An SIR epidemic model on a population with random network and household structure and several types of individuals

Frank Ball[*]   David Sirl[†]

22nd July 2010

Abstract

We consider a stochastic SIR (susceptible → infective → removed) epidemic model with several types of individuals. Infectious individuals can make infectious contacts on two levels, within their own 'household' and with their neighbours in a random graph representing additional social contacts. This random graph is an extension of the well-known configuration model to allow for several types of individuals. We give a strong approximation theorem which leads to a threshold theorem for the epidemic model and a method for calculating the probability of a major outbreak given few initial infectives. A multitype analogue of a theorem of Ball *et al.* (2009) heuristically motivates a method for calculating the expected size of such a major outbreak. We also consider vaccination and give some short numerical illustrations of our results.

Keywords: coupling; final outcome; households; local and global contacts; multitype branching process; multitype epidemic process; multitype random graph; threshold theorem.

2000 AMS Classification: Primary 92D30; Secondary 60J80, 05C80.

[*]School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom. frank.ball@nottingham.ac.uk

[†]School of Mathematical Sciences, University of Nottingham, University Park, Nottingham, NG7 2RD, United Kingdom. david.sirl@nottingham.ac.uk

1 Introduction

Stochastic epidemic models which allow for departures from the classical model of the spread of an SIR (susceptible → infective → removed) infection amongst homogeneous individuals in a homogeneously mixing population have become increasingly common in recent times in recognition of the fact that such assumptions are very rarely realistic. One way of incorporating heterogeneity into the population structures is to allow several *types* of individuals (reflecting, for example, sex or age groups) in the population; here the contact rates between pairs of individuals depend on the types of the two individuals involved (Watson 1972). A further method of increasing the heterogeneity in the population is to introduce *local* contacts of some kind, to model interaction between individuals who necessarily have relatively frequent contact, whilst maintaining homogeneously-mixing *global* contacts to model much rarer chance encounters with individuals from the population at large (Ball and Neal 2002, 2003). A common model for these local contacts is to partition the population into *households* and then allow local contacts only between individuals in the same household (Bartoszyński 1972, Becker and Dietz 1995, Ball *et al.* 1997). More recently, considerable attention has been given to analysing the spread of infection on populations with structure modelled by a random graph (Andersson 1998, Britton *et al.* 2008), including the case where the random graph has a specified degree distribution (Newman 2002, Kenah and Robins 2007). Some of these models have been combined to give, for example, the multitype households model of Ball and Lyne (2001), the model of Ball and Neal (2008) incorporating random networks and homogeneously mixing contacts and the network and households model of Ball *et al.* (2009, 2010). Other modifications and extensions of random network models are considered by, for example, Trapman (2007), Miller (2009) and Gleeson and Melnik (2009).

In this paper we introduce a new model which incorporates three ways of departing from the classical assumptions involving homogeneity. It is a network and household model with several types of individual. Most obviously it can be viewed as a generalisation of Ball *et al.* (2009, 2010) allowing for several types of individuals or a modification of Ball and Lyne (2001) in which global contacts are made through the edges of a random graph (with specified degree distributions) rather than by (type-stratified) homogeneous mixing. (Also, the special case of our model where all households are of size 1 yields a multitype version of the standard network model discussed by Newman (2002) and Kenah and Robins (2007).) This requires a multitype version of the configuration model for random graphs with specified degree distribution, which has some resemblance to the bipartite configuration model

of Newman (2002, Section V). Our main result is a strong approximation theorem establishing almost sure convergence of the 'epidemic process' to a branching process as the population size tends to infinity in a specified way. This provides a rigorous basis for a threshold parameter and the calculation of the probability that an epidemic initiated by few initial infectives leads to a major outbreak. This result complements the results of Ball *et al.* (2009) for the single-type version of our model, where the authors establish weaker (i.e. convergence in distribution) approximations to both the epidemic and an associated process which gives a rigorous basis for calculating the expected proportion of the population infected by a major outbreak. We assume that a multitype analogue of this result holds and give details of how this allows the expected relative final size (i.e. the expected proportion of initial susceptibles ultimately infected) of a major outbreak to be computed. We expect that the single-type argument of Ball *et al.* (2009) extends to the present multitype situation but that the proof will be long and detailed, without containing new ideas, so we do not attempt it here. We also consider vaccination in our model; we verify an intuitively, but not mathematically, obvious result showing that a wide class of vaccinnation models can be accommodated by simply introducing more types of individuals having appropriate degree distributions and contact rates.

The rest of the paper is organised as follows. In Section 2 we describe our model and the asymptotic regime under which we analyse it. Then in Section 3 we outline the branching process approximation we use and state the strong approximation theorem before describing how to calculate the main quantities of interest associated with our model. In Section 4 we consider the analysis of our model when vaccination is added to the basic model. In Section 5 we give some brief numerical results illustrating our results and then in Section 6 we give proofs concerning the properties of the random network and also of the strong approximation theorem. Lastly we make some concluding comments in Section 7.

2 Model

We assume that the population of interest contains J types of individuals, labelled $1, 2, \ldots, J$. Let

$$\mathcal{N} = \{\boldsymbol{n} = (n_1, n_2, \ldots, n_J) \in \mathbb{Z}_+^J : |\boldsymbol{n}| \geq 1\},$$

where $|\boldsymbol{n}| = \sum_{j=1}^J n_j$, and suppose that, for each $\boldsymbol{n} \in \mathcal{N}$, the population contains $m_{\boldsymbol{n}}$ households of category $\boldsymbol{n}$, a household of category $\boldsymbol{n}$ being one with n_j individuals of type j (for all $j \in \mathcal{J} = \{1, 2, \ldots, J\}$). (We use $\mathbb{Z}_+$ to

denote the non-negative integers, including zero.) Denote by $m = \sum_{\boldsymbol{n} \in \mathcal{N}} m_{\boldsymbol{n}}$ the total number of households and, for $j \in \mathcal{J}$, let $N_j = \sum_{\boldsymbol{n} \in \mathcal{N}} n_j m_{\boldsymbol{n}}$ denote the total number of type-j individuals in the population. We assume that all N_j $(j \in \mathcal{J})$ are finite, so $m_{\boldsymbol{n}}$ is zero for all but finitely many $\boldsymbol{n}$.

