{3+\delta}} < \infty$; and (ii) in the model with random degrees they converge if the moment-generating function $M_{D^2}(\theta) = \mathbb{E}\left[\exp\left(\theta D^2\right)\right]$ of D^2 is finite for some $\theta_0 > 0$. Note that the latter condition implies that $\mathbb{E}[D^\alpha] < \infty$ for all $\alpha \geq 0$.

In this context, we note that, for the model with prescribed degrees, the weakest conditions obtained on the moments of the degree distribution for convergence of the scaled stochastic epidemic on to its deterministic limit are given by Janson et al. (2014), who require uniform boundedness of the second moment of $\mathbf{D}^{(N)}$. However that paper, and the other related papers cited in the second paragraph of Sect. 1.1, (i) are concerned with the entire time course of the epidemic; (ii) assume that either the epidemic starts with a positive fraction of the population infected in the limit as $N \to \infty$, or if that limiting fraction is zero then the convergence is for epidemics which take off and involves a random time translation describing when the epidemic becomes suitably established; and (iii) consider the evolution of the proportion of the population that is susceptible, infective or recovered. By contrast, this paper is concerned with epidemics initiated by few infectives and considers the number, rather than proportion, of infectives during the early phase of such an epidemic. Under the coupling mentioned above, in the limit as $N \to \infty$, if an epidemic takes off then the duration of its early (exponentially growing) phase tends almost surely to infinity.

The limiting branching process may be described by a continuous-time multitype Markov branching process, with the type of an infective corresponding to its effective degree. Let $\tilde{D}$ denote the (size-biased) degree of a typical neighbour of a typical individual in the network and let $\tilde{p}_k = \mathbb{P}(\tilde{D} = k)$ $(k = 1, 2, \ldots)$. Then $\tilde{p}_k = \mu_D^{-1} k p_k$, where $\mu_D = \mathbb{E}[D]$, since when a stub is paired it is k times as likely to be paired with a stub from a given individual having degree k than it is with a stub from a given individual having degree 1. Under the branching process approximation, the effective degree of a newly infected individual is distributed according to $\tilde{D} - 1$, since one of that individual's stubs is 'used up' when it is infected. Note for future reference that $\mu_{\tilde{D}} = \mu_D^{-1} \mu_{D^2}$ and, more generally, $\mu_{f(\tilde{D})} = \mu_D^{-1} \mu_{Df(D)}$ for any real-valued function f. (For a random variable, X say, we use μ_X to denote its expectation $\mathbb{E}[X]$. Thus, for example, $\mu_{Df(D)} = \mathbb{E}[Df(D)]$.)

2.4 Explicit form for the multitype branching process

We now assume that there is a maximum degree $k_{\max}$. We show in Sect. 6 that our results for moments of the branching process extend to the case of no maximum degree size, subject to suitable moment conditions on D. Thus the type space for the branching process is $\mathcal{K} = \{0, 1, \ldots, k_{\max}\}$. Note that only initial infectives can have type $k_{\max}$. For $k \in \mathcal{K}$, an individual of type k dies if either its infectious period comes to an end or it transmits infection down one of its unattached stubs, whichever happens first. If the former happens first then the individual has no offspring, otherwise it has two offspring, namely an individual of type $k - 1$ and an individual whose type is distributed according to $\tilde{D} - 1$. Note that an individual of type 0 necessarily has no offspring when it dies. Thus, for $k \in \mathcal{K}$, the lifetime of an individual of type k has an exponential distribution with rate

$$\omega_k = \gamma + \tau k, \tag{2.1}$$

and when it dies its offspring is distributed as follows (recall that $\tilde{p}_{k_{\max}+1} = 0$):

$$\mathbb{P}\left(\text{Offspring} = \varnothing | \text{Parent type} = k\right) = \frac{\gamma}{\gamma + \tau k},$$

$$\mathbb{P}\left(\text{Offspring} = \{k - 1, l\} | \text{Parent type} = k\right) = \frac{\tau k \tilde{p}_{l+1}}{\gamma + \tau k} \quad (l = 0, 1, \ldots, k_{\max} - 1).$$

$$\tag{2.2}$$

The joint probability-generating function (PGF) for offspring of a type-k individual is therefore

$$P_k(\mathbf{s}) = \frac{1}{\omega_k} \left(\gamma + \tau k s_{k-1} \sum_{l=0}^{k_{\max}-1} \tilde{p}_{l+1} s_l \right), \tag{2.3}$$

where $\mathbf{s} = (s_k)$. In general we will write $\mathbf{v} = (v_k) = (v_0, v_1, \ldots, v_{k_{\max}})^\top$ for a column vector in $\mathbb{R}^{k_{\max}+1}$, where $^\top$ denotes transpose. Verbally, we will call v_0 the 0th element of such a vector, v_1 the first element etc. For $k \in \mathcal{K}$, let $\partial P_k(\mathbf{s})$ be the column vector whose ith element is $\frac{\partial P_k(\mathbf{s})}{\partial s_i}$ and let $\partial^2 P_k(\mathbf{s})$ be the matrix whose (i, j)th element is $\frac{\partial^2 P_k(\mathbf{s})}{\partial s_i \partial s_j}$. We note for future reference that, for $i, j, k \in \mathcal{K}$,

$$(\partial P_k(\mathbf{1}))_i = \frac{\tau k}{\gamma + \tau k} \left(\tilde{p}_{i+1} + \delta_{k-1,i} \right),$$

$$\left[\partial^2 P_k(\mathbf{1}) \right]_{i,j} = \frac{\tau k}{\gamma + \tau k} \left(\tilde{p}_{i+1} \delta_{k-1,j} + \tilde{p}_{j+1} \delta_{k-1,i} \right), \tag{2.4}$$

where $\mathbf{1}$ is the length-$(k_{\max} + 1)$ column vector of ones.

For $t \geq 0$, let $\mathbf{Z}(t) = (Z_i(t))$, where $Z_i(t)$ denotes the number of individuals of type i alive at time t, and let $Z(t) = Z_0(t) + Z_1(t) + \cdots + Z_{k_{\max}}(t) = \mathbf{1}^\top \mathbf{Z}(t)$ denote the total number individuals alive at time t. For $k \in \mathcal{K}$, we use the notation $\{\mathbf{Z}^{(k)}(t) : t \geq 0\}$,

where $\mathbf{Z}^{(k)}(t) = (Z_i^{(k)}(t))$, to denote a process starting with a single individual, whose type is k, at time 0 (i.e. where $Z_i^{(k)}(0) = \delta_{i,k}, i \in \mathcal{K}$). Further, $Z^{(k)}(t) = \mathbf{1}^\top \mathbf{Z}^{(k)}(t)$ denotes the total number of individuals at time t in such a process.

3 Behaviour of means

In the next three sections we consider the behaviour of the mean, variance and covariance function of the total number of individuals over time in the branching process which approximates the initial phase of an epidemic. For $t \geq 0$ and $i, j, k \in \mathcal{K}$, let

$$\mathbf{M}(t) = [m_{i,j}(t)], \quad \text{where} \quad m_{i,j}(t) = \mathbb{E}\left[Z_j^{(i)}(t) \right],$$

$$\mathbf{m}^{(k)}(t) = \mathbb{E}\left[\mathbf{Z}^{(k)}(t) \right] = \mathbf{M}(t)^\top \mathbf{u}_k,$$

$$m^{(k)}(t) = \mathbb{E}\left[Z^{(k)}(t) \right] = \mathbf{1}^\top \mathbf{m}^{(k)}(t) = \mathbf{u}_k^\top \mathbf{M}(t) \mathbf{1},$$

where $\mathbf{u}_k$ is a length-$(k_{\max} + 1)$ column vector with kth element equal to 1 and other elements equal to 0. A standard argument using the Kolmogorov forward equation (see e.g. Dorman et al. 2004, Sect. 7 and recall that $\tilde{p}_{k_{\max}+1} = 0$) then yields that

$$\frac{\mathrm{d}}{\mathrm{d}t} \mathbf{M}(t) = \mathbf{M}(t) \boldsymbol{\Omega}, \qquad \mathbf{M}(0) = \mathbf{I}, \tag{3.1}$$

where $\mathbf{I}$ denotes the $(k_{\max} + 1) \times (k_{\max} + 1)$ identity matrix and $\boldsymbol{\Omega} = [\Omega_{l,k}]$ is the $(k_{\max} + 1) \times (k_{\max} + 1)$ matrix with elements given by

$$\Omega_{l,k} = \tau l \left(\tilde{p}_{k+1} + \delta_{l,k+1} \right) - (\gamma + \tau l)\delta_{l,k} \qquad (l, k \in \mathcal{K}).$$

The solution to (3.1) is then straightforwardly given by

$$\mathbf{M}(t) = \mathrm{e}^{\boldsymbol{\Omega} t} = \sum_{l=0}^{\infty} \frac{t^l \boldsymbol{\Omega}^l}{l!}. \tag{3.2}$$

We show in "Appendix 2" that the eigenvalues of $\boldsymbol{\Omega}$ are

$$\lambda_i = \begin{cases} -\gamma - i\tau & \text{for } i \in \{0, 2, 3, \ldots, k_{\max}\}, \\ \tau \left(\left(\sum_{l=0}^{k_{\max}} l \tilde{p}_{l+1} \right) - 1 \right) - \gamma & \text{for } i = 1. \end{cases} \tag{3.3}$$

We denote the dominant eigenvalue, λ_1, by r, so

$$r = \tau \left(\left(\sum_{l=0}^{k_{\max}} l \tilde{p}_{l+1} \right) - 1 \right) - \gamma = \tau \mu_{\tilde{D}-2} - \gamma. \tag{3.4}$$

If $r \leq 0$, the branching process $\{\mathbf{Z}(t) : t \geq 0\}$ goes extinct almost surely. If $r > 0$, then r gives the asymptotic exponential growth rate of $\{Z(t) : t \geq 0\}$ (and also of $\{Z_i(t) : t \geq 0\}$ for $i \in \mathcal{K} \setminus \{k_{\max}\}$) when $\{\mathbf{Z}(t) : t \geq 0\}$ does not go extinct.

For $i \in \mathcal{K}$, let $\mathbf{w}_i^\top = (w_{i,k})$ be a left eigenvector of $\boldsymbol{\Omega}$ corresponding to the eigenvalue λ_i, so

$$\mathbf{w}_i^\top \boldsymbol{\Omega} = \lambda_i \mathbf{w}_i^\top. \tag{3.5}$$

The Perron–Frobenius theory implies that $\mathbf{w}_1$ can be chosen so that all of its elements are positive and $\mathbf{w}_1^\top \mathbf{1} = 1$. The left-eigenvector $\mathbf{w}_1$ then yields a probability distribution which gives the asymptotic relative frequencies of the different types, as $t \to \infty$, when $\{\mathbf{Z}(t) : t \geq 0\}$ does not go extinct.

Expanding (3.5) in components yields

$$\sum_{l=0}^{k_{\max}} w_{1,l} \left(\tau l \left(\tilde{p}_{k+1} + \delta_{l,k+1} \right) - (\gamma + \tau l)\delta_{l,k} \right) = r w_{1,k} \quad (k \in \mathcal{K}). \tag{3.6}$$

Let $w(s) = \sum_{l=0}^{k_{\max}} s^l w_{1,l}$ $(s \geq 0)$ denote the (probability-)generating function of $\mathbf{w}_1$. Multiplying (3.6) by s^k and summing over k yields

$$\tau f_{\tilde{D}-1}(s)\mu_W + \tau(1-s)w'(s) = (r + \gamma)w(s), \tag{3.7}$$

where $f_{\tilde{D}-1}(s) = \sum_{k=1}^{k_{\max}} \tilde{p}_k s^{k-1}$ is the PGF of $\tilde{D} - 1$ and $\mu_W = \sum_{k=0}^{k_{\max}} k w_{1,k}$ is the mean of the distribution $\mathbf{w}_1$. Setting $s = 1$ in (3.7) and using (3.4) yields

$$\mu_W = \frac{r + \gamma}{\tau} = \mu_{\tilde{D}-2}. \tag{3.8}$$

Note that (3.8) has a simple intuitive explanation. For large t, a typical individual gives birth at rate $\sum_{l=0}^{k_{\max}} w_{1,l} l \tau = \tau \mu_W$ and dies completely (i.e. without producing any offspring) at rate γ, so the population growth rate $r = \tau \mu_W - \gamma$ and (3.8) follows using (3.4) .

For $i, k \in \mathbb{Z}^+$, let $k_{[i]} = k(k-1)\dots(k-i+1)$ denote a falling factorial, with the convention $k_{[0]} = 1$. For $i = 0, 1, \dots$, let $\mu_W^{[i]} = \sum_{k=0}^{k_{\max}} k_{[i]} w_{1,k}$ be the ith factorial-moment of the distribution $\mathbf{w}_1$, so $\mu_W^{[0]} = 1$ and $\mu_W^{[1]} = \mu_W$. Note that $\mu_W^{[i]} = w^{(i)}(1)$ $(i = 0, 1, \dots)$, where $w^{(i)}(s)$ denotes the ith derivative of $w(s)$. Repeated differentiation of (3.7) yields

$$\mu_W^{[i]} = \frac{\mu_{\tilde{D}-2}}{\mu_{\tilde{D}-2+i}} \mu_{\tilde{D}-1}^{[i]}, \tag{3.9}$$

where $\mu_{\tilde{D}-1}^{[i]}$ is the ith factorial-moment of $\tilde{D} - 1$. Note that $\mu_{\tilde{D}-1}^{[i]} = 0$ for $i \geq k_{\max}$. It then follows, using the inversion formula which expresses the probability mass function of a non-negative integer-valued random variable in terms of its factorial-moments (see e.g. Daley and Vere-Jones 1988, p. 117), that

$$w_{1,k} = \begin{cases} \sum_{i=k}^{k_{\max}-1}(-1)^{i-k}\binom{i}{k}\frac{\mu_{\tilde{D}-2}\mu_{\tilde{D}-1}^{[i]}}{i!\mu_{\tilde{D}-2+i}} & \text{if } k = 0, 1, \dots, k_{\max} - 1, \\ 0 & \text{if } k = k_{\max}. \end{cases} \tag{3.10}$$

Observe that $w_{1,k_{max}} = 0$ since only initial infectives can have type k_{max}. Observe also that $\mathbf{w}_1$ is determined just by the degree distribution of the network and is invariant to the epidemic parameters γ and τ.

For $t \geq 0$, let $\mathbf{m}(t) = (m^{(i)}(t)) = (\boldsymbol{M}(t)\mathbf{1})^\top$. Thus the kth element of $\mathbf{m}(t)$ contains the mean total population size at time t given that the process starts with a single individual whose type is k. We derive a simple expression for $\mathbf{m}(t)$. The following proposition is useful.

Proposition 1 *For a matrix M and vectors $\mathbf{x}, \mathbf{y}$ such that $M\mathbf{x} = a\mathbf{x} + b\mathbf{y}$ and $M\mathbf{y} = c\mathbf{y}$, where a, b and c are scalars satisfying $a \neq c$,*

$$e^{Mt}\mathbf{x} = e^{at}\mathbf{x} + \frac{b}{a-c}\left(e^{at} - e^{ct}\right)\mathbf{y} \quad and \quad e^{Mt}\mathbf{y} = e^{ct}\mathbf{y}. \tag{3.11}$$

Proof The second identity follows straightforwardly from the definition of the matrix exponential and the fact that $\mathbf{y}$ is a right eigenvector with eigenvalue c. For the first identity,

$$e^{Mt}\mathbf{x} = \sum_{i=0}^{\infty} \frac{1}{i!}(Mt)^i\mathbf{x}$$

$$= \sum_{i=0}^{\infty} \frac{t^i}{i!}\left(a^i\mathbf{x} + \sum_{j=0}^{i-1} bc^{i-1-j}a^j\mathbf{y}\right)$$

$$= e^{at}\mathbf{x} + \sum_{i=0}^{\infty} \frac{t^i}{i!}bc^{i-1}\frac{\left(\frac{a}{c}\right)^i - 1}{\left(\frac{a}{c}\right) - 1}\mathbf{y}$$

$$= e^{at}\mathbf{x} + \frac{b}{a-c}\left(e^{at} - e^{ct}\right)\mathbf{y}. \tag{3.12}$$

$\square$

Let $\mathbf{n} = (0, 1, \ldots, k_{max})^\top$. Observe that

$$\boldsymbol{\Omega}\mathbf{1} = \tau\mathbf{n} - \gamma\mathbf{1} \quad \text{and} \quad \boldsymbol{\Omega}\mathbf{n} = r\mathbf{n}, \tag{3.13}$$

so using Proposition 1 with $M = \boldsymbol{\Omega}, \mathbf{x} = \mathbf{1}, \mathbf{y} = \mathbf{n}, a = -\gamma, b = \tau$ and $c = r$, and recalling from (3.4) that $r + \gamma = \mu_{\tilde{D}-2}\tau$, we have

$$e^{\boldsymbol{\Omega}x}\mathbf{1} = \mu_{\tilde{D}-2}^{-1}\left(e^{rx} - e^{-\gamma x}\right)\mathbf{n} + e^{-\gamma x}\mathbf{1} \quad \text{and} \quad e^{\boldsymbol{\Omega}x}\mathbf{n} = e^{rx}\mathbf{n}. \tag{3.14}$$

Thus,

$$\mathbf{m}(t) = \mu_{\tilde{D}-2}^{-1}\left(e^{rt} - e^{-\gamma t}\right)\mathbf{n} + e^{-\gamma t}\mathbf{1} \tag{3.15}$$

and

$$\lim_{t\to\infty} e^{-rt}\mathbf{m}(t) = \mu_{\tilde{D}-2}^{-1}\mathbf{n}. \tag{3.16}$$

While it is well known that asymptotically the mean prevalence grows exponentially with rate constant r, i.e. that prevalence $\propto e^{rt}$, these results allow us to see from (3.15) that the rate of convergence to this asymptotic behaviour is $r + \gamma$, and from (3.16) that the constant of proportionality is the degree of the initially infected individual divided by $\mu_{\tilde{D}-2}$.

We also consider the relationship between the equations above and the diverse ODE approaches to the mean behaviour of the full network epidemic model. Miller and Kiss (2014) consider several such approaches; their notation can be related to ours by defining

$$I(t) = \mathbf{m}^{(k)}(t)^{\top}\mathbf{1}, \qquad \lambda(t) = \mathbf{m}^{(k)}(t)^{\top}\mathbf{n}. \tag{3.17}$$

Substituting (3.17) and (3.13) into (3.1) gives

$$\frac{\mathrm{d}I}{\mathrm{d}t} = \tau\lambda - \gamma I, \qquad \frac{\mathrm{d}\lambda}{\mathrm{d}t} = r\lambda. \tag{3.18}$$

This pair of equations can be derived from various models considered by Miller and Kiss (2014, c.f. Sect. 3.4.1) assuming an initially small infectious population and negligible susceptible depletion. Therefore, our results suggest that the ODE approaches to mean behaviour do not require correction as the infectious population becomes extremely small, and the typical assumption that $1 \ll I(t = 0) \ll N$ for the ODE system to hold may be too conservative, with $N \gg 1$ being all that is required.

4 Variance

The variance in infectious prevalence during the exponentially growing phase of an epidemic was considered in Graham and House (2014), but using the diffusion limit and an argument about the neighborhood around an infective. A branching process limit lets us be more explicit. For $k \in \mathscr{K}$, let $\mathbf{u}_k$ denote the length-$(k_{\max} + 1)$ column vector whose element corresponding to type k is 1 and all other elements are 0, so $\mathbf{u}_k = (\delta_{i,k})$. For $t \geq 0$ and $k \in \mathscr{K}$, let $\sigma_{ij}^{(k)}(t) = \mathrm{cov}\left(Z_i^{(k)}(t), Z_j^{(k)}(t)\right)$ $(i, j \in \mathscr{K})$.

A matrix integrating factor argument gives

$$\begin{aligned}
\boldsymbol{V}^{(k)}(t) = \left[\sigma_{ij}^{(k)}(t)\right] &= \mathbb{E}\left[\mathbf{Z}^{(k)}(t)\mathbf{Z}^{(k)}(t)^{\top}\right] - \mathbb{E}\left[\mathbf{Z}^{(k)}(t)\right]\mathbb{E}\left[\mathbf{Z}^{(k)}(t)^{\top}\right] \\
&= \int_0^t e^{\boldsymbol{\Omega}^{\top}(t-u)}\boldsymbol{B}_k(u)e^{\boldsymbol{\Omega}(t-u)}\mathrm{d}u,
\end{aligned} \tag{4.1}$$

where

$$\begin{aligned}
\boldsymbol{B}_k(t) &= \sum_{l=0}^{k_{\max}}\left(e^{\boldsymbol{\Omega}t}\right)_{k,l}\boldsymbol{C}_l, \\
\boldsymbol{C}_k &= \omega_k\left(\mathfrak{d}^2 P_k(\mathbf{1}) + \mathrm{diag}(\mathbf{f}_k) - \mathbf{u}_k\mathbf{f}_k^{\top} - \mathbf{f}_k\mathbf{u}_k^{\top} + \mathbf{u}_k\mathbf{u}_k^{\top}\right), \\
\mathbf{f}_k &= \mathfrak{d} P_k(\mathbf{1}).
\end{aligned} \tag{4.2}$$

See Dorman et al. (2004, Sect. 9), and also Athreya and Ney (1972, p. 203), for details.[1]

For $t \geq 0$ and $k \in \mathcal{K}$, let $v^{(k)}(t)$ denote the variance of the total population size at time t given that the process starts with a single individual, whose type is k. Then $v^{(k)}(t) = \mathbf{1}^\top V^{(k)}(t)\mathbf{1}$ and it follows using (4.1) that

$$
\begin{aligned}
v^{(k)}(t) &= \int_0^t (e^{\mathbf{\Omega}(t-u)}\mathbf{1})^\top B_k(u)(e^{\mathbf{\Omega}(t-u)}\mathbf{1}) \mathrm{d}u \\
&= \int_0^t \left(\mu_{\tilde{D}-2}^{-1} \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) \right)^2 \mathbf{n}^\top B_k(u)\mathbf{n} \; \mathrm{d}u \\
&\quad + 2\int_0^t \mu_{\tilde{D}-2}^{-1} \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{-\gamma(t-u)} \mathbf{1}^\top B_k(u)\mathbf{n} \; \mathrm{d}u \\
&\quad + \int_0^t e^{-2\gamma(t-u)} \mathbf{1}^\top B_k(u)\mathbf{1} \; \mathrm{d}u,
\end{aligned} \tag{4.3}
$$

where we have used the first equation in (3.14) in deriving the last line. This quantity has an exact but rather complex closed-form solution, which we give below and derive in "Appendix 3".

Let $\mathbf{n}_2 = (0^2, 1^2, \ldots, k_{\max}{}^2)^\top$ and, for $t \geq 0$, let $\mathbf{v}(t) = (v^{(i)}(t))$. Then

$$
\mathbf{v}(t) = \alpha_0(t)\mathbf{1} + \alpha_1(t)\mathbf{n} + \alpha_2(t)\mathbf{n}_2, \tag{4.4}
$$

where

$$
\begin{aligned}
\alpha_0(t) &= \gamma I_2(t), \\
\alpha_1(t) &= \gamma \mu_{\tilde{D}-2}^{-1} \left[I_1(t) - I_2(t) + 2I_3(t) + \mu_{\tilde{D}-2}^{-1}\mu_{\tilde{D}}^{-1}\mu_{(\tilde{D}-1)^2+1}(I_4(t) - I_5(t)) \right] \\
&\quad + \tau \left[I_1(t) + 2I_3(t) + \mu_{\tilde{D}-2}^{-2}\mu_{(\tilde{D}-2)^2} I_4(t) \right], \\
\alpha_2(t) &= \gamma \mu_{\tilde{D}-2}^{-2} I_5(t),
\end{aligned} \tag{4.5}
$$

with

$$
\begin{aligned}
I_1(t) &= \frac{e^{rt} - e^{-2\gamma t}}{r + 2\gamma}, \\
I_2(t) &= \frac{e^{-\gamma t}\left(1 - e^{-\gamma t}\right)}{\gamma}, \\
I_3(t) &= \frac{e^{rt}\left(1 - e^{-\gamma t}\right)}{\gamma} - I_1(t), \\
I_4(t) &= \frac{e^{rt}\left(e^{rt} - 1\right)}{r} - 2I_3(t) - I_1(t), \\
I_5(t) &= \frac{e^{2rt} - e^{-(\gamma+2\tau)t}}{2r + 2\tau + \gamma} - 2\frac{e^{-\gamma t}\left(e^{rt} - e^{-2\tau t}\right)}{r + 2\tau} + \frac{e^{-\gamma t}\left(e^{-2\tau t} - e^{-\gamma t}\right)}{\gamma - 2\tau}.
\end{aligned} \tag{4.6}
$$

---

[1] There is a small error in the latter—in the expression for $b_{jk}^{(i)}$ on p. 203 of Athreya and Ney (1972), $\delta_{jk} - b_{ij}\delta_{ik} - b_{ik}\delta_{ij} + \delta_{ij}\delta_{ik}$ should be replaced by $b_{ij}\delta_{jk} - b_{ik}\delta_{ij} - b_{ij}\delta_{ik}$.

In (4.6), if a denominator is zero then the expression is given by the limit as that denominator tends to zero. For example, if $\gamma = 0$ then $I_2(t) = te^{-\gamma t}$, and if $\gamma = 2\tau$, the final term in $I_5(t)$ is replaced by $te^{-2\gamma t}$.

Equation (4.4) leads to a rather complex expression for $\mathbf{v}(t)$ in terms of elementary functions. However, its asymptotic form as $t \to \infty$ is much simpler. Note that $\lim_{t\to\infty} e^{-2rt} I_k(t) = 0$, for $k = 1, 2, 3$, $\lim_{t\to\infty} e^{-2rt} I_4(t) = \frac{1}{r}$ and $\lim_{t\to\infty} e^{-2rt} I_5(t) = \frac{1}{2r+2\tau+\gamma} = \frac{1}{2\mu_{\tilde{D}-1}\tau-\gamma}$. Substituting these limits into (4.4) and (4.5) yields

$$\lim_{t\to\infty} e^{-2rt} \mathbf{v}(t) = \frac{1}{\mu_{\tilde{D}-2}^2 \left(2\mu_{\tilde{D}-1}\tau - \gamma\right)} \left[\frac{2\tau\mu_{\tilde{D}-1}\left(\mu_{(\tilde{D}-2)^2}\tau + \gamma\right)}{r}\mathbf{n} + \gamma\mathbf{n}_2 \right]. \tag{4.7}$$

Note that both the asymptotic and exact expressions for $\mathbf{v}(t)$ depend on the degree distribution D only through its first three moments.

It follows from (3.16) and (4.7) that, for $k \in \mathcal{K}$,

$$\lim_{t\to\infty} \frac{\mathrm{var}\left(Z^{(k)}(t)\right)}{\mathbb{E}\left[Z^{(k)}(t)\right]^2} = \frac{1}{2\mu_{\tilde{D}-1}\tau - \gamma} \left[\gamma + \frac{2\tau\mu_{\tilde{D}-1}\left(\mu_{(\tilde{D}-2)^2}\tau + \gamma\right)}{kr} \right]. \tag{4.8}$$

We note two features of these results. First, the Eqs. (4.4) and (4.5) involve many rates that are linear combinations of r, τ and γ, with the dominant being $2r$ and the subdominant being r. This leads to complex real-time behaviour as the system approaches its asymptotic limit. In the diffusion limit, only the dominant and subdominant rates are present, leading to the same overall rate of convergence r, but other rates are not present (Graham and House 2014). Secondly, the dependence of the variance on initial conditions is not simple proportionality, meaning that (4.7) contains terms proportional to both $\mathbf{n}$ and $\mathbf{n}_2$ (unless $\gamma = 0$) and the right-hand side of (4.8) depends on k.

5 Covariance function

For $t, s \geq 0$ and $k \in \mathcal{K}$, let $\sigma^{(k)}(t, s) = \mathrm{cov}\left(Z^{(k)}(t), Z^{(k)}(s)\right)$ denote the covariance of the total population sizes at times t and s in the branching process which approximates the early phase of an epidemic, given that the process starts with a single individual, whose type is k. We assume without loss of generality that $t \leq s$; although this choice does not respect alphabetical order, the majority of results that follow take t as an argument rather than s, and are therefore more easily read as functions of time. Then

$$\sigma^{(k)}(t, s) = \mathbb{E}\left[\mathrm{cov}\left(Z^{(k)}(t), Z^{(k)}(s)|\mathbf{Z}^{(k)}(t)\right)\right]$$
$$+ \mathrm{cov}\left(\mathbb{E}\left[Z^{(k)}(t)|\mathbf{Z}^{(k)}(t)\right], \mathbb{E}\left[Z^{(k)}(s)|\mathbf{Z}^{(k)}(t)\right]\right). \tag{5.1}$$

The first term on the right hand side of (5.1) is zero, since $Z^{(k)}(t)$ is non-random given $\mathbf{Z}^{(k)}(t)$. Now, $\mathbb{E}\left[Z^{(k)}(t)|\mathbf{Z}^{(k)}(t)\right] = \mathbf{1}^\top \mathbf{Z}^{(k)}(t)$ and $\mathbb{E}\left[Z^{(k)}(s)|\mathbf{Z}^{(k)}(t)\right] = \mathbf{Z}^{(k)}(t)^\top M(s-t)\mathbf{1}$, so

$$
\begin{aligned}
\sigma^{(k)}(t,s) &= \mathrm{cov}\left(\mathbf{1}^\top \mathbf{Z}^{(k)}(t), \mathbf{Z}^{(k)}(t)^\top M(s-t)\mathbf{1}\right) \\
&= \mathbf{1}^\top V^{(k)}(t)M(s-t)\mathbf{1} \\
&= \mu_{\tilde{D}-2}^{-1}\left(e^{r(s-t)} - e^{-\gamma(s-t)}\right)\mathbf{1}^\top V^{(k)}(t)\mathbf{n} + e^{-\gamma(s-t)}v^{(k)}(t), \quad (5.2)
\end{aligned}
$$

using (3.2), the first equation in (3.14) and noting that $\mathbf{1}^\top V^{(k)}(t)\mathbf{1} = v^{(k)}(t)$. This leads to an exact, closed-form expression for the covariance function in terms of elementary functions, which we state below and derive in "Appendix 4".

For $t, s \geq 0$, let $\sigma(t,s) = \left(\sigma^{(0)}(t,s), \sigma^{(1)}(t,s), \ldots, \sigma^{(k_{\max})}(t,s)\right)^\top$. Then

$$
\sigma(t,s) = \mu_{\tilde{D}-2}^{-1}\left(e^{r(s-t)} - e^{-\gamma(s-t)}\right)(\beta_1(t)\mathbf{n} + \beta_2(t)\mathbf{n}_2) + e^{-\gamma(s-t)}\mathbf{v}(t), \quad (5.3)
$$

where

$$
\begin{aligned}
\beta_1(t) &= \gamma\mu_{\tilde{D}-2}^{-1}\left[\mu_{\tilde{D}}^{-1}\mu_{(\tilde{D}-1)^2+1}(I_7(t) - I_8(t)) + \mu_{\tilde{D}-2}I_6(t)\right] \\
&\quad + \tau\mu_{\tilde{D}-2}^{-1}\left[\mu_{(\tilde{D}-2)^2}I_7(t) + \mu_{\tilde{D}-2}^2I_6(t)\right], \\
\beta_2(t) &= \gamma\mu_{\tilde{D}-2}^{-1}I_8(t), \quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad\quad (5.4)
\end{aligned}
$$

with

$$
\begin{aligned}
I_6(t) &= \frac{e^{rt}\left(1 - e^{-\gamma t}\right)}{\gamma}, \\
I_7(t) &= \frac{e^{rt}\left(e^{rt} - 1\right)}{r} - I_6(t), \\
I_8(t) &= \frac{e^{2rt} - e^{-(\gamma+2\tau)t}}{2r + 2\tau + \gamma} - \frac{e^{-\gamma t}\left(e^{rt} - e^{-2\tau t}\right)}{r + 2\tau}. \quad (5.5)
\end{aligned}
$$

As at (4.6), an appropriate limit is taken if a denominator in (5.5) is zero.

The covariance function takes a simple form in the limit as t and $s \to \infty$. Note that $\lim_{t\to\infty} e^{-2rt}I_6(t) = 0$, $\lim_{t\to\infty} e^{-2rt}I_7(t) = \frac{1}{r}$ and $\lim_{t\to\infty} e^{-2rt}I_8(t) = \frac{1}{2\mu_{\tilde{D}-1}\tau-\gamma}$. Substituting these limits into (5.3) yields that, for any $s \geq 0$,

$$
\lim_{t\to\infty} e^{-2rt}\sigma(t, t+s) = e^{rs}\lim_{t\to\infty} e^{-2rt}\mathbf{v}(t). \quad (5.6)
$$

It follows that, for $k \in \mathscr{K}$ and $s > 0$,

$$
\lim_{t\to\infty} \mathrm{corr}\left(Z^{(k)}(t), Z^{(k)}(t+s)\right) = 1,
$$

where corr denotes correlation. This is not surprising since it is well known that

$$\mathrm{e}^{-rt} Z^{(k)}(t) \xrightarrow{\text{a.s.}} W^{(k)} \quad \text{as } t \to \infty,$$

where $\xrightarrow{\text{a.s.}}$ denotes almost sure convergence (i.e. convergence with probability 1) and $W^{(k)}$ is a non-negative random which satsifies $W^{(k)} = 0$ if and only if the branching process goes extinct; see e.g. Athreya and Ney (1972, Theorem V.7.2).