The global structure of our population is defined by a random graph model which is a multitype version of the 'configuration model'. Here we specify not just the number of neighbours of an individual but the number of neighbours of each type. An individual's degree is thus specified by a vector $\boldsymbol{d} = (d_1, d_2, \ldots, d_J)$, where d_i is the number of type-i global neighbours it has. We can assume either that the degrees are prescribed in advance (cf. Molloy and Reed (1995)) or that they are random (cf. Newman *et al.* (2001, 2002)) and the (joint) distribution of $\boldsymbol{d}$ depends on the type of the individual concerned. For clarity of presentation, unless specified otherwise, we adopt the latter approach and assume that, for a type-j individual, this distribution is the same as that of the random variable $\boldsymbol{D}^{(j)} = (D_1^{(j)}, D_2^{(j)}, \ldots, D_J^{(j)})$, with $\mathrm{P}(\boldsymbol{D}^{(j)} = \boldsymbol{d}) = p_{\boldsymbol{d}}^{(j)}$ for all $\boldsymbol{d} \in \mathbb{Z}_+^J$. We assume that the means $\mu_i^{(j)} = \mathrm{E}[D_i^{(j)}]$ are finite for all $i, j \in \mathcal{J}$.

To construct a realisation of the network of possible global contacts we assign to each type-j individual a number of 'type $j \to i$ half-edges' $(i \in \mathcal{J})$ according to independent realisations of $\boldsymbol{D}^{(j)}$. For each $j \in \mathcal{J}$ the $j \to j$ half-edges are paired uniformly at random, ignoring the left over half-edge in the event that the there is an odd number of these edges (just as in the standard configuration model). For each pair $i, j \in \mathcal{J}$ with $i < j$ we take an $i \to j$ half-edge chosen uniformly at random and pair it with a $j \to i$ half-edge chosen uniformly at random, repeating this with the remaining half-edges until one type of half-edge is exhausted. We ignore any leftover half-edges in the event that the numbers of $i \to j$ and $j \to i$ half-edges are unequal.

In order that, when pairing up half-edges to form this network of possible global contacts, there are similar numbers of type $i \to j$ and $j \to i$ half-edges we must insist that

$$\nu_i \mu_j^{(i)} = \nu_j \mu_i^{(j)} \tag{1}$$

for all $i, j \in \mathcal{J}$, where $\nu_i = \sum_{\boldsymbol{n} \in \mathcal{N}} n_i \rho_{\boldsymbol{n}} / \sum_{\boldsymbol{n}' \in \mathcal{N}} |\boldsymbol{n}'| \rho_{\boldsymbol{n}'}$ is the proportion of individuals in the population of type i (assumed to be strictly positive for all $i \in \mathcal{J}$). It is also important to note that there may be 'imperfections' in the random graph of global contacts, in the form of self-loops and parallel edges as well as leftover half-edges. We also need to take account of the possibility of 'parallel edges' between households and global contacts between members of the same household. In Section 6.1 we show that, provided the random variables $\boldsymbol{D}^{(j)}$ $(j \in \mathcal{J})$ have finite second moments, the proportion of half-edges in the random graph that are involved in such imperfections, or are

leftover, converges in probability to 0 as $m \to \infty$.

We define the following SIR (susceptible $\to$ infective $\to$ removed) type epidemic model which evolves on this population structure. Initially all individuals are susceptible except a single initial infective, chosen from the population uniformly at random. (In Section 7 we comment on other possible initial conditions.) An infected type-i individual stays infectious for a random period of time distributed as $I^{(i)}$ (specified by its Laplace transform $\phi^{(i)}(\theta) = \mathrm{E}[\mathrm{e}^{-\theta I^{(i)}}]$, $\theta \geq 0$) before it becomes removed. During its infectious period it makes infectious contacts with any given local (within-household) neighbour of type-j at the points of a Poisson process of rate $\lambda_{ij}^{(L)}$ and with any given type-j global neighbour at the points of a Poisson process of rate $\lambda_{ij}^{(G)}$. If an individual so contacted is susceptible then it becomes infected, otherwise nothing happens. All Poisson processes and infectious periods are assumed to be mutually independent (and independent of the population structure). The epidemic continues until no infectious individual remains in the population. For convenience we write $\Lambda^{(L)} = (\lambda_{ij}^{(L)}, i, j \in \mathcal{J})$ and $\Lambda^{(G)} = (\lambda_{ij}^{(G)}, i, j \in \mathcal{J})$. We refer to this model as the multitype network and household model (NHM).

Our analysis of this model is valid as the number of households, m, goes to infinity in such a way that $m_{\boldsymbol{n}}/m \to \rho_{\boldsymbol{n}}$, for some proper distribution $\rho = (\rho_{\boldsymbol{n}}, \boldsymbol{n} \in \mathcal{N})$. The distribution ρ must satisfy $\sum_{\boldsymbol{n} \in \mathcal{N}} |\boldsymbol{n}| \rho_{\boldsymbol{n}} < \infty$ and there are also some further technical conditions on the household category and degree distributions (cf. assumptions (A1) and (A2) in Section 6.2.1) that must be satisfied. Under our assumption that the degree sequences of individuals are independent and, conditional on the individual's type, identically distributed (IID), these conditions are satisfied when either (i) the household categories are prescribed in advance and there is a maximum household size $n_{\max}$ (i.e. for all m, we have $m_{\boldsymbol{n}} = 0$ for all $\boldsymbol{n}$ with $|\boldsymbol{n}| > n_{\max}$) or (ii) the category of each household is IID with distribution given by ρ. Of course assumptions (A1) and (A2) can be satisfied with appropriate prescribed degree sequences. What matters for our purposes is that as $m \to \infty$, $p_{\boldsymbol{d}}^{(j)}$ ($j \in \mathcal{J}$, $\boldsymbol{d} \in \mathbb{Z}_+^J$) is the the asymptotic proportion of type-j individuals with degree sequence $\boldsymbol{d}$. However, note that the invariance of our conclusions to whether the household categories and degrees are prescribed or IID is not true for all final size properties of our model; for example, the asymptotic variance of the size of a major outbreak is sensitive to these assumptions. For ease of presentation we henceforth assume that both the household categories and the degree sequences are IID from the appropriate distributions.

The assumptions (A1) and (A2) are sufficient for the proof of the strong approximation result presented in this paper, and in our framework of ran-

dom household categories $\boldsymbol{H}$ and degree sequences $\boldsymbol{D}^{(j)}$ ($j \in \mathcal{J}$) they require $\mathrm{E}[\|\boldsymbol{H}\|]$ and all $\mathrm{E}[\|\boldsymbol{D}^{(j)}\|]$s to be finite. The (multitype versions of the) proofs of the final size results in Ball *et al.* (2009) require $\mathrm{E}[\|\boldsymbol{H}\|^2]$ and all $\mathrm{E}[\|\boldsymbol{D}^{(j)}\|^2]$s to be finite, though we strongly suspect (and have some numerical evidence) that these second moment assumptions can be relaxed.