6 Unbounded degree distributions

The above results have all assumed that there is a maximum degree $k_{\max}$. Suppose that is not the case, so the branching process has countably many types. For $t \geq 0$, let $\mathbf{Z}(t) = (Z_0(t), Z_1(t), \ldots)^\top$, where $Z_i(t)$ denotes the number of individuals of type i alive at time t, and let $Z(t) = \sum_{i=0}^{\infty} Z_i(t)$ denote the total number individuals alive at time t. (For ease of notation we drop explict reference to the type of the initial individual.) For $k_{\max} = 1, 2, \ldots$, let $\{\mathbf{Z}(t, k_{\max}) : t \geq 0\}$ denote the branching process derived from $\{\mathbf{Z}(t) : t \geq 0\}$ by ignoring all individuals having type strictly greater than $k_{\max}$ and any offspring of such individuals. For $t \geq 0$, let $Z(t, k_{\max}) = \mathbf{1}^\top \mathbf{Z}(t, k_{\max})$ be the total number of individuals alive in $\{\mathbf{Z}(t, k_{\max}) : t \geq 0\}$ at time t. Now, for any $t \geq 0$, $Z(t, k_{\max})$ is monotonically increasing in $k_{\max}$ and converges almost surely to $Z(t)$ as $k_{\max} \to \infty$. Thus, by the monotone convergence theorem, $\mathbb{E}[Z(t)] = \lim_{k_{\max} \to \infty} \mathbb{E}[Z(t, k_{\max})]$.

The process $\{\mathbf{Z}(t, k_{\max}) : t \geq 0\}$ behaves like the branching process described in Sect. 2.3 but with infection rate τ replaced by $\tau(k_{\max}) = \tau \mathbb{P}\left(\tilde{D} \leq k_{\max} + 1\right)$, and size-biased degree distribution $\tilde{D}$ replaced by $\tilde{D}(k_{\max})$, where

$$\mathbb{P}\left(\tilde{D}(k_{\max}) = k\right) = \begin{cases} \dfrac{\tilde{p}_k}{\mathbb{P}(\tilde{D} \leq k_{\max} + 1)} & \text{if } k = 1, 2, \ldots, k_{\max} + 1, \\ 0 & \text{if } k = k_{\max} + 2, k_{\max} + 3, \ldots. \end{cases}$$

The presence of $k_{\max} + 1$ rather than $k_{\max}$ is because contacts with individuals having degree strictly greater than $k_{\max} + 1$ are ignored, as they yield individuals with effective degree (and hence type) strictly greater than $k_{\max}$. Now $\tau(k_{\max}) \to \tau$ and $\mathbb{E}\left[\tilde{D}(k_{\max})\right] \to \mathbb{E}\left[\tilde{D}\right]$ as $k_{\max} \to \infty$, so the expression (3.15) for the mean total population size at time t continues to hold in the unbouded degree case, provided that $\mathbb{E}\left[\tilde{D}\right] < \infty$, or equivalently that $\mathbb{E}\left[D^2\right] < \infty$. A similar argument shows that the expressions for the variance of $Z(t)$ and the covariance of $Z(t)$ and $Z(s)$, derived in Sects. 4 and 5, respectively, continue to hold provided $\mathbb{E}\left[\tilde{D}^2\right] < \infty$, or equivalently $\mathbb{E}\left[D^3\right] < \infty$.

7 Probability of extinction

For $t \geq 0$ and $k \in \mathcal{K}$, let $\pi_k(t) = \mathbb{P}\left(Z^{(k)}(t) = 0\right)$ be the probability that the branching process is extinct at time t given that it starts with one individual of type k. Then in general

$$\frac{d}{dt}\pi_k(t) = -\omega_k\pi_k(t) + \omega_k P_k(\boldsymbol{\pi}(t)), \tag{7.1}$$

where $\boldsymbol{\pi}(t) = (\pi_i(t))$. For our specific model, using (2.1) and (2.3), we have

$$\frac{d}{dt}\pi_k(t) = -(\gamma + \tau k)\pi_k(t) + \gamma + \tau k\pi_{k-1}(t)\sum_{l=0}^{k_{\max}-1}\tilde{p}_{l+1}\pi_l(t). \tag{7.2}$$

These equations are not amenable to closed-form solution. Note, however, that studies of time to extinction for network epidemics—e.g. Holme (2013)—have tended to be based on Monte Carlo methods, but (7.2) could provide a complementary approach that is numerically cheaper and more analytically tractable.

We will now consider three regimes in which asymptotic methods can be used to bound the real-time behaviour of the probabilities of extinction. In particular, we will see that early real-time behaviour is bounded by the death rates ω_k, while late-time behaviour is bounded by the asymptotic real-time growth rate r provided $r > -\gamma$.

7.1 Late behaviour of the subcritical case

Suppose that $r < 0$, so the branching process is subcritical. For $t \geq 0$ and $k \in \mathbb{N}_0 = \{0, 1, \ldots\}$, we will work with the probability of survival $q_k(t) = 1 - \pi_k(t) = \mathbb{P}\left(Z^{(k)}(t) > 0\right)$. Now $\mathbb{P}\left(Z^{(k)}(t) > 0\right) \leq \mathbb{E}\left[Z^{(k)}(t)\right]$, so using (3.15) a simple upper bound for $q_k(t)$, valid also in the unbounded degree setting using the results in Sect. 6, is

$$q_k(t) \leq k\mu_{\tilde{D}-2}^{-1}\left(e^{rt} - e^{-\gamma t}\right) + e^{-\gamma t}. \tag{7.3}$$

Note that $\mu_{\tilde{D}} < \infty$ is a necessary condition for $r < 0$. Under the stronger condition that $\mu_{\tilde{D}^2} < \infty$, Windridge (2015) gives an exponential approximation, for large t, to a quantity closely related to $q_k(t)$. He assumes that what we call type-0 individuals are dead. For $k = 1, 2, \ldots$, let $\hat{q}_k(t) = \mathbb{P}\left(\sum_{i=1}^{\infty} Z_i^{(k)}(t) > 0\right)$. Then, Windridge shows that there exists a constant $\hat{c} \in (0, 1]$ such that, for any $a < \min\{\tau, -r\}$,

$$\hat{q}_k(t) = \hat{c}ke^{rt}\left(1 + O(ke^{-at})\right) \quad \text{as } t \to \infty, \tag{7.4}$$

for any $k \geq 1$. The constant $\hat{c} = \lim_{t\to\infty} e^{-rt}\hat{q}_1(t)$. Note that for some practical purposes, $\hat{q}_k(t)$ may be of more interest than $q_k(t)$, since type-0 individuals are unable to transmit infection. In particular, in "Appendix 5" we sketch the argument that for the case where $r > -\gamma$ an analogous result to (7.4) holds, i.e. , for $k \geq 1$,

$$q_k(t) \sim cke^{rt} \quad \text{as } t \to \infty, \quad \text{where } c = \lim_{t\to\infty} e^{-rt}q_1(t) > 0. \tag{7.5}$$

(For real-valued functions, f and g say, $f(t) \sim g(t)$ as $t \to \infty$ if $\lim_{t\to\infty} f(t)/g(t) = 1$.)

For the case where $r < -\gamma$ (so, from (3.4), $\mu_{\tilde{D}} < 2$) we show in "Appendix 5" that if $\mu_{\tilde{D}^2} < \infty$ then, for $k \geq 0$,

$$q_k(t) \sim (1 - k\mu_{\tilde{D}-2}^{-1})e^{-\gamma t} \quad \text{as } t \to \infty. \tag{7.6}$$

Note that in this case the asymptotic behaviour of the survival probability $q_k(t)$ is independent of the infection rate τ. The case $r < -\gamma$ could occur, for example, at the end of an epidemic where $\tau \gg \gamma$. Such an epidemic would consist primarily of transmission events at early times, with the late behaviour dominated by recovery events.

7.2 Late behaviour of the supercritical case

An approximation to $q_k(t)$ in the supercritical case ($r > 0$) can be obtained by exploiting the fact that a supercritical branching process conditioned on extinction is probabilistically equivalent to a subcritical branching process. For $k \in \mathbb{N}_0$, let $T^{(k)} = \inf\{t \geq 0 : Z^{(k)}(t) = 0\}$ denote the extinction time of the branching process given that it starts with one individual of type k, where $T^{(k)} = \infty$ if the branching process survives forever, and let $\pi_k = \mathbb{P}\left(T^{(k)} < \infty\right) = \pi_k(\infty)$ be the probability that the branching process ultimately goes extinct. Then,

$$q_k(t) = 1 - \pi_k + \pi_k \mathbb{P}\left(T^{(k)} > t \mid T^{(k)} < \infty\right). \tag{7.7}$$

Let $\{\tilde{\mathbf{Z}}^{(k)}(t) : t \geq 0\}$ be distributed as $\{\mathbf{Z}^{(k)}(t) : t \geq 0 \mid T^{(k)} < \infty\}$. Then it follows from Waugh (1958, Sect. 5), that $\{\tilde{\mathbf{Z}}^{(k)}(t) : t \geq 0\}$ is also a continuous-time multitype Markov branching process, in which the lifetime of a typical type-k individual has an exponential distribution with rate $\gamma + \tau k$, as at (2.1), but when it dies its offspring is now distributed as follows:

$$\mathbb{P}\left(\text{Offspring} = \varnothing \mid \text{Parent type} = k\right) = \frac{1}{\pi_k} \frac{\gamma}{\gamma + \tau k},$$

$$\mathbb{P}\left(\text{Offspring} = \{k-1, l\} \mid \text{Parent type} = k\right) = \frac{\pi_{k-1}\pi_l}{\pi_k} \frac{\tau k \tilde{p}_{l+1}}{\gamma + \tau k} \quad (l \in \mathbb{N}_0). \tag{7.8}$$

Suppose now that there is a maximum degree size $k_{\max}$. Let $\tilde{\boldsymbol{\Omega}} = [\tilde{\Omega}_{l,k}]$ be the $(k_{\max} + 1) \times (k_{\max} + 1)$ matrix with elements given by

$$\tilde{\Omega}_{l,k} = \frac{\pi_{l-1}\pi_k}{\pi_l} \tau l \left(\tilde{p}_{k+1} + \delta_{l,k+1}\right) - (\gamma + \tau l)\delta_{l,k} \quad (l, k \in \mathcal{K}).$$

Then, recalling (3.2),

$$\mathbb{E}\left[\tilde{Z}^{(k)}(t)\right] = \mathbf{u}_k^\top e^{\tilde{\boldsymbol{\Omega}}t} \mathbf{1}, \tag{7.9}$$

where $\tilde{Z}^{(k)}(t) = \tilde{Z}_0^{(k)}(t) + \tilde{Z}_1^{(k)}(t) + \ldots + \tilde{Z}_{k_{\max}}^{(k)}(t)$.

Let $\tilde{r}$ denote the dominant eigenvalue of $\tilde{\boldsymbol{\Omega}}$ and note that $\tilde{r} < 0$. For $t \geq 0$ and $k = 0, 1, \ldots, k_{\max}$, let $\tilde{q}_k(t) = \mathbb{P}\left(\tilde{Z}^{(k)}(t) > 0\right)$ be the probability that the branching process $\{\tilde{\mathbf{Z}}^{(k)}(t) : t \geq 0\}$ is not extinct at time t given that it starts with one individual of type k. Then we expect that arguments similar to those used in the proof of Heinzmann (2009, Theorem 3.1), will show that there exists constants $\tilde{c}_1, \tilde{c}_2, \ldots, \tilde{c}_{k_{\max}}$, satisfying $0 < \tilde{c}_k < \infty$, such for $k = 1, 2, \ldots, k_{\max}$,

$$\tilde{q}_k(t) = \tilde{c}_k \mathrm{e}^{\tilde{r}t}\left(1 + o(\mathrm{e}^{-\tilde{\gamma}t})\right) \quad \text{as } t \to \infty, \tag{7.10}$$

for any $\tilde{\gamma} > 0$. It then follows using (7.7) that

$$q_k(t) = 1 - \pi_k + \pi_k \tilde{c}_k \mathrm{e}^{\tilde{r}t}\left(1 + o(\mathrm{e}^{-\tilde{\gamma}t})\right) \quad \text{as } t \to \infty. \tag{7.11}$$

Heinzmann (2009, Theorem 3.1), cannot be applied directly as it assumes that the matrix $\tilde{\boldsymbol{\Omega}}$ is irreducible. We do not consider it here but we expect that Heinzmann's proof can be extended to our situation. If we assume that type-0 individuals are dead and only consider initial individuals of types $1, 2, \ldots, k_{\max} - 1$ (recall that only initial infectives can have type $k_{\max}$) then $\tilde{\boldsymbol{\Omega}}$ becomes a $(k_{\max} - 1) \times (k_{\max} - 1)$ irreducible matrix. Heinzmann (2009, Theorem 3.1), then yields (7.10); note that now π_k is replaced by $\bar{\pi}_k = 1 - \lim_{t \to \infty} \bar{q}_k(t)$ $(k = 1, 2, k_{\max} - 1)$ in the definition of $\tilde{\boldsymbol{\Omega}}$ and $\tilde{q}_k(t) = \tilde{q}_k(t) = \mathbb{P}\left(\sum_{i=1}^{k_{\max}-1} \tilde{Z}_i^{(k)}(t) > 0\right)$. The approximation (7.11) then holds with $q_k(t)$ and π_k replaced by $\bar{q}_k(t)$ and $\bar{\pi}_k$, repsectively. Moreover, if we then let $\tilde{\mathbf{f}}_1^\top$ and $\tilde{\mathbf{b}}_1$ be left and right eigenvectors of $\tilde{\boldsymbol{\Omega}}$ corresponding to the eigenvalue $\tilde{r}$, satisfying $\tilde{\mathbf{f}}_1^\top \tilde{\mathbf{b}}_1 = 1$, then $\tilde{c}_k = \left(\mathbf{u}_k^\top \tilde{\mathbf{b}}_1\right) h^*$, where $h^* = \lim_{t \to \infty} \mathrm{e}^{-\tilde{r}t} \tilde{\mathbf{f}}_1^\top \tilde{\mathbf{q}}(t)$ and $\tilde{\mathbf{q}}(t) = \left(\tilde{q}_i(t), \tilde{q}_2(t), \ldots, \tilde{q}_{k_{\max}-1}(t)\right)^\top$. Unfortunately, unlike with $\boldsymbol{\Omega}$, there do not appear to be closed-form expressions for $\tilde{r}$ and its associated eigenvectors.

7.3 Early behaviour and matched asymptotics

Matched asymptotics is a standard technique in mathematical biology for writing down approximations to non-linear models that match known asymptotic behaviour (Murray 2002, 2003). While numerical solution of the ODEs (7.2) is efficient (as we have noted above) we now obtain a crude approximation to the full system that takes a closed form in terms of elementary functions.