3 Heuristics and main results

In this section we give heuristic arguments for the results that allow us to calculate (i) a threshold parameter for our model that determines whether or not major outbreaks, that infect a significant proportion of the population, can occur; and then (ii) the probability and (iii) expected proportion of the population infected by such outbreaks. In what follows we use the following notational conventions. For vectors of suitable lengths $\boldsymbol{x} = (x_1, x_2, \ldots, x_\ell)$ and $\boldsymbol{y} = (y_1, y_2, \ldots, y_\ell)$, we define $\boldsymbol{x}! = \prod_{i=1}^{\ell} x_i!$, $\boldsymbol{x}^{\boldsymbol{y}} = \prod_{i=1}^{\ell} x_i^{y_i}$ and the falling factorial $\boldsymbol{x}_{[\boldsymbol{y}]} = \prod_{i=1}^{\ell} x_i!/(x_i - y_i)!$. We say that $\boldsymbol{x} \leq \boldsymbol{y}$ if the inequality holds componentwise, and that $\boldsymbol{x} < \boldsymbol{y}$ if in addition at least one of the componentwise inequalities is strict. The unit vector with all entries zero except for a 1 in the i-th entry is denoted by $\boldsymbol{e}_i$ and $\boldsymbol{\infty}$ denotes a vector of suitable dimensions will all entries ∞. We also adopt the convention that summations over vector indices are rectangular, i.e. $\sum_{\boldsymbol{i}=\boldsymbol{x}}^{\boldsymbol{y}} = \sum_{i_1=x_1}^{y_1} \cdots \sum_{i_\ell=x_\ell}^{y_\ell}$. In addition, we denote the probability generating function (PGF) of the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_\ell)$ taking values in the range $\mathcal{R} \subseteq \mathbb{Z}_+^\ell$ by $f_{\boldsymbol{X}}(\boldsymbol{s}) = \sum_{\boldsymbol{x} \in \mathcal{R}} \mathrm{P}(\boldsymbol{X} = \boldsymbol{x})\boldsymbol{s}^{\boldsymbol{x}}$, $\boldsymbol{s} \in [0,1]^\ell$ and, lastly, for any $\boldsymbol{k} = (k_1, k_2, \ldots, k_\ell) \in \mathbb{Z}_+^\ell$, $f^{(\boldsymbol{k})}(\boldsymbol{s})$ denotes the mixed derivative of $f(\boldsymbol{s})$ of order k_j with respect to s_j ($j \in \{1, 2, \ldots, \ell\}$).

3.1 Early stages of an epidemic

3.1.1 Branching process approximation

We analyse the early stages of an epidemic by comparing it with a (multitype) branching process (BP) which approximates the proliferation of infected households. Because we are concerned only with the final outcome properties of our model, we can think of the epidemic evolving in the following manner (cf. Pellis *et al.* (2008)). Firstly we run the local (within-household) epidemic initiated by the initial infective, then let the individuals so infected (including the initial case) make their global contacts (through the random network). With high probability the individuals so infected are all in previously uninfected households (because of the way the network is

constructed), and in each of these households we repeat this sequence of local epidemic followed by global infections emanating from those infected in the local epidemic. We think of the household containing the initial infective as the ancestor (in the 'zero-th' generation). Then the households infected by an individual itself infected in the within-household epidemic in a household in generation n are in generation $n + 1$. In each within-household epidemic we call the initial case the primary individual and any locally infected cases secondary individuals.

It is important to note that the degree distribution of a type-j primary individual is different to that of a type-j secondary individual. The distribution of the number of susceptible neighbours of such a secondary individual is the same as that of $\boldsymbol{D}^{(j)}$; this is also true of the primary individual in generation zero, the initial case in the epidemic. However, the primary case in a subsequently infected household, supposing it is a type-j individual and was (globally) infected by a type-i individual, has the size-biased degree distribution $\tilde{\boldsymbol{D}}^{(ji)}$, where, for $\boldsymbol{d} \in \mathbb{Z}_+^J$, $\mathrm{P}(\tilde{\boldsymbol{D}}^{(ji)} = \boldsymbol{d}) = d_i p_{\boldsymbol{d}}^{(j)}/\mu_i^{(j)}$. The distribution of the number of susceptible neighbours of such a primary individual in the early stages of an epidemic is thus the same as that of $\boldsymbol{D}^{(ji)} = \tilde{\boldsymbol{D}}^{(ji)} - \boldsymbol{e}_i$. It is easy to show that the probability generating functions $f_{\boldsymbol{D}^{(ji)}}(\boldsymbol{s})$ and $f_{\boldsymbol{D}^{(j)}}(\boldsymbol{s})$ satisfy $f_{\boldsymbol{D}^{(ji)}}(\boldsymbol{s}) = f_{\boldsymbol{D}^{(j)}}^{(\boldsymbol{e}_i)}(\boldsymbol{s})/\mu_i^{(j)}$. In addition, we define $\mu_k^{(ji)} = \mathrm{E}[\boldsymbol{D}_k^{(ji)}]$. It follows easily that

$$\mu_k^{(ji)} = \frac{f_{\boldsymbol{D}^{(j)}}^{(\boldsymbol{e}_i + \boldsymbol{e}_k)}(\boldsymbol{s})|_{\boldsymbol{s}=1}}{\mu_i^{(j)}} = \begin{cases} \mathrm{E}[D_i^{(j)} D_k^{(j)}]/\mu_i^{(j)} & \text{if } i \neq k, \\ \mathrm{E}[D_i^{(j)}(D_i^{(j)} - 1)]/\mu_i^{(j)} & \text{if } i = k. \end{cases} \tag{2}$$

The dependence of the primary individual's degree distribution on both its type and also that of the individual who infected it means that the households in the branching process approximation must be typed according to both of these features. Thus, in the multitype BP of infected households there are (for the moment ignoring the first generation) J^2 types, a type-(i, i') household ($i, i' \in \mathcal{J}$) being one where the primary individual is of type i' and was (globally) infected by a type i individual. (Note that there may be redundant types in this BP; specifically, if $\lambda_{ii'}^{(G)} \mu_{i'}^{(i)} = 0$ then type-(i, i') households cannot appear in the BP. For ease of presentation we assume that there are no such redundant types, though our arguments and results are easily modified if there are such redundant types.) Denote by $_{ii'}\tilde{C}_{jj'}$ the number of type-j' individuals infected by type-j individuals that were themselves infected in the local epidemic (including the primary individual) in a typical type-(i, i') household; i.e. the number of type-(j, j') households emanating from a type-(i, i') household. The (joint) distributions of the random variables $_{ii'}\tilde{\boldsymbol{C}} = (_{ii'}\tilde{C}_{jj'}, j, j' \in \mathcal{J})$ are the offspring distributions for the sec-

ond and subsequent generations of the BP approximation. Similarly, write $_iC = (_iC_{jj'}, j, j' \in \mathcal{J})$ for the number of type-(j, j') households emanating from the household of the initial infective in the case that the initial infective is a type-i individual. The distribution of the appropriate (depending on the type of the initial infective) $_iC$ is the offspring distribution of the first generation in the BP approximation.

In addition to assuming that there are no redundant types in the branching process approximation of the epidemic we need to make further assumptions concerning the structure of the process. These assumptions are defined in terms of the $J^2 \times J^2$ mean matrix $\hat{M} = (\hat{m}_{(ii')(jj')}, (i, i'), (j, j') \in \mathcal{J}^2)$ with entries $\hat{m}_{(ii')(jj')} = \mathrm{E}[_{ii'}\tilde{C}_{jj'}]$ giving, in the second and subsequent generations of the BP, the mean number of offspring of each type (j, j') that arise from each type-(i, i') individual. The BP is called *irreducible* if for every $(i, i'), (j, j') \in \mathcal{J}^2$ the $(i, i'), (j, j')$-th entry of the matrix $\hat{M}^n$ is strictly positive for some $n \in \mathbb{Z}_+$. If such an n can be chosen independently of $(i, i'), (j, j') \in \mathcal{J}^2$ then the BP is said to be *positively regular* (otherwise it is *periodic*). At this point we also explicitly rule out the possibility that every individual in the branching process approximation gives rise to precisely one offspring; though from the epidemic modelling viewpoint this eventuality requires a quite exceptional and somewhat unrealistic combination of model parameters.

Our main result concerning this branching process approximation is a strong approximation theorem for the number of infected households in the epidemic. For an epidemic $E^{(m)}$ on m households, let $\hat{E}_{ij}^{(m)}$ be the total number of type-(i, j) households infected, not including the initially infected household, and write $\hat{\boldsymbol{E}}^{(m)} = (\hat{E}_{ij}^{(m)}, (i, j) \in \mathcal{J}^2)$. Also let $Z^{(1)}$ be the branching process described above and let $\hat{\boldsymbol{Z}}^{(1)} = (\hat{Z}_{ij}^{(1)}, (i, j) \in \mathcal{J}^2)$ be the total progeny (of each type) for $Z^{(1)}$, not including the ancestor. (This BP is denoted $Z^{(1)}$ since it is actually the first component of a 'two-type' BP $Z = (Z^{(1)}, Z^{(2)})$ defined in Section 6.2.2.) In Section 6.2 we show how these processes can be constructed on a common probability space $(\Omega, \mathcal{F}, \mathrm{P})$. Let $F_{\mathrm{EXT}} \subseteq \Omega$ be the set upon which $Z^{(1)}$ becomes extinct (i.e. all entries of $\hat{\boldsymbol{Z}}^{(1)}$ are finite).

Theorem 1. *Under mild assumptions on the household category distribution and the degree distributions,*

(i) *for* P-*almost all* $\omega \in F_{\mathrm{EXT}}$, $\hat{\boldsymbol{E}}^{(m)}(\omega) \to \hat{\boldsymbol{Z}}^{(1)}(\omega)$ *as* $m \to \infty$;

(ii) *for* P-*almost all* $\omega \in \Omega \setminus F_{\mathrm{EXT}}$, $\hat{\boldsymbol{E}}^{(m)}(\omega) \to \infty$ *as* $m \to \infty$, *assuming* $\hat{M}$ *is irreducible.*

8

Remarks. 1. The mild assumptions on the household category and degree distributions are the assumptions (A1) and (A2) referred to in the previous Section. They are specified in Section 6.2, where we also discuss their verification.

2. We refer to the approximating branching process $Z^{(1)}$ as the 'forward process' associated with a model, since it approximates the forward evolution of the epidemic (cf. the 'backward process' introduced in Section 3.2).

In light of this result it makes sense to associate non-extinction of the branching process $Z^{(1)}$ with a major outbreak in the epidemic model. Set R_* to be the dominant eigenvalue of $\hat{M}$. It follows that if $R_* \leq 1$ then, regardless of any other classification of $\hat{M}$, the probability of a major outbreak is zero (see Mode (1971, Theorems 1.7.1(i), 2.2.1 and 2.3.1.)) If $R_* > 1$ we shall calculate the probability and expected relative final size of a major outbreak, subject to $\hat{M}$ being positively regular.

3.1.2 Threshold parameter and probability of a major outbreak

We now examine in detail calculation of the threshold parameter R_* for our epidemic model, the largest eigenvalue of the $J^2 \times J^2$ mean matrix $\hat{M}$. It is straightforward to show that the entries $\hat{m}_{(ii')(jj')} = \mathrm{E}[_{ii'}\tilde{C}_{jj'}]$ $((i, i'), (j, j') \in \mathcal{J}^2)$ of the mean matrix can be written as

$$\hat{m}_{(ii')(jj')} = \sum_{\boldsymbol{n} \in \mathcal{N}} \tilde{\rho}_{\boldsymbol{n}}^{(i')} \left(\delta_{i'j} \mu_{j'}^{(i'i)} + (\mathrm{E}[\boldsymbol{T}_{\boldsymbol{n}}^{\boldsymbol{e}_{i'}}])_j \mu_{j'}^{(j)} \right) p_{jj'}^{(G)}, \tag{3}$$

where $\tilde{\rho}_{\boldsymbol{n}}^{(i)} = n_i \rho_{\boldsymbol{n}} / \sum_{\boldsymbol{n'} \in \mathcal{N}} n_i' \rho_{\boldsymbol{n'}}$ is the probability that a type-i individual chosen uniformly at random from all type-i individuals is in a household of category $\boldsymbol{n}$, $\boldsymbol{T}_{\boldsymbol{n}}^{\boldsymbol{a}}$ is the final size (excluding initial infectives) of a standard J-type SIR epidemic with $\boldsymbol{n}$ individuals, $\boldsymbol{a}$ of which are initially infective, $p_{ij}^{(G)} = 1 - \phi^{(i)}(\lambda_{ij}^{(G)})$ is the marginal probability that an infectious type i individual contacts a given type j global neighbour and δ_{ij} is the Kronecker delta (equal to 1 if $i = j$ and 0 otherwise). This expression is derived by conditioning on the category of the household the primary type-i' individual is in; then considering the expected number of type-j individuals infected in the local epidemic multiplied by the expected number of type-j' global neighbours of each such type-j individual (treating the primary and secondary individuals seperately) and finally multiplying by the probability that the type-j to type-j' infection occurs. The expectation $\mathrm{E}[\boldsymbol{T}_{\boldsymbol{n}}^{\boldsymbol{a}}]$ can be evaluated numerically

using formulae derived from Theorem 3.5 of Ball (1986) in the same way that equations (2.25) and (2.26) are derived from Theorem 2.6 of that paper.