First note that for $k \in \mathscr{K}$, $\pi_k(0) = 0$ and $\pi_k(t)$ is monotonically increasing with t. If we neglect the quadratic terms in π in (7.2) then, since these are only positive and increasing over time, we get a lower bound for the extinction probabilities:

$$\pi_k(t) \geq \pi_k^{(0)}(t) = \frac{\gamma}{\omega_k}\left(1 - \mathrm{e}^{-\omega_k t}\right). \tag{7.12}$$

Note that in standard matched asymptotics, we would identify a small parameter from a ratio of rate constants as the basis for a systematic approximation scheme (Murray 1984); an alternative would be to approximate systematically by, for example, letting

$\pi_k^{(1)} = \pi_k - \pi_k^{(0)}$, substituting into (7.2) and neglecting quadratic terms to give a linear set of equations for the next order of approximation. Here we consider only the lowest order approximation, and hence define an 'internal' solution for the survival probability as:

$$q_k^{(I)}(t) = 1 - \pi_k^{(0)}(t) = \frac{1}{\omega_k}\left(\tau k + \gamma e^{-\omega_k t}\right).$$
(7.13)

Next, supposing we are in the subcritical case so that our result (7.5) holds. We will call this the 'external' solution

$$q_k^{(E)}(t) = cke^{rt}.$$
(7.14)

To fix the constant c, we match the late behaviour of the internal solution with the early behaviour of the external solution:

$$\overline{q} = \lim_{t\to\infty} q_k^{(I)}(t) = \lim_{t\to 0} q_k^{(E)}(t) \quad \Rightarrow \quad q_k^{(E)}(t) = \frac{\tau k}{\omega_k}e^{rt}.$$
(7.15)

Finally, the matched asymptotic solution is

$$q_k^{(A)}(t) = q_k^{(I)}(t) + q_k^{(E)}(t) - \overline{q} = \frac{1}{\omega_k}\left(\tau k e^{rt} + \gamma e^{-\omega_k t}\right).$$
(7.16)

We compared this approximation as well as the internal and external solutions to the exact solution $q_k(t)$, with results shown in Fig. 1. As advertised, this is a relatively crude approximation, but is expressed in terms of elementary functions and satisfies known asymptotic limits.
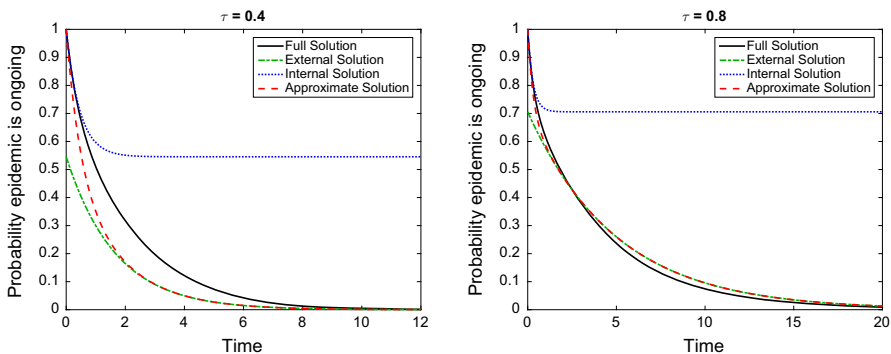
Fig. 1 Extinction probability for a subcritical epidemic compared to approximations. Results are for a 3-regular graph with $\gamma = 1$ and values of τ indicated in the figure titles. The internal and external solutions each fail severely at certain points, but the approximate solution crudely captures the overall behaviour (colour figure online)

8 Fluctuations in the emerging phase of a major outbreak

We now consider the early behaviour of supercritical epidemics that take off (i.e. do not go extinct early on but ultimately end owing to long-term depletion of susceptibles). The early stages of such an epidemic are approximated by the branching process defined in Sect. 2.3 but conditioned on non-extinction. It is straightforward to adapt the results on means and variances in Sects. 3 and 4 to condition on $Z^{(k)}(t) > 0$. Elementary calculation shows that, for $t \geq 0$ and $k \in \mathbb{N}_0$,

$$\mathbb{E}\left[Z^{(k)}(t)\Big|Z^{(k)}(t) > 0\right] = \frac{\mathbb{E}\left[Z^{(k)}(t)\right]}{q_k(t)},$$

$$\mathrm{var}\left(Z^{(k)}(t)\Big|Z^{(k)}(t) > 0\right) = \frac{\mathrm{var}\left(Z^{(k)}(t)\right)}{q_k(t)} - \pi_k(t)\left(\frac{\mathbb{E}\left[Z^{(k)}(t)\right]}{q_k(t)}\right)^2.$$

Expressions for $\mathbb{E}\left[Z^{(k)}(t)\Big|Z^{(k)}(t) > 0\right]$ and $\mathrm{var}\left(Z^{(k)}(t)\Big|Z^{(k)}(t) > 0\right)$ then follow using (3.15) and (4.4), respectively, though there is no closed-form formula for $q_k(t)$ or $\pi_k(t)$. Note that, assuming $r > 0$ so $\pi_k < 1$,

$$\lim_{t \to \infty} \frac{\mathrm{var}\left(Z^{(k)}(t)\Big|Z^{(k)}(t) > 0\right)}{\mathbb{E}\left[Z^{(k)}(t)\Big|Z^{(k)}(t) > 0\right]^2} = (1 - \pi_k)\lim_{t \to \infty} \frac{\mathrm{var}\left(Z^{(k)}(t)\right)}{\mathbb{E}\left[Z^{(k)}(t)\right]^2} - \pi_k,$$

which depends on the degree k of the initial infective.

The diffusion approximation studied in Graham and House (2014) corresponds to the case where the number of infectives at time $t = 0$ is large. Return to the case where there is a maximum degree $k_{\max}$ and suppose that the branching process does not go extinct. Then it follows from Athreya and Ney (1972, Theorem V.7.2), that, for any $k \in \mathcal{K}$,

$$\frac{\mathbf{Z}^{(k)}(t)}{Z^{(k)}(t)} \xrightarrow{\text{a.s.}} \mathbf{w}_1 \quad \text{as } t \to \infty,$$

where $\mathbf{w}_1$, given by (3.10), is a left eigenvector of $\boldsymbol{\Omega}$ corresponding to the dominant eigenvalue $\lambda_1 = r$. Thus if an epidemic takes off and is still in its exponentiallly growing phase then the relative frequencies of the different types of infectives will be close to $\mathbf{w}_1$. Hence, we now assume that the initial number of individuals in the branching process $Z(0) = K$, where K is large, and that $Z_i(0) \approx w_{1,i}K$ for $i \in \mathcal{K}$. Label the initial individuals $1, 2, \ldots, K$. Then, for $t \geq 0$, the total population size is $Z(t) = \hat{Z}_1(t) + \hat{Z}_2(t) + \ldots + \hat{Z}_K(t)$, where $\hat{Z}_i(t)$ denotes the total number of descendants of the initial individual i that are alive at time t, including i itself if it is still alive. Thus, $\mathbb{E}[Z(t)] = \sum_{i=1}^{K} \mathbb{E}\left[\hat{Z}_i(t)\right]$, for all $t \geq 0$, and, since the processes $\{\hat{Z}_i(t) : t \geq 0\}$ $(i = 1, 2, \ldots, K)$ are mutually independent, $\mathrm{var}(Z(t)) = \sum_{i=1}^{K} \mathrm{var}\left(\hat{Z}_i(t)\right)$, for all $t \geq 0$, and $\mathrm{cov}(Z(t), Z(s)) = \sum_{i=1}^{K} \mathrm{cov}\left(\hat{Z}_i(t), \hat{Z}_i(s)\right)$, for all $t, s \geq 0$.

Note that (3.9) implies that

$$\mathbf{w}_1^\top \mathbf{n} = \sum_{i=0}^{k_{max}} i w_{1,i} = \mu_{\tilde{D}-2} \quad \text{and} \quad \mathbf{w}_1^\top \mathbf{n}_2 = \sum_{i=0}^{k_{max}} i^2 w_{1,i} = \frac{\mu_{\tilde{D}-2}\mu_{(\tilde{D}-1)^2+1}}{\mu_{\tilde{D}}}.$$

Assuming that the above approximation is exact, then, for $t \geq 0$, it follows from (3.15) that

$$\mathbb{E}[Z(t)] = K\mathbf{w}_1^\top \mathbf{m}(t) = Ke^{rt} \tag{8.1}$$

and, after a little algebra, it follows from (4.4) that

$$\begin{aligned}
\text{var}\,(Z(t)) &= K\mathbf{w}_1^\top \mathbf{v}(t) \\
&= K\left(\gamma\left[I_9(t) + \mu_{\tilde{D}}^{-1}\mu_{\tilde{D}-2}^{-1}\left(\sigma_{\tilde{D}}^2 + 2\right)I_4(t)\right]\right. \\
&\quad \left. + \tau\mu_{\tilde{D}-2}^{-1}\left[\mu_{\tilde{D}-2}^2 I_9(t) + \sigma_{\tilde{D}}^2 I_4(t)\right]\right),
\end{aligned} \tag{8.2}$$

where $\sigma_{\tilde{D}}^2 = \text{var}(\tilde{D})$ and

$$I_9(t) = \frac{e^{rt}\left(e^{rt} - 1\right)}{r}.$$

Comparison of (8.1) and (8.2) with the diffusion-based result of Graham and House (2014) in the limit of large t gives agreement when $\gamma = 0$ (i.e. for the SI model) but not for $\gamma > 0$. We believe that this is due to the fact that the diffusion model was only four dimensional, so a heuristic argument (given in Sect. 3.3 of Graham and House 2014, which gave results that were in good agreement with simulation) about the neighbourhood of an infective node had to be made, in contrast to the approach here that deals with each effective degree explicitly and so has $k_{max} + 1$ dimensions. The argument about the neighbourhood around an infective tries to account for correlations caused by variability in recovery times, and so if $\gamma \to 0$ then these correlations do not exist.

Recent work by Constable and McKane (2014) considered the reduction of high-dimensional stochastic models to low-dimensional diffusions and this approach was shown to be asymptotically exact for some systems in the small-noise limit by Parsons and Rogers (2015). It is an open question whether the argument in Sect. 3.3 of Graham and House (2014) could be justified rigorously by a similar argument, however we note that a branching-process approach makes fewer assumptions than a low-dimensional diffusion limit and so will be more generally applicable.

Considering further results that can be obtained, it follows from (5.3) and a little algebra that, for $0 \leq t \leq s$,

$$\begin{aligned}
\text{cov}\,(Z(t), Z(s)) &= e^{-\gamma(s-t)}\text{var}\,(Z(t)) + K\mu_{\tilde{D}-2}^{-1}\left(e^{r(s-t)} - e^{-\gamma(s-t)}\right) \\
&\quad \times \left\{\gamma\mu_{\tilde{D}}^{-1}\left[\mu_{(\tilde{D}-1)^2+1}I_9(t) - \left(\sigma_{\tilde{D}}^2 + 2\right)I_6(t)\right]\right. \\
&\quad \left. + \tau\left[\mu_{\tilde{D}-2}^2 I_9(t) - \sigma_{\tilde{D}}^2 I_6(t)\right]\right\}.
\end{aligned} \tag{8.3}$$

It seems plausible that these results extend to the case when there is no maximal degree but that would involve results for countably infinite matrices which we do not consider here.

Recall that the processes $\{\hat{Z}_i(t) : t \geq 0\}$ $(i = 1, 2, \ldots, K)$ are mutually independent. It follows using the central limit theorem that, for sufficiently large K, the process $\{Z(t) : t \geq 0\}$, which approximates the prevalence of infection during the early growth of an epidemic, is approximately Gaussian with mean function given by (8.1) and covariance function given by (8.3).

9 Numerical examples

9.1 Forward simulations

We conducted a series of numerical experiments to provide specific examples of the general results presented here. $M = 10^4$ Monte Carlo simulations were performed on three different configuration model networks, each of size $N = 10^4$, and with the degree distributions shown in Fig. 2 Row (a). (Note that each Monte Carlo simulation consisted of first simulating a network and then simulating a single epidemic on it.) Two different scenarios were considered. In the first – most commonly considered in the literature when simulations are compared to analytic approaches – time $= 0$ was defined as the first point when prevalence is at a given level, K. In our simulations we took $K = 100$, but in general K should take a value where the probability of extinction has become negligible, but the depletion of the susceptible population has not had a significant effect on the epidemic dynamics. In the second, each epidemic was started from one node, picked uniformly at random, so the probability of extinction played a major role. This scenario is less commonly considered when comparing real-time simulated epidemics to differential equation models because the latter are typically designed to hold when the epidemic is already established.

Since analytic results for the probabilities of extinction $\pi_k(t)$ are not available, the branching process results required numerical integration of ordinary differential equations (in our case using Runge–Kutta methods). We stress that the computational effort required to do this is much less than that involved in performing Monte Carlo simulations, and has the benefit of not depending on N.

The results for the first approach (restarting time at the first time prevalence reaches 100) are given in Fig. 2. Row (b) shows sample trajectories (which all agree on prevalence at time 0). Row (c) shows the simulated mean after time 0 on a logarithmic scale, which initially grows at the constant rate predicted by the branching process model, and then reduces as the susceptible population is depleted. Row (d) shows the variance, which has not converged to its asymptotic growth regime by the time prevalence is equal to 100, an effect that is captured by the branching process model. The variance does not take its largest value at the peak prevalence, but instead has local maxima before and after the peak.

Figure 3 shows the results for the second approach in which there is one randomly chosen initial infective at time 0. Row (a) shows some sample trajectories.

(a)

(b)

(c)

(d)
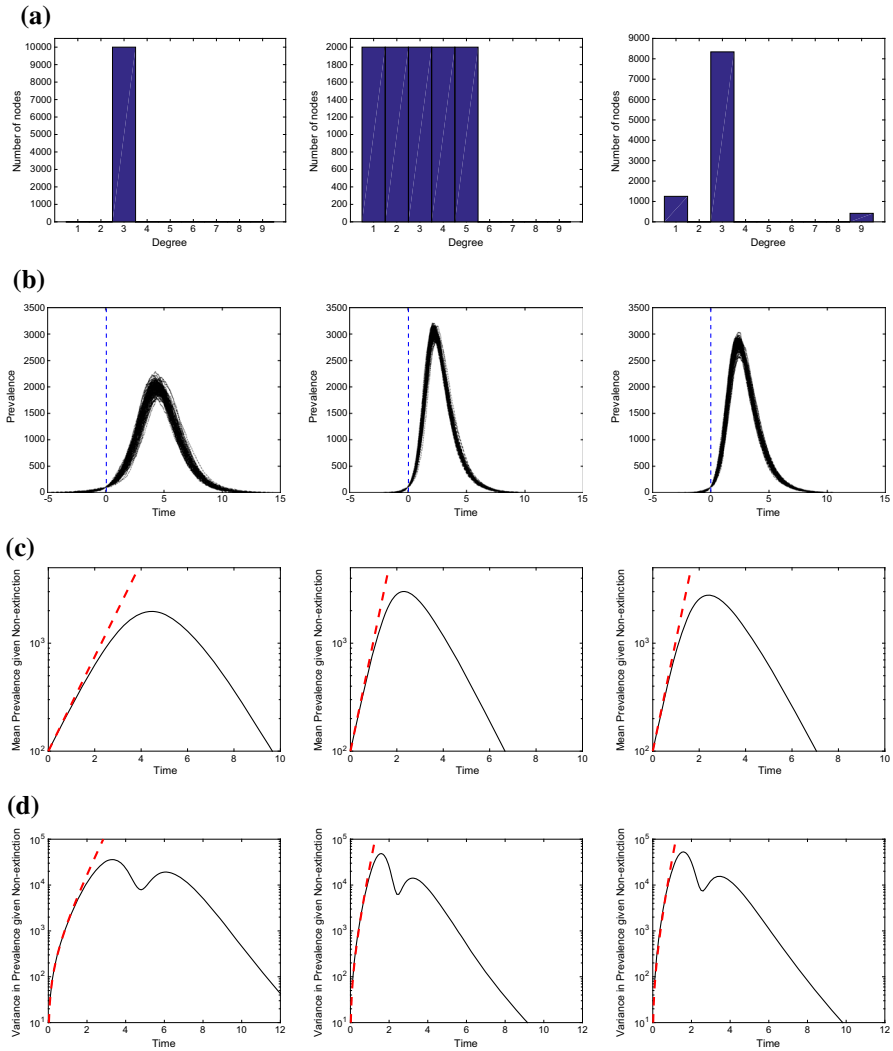
Fig. 2 Epidemic simulations that set time $= 0$ when prevalence is equal to 100. Parameters are $\tau = 2$, $\gamma = 1$ throughout. **a** Degree distribution histograms, **b** 100 sample trajectories, **c** Mean prevalence. *Black solid* simulations; *Red dashed* branching process **d** Variance in prevalence. *Black solid* simulations; *Red dashed* branching process (colour figure online)