Determining the probability of a major outbreak requires knowledge of the PGFs of the offspring distributions, which we write as

$$\tilde{G}_{ii'}(\boldsymbol{s}) = f_{\,_{ii'}\tilde{\boldsymbol{C}}}(\boldsymbol{s}) = \mathrm{E}\left[\boldsymbol{s}^{\,_{ii'}\tilde{\boldsymbol{C}}}\right] \qquad (\boldsymbol{s} \in [0,1]^{J^2},\ (i,i') \in \mathcal{J}^2) \qquad (4)$$

and

$$G_i(\boldsymbol{s}) = f_{\,_i\boldsymbol{C}}(\boldsymbol{s}) = \mathrm{E}\left[\boldsymbol{s}^{\,_i\boldsymbol{C}}\right] \qquad (\boldsymbol{s} \in [0,1]^{J^2},\ i \in \mathcal{J}). \qquad (5)$$

Deriving expressions for these PGFs in terms of the parameters of our model is made difficult by dependencies between entries in each $_{ii'}\tilde{\boldsymbol{C}}$ and $_i\boldsymbol{C}$. These dependencies arise because an individual making more (respectively fewer) global infectious contacts with one type of individual suggests that it has a longer (resp. shorter) infectious period, which in turn suggests it may have made more (resp. fewer) global infectious contacts with other types. This longer (resp. shorter) infectious period also suggests the infection of more (resp. fewer) individuals within the household and thus more (resp. fewer) global infections enanating from other individuals too. These difficulties can be overcome by using the 'final state random variable' framework of Ball and O'Neill (1999), but this is rather involved so we do not present it here. However, these dependencies do not arise in the case where the infectious period is fixed, i.e. for every $j \in \mathcal{J}$ we have $\mathrm{P}(I^{(j)} = \iota_j) = 1$ for some strictly positive constant ι_j. Further analysis of this special case is given in Section 3.3.

Assume that we can calculate these PGFs and write $\tilde{\boldsymbol{G}}(\boldsymbol{s}) = (\tilde{G}_{ii'}(\boldsymbol{s}), (i,i') \in \mathcal{J}^2)$. Then, assuming that $\hat{M}$ is positively regular, the probability that an epidemic initiated by a single individual chosen uniformly at random from type-i individuals in the population is asymptotically $p_{\mathrm{maj}}^{(i)} = 1 - G_i(\boldsymbol{\sigma})$, where $\boldsymbol{\sigma}$ is the smallest solution of $\tilde{\boldsymbol{G}}(\boldsymbol{s}) = \boldsymbol{s}$ in $[0,1]^{J^2}$ (see, for example, Mode (1971, Section 1.7.1)). If, as we have assumed, the initial individual is chosen at random from the population then the major outbreak probability is $p_{\mathrm{maj}} = \sum_{i=1}^J \nu_i p_{\mathrm{maj}}^{(i)}$. (If $R_* \leq 1$ then $\boldsymbol{\sigma} = \boldsymbol{1}$ is the only solution of $\tilde{\boldsymbol{G}}(\boldsymbol{s}) = \boldsymbol{s}$ in $[0,1]^{J^2}$ and thus $p_{\mathrm{maj}}^{(i)} \equiv 0$; of course, if $R_* > 1$ then $p_{\mathrm{maj}}^{(i)} > 0$ for all $i \in \mathcal{J}$.)

3.2 Final outcome of a major outbreak

3.2.1 Backward process and expected relative final size of a major outbreak

We now consider the expected proportion of individuals of a given type that are ultimately infected in the event of a major outbreak. An exchangability

argument shows that this expected proportion is equal to the probability that a randomly chosen individual (of the specified type) that is initially in an uninfected household is ultimately infected by a major outbreak. This probability can be determined by considering (the size of) the *susceptibility set* of the randomly chosen individual. An individual's susceptibility set is a random set of individuals in the population; it is the set of individuals that, were they to be infected, would lead to the individual of interest becoming infected. More concretely, for every individual i in the population; suppose it is infected, choose its infectious period from the appropriate infectious period distribution and determine, using independent Poisson processes with the appropriate rates, which of its (local and global) neighbours it would infect if it were to become infected. Now, construct a directed graph in which nodes represent individuals and an arc from individual i to individual j is present if and only if individual i would infect individual j if it were to become infected. The susceptibility set of individual i then comprises all individuals j from which there is a (directed) path to i in this (random directed) graph, including i itself. An individual's *local* susceptibility set, where we consider only local (within-household) contacts is also important. The notion of susceptibility sets has proved a fruitful framework within which to study the final outcome of epidemics where individuals interact in more than one way; see, for example, Ball and Lyne (2001), Ball and Neal (2002, 2008) and Ball *et al.* (2009, 2010).

The details of the connection between the size of an individual's susceptibility set and the probability it becomes infected are very involved and we do not set out to prove any such results here. A very brief outline is as follows. Run the (forward) epidemic process until $\log m$ households are infected (in a major outbreak at least this many households are infected with probability tending to 1 as $m \rightarrow \infty$). Then approximate the size (in terms of number of households intersected) of an individual's susceptibility set by a suitable branching process, say Y. It follows that, in the limit as $m \rightarrow \infty$, (i) if Y goes extinct then the susceptibility set intersects an infected household with probability 0 and (ii) if Y avoids extinction then the size of the susceptibility set is of exact order m and thus the susceptibility set intersects an infected household with probability 1.

Complete details of this argument in the case where every individual is of the same type are given in Section 6 of Ball *et al.* (2009). In particular, we assume that a multitype analogue of Corollary 6.2 of that paper holds. Under this assumption, it remains to determine (the PGFs of) the offspring distributions of the branching process which approximates the generation-wise 'growth' of an individual's susceptibility set; cf. Sections 3.2 and 4.2 of Ball *et al.* (2009).

11

The branching process approximation for the size of an individual's susceptibility set, in terms of the number of households it intersects, is similar to that for the size of an epidemic starting with a single infective. We think of the household containing the individual of interest as the ancestor (comprising the zero-th generation). Then generation $n+1$ consists of the households containing an individual that makes global infectious contact with a member of the local susceptibility set of a primary individual in a household in generation n. Similarly to before, the primary individual in a household is the one that joins the susceptibility set by making a global contact and those who join the susceptibility set by virtue of local contacts are called secondary. We refer to this approximating branching process as the 'backward process' associated with the model, since in some sense it approximates the backwards evolution of the epidemic.

The random variables we must analyse to determine the offspring distributions for the branching process are ${}_{ii'}\tilde{\boldsymbol{B}} = ({}_{ii'}\tilde{B}_{jj'},\ j,j' \in \mathcal{J})$, where ${}_{ii'}\tilde{B}_{jj'}$ is the number of type-j' individuals that make global contact with a type-j member of the local susceptibility set of a type-i' individual that itself has made global contact with a type-i individual. For the first generation we need to consider the random variables ${}_{i}\boldsymbol{B} = ({}_{i}B_{jj'},\ j,j' \in \mathcal{J})$, ${}_{i}B_{jj'}$ being the number of type-j' individuals that make global contact with a type-j member of the local susceptibility set of a type-i individual chosen uniformly at random from the population. (In either case, during the 'early growth' of the susceptibility set, the type-j' individuals that join the susceptibility set are, with probability tending to 1 as $m \to \infty$, in households previously unassociated with the susceptibility set, thus ensuring the branching property of the process.) Our aim, therefore, is to derive expressions for the PGFs

$$\tilde{H}_{ii'}(\boldsymbol{s}) = f_{{}_{ii'}\tilde{\boldsymbol{B}}}(\boldsymbol{s}) = \mathrm{E}\left[\boldsymbol{s}^{{}_{ii'}\tilde{\boldsymbol{B}}}\right] \qquad (\boldsymbol{s} \in [0,1]^{J^2},\ (i,i') \in \mathcal{J}^2)$$

and

$$H_i(\boldsymbol{s}) = f_{{}_{i}\boldsymbol{B}}(\boldsymbol{s}) = \mathrm{E}\left[\boldsymbol{s}^{{}_{i}\boldsymbol{B}}\right] \qquad (\boldsymbol{s} \in [0,1]^{J^2},\ i \in \mathcal{J}).$$

In order to derive these PGFs it is fruitful to consider first the multitype standard network model (SNM); the special case of our model where all households are of size 1 (i.e. $\rho_{\boldsymbol{n}} > 0$ only when $|\boldsymbol{n}| = 1$). Because 'households' here contain just a single individual, the quantities ${}_{ii'}\tilde{B}_{jj'}$ and ${}_{i'}B_{jj'}$ are non-zero only when $i' = j$. For $(i,i') \in \mathcal{J}^2$, write ${}_{ii'}\tilde{\boldsymbol{B}}_{i'} = ({}_{ii'}\tilde{B}_{i'j},\ j \in \mathcal{J})$ and for $i \in \mathcal{J}$, write ${}_{i}\boldsymbol{B}_i = ({}_{i}B_{ij},\ j \in \mathcal{J})$. For $\boldsymbol{s} \in [0,1]^J$, define the PGFs $\tilde{h}_{ii'}(\boldsymbol{s}) = \mathrm{E}[\boldsymbol{s}^{{}_{ii'}\tilde{\boldsymbol{B}}_{i'}}]$ (for $(i,i') \in \mathcal{J}^2$) and $h_i(\boldsymbol{s}) = \mathrm{E}[\boldsymbol{s}^{{}_{i}\boldsymbol{B}_i}]$ (for $i \in \mathcal{J}$); these specify the offspring distributions for households in generation 1 and subsequent generations, and for the initial household, respectively, in the multitype SNM.

Recall that the marginal probability that a given type-i individual makes global contact with a given type-j global neighbour is $p_{ij}^{(G)} = 1 - \phi^{(i)}(\lambda_{ij}^{(G)})$. Consider firstly the PGF $\tilde{h}_{ii'}(\boldsymbol{s})$, the joint PGF of the number of individuals of the different types that make (global) contact with a type-i' individual who we know has already made contact with a type-i (global) neighbour. Conditioning on the number of neighbours of the type-i' individual under consideration that are not already in the susceptibility set, it follows that $({}_{ii'}\tilde{B}_{i'j} \mid D_j^{(i'i)}) \sim \mathrm{Bin}(D_j^{(i'i)}, p_{ji'}^{(G)})$; the contacts being independent because (unlike in the forward process) they all originate from different individuals. We therefore have

$$\tilde{h}_{ii'}(\boldsymbol{s}) = f_{\boldsymbol{D}^{(i'i)}}(1 - p_{1i'}^{(G)}(1 - s_1), \ldots, 1 - p_{Ji'}^{(G)}(1 - s_J)).$$

Similar reasoning for the first generation shows that

$$h_i(\boldsymbol{s}) = f_{\boldsymbol{D}^{(i)}}(1 - p_{1i}^{(G)}(1 - s_1), \ldots, 1 - p_{Ji}^{(G)}(1 - s_J)).$$

Returning now to the full model including households, consider the PGF $\tilde{H}_{ii'}(\boldsymbol{s})$. First we condition on the household category of the type-i' primary individual, so that $\tilde{H}_{ii'}(\boldsymbol{s}) = \sum_{\boldsymbol{n} \in \mathcal{N}} \tilde{\rho}_{\boldsymbol{n}}^{(i')} \tilde{H}_{ii'}^{(\boldsymbol{n})}(\boldsymbol{s})$, where $\tilde{H}_{ii'}^{(\boldsymbol{n})}(\boldsymbol{s})$ is the PGF conditional on the household category being $\boldsymbol{n}$. Then, by conditioning on the size of the primary individual's local susceptibility set we find that

$$\tilde{H}_{ii'}^{(\boldsymbol{n})}(\boldsymbol{s}) = \tilde{h}_{ii'}(\boldsymbol{s}^{(i')}) f_{\boldsymbol{S}_{\boldsymbol{n}}^{\boldsymbol{e}_{i'}}}(h_1(\boldsymbol{s}^{(1)}), \ldots, h_J(\boldsymbol{s}^{(J)})),$$

where $\boldsymbol{S}_{\boldsymbol{n}}^{\boldsymbol{a}}$ is the the size of the susceptibility set of a group of $\boldsymbol{a}$ individuals in a standard multitype SIR epidemic with $\boldsymbol{n}$ individuals, not counting the $\boldsymbol{a}$ individuals whose susceptibility set is under consideration and we write $\boldsymbol{s}^{(i)} = (s_{ii'}, i' \in \mathcal{J})$, so that $\boldsymbol{s} = (\boldsymbol{s}^{(i)}, i \in \mathcal{J})$. This latter quantity can be calculated from the following multitype analogue of Lemma 3.1 of Ball (2000) (also Lemma 3.1 of Ball and Neal (2002)). We omit the proof since it follows exactly the same argument as in the single-type case.