Row (b) shows the extinction probabilities, which are accurately captured in the branching process model until very close to the end of the epidemic when prevalence is low and extinction becomes likely again. Row (c) shows the mean, which does not start growing at a constant rate with the convergence rate accurately captured in the branching process model. Row (d) shows the variance and convergence onto its asymptotic value; in this case there is a single maximum just before the peak in prevalence.
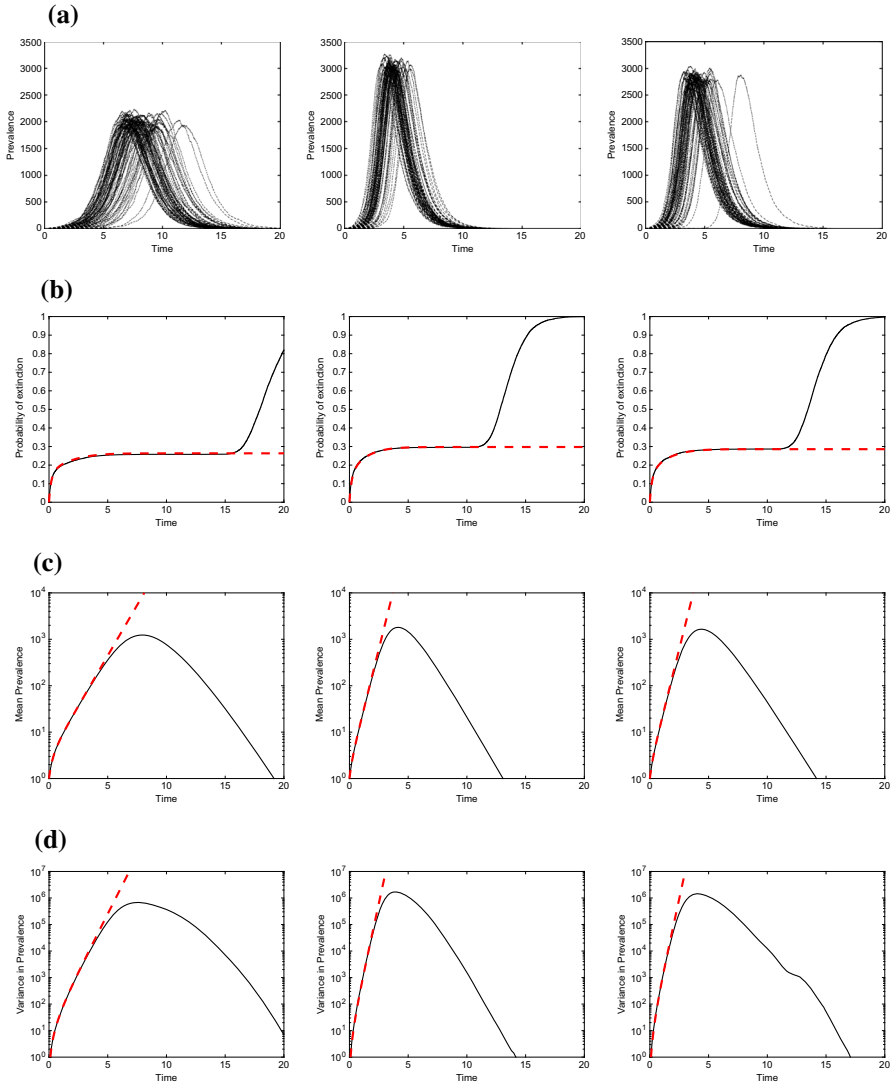
Fig. 3 Epidemic simulations starting from one node selected uniformly at random. Parameters are $\tau = 2$, $\gamma = 1$ throughout. Degree distributions are as for Fig. 2 above. **a** 100 sample trajectories, **b** Extinction probabilities, *Black solid* simulations; *Red dashed* branching process, **c** Mean prevalence. *Black solid* simulations; *Red dashed* branching process, **d** Variance in prevalence. *Black solid* simulations; *Red dashed* branching process (colour figure online)

Another important point is that while mean numbers infected are comparable between Figs. 2 and 3, the variability in the time for the epidemic to take off, as well as the contribution from extinct epidemics, makes the real-time variance in Fig. 3 orders of magnitude larger than in Fig. 2.

9.2 Statistical inference

In order to demonstrate the potential use of the real-time effective degree branching process model for statistical inference, we carried out a simulation study. Here we simulated one epidemic that took off on a configuration model network of size $N = 10^6$ with degree distribution $D^{(3)}$ as in the right-hand column of Fig. 2 ($d_1^{(3)} = 1/8$, $d_3^{(3)} = 5/6, d_9^{(3)} = 1/24$) and true rates $\tau_0 = 2, \gamma_0 = 1$. Letting $I(t)$ be the prevalence of infection in the network model, we set time $t = 0$ when $I(t) = 100$ for the first time and make 40 evenly-spaced observations (with gap $\delta t = 0.05$ between each) of $I(t)$ up to $t_{\text{end}} = 2$.

We then define an approximate likelihood based on the methods of Ross et al. (2006), in which a Gaussian process approximation based on known first and second moments is used, which will be more accurate for larger N, larger $I(0)$ and smaller δt. There should be, however, no *a priori* obstacle to fitting our model to data on smaller populations even with incomplete data, for example by using Markov chain Monte Carlo methods to perform multiple imputation as suggested by O'Neill and Roberts (1999).

Explicitly, we let the probability density function f for sequential observations be given by

$$f(I(t + \delta t)|I(t)) = \mathcal{N}(\mathbb{E}[Z(t + \delta t)|Z(t) = I(t)], \text{var}(Z(t + \delta t)|Z(t) = I(t)), \tag{9.1}$$

where $\mathcal{N}(m, V)$ is the probability density function of a normal distribution with mean m and variance V, and the expectation and variance of $Z(t+\delta t)|Z(t) = I(t)$ are given by the results of Sect. 8 above. The likelihood is then

$$L = \prod_{t \in \{0, \delta t, ..., t_{\text{end}} - \delta t\}} f(I(t + \delta t)|I(t)). \tag{9.2}$$

We consider values of this likelihood across the range of rate constant parameters τ and γ under two different degree distributions: the correct one, $D^{(3)}$, and a misspecified degree distribution $D^{(1)}$, which is the one used in the left-hand column of Fig. 2 ($d_3^{(1)} = 1$).

Figure 4 shows the first quarter of the simulated epidemic together with the Gaussian process approximation, as well as likelihood surfaces for the correct and misspecified degree distributions. Performing maximum likelihood estimation using MATLAB's mle() function allows us to obtain point estimates for parameters $\hat{\tau}$ and $\hat{\gamma}$, as well as asymptotic 95% confidence intervals and the parameter covariance matrix $\hat{C}$ from the inverse Hessian. We quote results to 2 significant figures; the asymptotic approximations also give very slightly negative lower confidence intervals for $\hat{\gamma}$ which we round up to 0. For the correct degree distribution we obtain
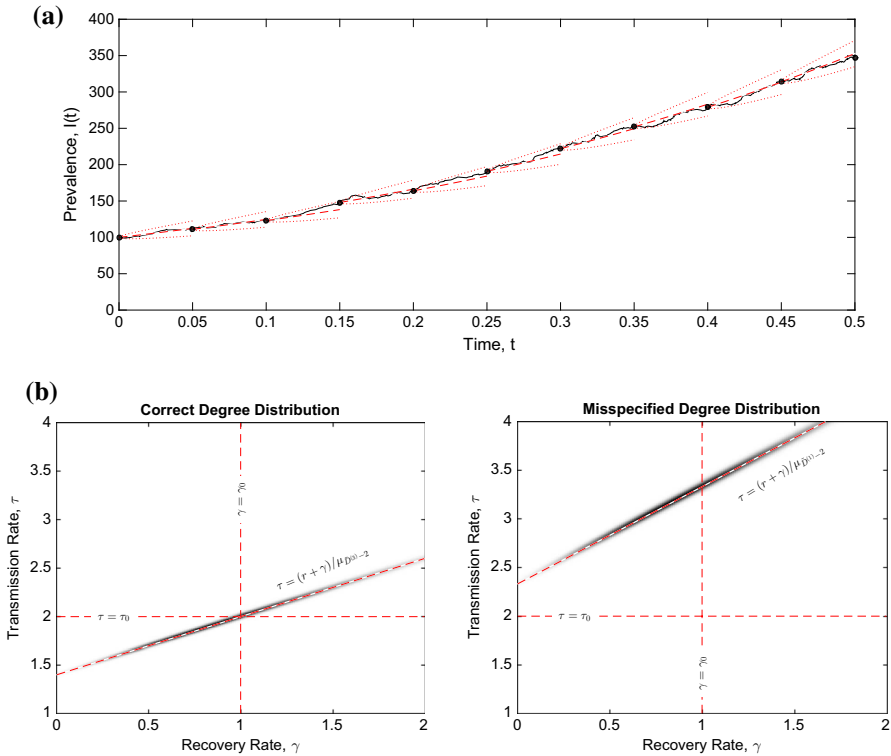
Fig. 4 Simulation study. The *top plot* **a** shows the first quarter of the timepoints **a** Simulated epidemic, **b** Likelihood surfaces (observations as *black dots*, full trajectory as *black solid line*, Gaussian approximation mean as *red dashed line*, Gaussian approximation 95% prediction interval as *red dotted line*). The *bottom plots* **b** show likelihood surfaces for a Gaussian approximation model as described in Sect. 9.2. True parameters are $\tau_0 = 2$, $\gamma_0 = 1$. Correct degree distribution $D^{(3)}$ is as given in the third columns of Figs. 2 and 3 above, and the misspecified distribution $D^{(1)}$ is the distribution from the first column. Likelihood at a point is proportional to the intensity of shading, and with three straight lines, corresponding to the true values of τ, γ and r, are shown each plot as dashed red lines (color figure online)

$$\begin{pmatrix} \hat{\tau}^{(3)} \\ \hat{\gamma}^{(3)} \end{pmatrix} = \begin{pmatrix} 1.9 & [1.4, 2.4] \\ 0.8 & [0, 1.7] \end{pmatrix}, \qquad \hat{C}^{(3)} = \begin{pmatrix} 0.069 & 0.11 \\ 0.11 & 0.19 \end{pmatrix}, \tag{9.3}$$

and for the misspecified degree distribution we obtain

$$\begin{pmatrix} \hat{\tau}^{(1)} \\ \hat{\gamma}^{(1)} \end{pmatrix} = \begin{pmatrix} 3.2 & [2.3, 4.0] \\ 0.8 & [0, 1.7] \end{pmatrix}, \qquad \hat{C}^{(1)} = \begin{pmatrix} 0.0045 & 0.011 \\ 0.011 & 0.030 \end{pmatrix}. \tag{9.4}$$

This shows that knowledge of the correct distribution allows both τ and γ to be estimated; although as would be expected the early asymptotic growth rate r is much more closely constrained by simulated data than other directions in parameter space. It also shows that misspecification of the degree distribution allows r to be identified, but biases the estimate of, in this case, τ.

10 Concluding comments

10.1 Summary of results

In this paper, we have provided explicit closed-form expressions for the real-time mean, variance and covariance function for disease prevalence during the early stages of the Markovian SIR model on a configuration model network, as well as deriving differential equations for the probabilities of extinction over time that are relatively numerically cheap to solve. These allow for a more explicit treatment of e.g. rates of convergence to asymptotic behaviour than has previously been possible.

10.2 Future directions

We believe that the methods of real-time, multitype branching processes could be more widely applied in infectious disease epidemiology, since they provide results concerning extinction and variance that are not available using deterministic differential equation models. For example, the effective-degree based methodology presented here may be extended to include degree correlation (e.g. in the sense of Newman 2002) by keeping track of the actual, as well as effective, degrees of individuals, though the type space becomes larger and explicit analytic results are unlikely to be available. We note that there is increasing interest in the eradication of infections (e.g. Klepac et al. 2013) and that arguably calculating extinction probabilities and variability in outbreak sizes is of equal or greater importance in this context than calculation of mean behaviour.

The explicit closed-form expressions derived have the potential to enhance statistical work on epidemic prevalence curves. In particular, many empirically observed epidemics of human pathogens exhibit more variability around the trend than simple models would predict (see Black et al. 2014, particularly Section 1, for a discussion of this), which can bias parameter estimation if an insufficiently variable model is used. Application of our methods to real data would be an interesting extension of our work.

The possibility of a more general non-Markovian stochastic epidemic being approximated by an appropriate real-time branching process is raised by the results of Barbour and Reinert (2013) and it would be interesting to investigate whether our analysis could be adapted to this scenario.

Finally, there is the question of low-dimensional PGF-based modelling of the whole network epidemic that incorporates stochasticity accurately. For example, the work of Miller (2014) considered accounting for early fluctuations and Graham and House (2014) considered a diffusion approximation once early fluctuations were negligible, but the results presented here as well as those of Barbour and Reinert (2013) suggest that a more unified low-dimensional stochastic approach that explicitly models early fluctuations may be possible.

Appendix 1: Convergence of moments

We determine sufficient conditions for the first two moments of the number of infectives in the epidemic E_N among a population of N individuals to converge to the corresponding moments of the limiting branching process $\mathscr{B}$. For ease of exposition, in E_N we assume that at time $t = 0$ there is one infective and the remaining $N - 1$ individuals are susceptible. The initial infective is chosen by sampling a stub uniformly at random from all stubs used to form the network, with the individual attached to that stub being the initial infective. The arguments are easily extended to other choices of initial infective(s).

In the independent and identically distributed (i.i.d.) degree case, we assume that a single sequence $D_1, D_2, \ldots$ of i.i.d. copies of D is used to construct a sequence of epidemics (E_N), where, for $N = 1, 2, \ldots$, the epidemic E_N is constructed using $D_1, D_2, \ldots, D_N$. In the prescribed degree case (see Sect. 2.1), recall that $p_k^{(N)}$ ($k = 0, 1, \ldots$) denotes the empirical degree distribution in the epidemic E_N and, for $f : \mathbb{R} \to \mathbb{R}$, let $\mu_{f(D^{(N)})} = \sum_{k=0}^{\infty} p_k^{(N)} f(k)$.

For $N = 1, 2, \ldots$ and $t \geq 0$, let $Y_N(t)$ be the number of infectives in E_N at time t and let $Z(t)$ be the number of individuals in the limiting branching process $\mathscr{B}$. Then arguing as in the proof of (Ball and Neal 2008, Theorem A.1), shows that in the i.i.d. degree case, if $\mu_D < \infty$ then the sequence of epidemics (E_N) and the limiting branching process $\mathscr{B}$ can be constructed on a common probability space so that, with probability one, for any $t > 0$, $Y_N(u)$ and $Z(u)$ coincide for all $u \in [0, t]$ for all sufficiently large N. The same conclusion holds in the prescribed degree case provided $p_k^{(N)} \to p_k$ ($k = 0, 1, \ldots$) and $\mu_D^{(N)} \to \mu_D$ as $N \to \infty$, where $\sum_{k=0}^{\infty} p_k = 1$ and $\mu_D < \infty$. Thus, under these conditions, in both cases, for any $t \geq 0$, $Y_N(t)$ converges almost surely to $Z(t)$ as $N \to \infty$. We obtain further conditions, under which, for fixed $t \geq 0$, the sequence $(Y_N(t)^2)$ is uniformly integrable, which then (e.g. Grimmett and Stirzaker 2001, Chap. 7, Sect. 10) implies immediately that $\lim_{N \to \infty} \mathbb{E}[Y_N(t)] = \mathbb{E}[Z(t)], \lim_{N \to \infty} \mathbb{E}[Y_N(t)^2] = \mathbb{E}[Z(t)^2]$ and, for any $0 \leq s \leq t$, $\lim_{N \to \infty} \text{cov}(Y_N(s), Y_N(t)) = \text{cov}(Z(s), Z(t))$. To show that $(Y_N(t)^2)$ is uniformly integrable it is sufficient to show that the sequence $(\mathbb{E}[Y_N(t)^{2+\delta}])$ is bounded above for some $\delta > 0$.

For ease of exposition, we use notation from the i.i.d. degree case. The construction in Ball and Neal (2008) involves for each N constructing a realisation of a branching process, $\mathscr{B}_N$ say, which is defined analogously to $\mathscr{B}$ but using the empirical distribution of $D_1, D_2, \ldots, D_N$ rather than the distribution of D. In $\mathscr{B}_N$, for each birth a stub is chosen independently and uniformly from all the $D_1 + D_2 + \cdots + D_N$ stubs and the degree of the individual that the chosen stub belongs to gives the degree of the

individual born at that birth. The process of infectives in E_N follows $\mathscr{B}_N$ except when (i) a sampled stub has previously been chosen, in which case stubs are resampled until one that has not been chosen previously is obtained, or (ii) a sampled stub has not been chosen previously but is attached to an individual that has already been infected, in which case the corresponding birth and all descendants of that individual in $\mathscr{B}_N$ are ignored in E_N. Note that (i) implies that $\mathscr{B}_N$ need not be an almost sure upper bound for the process of infectives in E_N. The branching processes $\mathscr{B}_N$ ($N = 1, 2, \ldots$) and $\mathscr{B}$ are coupled so that with probability one, for any fixed $t > 0$, $\mathscr{B}_N$ and $\mathscr{B}$ coincide over $[0, t]$ for all sufficiently large N.