Lemma 1. *The random variable $\boldsymbol{S}_{\boldsymbol{n}}^{\boldsymbol{a}}$ has mass function*

$$\mathrm{P}(\boldsymbol{S}_{\boldsymbol{n}}^{\boldsymbol{a}} = \boldsymbol{l}) = (\boldsymbol{n} - \boldsymbol{a})_{[\boldsymbol{l}]} \boldsymbol{q}_{\boldsymbol{a}+\boldsymbol{l}}^{\boldsymbol{n}-\boldsymbol{a}-\boldsymbol{l}} \beta_{\boldsymbol{a}\boldsymbol{l}} \qquad (\boldsymbol{0} \leq \boldsymbol{l} \leq \boldsymbol{n} - \boldsymbol{a}),$$

where $\boldsymbol{q_j} = (\phi^{(i)}(\sum_{k \in \mathcal{J}} \lambda_{ik}^{(L)} j_k), i \in \mathcal{J})$ and $(\beta_{\boldsymbol{a}\boldsymbol{j}}, \boldsymbol{0} \leq \boldsymbol{a} \leq \boldsymbol{j})$ is defined by

$$\sum_{\boldsymbol{i}=\boldsymbol{0}}^{\boldsymbol{k}} \boldsymbol{k}_{[\boldsymbol{i}]} \boldsymbol{q}_{\boldsymbol{j}+\boldsymbol{i}}^{\boldsymbol{k}-\boldsymbol{i}} \beta_{\boldsymbol{a}\boldsymbol{i}} = 1 \qquad (\boldsymbol{0} \leq \boldsymbol{k} \leq \boldsymbol{a}).$$

13

Remark. In fact $\beta_{\boldsymbol{aj}} = \mathrm{P}(\boldsymbol{S^a_{a+j}} = \boldsymbol{j})$, cf. Ball and Neal (2002, Lemma 3.1).

Turning to the first generation of the process, essentially the same calculations show that $H_i(\boldsymbol{s}) = \sum_{\boldsymbol{n} \in \mathcal{N}} \tilde{\rho}^{(i)}_{\boldsymbol{n}} H^{(\boldsymbol{n})}_i(\boldsymbol{s})$, where

$$H^{(\boldsymbol{n})}_i(\boldsymbol{s}) = h_i(\boldsymbol{s}^{(i)}) f_{\boldsymbol{S^{e_{i'}}_n}}(h_1(\boldsymbol{s}^{(1)}), \ldots, h_J(\boldsymbol{s}^{(J)})).$$

Thus, writing $\tilde{\boldsymbol{H}}(\boldsymbol{s}) = (\tilde{H}_{ii'}(\boldsymbol{s}), i, i' = 1, 2, \ldots, J)$ and assuming that the mean matrix $\tilde{M}$ associated with the second and subsequent generations is positively regular, a type-i individual not in the initially infected household is ultimately infected in a major outbreak with probability $z^{(i)} = 1 - H_i(\boldsymbol{\xi})$, where $\boldsymbol{\xi}$ is the smallest solution of $\tilde{\boldsymbol{H}}(\boldsymbol{s}) = \boldsymbol{s}$ in $[0, 1]^{J \times J}$.

3.2.2 Common criticality of the forward and backward processes

Let $\tilde{M} = (\tilde{m}_{(ii')(jj')}, (i, i'), (j, j') \in \mathcal{J}^2)$ denote the mean matrix (of the second and subsequent generations) of the backward process. Then, by similar reasoning to that in the derivation of (3), we have, for all $(i, i'), (j, j') \in \mathcal{J}^2$,

$$\tilde{m}_{(ii')(jj')} = \sum_{\boldsymbol{n} \in \mathcal{N}} \tilde{\rho}^{(i')}_{\boldsymbol{n}} \left(\delta_{i'j} \mu^{(i'i)}_{j'} + (\mathrm{E}\boldsymbol{S^{e_{i'}}_n})_j \mu^{(j)}_{j'} \right) p^{(G)}_{j'j}. \tag{6}$$

We now show that R_* is also the dominant eigenvalue of $\tilde{M}$. In our notation, equation (5.5) of Ball and Lyne (2001) is

$$(\mathrm{E}\boldsymbol{S^{e_i}_n})_j = n_j n_i^{-1} (\mathrm{E}\boldsymbol{T^{e_j}_n})_i \qquad (i, j \in \mathcal{J}, \boldsymbol{n} \in \mathcal{N}_i), \tag{7}$$

where $\mathcal{N}_i = \{\boldsymbol{n} \in \mathcal{N} : n_i > 0\}$. (The within-household epidemic of their model is exactly the same as in ours.) Also note that it follows from the first expression for $\mu^{(ji)}_k$ in (2) and the balance equation (1) that the size-biased means satisfy the further balance equation

$$\mu^{(ik)}_l = \mu^{(il)}_k \frac{\nu_l \mu^{(l)}_i}{\nu_i \mu^{(i)}_k} \qquad (i, k, l \in \mathcal{J}). \tag{8}$$

For $(i, i'), (j, j') \in \mathcal{J}^2$, define the $J^2 \times J^2$ matrices $U = (u_{(ii')(jj')})$, $V = \mathrm{diag}(v_{(jj')})$, $W = (w_{(ii')(jj')})$, $P = \mathrm{diag}(p^{(G)}_{jj'})$, $N = \mathrm{diag}(n_{(jj')})$, where

$$u_{(ii')(jj')} = \sum_{\boldsymbol{n} \in \mathcal{N}} \tilde{\rho}^{(i')}_{\boldsymbol{n}} (\mathrm{E}\boldsymbol{T^{e_{i'}}_n})_j,$$