Let $D_{(1)}, D_{(2)}, \ldots, D_{(N)}$ be the order statistics of $D_1, D_2, \ldots, D_N$, i.e. $D_1, D_2, \ldots, D_N$ arranged in increasing order. For $\epsilon \in (0, 1)$, let $\mathscr{B}_{N,\epsilon}$ be the branching process that is defined analagously to $\mathscr{B}$, but using the empirical distribution of $D_{([N\epsilon]+1)}, D_{([N\epsilon]+2)}, \ldots, D_{(N)}$. (For $x \in \mathbb{R}$, $[x]$ denotes the greatest integer $\leq x$.) For $t \geq 0$, let $Z_{N,\epsilon}(t)$ be the number of individuals alive in $\mathscr{B}_{N,\epsilon}$ at time t and let $T_{N,\epsilon}(t)$ denote the total progeny of $\mathscr{B}_{N,\epsilon}$ by time t, including the initial ancestor. Then $Y_N(t) \overset{st}{\leq} Z_{N,\epsilon}(t)$, provided that $T_{N,\epsilon}(t) \leq N\epsilon$, where $\overset{st}{\leq}$ denotes stochastically smaller than. (Up until $[N\epsilon]$ infections have occurred in E_N, the empirical distribution of the degrees of unsampled stubs, where the degree of a stub is the degree of the individual to which it is attached, is stochastically smaller than that of the stubs belonging to the top $N - [N\epsilon]$ individuals when ordered by degree.) As $Y_N(t)$ is at most N, it follows that

$$\mathbb{E}\left[Y_N(t)^{2+\delta}\right] \leq \mathbb{E}\left[Z_{N,\epsilon}(t)^{2+\delta}\right] + N^{2+\delta}\mathbb{P}\left(T_{N,\epsilon}(t) > N\epsilon\right).$$

By Markov's inequality,

$$\mathbb{P}\left(T_{N,\epsilon}(t) > N\epsilon\right) \leq \frac{1}{(N\epsilon)^{2+\delta}}\mathbb{E}\left[T_{N,\epsilon}(t)^{2+\delta}\right].$$

Also, $Z_{N,\epsilon}(t) \overset{st}{\leq} T_{N,\epsilon}(t)$, so

$$\mathbb{E}\left[Y_N(t)^{2+\delta}\right] \leq \left(1 + \epsilon^{-(2+\delta)}\right)\mathbb{E}\left[T_{N,\epsilon}(t)^{2+\delta}\right]. \tag{10.1}$$

We now bound $\mathbb{E}\left[T_{N,\epsilon}(t)^{2+\delta}\right]$, for $\delta \in (0, 1)$. Note that this moment is smaller than the corresponding moment for the branching process in which $\gamma = 0$ and individuals retain their original effective degree throughout their lifetime. Moreover, by rescaling the time axis, we can assume without loss of generality that $\tau = 1$. Thus consider a multitype Markov birth process, with types $0, 1, \ldots, J$, in which an individual of type i gives birth at rate i and the types of successive births are i.i.d. with probability mass function $\hat{p}_j$ ($j = 0, 1, \ldots, J$). For $i = 0, 1, \ldots, J$, let $\hat{T}_i(t)$ denote the total number of individuals alive in this process at time t given that at time 0 there is one indivdiual, whose type is i. For $\alpha > 0$ and $t \geq 0$, let $\hat{\mu}_i^{(\alpha)}(t) = \mathbb{E}\left[\hat{T}_i(t)^\alpha\right]$ ($i = 0, 1, \ldots, J$) and let $\hat{\mu}^{(\alpha)}(t) = \sum_{i=0}^{J} \hat{p}_i \hat{\mu}_i^{(\alpha)}(t)$. It is possible using a backward argument to derive explicit expressions for $\hat{\mu}_i^{(k)}(t)$ for $k = 1, 2, \ldots$, though the algebra

soon becomes very tedious. As our aim is to bound $\hat{\mu}^{(2+\delta)}(t)$, we simply derive bounds for $\hat{\mu}_i^{(1)}(t)$, $\hat{\mu}_i^{(2)}(t)$ ($i = 0, 1, \ldots, J$) and finally $\hat{\mu}^{(2+\delta)}(t)$. Moreover, our bounds are deliberately coarse to facilitate easy application to $\mathbb{E}\left[T_{N,\epsilon}(t)^{2+\delta}\right]$. For $f : \mathbb{R} \to \mathbb{R}$, let $\hat{\mu}_{f(D)} = \sum_{i=0}^{J} f(i)\hat{p}_i$.

For $i = 0, 1, \ldots, J$, the backward equation for $\hat{\mu}_i^{(1)}(t)$ is

$$\frac{d\hat{\mu}_i^{(1)}}{dt} = i\hat{\mu}^{(1)}, \qquad \hat{\mu}_i^{(1)}(0) = 1. \tag{10.2}$$

Multiplying (10.2) by $\hat{p}_i$ and summing over $i = 0, 1, \ldots, J$, yields

$$\frac{d\hat{\mu}^{(1)}}{dt} = \hat{\mu}_D \hat{\mu}^{(1)}, \qquad \hat{\mu}^{(1)}(0) = 1.$$

Thus

$$\hat{\mu}^{(1)}(t) = e^{\hat{\mu}_D t}, \tag{10.3}$$

which on substituting into (10.2) yields $\hat{\mu}_i^{(1)} = 1 + i \int_0^t e^{\hat{\mu}_D s} ds$, so

$$\hat{\mu}_i^{(1)} \leq i(1+t)e^{\hat{\mu}_D t} \qquad (i = 1, 2, \ldots, J). \tag{10.4}$$

For $i = 0, 1, \ldots, J$, the backward equation for $\hat{\mu}_i^{(2)}(t)$ is

$$\frac{d\hat{\mu}_i^{(2)}}{dt} = 2i\hat{\mu}_i^{(1)}\hat{\mu}^{(1)} + i\hat{\mu}^{(2)}, \qquad \hat{\mu}_i^{(2)}(0) = 1. \tag{10.5}$$

Substituting from (10.3) and (10.4), and then multiplying (10.5) by $\hat{p}_i$ and summing over $i = 0, 1, \ldots, J$, yields

$$\frac{d\hat{\mu}^{(2)}}{dt} \leq 2\hat{\mu}_{D^2}(1+t)e^{2\hat{\mu}_D t} + \hat{\mu}_D \hat{\mu}^{(2)}, \qquad \hat{\mu}^{(2)}(0) = 1,$$

whence

$$\hat{\mu}^{(2)}(t) \leq e^{\hat{\mu}_D t} + \int_0^t e^{\hat{\mu}_D(t-s)} 2\hat{\mu}_{D^2}(1+s)e^{2\hat{\mu}_D s} ds$$
$$\leq \left(1 + 2\hat{\mu}_{D^2} t(1+t)\right) e^{2\hat{\mu}_D t}.$$

Substituting this bound into (10.5) and noting that the right-hand side of (10.5) is increasing in t yields

$$\hat{\mu}_i^{(2)}(t) \leq i^2 g(t, \hat{\mu}_{D^2}) e^{2\hat{\mu}_D t} \qquad (i = 1, 2, \ldots, J), \tag{10.6}$$

where

$$g(t, \hat{\mu}_{D^2}) = 1 + t\left[1 + 2(1 + t)(1 + \hat{\mu}_{D^2}t)\right].$$ (10.7)

Let $\delta \in (0, 1)$. For $i = 0, 1, \ldots, J$, the backward equation for $\hat{\mu}_i^{(2+\delta)}(t)$ is

$$\frac{d\hat{\mu}_i^{(2+\delta)}}{dt} = -i\hat{\mu}_i^{(2+\delta)}(t) + i\mathbb{E}\left[\left(\hat{T}_i(t) + \hat{T}(t)\right)^{2+\delta}\right], \quad \hat{\mu}_i^{(2+\delta)}(0) = 1,$$ (10.8)

where $\hat{T}_i(t)$ and $\hat{T}(t)$ are independent and $\hat{T}(t)$ is distributed as a mixture of $\hat{T}_0(t), \hat{T}_1(t), \ldots, \hat{T}_J(t)$ with mixing probabilities $\hat{p}_0, \hat{p}_1, \ldots, \hat{p}_J$.

Let a and b be nonnegative real numbers. Application of the mean value theorem yields that

$$(a + b)^{2+\delta} \leq a^{2+\delta} + (2 + \delta)b(a + b)^{1+\delta}.$$

Further, $(a + b)^{1+\delta} \leq 2^\delta \left(a^{1+\delta} + b^{1+\delta}\right)$, so

$$(a + b)^{2+\delta} \leq a^{2+\delta} + 2^\delta(2 + \delta)\left(ba^{1+\delta} + b^{2+\delta}\right).$$

Setting $a = \hat{T}_i(t)$ and $b = \hat{T}(t)$ in this inequality, taking expectations exploiting the independence of $\hat{T}_i(t)$ and $\hat{T}(t)$, substituting into (10.8) and noting that $2^\delta(2+\delta) < 6$ gives

$$\frac{d\hat{\mu}_i^{(2+\delta)}}{dt} \leq 6i\left[\hat{\mu}^{(1)}(t)\hat{\mu}_i^{(1+\delta)}(t) + \hat{\mu}^{(2+\delta)}(t)\right], \quad \hat{\mu}_i^{(2+\delta)}(0) = 1.$$ (10.9)

Now

$$\hat{\mu}_i^{(2)}(t) = \mathbb{E}\left[\left(\hat{T}_i(t)^{1+\delta}\right)^{\frac{2}{1+\delta}}\right] \geq \left(\hat{\mu}_i^{(1+\delta)}(t)\right)^{\frac{2}{1+\delta}},$$

by Jensen's inequality, so

$$\hat{\mu}_i^{(1+\delta)}(t) \leq \left(\hat{\mu}_i^{(2)}(t)\right)^{\frac{1+\delta}{2}} \leq i^{1+\delta}g(t, \hat{\mu}_{D^2})e^{(1+\delta)\hat{\mu}_D t},$$ (10.10)

using (10.6) and noting that $g(t, \hat{\mu}_{D^2}) \geq 1$. Substituting (10.10) into (10.9), multiplying by $\hat{p}_i$ and summing over $i = 0, 1, \ldots, J$, yields

$$\frac{d\hat{\mu}^{(2+\delta)}}{dt} \leq 6\hat{\mu}_{D^{2+\delta}}g(t, \hat{\mu}_{D^2})e^{(2+\delta)\hat{\mu}_D t} + 6\hat{\mu}_D\hat{\mu}^{(2+\delta)}.$$

Hence,

$$\hat{\mu}^{(2+\delta)} \leq e^{6\hat{\mu}_D t} + \int_0^t e^{6\hat{\mu}_D(t-s)}6\hat{\mu}_{D^{2+\delta}}g(s, \hat{\mu}_{D^2})e^{(2+\delta)\hat{\mu}_D s}ds$$

$$\leq e^{6\hat{\mu}_D t}\left[1 + 6\hat{\mu}_{D^{2+\delta}}tg(t, \hat{\mu}_{D^2})\right].$$ (10.11)

We return to epidemics on networks and introduce some more notation. For $\epsilon \in (0, 1)$, let $k_0(\epsilon) = \min\{k : p_0 + p_1 + \ldots, p_k > \epsilon\}$ and, for $k = 0, 1, \ldots$, let

$$p_k(\epsilon) = \begin{cases} 0 & \text{if } k < k_0(\epsilon), \\ 1 - \frac{1}{1-\epsilon} \sum_{k=k_0(\epsilon)}^{\infty} p_k & \text{if } k = k_0(\epsilon), \\ \frac{p_k}{1-\epsilon} & \text{if } k > k_0(\epsilon). \end{cases}$$

For $f : \mathbb{R} \to \mathbb{R}$, let $\mu_{f(D(\epsilon))} = \sum_{k=0}^{\infty} p_k(\epsilon) f(k)$. As above, we assume without loss of generality that $\tau = 1$.

Consider first the model with prescribed degrees. Let $D_{(1)}^{(N)}, D_{(2)}^{(N)}, \ldots, D_{(N)}^{(N)}$ be the order statistics of $D_1^{(N)}, D_2^{(N)}, \ldots, D_N^{(N)}$. For $\epsilon \in (0, 1)$, let $p_k^{(N)}(\epsilon) = (N - [N\epsilon])^{-1} \sum_{i=N-[N\epsilon]}^{N} \delta_{k, D_{(i)}^{(N)}}$ $(k = 0, 1, \ldots)$ and, for $f : \mathbb{R} \to \mathbb{R}$, let

$$\mu_{f(D^{(N)}(\epsilon))} = \sum_{k=0}^{\infty} p_k^{(N)}(\epsilon) f(k) = \frac{1}{N - [N\epsilon]} \sum_{k=[N\epsilon]+1}^{N} f\left(D_{(k)}^{(N)}\right)^{\alpha}.$$

Let $\tilde{p}_k^{(N)}(\epsilon) = k p_k^{(N)}(\epsilon) / \mu_{D(\epsilon)}^{(N)}$ $(k = 1, 2, \ldots)$. For $f : \mathbb{R} \to \mathbb{R}$, let

$$\mu_{f(\tilde{D}^{(N)}(\epsilon))} = \sum_{k=1}^{\infty} \tilde{p}_k^{(N)}(\epsilon) f(k) = \frac{\mu_{D^{(N)}(\epsilon) f(D^{(N)}(\epsilon))}}{\mu_{D^{(N)}(\epsilon)}}.$$

Fix the population size N and $\epsilon \in (0, 1)$. In the birth process used to bound the right-hand side of (10.1), the types of individuals born are distributed according to $\hat{p}_i = \tilde{p}_{i+1}^{(N)}(\epsilon)$ $(i = 0, 1, \ldots, J)$, where $J = \max\left\{D_{(k)}^{(N)} : k = 1, 2, \ldots, N\right\}$. Thus, for $\delta \in (0, 1)$, it follows using (10.1) and (10.11) that

$$\mathbb{E}\left[Y_N(t)^{2+\delta}\right] \le h_1(N, \epsilon, t),$$

where

$$h_1(N, \epsilon, t) = \left(1 + \epsilon^{-(2+\delta)}\right)\left[1 + 6\mu_{(\tilde{D}^{(N)}(\epsilon)-1)^{2+\delta}} t g\left(t, \mu_{(\tilde{D}^{(N)}(\epsilon)-1)^2}\right) \exp\left(6\mu_{\tilde{D}^{(N)}(\epsilon)-1} t\right)\right].$$

Suppose that there exists $\delta \in (0, 1)$ such that $\mu_{D^{3+\delta}} < \infty$ and $\mu_{(D^{(N)})^{3+\delta}} \to \mu_{D^{3+\delta}}$ as $N \to \infty$. It is easily verified that these conditions imply that, for each $\epsilon \in (0, 1)$ and any $\alpha \in [0, 3+\delta]$, $\mu_{D(\epsilon)^{\alpha}} < \infty$ and $\mu_{(D^{(N)}(\epsilon))^{\alpha}} \to \mu_{D(\epsilon)^{\alpha}}$ as $N \to \infty$. Hence, for any $\alpha \in [0, 2+\delta]$, $\mu_{(\tilde{D}^{(N)}(\epsilon)-1)^{\alpha}} \to \mu_{(\tilde{D}(\epsilon)-1)^{\alpha}}$ as $N \to \infty$, where $\mu_{(\tilde{D}(\epsilon)-1)^{\alpha}} < \infty$. It follows that $h_1(N, \epsilon, t) \to h_1(\epsilon, t)$ as $N \to \infty$, where

$$h_1(\epsilon, t) = \left(1 + \epsilon^{-(2+\delta)}\right)\left[1 + 6\mu_{(\tilde{D}(\epsilon)-1)^{2+\delta}} t g\left(t, \mu_{(\tilde{D}(\epsilon)-1)^2}\right) \exp\left(6\mu_{\tilde{D}(\epsilon)-1} t\right)\right] < \infty.$$

Thus the sequence $\left(\mathbb{E}\left[Y_N(t)^{2+\delta}\right]\right)$ is bounded and, for any $t \ge 0$ and any $\alpha \in [0, 2]$, $\mathbb{E}[Y_N(t)^{\alpha}] \to \mathbb{E}[Z(t)^{\alpha}]$ as $N \to \infty$.

Turn now to the model with i.i.d. degrees. Recall that we construct a sequence of epidemics (E_N) from a single sequence $D_1, D_2, \ldots$ of i.i.d. copies of D. For $N = 1, 2, \ldots$, let $D_k^{(N)} = D_k$ $(k = 1, 2, \ldots, N)$. Using the formulae derived previously for the prescribed degree case, but noting that now the degrees are random, by conditioning on the degree sequence $D_1, D_2, \ldots$ we obtain that

$$\mathbb{E}\left[Y_N(t)^{2+\delta}\right] \leq h_2(N, \varepsilon, t), \tag{10.12}$$

where

$$h_2(N, \varepsilon, t) = \mathbb{E}\left[\left(1 + \varepsilon^{-(2+\delta)}\right)\left\{1 + 6\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^{2+\delta}}\, tg\left(t, \mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^2}\right)\exp\left(6\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right\}\right].$$

Fix $\varepsilon \in (0, 1)$. Recalling the definition (10.7) of the function g, to obtain an upper bound for $h(N, \varepsilon, t)$, it is sufficient to obtain upper bounds for

$$\mathbb{E}\left[\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^{2+\delta}}\exp\left(6\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right] \qquad \text{and}$$

$$\mathbb{E}\left[\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^{2+\delta}}\,\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^2}\exp\left(6\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right].$$

Now $\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^2} \leq \mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^{2+\delta}}$ and, by Jensen's inequality, $\left(\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^{2+\delta}}\right)^2 \leq \mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^{2(2+\delta)}}$, so, since $\delta \in (0, 1)$, it is sufficient to obtain an upper bound for $\mathbb{E}\left[\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^6}\exp\left(6\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right]$. Further, using the Cauchy-Schwarz inequality,

$$\mathbb{E}\left[\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^6}\exp\left(6\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right] \leq \sqrt{\mathbb{E}\left[\mu^2_{(\tilde{D}^{(N)}(\varepsilon)-1)^6}\right]\mathbb{E}\left[\exp\left(12\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right]}.$$

Let $M_{D^2}(\theta) = \mathbb{E}\left[\exp\left(\theta D^2\right)\right]$ $(\theta \in \mathbb{R})$ be the moment-generating function of D^2 and suppose that there exists $\theta_0 > 0$ such that $M_{D^2}(\theta_0) < \infty$. Note that this implies that $\mathbb{E}[D^\alpha] < \infty$ for all $\alpha \geq 0$.

Assume first that $p_0 = 0$, so $\mu_{D^{(N)}(\varepsilon)} \geq 1$ almost surely. Now

$$
\begin{aligned}
\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^6} &= \frac{\mu_{D^{(N)}(\varepsilon)(D^{(N)}(\varepsilon)-1)^6}}{\mu_{D^{(N)}(\varepsilon)}} \\
&\leq \mu_{D^{(N)}(\varepsilon)(D^{(N)}(\varepsilon)-1)^6} \qquad \text{almost surely} \\
&= \frac{1}{N - [N\varepsilon]}\sum_{k=[N\varepsilon]+1}^{N} D_{(k)}^{(N)}\left(D_{(k)}^{(N)} - 1\right)^6 \\
&\leq \frac{1}{N(1 - \varepsilon)}\sum_{k=1}^{N} D_k^7, \tag{10.13}
\end{aligned}
$$

since $D_k^{(N)} = D_k$ $(k = 1, 2, \ldots, N)$. Thus, since $D_1, D_2, \ldots, D_N$ are i.i.d.,

$$\mathbb{E}\left[\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)^6}^2\right] \leq \frac{1}{N^2(1-\varepsilon)^2}\mathbb{E}\left[\left(\sum_{k=1}^{N} D_k^7\right)^2\right]$$

$$= \frac{1}{N^2(1-\varepsilon)^2}\left[N\mu_{D^{14}} + N(N-1)\mu_{D^7}^2\right]$$

$$\leq \frac{1}{(1-\varepsilon)^2}\left(\mu_{D^{14}} + \mu_{D^7}^2\right).$$

A similar argument to (10.13) yields $\mu_{(\tilde{D}^{(N)}(\varepsilon)-1)} \leq \frac{1}{N(1-\varepsilon)}\sum_{k=1}^{N} D_k^2$, so

$$\mathbb{E}\left[\exp\left(12\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right] \leq \mathbb{E}\left[\exp\left(\frac{12t}{N(1-\varepsilon)}\sum_{k=1}^{N} D_k^2\right)\right] = \left[M_{D^2}\left(\frac{12t}{N(1-\varepsilon)}\right)\right]^N.$$

Fix $t \geq 0$. Then $\mathbb{E}\left[\exp\left(12\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right] < \infty$ for $N \geq N(\varepsilon, t)$, where $N(\varepsilon, t) = \frac{12t}{(1-\varepsilon)\theta_0}$. Now $M_{D^2}(\theta) = 1 + \mu_{D^2}\theta + o(\theta)$ as $\theta \to 0$, so

$$\lim_{N\to\infty}\mathbb{E}\left[\exp\left(12\mu_{\tilde{D}^{(N)}(\varepsilon)-1}t\right)\right] \leq \lim_{N\to\infty}\left[1 + \frac{12t}{N(1-\varepsilon)}\mu_{D^2} + o\left(\frac{1}{N}\right)\right]^N$$

$$= \exp\left(\frac{12t\mu_{D^2}}{1-\varepsilon}\right) < \infty.$$

The above arguments show that there exists $h_2(\varepsilon, t) < \infty$ such that $h_2(N, \varepsilon, t) < h_2(\varepsilon, t)$ for all $N \geq N(\varepsilon, t)$. Now $Y_N(t) \leq N$ for all N, so $\mathbb{E}\left[Y_N(t)^{2+\delta}\right] \leq N(\varepsilon, t)^{2+\delta}$ for $N < N(\varepsilon, t)$. Thus, recalling (10.12), the sequence $\left(\mathbb{E}\left[Y_N(t)^{2+\delta}\right]\right)$ is bounded and, for any $\alpha \in [0, 2]$, $\mathbb{E}\left[Y_N(t)^\alpha\right] \to \mathbb{E}\left[Z(t)^\alpha\right]$ as $N \to \infty$.

Suppose now that $p_0 > 0$. Then $D \stackrel{st}{\leq} D'$, where D' has distribution given by $\mathbb{P}(D' = 1) = p_0 + p_1$ and $\mathbb{P}(D' = k) = p_k$ $(k = 2, 3, \ldots)$. It follows that for fixed population size N, fixed $\varepsilon \in (0, 1)$ and any $t \geq 0$, $T_{N,\varepsilon}(t) \stackrel{st}{\leq} T'_{N,\varepsilon}(t)$, where $T'_{N,\varepsilon}(t)$ is the total progeny at time t of the branching process defined analagously to $\mathcal{B}_{N,\varepsilon}$ but using the empirical distribution of $D'_1, D'_2, \ldots, D'_N$, where $D'_1, D'_2, \ldots, D'_N$ are i.i.d. copies of D'. Further, $M_{D'^2}(\theta_0) < \infty$ if $M_{D^2}(\theta_0) < \infty$ and the above argument can be used to show that the sequence $\left(\mathbb{E}\left[Y_N(t)^{2+\delta}\right]\right)$ is bounded.

The above argument is easily adapted to show that in the prescribed degree case $\lim_{N\to\infty}\mathbb{E}\left[Y_N(t)\right] = \mathbb{E}\left[Z(t)\right]$ under the weaker condition that there exists $\delta > 0$ such that $\mu_{D^{2+\delta}} < \infty$ and $\mu_{(D^{(N)})^{2+\delta}} \to \mu_{D^{2+\delta}}$ as $N \to \infty$. Moreover, although we have not worked through all of the details, it seems likely that the argument can also be adapted to prove that, for any $\alpha > 1$, if there exists $\delta > 1$ such that $\mu_{D^{\alpha+\delta}} < \infty$ and $\mu_{(D^{(N)})^{\alpha+\delta}} \to \mu_{D^{\alpha+\delta}}$ as $N \to \infty$ then $\lim_{N\to\infty}\mathbb{E}\left[Y_N(t)^\alpha\right] = \mathbb{E}\left[Z(t)^\alpha\right]$ for all $t \geq 0$. Again we have not worked through all of the details but in the i.i.d. degree case it seems likely that the above condition on $M_{D^2}(\theta)$ guarantees that

$\lim_{N\to\infty} \mathbb{E}[Y_N(t)^\alpha] = \mathbb{E}[Z(t)^\alpha]$ for all α, $t \geq 0$. Finally, in the i.i.d. degree case it seems likely that weaker conditions will suffice when the limiting branching process $\mathscr{B}$ is subcritical, i.e. when $r < 0$, as in that case the exponential functions appearing in $\mathbb{E}[T_{N,\varepsilon}(t)^{2+\delta}]$, prior to taking expectations, will all have negative arguments.

Appendix 2: Eigenvalues of Ω

Let $A = [a_{l,k}] = \Omega - \lambda I$. Observe that $a_{0,0} = -(\lambda + \gamma)$, $a_{0,k} = 0$ for $k = 1, 2, \ldots, k_{\max}$, $a_{l,k_{\max}} = 0$ for $l = 1, 2, \ldots, k_{\max} - 1$ and $a_{k_{\max},k_{\max}} = -(k_{\max}\tau + \lambda + \gamma)$. Thus, expanding the determinant $|A|$ along the 0th row and then the cofactor $A_{0,0}$ down the last column yields

$$|A| = (\lambda + \gamma)(k_{\max}\tau + \lambda + \gamma)|B|, \qquad (10.14)$$

where

$$B = \begin{bmatrix} \tau(\tilde{p}_2 - 1) - \lambda - \gamma & \tau\tilde{p}_3 & \cdots & \tau\tilde{p}_{k_{\max}} \\ 2\tau(\tilde{p}_2 + 1) & 2\tau(\tilde{p}_3 - 1) - \lambda - \gamma & \cdots & 2\tau\tilde{p}_{k_{\max}} \\ \vdots & \vdots & \ddots & \vdots \\ (k_{\max} - 1)\tau\tilde{p}_2 & (k_{\max} - 1)\tau\tilde{p}_3 & \cdots & (k_{\max} - 1)\tau(\tilde{p}_{k_{\max}} - 1) - \lambda - \gamma \end{bmatrix}.$$

More precisely, B is the $(k_{\max} - 1) \times (k_{\max} - 1)$ matrix with elements

$$b_{l,k} = \tau l \left(\tilde{p}_{k+1} + \delta_{l,k+1}\right) - (\gamma + \tau l + \lambda)\delta_{l,k} \qquad (l, k = 1, 2, \ldots, k_{\max} - 1).$$

Subtracting $l\times$ the first row of B from the lth row of B, for $l = 2, 3, \ldots, k_{\max} - 1$, now gives $|B| = |C|$, where

$$C = \begin{bmatrix} \tau(\tilde{p}_2 - 1) - \lambda - \gamma & \tau\tilde{p}_3 & \cdots & \tau\tilde{p}_{k_{\max}} \\ 2(2\tau + \lambda + \gamma) & -2\tau - \lambda - \gamma & \cdots & 0 \\ 3(\tau + \lambda + \gamma) & 3\tau & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ (k_{\max} - 1)(\tau + \lambda + \gamma) & 0 & \cdots & -(k_{\max} - 1)\tau - \lambda - \gamma \end{bmatrix}$$

has elements given by

$$c_{1,k} = \tau\tilde{p}_{k+1} - (\gamma + \tau + \lambda)\delta_{1,k} \qquad (k = 1, 2, \ldots, k_{\max} - 1),$$
$$c_{l,k} = \tau l\delta_{l,k+1} - (\gamma + \tau l + \lambda)\delta_{l,k} + l(\gamma + \tau + \lambda)\delta_{1,k}$$
$$(l = 2, 3, \ldots, k_{\max} - 1; k = 1, 2, \ldots, k_{\max} - 1).$$

In particular, $c_{1,1} = \tau(\tilde{p}_2 - 1) - \gamma - \lambda$, $c_{1,k} = \tau\tilde{p}_{k+1}$ for $k = 2, 3, \ldots, k_{\max} - 1$, and $c_{l,k} = 0$ for $2 \le l < k \le k_{\max} - 1$. Thus, adding $k\times$ the kth column of C to the first column of C, for $k = 2, 3, \ldots, k_{\max} - 1$, yields $|C| = |D|$, where

$$D = \begin{bmatrix} \tau\left(\left(\sum_{l=0}^{k_{\max}} l\tilde{p}_{l+1}\right) - 1\right) - \lambda - \gamma & \tau\tilde{p}_3 & \cdots & \tau\tilde{p}_{k_{\max}} \\ 0 & -2\tau - \lambda - \gamma & \cdots & 0 \\ 0 & 3\tau & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & -(k_{\max} - 1)\tau - \lambda - \gamma \end{bmatrix}.$$

Note that $d_{1,1} = \tau\left[\left(\sum_{l=0}^{k_{\max}} l\tilde{p}_{l+1}\right) - 1\right] - \gamma - \lambda$ and, for $l \ge 2$, that $d_{l,1} = 0$, $d_{l,l} = -(\gamma + \tau + \lambda)$ and $d_{l,k} = 0$ for $k > l$. Thus expanding $|D|$ down the first column gives

$$|D| = \left\{\tau\left[\left(\sum_{l=0}^{k_{\max}} l\tilde{p}_{l+1}\right) - 1\right] - \gamma - \lambda\right\}(-1)^{k_{\max} - 2} \prod_{l=2}^{k_{\max} - 1} (l\tau + \gamma + \lambda).$$

Recalling (10.14) and $|B| = |C| = |D|$, it follows that the eigenvalues of $\mathbf{\Omega}$ are given by (3.3).

Appendix 3: Derivation of variance

Recall the definition of C_k at (4.2). Note, using (2.4), that for $k \in \mathcal{K}$,

$$\mathbf{1}^\top C_k \mathbf{1} = \tau k + \gamma,$$
$$\mathbf{1}^\top C_k \mathbf{n} = \mathbf{n}^\top C_k \mathbf{1} = (r + 2\gamma)k,$$
$$\mathbf{n}^\top C_k \mathbf{n} = \tau\mu_{(\tilde{D}-2)^2}k + \gamma k^2. \tag{10.15}$$

Now let $\mathbf{c}_{11}$ be a column vector whose kth element is $\mathbf{1}^\top C_k \mathbf{1}$, and define $\mathbf{c}_{1n}$ and $\mathbf{c}_{nn}$ similarly, using $\mathbf{1}^\top C_k \mathbf{n}$ and $\mathbf{n}^\top C_k \mathbf{n}$, respectively. Noting that $r + 2\gamma = \gamma + \tau\mu_{\tilde{D}-2}$, in this more compact notation (10.15) becomes

$$\mathbf{c}_{11} = \tau\mathbf{n} + \gamma\mathbf{1},$$
$$\mathbf{c}_{1n} = \left(\gamma + \tau\mu_{\tilde{D}-2}\right)\mathbf{n},$$
$$\mathbf{c}_{nn} = \tau\mu_{(\tilde{D}-2)^2}\mathbf{n} + \gamma\mathbf{n}_2, \tag{10.16}$$

and (4.3) yields

$$\mathbf{v}(t) = \int_0^t \left(\mu_{\tilde{D}-2}^{-1}\left(e^{r(t-u)} - e^{-\gamma(t-u)}\right)\right)^2 e^{\mathbf{\Omega}u}\mathbf{c}_{nn} \, du$$
$$+ 2\int_0^t \mu_{\tilde{D}-2}^{-1}\left(e^{r(t-u)} - e^{-\gamma(t-u)}\right)e^{-\gamma(t-u)}e^{\mathbf{\Omega}u}\mathbf{c}_{1n} \, du$$
$$+ \int_0^t e^{-2\gamma(t-u)}e^{\mathbf{\Omega}u}\mathbf{c}_{11} \, du. \tag{10.17}$$

Now

$$\boldsymbol{\Omega}\mathbf{n}_2 = \tau\mu_{(\tilde{D}-1)^2+1}\mathbf{n} - (2\tau + \gamma)\mathbf{n}_2, \tag{10.18}$$

so, using Proposition 1, with $\boldsymbol{M} = \boldsymbol{\Omega}, \mathbf{x} = \mathbf{n}_2, \mathbf{y} = \mathbf{n}, a = -(2\tau + \gamma), b = \tau\mu_{(\tilde{D}-1)^2+1}$ and $c = r$, and recalling from (3.4) that $r + \gamma = \mu_{\tilde{D}-2}\tau$ so $a - c = -\tau(2 + \mu_{\tilde{D}-2}) = -\tau\mu_{\tilde{D}}$, we have

$$e^{\boldsymbol{\Omega}u}\mathbf{n}_2 = \mu_{\tilde{D}}^{-1}\mu_{(\tilde{D}-1)^2+1}\left(e^{ru} - e^{-(2\tau+\gamma)u}\right)\mathbf{n} + e^{-(2\tau+\gamma)u}\mathbf{n}_2. \tag{10.19}$$

Hence, using also (3.14),

$$
\begin{aligned}
e^{\boldsymbol{\Omega}u}\mathbf{c}_{11} &= \mu_{\tilde{D}-2}^{-1}\left(\gamma + \tau\mu_{\tilde{D}-2}\right)e^{ru}\mathbf{n} + \gamma e^{-\gamma u}\left(1 - \mu_{\tilde{D}-2}^{-1}\mathbf{n}\right), \\
e^{\boldsymbol{\Omega}u}\mathbf{c}_{1n} &= \left(\gamma + \tau\mu_{\tilde{D}-2}\right)e^{ru}\mathbf{n}, \\
e^{\boldsymbol{\Omega}u}\mathbf{c}_{nn} &= \left(\tau\mu_{(\tilde{D}-2)^2} + \gamma\mu_{\tilde{D}}^{-1}\mu_{(\tilde{D}-1)^2+1}\right)e^{ru}\mathbf{n} + \gamma e^{-(2\tau+\gamma)u}\left(\mathbf{n}_2 - \mu_{\tilde{D}}^{-1}\mu_{(\tilde{D}-1)^2+1}\mathbf{n}\right).
\end{aligned}
\tag{10.20}
$$

Let $I_1(t) = \int_0^t e^{-2\gamma(t-u)}e^{ru} \, du$, $I_2(t) = \int_0^t e^{-2\gamma(t-u)}e^{-\gamma u} \, du$, $I_3(t) = \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)}\right)e^{-\gamma(t-u)}e^{ru} \, du$, $I_4(t) = \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)}\right)^2 e^{ru} \, du$ and $I_5(t) = \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)}\right)^2 e^{-(2\tau+\gamma)u} \, du$. It is easily verified that these integrals are given by (4.6). Substituting (10.20) into (10.17) and using (4.6) yields (4.4).

Appendix 4: Derivation of covariance function

For $t \geq 0$, let $\mathbf{u}(t)$ be the column vector whose kth element is $\mathbf{1}^\top \boldsymbol{V}^{(k)}(t)\mathbf{n}$. Arguing as in the derivation of (10.17) yields

$$
\begin{aligned}
\mathbf{u}(t) = &\int_0^t \mu_{\tilde{D}-2}^{-1}\left(e^{r(t-u)} - e^{-\gamma(t-u)}\right)e^{r(t-u)}e^{\boldsymbol{\Omega}u}\mathbf{c}_{nn} \, du \\
&+ \int_0^t e^{-\gamma(t-u)}e^{r(t-u)}e^{\boldsymbol{\Omega}u}\mathbf{c}_{1n} \, du.
\end{aligned}
\tag{10.21}
$$

Using (10.20) now gives

$$\mathbf{u}(t) = \beta_1(t)\mathbf{n} + \beta_2(t)\mathbf{n}_2, \tag{10.22}$$

where

$$
\begin{aligned}
\beta_1(t) &= \gamma\mu_{\tilde{D}-2}^{-1}\left[\mu_{\tilde{D}}^{-1}\mu_{(\tilde{D}-1)^2+1}(I_7(t) - I_8(t)) + \mu_{\tilde{D}-2}I_6(t)\right] \\
&\quad + \tau\mu_{\tilde{D}-2}^{-1}\left[\mu_{(\tilde{D}-2)^2}I_7(t) + \mu_{\tilde{D}-2}^2 I_6(t)\right], \\
\beta_2(t) &= \gamma\mu_{\tilde{D}-2}^{-1}I_8(t),
\end{aligned}
\tag{10.23}
$$

with $I_6(t) = \int_0^t e^{-\gamma(t-u)} e^{r(t-u)} e^{ru} \, du$, $I_7(t) = \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{r(t-u)} e^{ru} \, du$ and $I_8(t) = \int_0^t \left(e^{r(t-u)} - e^{-\gamma(t-u)} \right) e^{r(t-u)} e^{-(2\tau+\gamma)u} \, du$. It is easily verified that these integrals are given by (5.5). The expression (5.3) for the covariance follows using (5.2), (10.21), (10.22) and (4.3).

Appendix 5: Late behaviour of subcritical survival probabilities

To bound the late probabilities of survival in the subcritical case, first note that due to the inability of type-0 individuals to transmit we have

$$q_0(t) = e^{-\gamma t}. \tag{10.24}$$

Recalling that $q_k(t) = 1 - \pi_k(t)$, it follows from (7.2), with $k_{\max} = \infty$, that for $k > 0$,

$$\frac{dq_k}{dt} = -(\gamma + \tau k)q_k + \tau k q_{k-1} + \tau k (1 - q_{k-1}) \sum_{l=0}^{\infty} \tilde{p}_{l+1} q_l. \tag{10.25}$$

This leads to the equation for $k = 1$ in the form

$$\frac{dq_1}{dt} = r q_1 + h(t), \tag{10.26}$$

where

$$h(t) = \underbrace{\tau e^{-\gamma t} \left(1 + \tilde{p}_1 - \sum_{k=0}^{\infty} \tilde{p}_{l+1} q_k(t) \right)}_{h_1(t)} + \underbrace{\tau \sum_{k=1}^{\infty} \tilde{p}_{l+1} \left(q_k(t) - k q_1(t) \right)}_{h_2(t)}. \tag{10.27}$$

We assume first that $r > -\gamma$. Integrating (10.26) gives

$$e^{-rt} q_1(t) = 1 - \int_0^t e^{-rt} h(u) \, du.$$

The limit $\lim_{t \to \infty} e^{-rt} q_1(t)$ therefore exists if r is within the region of convergence of the Laplace transform of h. Considering $h_1(t)$ in (10.27), since $q_k(t) \in [0, 1]$, for all k, we have that $\tilde{p}_1 \leq h_1(t) \leq \tau(1 + \tilde{p}_1) e^{-\gamma t}$, so the Laplace transform of h_1 converges by the assumption that $r > -\gamma$.

Now considering h_2, we follow Windridge (2015) and consider an initial individual with effective degree k, and stubs labelled by integer $i = 1, 2, \ldots, k$. Let T_i be the time that the individual and its progeny through stub i exist, and let T be the lifetime of the branching process. Then, using the Bonferroni inequalities as in Windridge (2015), for $k \geq 1$,

$$\{T > t\} = \bigcup_{i=1}^{k}\{T_i > t\}, \quad q_k(t) \le kq_1(t) \quad \text{and} \quad q_k(t) \ge kq_1(t) - k^2\mathbb{P}(T_1 > t, T_2 > t).$$

(10.28)

Therefore, we have that

$$0 \le h_2(t) \le \tau\mu_{(\tilde{D}-1)^2}\mathbb{P}(T_1 > t, T_2 > t).$$

Now, from (7.3) we have that for some constant κ,

$$q_k(t) \le k\kappa e^{rt}.$$

(10.29)

Writing R for the lifetime of the initial infective individual we have that

$$\mathbb{P}(T_1 > t, T_2 > t) = \mathbb{P}(R > t) + \mathbb{P}(T_1 > t, T_2 > t, R \le t).$$

We then recall that $\mathbb{P}(R > t) = q_0(t) = e^{-\gamma t}$, after which the argument follows closely that of Windridge (2015) (as in the derivation of (10.36) below) and we find that

$$\lim_{t\to\infty}\int_0^\infty e^{-rt}\mathbb{P}(T_1 > t, T_2 > t)\mathrm{d}t < \infty.$$

(10.30)

Thus the Laplace transform of h_2 converges at r, whence $\lim_{t\to\infty} e^{-rt}q_1(t)$ is equal to a finite constant, c say. Note that c is strictly positive since $c \ge \lim_{t\to\infty} e^{-rt}\hat{q}_1(t) = \hat{c} > 0$. Further, (10.30) implies that $\lim_{t\to\infty} e^{-rt}\mathbb{P}(T_1 > t, T_2 > t) = 0$, and it follows from the two inequalities in (10.28) that, $\lim_{t\to\infty} e^{-rt}q_k(t) = kc$, for $k \ge 0$, proving (7.5).

We consider now the case when $r < -\gamma$. For $t \ge 0$ and $k = 0, 1, \ldots$, write

$$q_k(t) = e^{-\gamma t} + \tilde{q}_k(t) \quad \text{and} \quad \tilde{u}_k(t) = e^{\gamma t}\tilde{q}_k(t),$$

(10.31)

so $\tilde{q}_k(t)$ is the probability that the branching process has survived to time t but the initial individual has not, and $\tilde{u}_0(t) = 0$ for all $t \ge 0$. It follows using (10.25) that

$$\frac{\mathrm{d}\tilde{u}_1}{\mathrm{d}t} = \tau\left[-\tilde{u}_1 + (1 - e^{-\gamma t})\left(1 + \sum_{l=1}^{\infty}\tilde{p}_{l+1}\tilde{u}_l\right)\right].$$

(10.32)

The Bonferroni inequalities yield that, for $k \ge 1$,

$$k\tilde{q}_1(t) - k^2\mathbb{P}(R < t, T_1 > t, T_2 > t) \le \tilde{q}_k(t) \le k\tilde{q}_1(t),$$

(10.33)

so

$$0 \le \tilde{u}_k(t) \le k\tilde{u}_1(t),$$

(10.34)

and (10.32) implies that

$$\frac{d\tilde{u}_1}{dt} \leq \tau \left(1 + \mu_{\tilde{D}-2} \tilde{u}_1 \right),$$

whence, for all $t \geq 0$, recalling that $\mu_{\tilde{D}-2} < 0$,

$$0 \leq \tilde{u}_1(t) \leq -\frac{1}{\mu_{\tilde{D}-2}} \left(1 - e^{\tau \mu_{\tilde{D}-2} t} \right) \leq -\frac{1}{\mu_{\tilde{D}-2}}. \tag{10.35}$$

Conditioning on the lifetime of the initial individual in the branching process,

$$\mathbb{P}(R \leq t, T_1 > t, T_2 > t) = \int_{u=0}^{t} \gamma e^{-\gamma u} \left[\int_{v=0}^{u} \tau e^{-\tau v} q_1(t-v) dv \right]^2 du$$

$$\leq \int_{u=0}^{t} \gamma e^{-\gamma u} \left[\int_{v=0}^{u} \tau e^{-\tau v} \left(-\frac{\mu_{\tilde{D}-1}}{\mu_{\tilde{D}-2}} \right) e^{-\gamma(t-v)} dv \right]^2 du,$$

since (10.31) and (10.35) imply that $q_1(t) \leq -\frac{\mu_{\tilde{D}-1}}{\mu_{\tilde{D}-2}} e^{-\gamma t}$. Elementary integration then shows that there exists $c' = c'(\mu, \tau, \mu_{\tilde{D}}) < \infty$ such that, for all $t \geq 0$,

$$e^{\gamma t} \mathbb{P}(R \leq t, T_1 > t, T_2 > t) \leq \begin{cases} c' e^{-\min\{\gamma, 2\tau\} t} & \text{if } \gamma \neq 2\tau, \\ c' t e^{-\gamma t} & \text{if } \gamma = 2\tau. \end{cases} \tag{10.36}$$

The differential equation (10.32) may be written in the form

$$\frac{d\tilde{u}_1}{dt} = \tau(1 + \tilde{u}_1) - \underbrace{\tau e^{-\gamma t} \left(1 + \sum_{l=1}^{\infty} \tilde{p}_{l+1} \tilde{u}_l \right)}_{h_3(t)} - \underbrace{\tau \sum_{l=2}^{\infty} \tilde{p}_{l+1} (l\tilde{u}_1 - \tilde{u}_l)}_{h_4(t)},$$

whence

$$\tilde{u}_1(t) = -\frac{1}{\mu_{\tilde{D}-2}} \left(1 - e^{\tau \mu_{\tilde{D}-2} t} \right) - \tau e^{\tau \mu_{\tilde{D}-2} t} \int_0^t e^{-\tau \mu_{\tilde{D}-2} u} \left(h_3(u) + h_4(u) \right) du. \tag{10.37}$$

Now (10.34) and (10.35) imply that $0 \leq h_3(t) \leq -\frac{1}{\mu_{\tilde{D}-2}} e^{-\gamma t}$, whence

$$\lim_{t \to \infty} e^{\tau \mu_{\tilde{D}-2} t} \int_0^t e^{-\tau \mu_{\tilde{D}-2} u} h_3(u) du = 0. \tag{10.38}$$

Further, it follows using (10.33) and (10.36) that $0 \leq h_4(t) \leq \mu_{(\tilde{D}-1)^2} c' e^{-\min\{\gamma, 2\tau\} t}$, where $e^{-\min\{\gamma, 2\tau\} t}$ is replaced by $t e^{-\gamma t}$ if $\gamma = 2\tau$, whence (10.38) also holds when $h_3(u)$ is replaced by $h_4(u)$. Letting $t \to \infty$ in (10.37) yields (7.6).

References

Athreya KB, Ney PE (1972) Branching processes. Springer, Berlin

Bailey NTJ (1957) The mathematical theory of epidemics. Griffin, London

Ball F, Donnelly P (1995) Strong approximations for epidemic models. Stoch Process Appl 55(1):1–21

Ball F, Neal P (2008) Network epidemic models with two levels of mixing. Math Biosci 212(1):69–87

Barbour A, Reinert G (2013) Approximating the epidemic curve. Electron J Probab 18(54):1–30

Black AJ, House T, Keeling MJ, Ross JV (2014) The effect of clumped population structure on the variability of spreading dynamics. J Theor Biol 359:45–53

Bohman T, Picollelli M (2012) SIR epidemics on random graphs with a fixed degree sequence. Random Struct Algorithms 41(2):179–214

Constable GWA, McKane AJ (2014) Fast-mode elimination in stochastic metapopulation models. Phys Rev E 89(3):032141

Daley DJ, Vere-Jones D (1988) An introduction to the theory of point processes. Probability and its applications. Springer, New York

Danon L, Ford AP, House T, Jewell CP, Keeling MJ, Roberts GO, Ross JV, Vernon MC (2011) Networks and the epidemiology of infectious disease. Interdiscip Perspect Infect Dis 2011:1–28

Decreusefond L, Dhersin JS, Moyal P, Tran VC (2012) Large graph limit for an SIR process in random network with heterogeneous connectivity. Annal Appl Probab 22(2):541–575

Dorman K, Sinsheimer J, Lange K (2004) In the garden of branching processes. SIAM Rev 46(2):202–229

Durrett R (2007) Random graph dynamics. Cambridge University Press, Cambridge

Eames KTD, Keeling MJ (2002) Modeling dynamic and network heterogeneities in the spread of sexually transmitted diseases. PNAS 99(20):13330–13335

Ethier SN, Kurtz TG (1986) Markov processes: characterization and convergence. Wiley series in probability and mathematical statistics. Wiley, Hoboken

Graham M, House T (2014) Dynamics of stochastic epidemics on heterogeneous networks. J Math Biol 68(7):1583–1605

Grimmett GR, Stirzaker DR (2001) Probability and random processes, 3rd edn. Oxford University Press, Oxford

Heesterbeek H, Anderson RM, Andreasen V, Bansal S, De Angelis D, Dye C, Eames KTD, Edmunds WJ, Frost SDW, Funk S, Hollingsworth TD, House T, Isham V, Klepac P, Lessler J, Lloyd-Smith JO, Metcalf CJE, Mollison D, Pellis L, Pulliam JRC, Roberts MG, Viboud C, Isaac Newton Institute IID collaboration (2015) Modelling infectious disease dynamics in the complex landscape of global health. Science 347(6227):aaa4339

Heinzmann D (2009) Extinction times in multitype Markov branching processes. J Appl Probab 46(1):296–307

Holme P (2013) Extinction times of epidemic outbreaks in networks. PLoS ONE 8(12):e84429

House T, Keeling MJ (2010) Insights from unifying modern approximations to infections on networks. J R Soc Interface 8(54):67–73

Janson S, Luczak M, Windridge P (2014) Law of large numbers for the SIR epidemic on a random graph with given degrees. Random Struct Algorithms 45(4):726–763

Klepac P, Metcalf CJE, McLean AR, Hampson K (2013) Towards the endgame and beyond: complexities and challenges for the elimination of infectious diseases. Philos Trans R Soc Lond B Biol Sci 368(1623): 20120137

Lindquist J, Ma J, Driessche P, Willeboordse FH (2010) Effective degree network disease models. J Math Biol 62(2):143–164

Miller JC (2011) A note on a paper by Erik Volz: SIR dynamics in random networks. J Math Biol 62(3):349–358

Miller JC (2014) Epidemics on networks with large initial conditions or changing structure. PLoS ONE 9(7):e101421

Miller JC, Kiss IZ (2014) Epidemic spread in networks: existing methods and current challenges. Math Model Nat Phenom 9(2):4–42

Miller JC, Slim AC, Volz EM (2012) Edge-based compartmental modelling for infectious disease spread. J R Soc Interface 9(70):890–906

Molloy M, Reed B (1995) A critical point for random graphs with a given degree sequence. Random Struct Algorithms 6:161–179

Murray JD (1984) Asymptotic analysis, applied mathematical sciences, vol 48. Springer, New York

Murray JD (2002) Mathematical biology I, 3rd edn. Springer, Berlin

Murray JD (2003) Mathematical biology II, 3rd edn. Springer, Berlin

Newman MEJ (2002) Assortative mixing in networks. Phys Rev Lett 89(20):208701

Newman MEJ (2002) Spread of epidemic disease on networks. Phys Rev E 66(1):016128

O'Neill PD, Roberts GO (1999) Bayesian inference for partially observed stochastic epidemics. J R Stat Soc A 162:121–129

Parsons TL, Rogers T (2015) Dimension reduction via timescale separation in stochastic dynamical systems. [arXiv:1510.07031]

Ross JV, Taimre T, Pollett PK (2006) On parameter estimation in population models. Theor Popul Biol 70(4):498–510

Volz EM (2008) SIR dynamics in random networks with heterogeneous connectivity. J Math Biol 56(3):293–310

Waugh WAO (1958) Conditioned Markov processes. Biometrika 45(1–2):241–249

Windridge P (2015) The extinction time of a subcritical branching process related to the SIR epidemic on a random graph. J Appl Probab 52(4):1195–1201