$$v_{(jj')} = \mu^{(j)}_{j'},$$

$$w_{(ii')(jj')} = \delta_{i'j} \mu^{(i'i)}_{j'} \text{and}$$

$$n_{(jj')} = \nu_j.$$

(Here, for example, $\text{diag}(v_{(jj')})$ is the $J^2 \times J^2$ diagonal matrix with diagonal entries given by $v_{(jj')}$ $(j, j') \in \mathcal{J}^2$.) Then it follows from (3) that $\hat{M} = (UV + W)P$. Noting that, for all $i, j \in \mathcal{J}$ and $\boldsymbol{n} \in \mathcal{N}_i$, $\tilde{\rho}_{\boldsymbol{n}}^{(i)} n_j n_i^{-1} = \tilde{\rho}_{\boldsymbol{n}}^{(j)} \nu_j \nu_i^{-1}$ and $(\mathrm{E}\boldsymbol{T}_{\boldsymbol{n}}^{\boldsymbol{e}_i})_j = 0$ if $n_i = 0$ or $n_j = 0$, it follows from (1), (6), (7) and (8) that $\tilde{M}^* = DP(UV + W)D^{-1}$. Here $D = NP$ is a diagonal matrix and the superscript * denotes the combination of transposition and index re-ordering defined by $M = (m_{(ii')(jj')}) \implies M^* = (m^*_{(ii')(jj')} = m_{(j'j)(i'i)})$. (The inverse D^{-1} exists as we have assumed that all ν_js are strictly positive and our assumption that there are no redundant types in the (forward) BP implies that all $p_{jj'}^{(G)}$s are strictly positive). This, together with the fact that, for matrices A and B of suitable dimensions, AB and BA have the same set of eigenvalues, establishes that $\hat{M}$ and $\tilde{M}$ have the same set of eigenvalues. In particular, R_* (defined as the dominant eigenvalue of $\hat{M}$) is also the dominant eigenvalue of $\tilde{M}$. This implies that the forward and backward processes share the same criticality so, assuming that both $\hat{M}$ and $\tilde{M}$ are positively regular, there exists or does not exist solutions $\boldsymbol{\sigma}$ and $\boldsymbol{\xi}$, respectively, of $\tilde{\boldsymbol{G}}(\boldsymbol{s}) = \boldsymbol{s}$ and $\tilde{\boldsymbol{H}}(\boldsymbol{s}) = \boldsymbol{s}$ with all entries strictly less than 1, i.e. all of the quantities $p_{\text{maj}}^{(i)}$ and $z^{(i)}$ $(i \in \mathcal{J})$ are zero or strictly positive, together.

We note especially that this similarity relies crucially on the balance equation (1). If (1) fails to hold then the mean matrices of the forward and backward processes are not necessarily similar. In this situation both $\rho(\hat{M}) < 1 < \rho(\tilde{M})$ and $\rho(\tilde{M}) < 1 < \rho(\hat{M})$ are possible, where $\rho(A)$ denotes the dominant eigenvalue of the matrix A. Also observe that the single type case of this establishes that the forward and backward processes in Ball *et al.* (2009, 2010) have the same criticality, a fact not proved in those papers.

Though it is not the purpose of this paper to give a detailed analysis of when $\hat{M}$ and/or $\tilde{M}$ are/is positively regular, we note that the relationship between $\hat{M}$ and $\tilde{M}$ can be shown to imply, together with our assumption that D is invertible, that $\hat{M}$ and $\tilde{M}$ are positively regular, irreducible or neither together.

3.3 Fixed infectious periods

As mentioned in Section 3.1, calculating the PGFs $\tilde{G}_{ii'}(\boldsymbol{s})$ and $G_i(\boldsymbol{s})$ (defined by (4) and (5)) associated with the forward process is in general very involved but these calculations simplify appreciably if we assume that the infectious period of every individual of the same type is the same. When the infectious periods are fixed then, conditional on the final outcome of the within-household epidemic, the numbers of global infections of different classes of individual made by those infected in the local epidemic are inde-

pendent. Thus every neighbour of an infected individual is infected by that individual independently with probability depending on the types of the two individuals and whether they are local or global neighbours. This similarity to the situation in the backward process, where this independence is a result of the fact that we consider global infections made by *other* individuals which all have mutually independent infectious periods, suggests the following connection between the forward and backward processes of two closely related multitype NHMs.

It involves no loss of generality to assume that $I^{(j)} \equiv 1$ for all $j \in \mathcal{J}$. (If $I^{(j)} \equiv \iota_j$ for all $j \in \mathcal{J}$ then setting $I^{(j)} \equiv 1$, $\lambda_{ij}^{(G)} = \lambda_{ij}^{(G)} \iota_i$ and $\lambda_{ij}^{(L)} = \lambda_{ij}^{(L)} \iota_i$ for all $i, j \in \mathcal{J}$ yields a multitype NHM with the same final outcome distribution.) Now, in the spread of the epidemic, an infected type-i individual makes infectious contact with each of its type-j neighbours independently, with probability $p_{ij}^{(L)}$ or $p_{ij}^{(G)}$ as appropriate, $p_{ij}^{(k)} = 1 - \mathrm{e}^{-\lambda_{ij}^{(k)}}$ ($k \in \{L, G\}$) being the marginal probability of such a local or global contact. Similarly, in the 'spread' of a susceptibility set each type-j neighbour of a type-i individual in the susceptibility set joins the susceptibility set by contacting the type-i individual independently with probability $p_{ji}^{(L)}$ or $p_{ji}^{(G)}$ as appropriate. We have thus proved the following proposition.

Proposition 1. *Let $\mathcal{E}$ denote the multitype NHM with unit infectious periods defined by the household category distribution ρ, degree distributions $(\boldsymbol{D}^{(j)}, j \in \mathcal{J})$ and contact rate matrices $\Lambda^{(G)}$ and $\Lambda^{(L)}$. Let $\mathcal{E}^*$ denote the multitype NHM with the same parameters as $\mathcal{E}$ except with contact rate matrices $^*\Lambda^{(G)} = (^*\lambda_{ij}^{(G)} = \lambda_{ji}^{(G)})$ and $^*\Lambda^{(L)} = (^*\lambda_{ij}^{(L)} = \lambda_{ji}^{(L)})$. Then the forward process associated with $\mathcal{E}$ and the backward process associated with $\mathcal{E}^*$ have the same offspring distributions.*

Remark. It is straightforward but tedious to verify Proposition 1 by directly checking that the PGFs of their offspring distributions are the same.

Since these branching process have the same offspring distributions, if their initial distributions are the same then the whole processes are equal in distribution. If this is the case then they have the same extinction probabilities and the following extension of the well-known result applicable to many single-type models that when the infectious period is fixed p_{maj} and z are the same is immediate.

Corollary 1. *With $\mathcal{E}$ and $\mathcal{E}^*$ defined as in Proposition 1, $p_{\mathrm{maj}}^{(j)}(\mathcal{E}) = z^{(j)}(\mathcal{E}^*)$ and $p_{\mathrm{maj}}^{(j)}(\mathcal{E}^*) = z^{(j)}(\mathcal{E})$ for all $j \in \mathcal{J}$.*

4 Vaccination

We now discuss vaccination in our multitype NHM. The framework we work within allows for the possibility that different types of individuals might be more or less likely to be vaccinated and also that vaccination might have different effects on individuals of different types.

The two aspects of vaccination that need to be specified are the vaccine action model, i.e. the effect of vaccination on an individual's susceptibility and infectivity, and the vaccine allocation model that describes which members of the population are actually vaccinated. Initially we restrict our attention to a vaccine action known as *non-random*, where vaccinated individuals of the same type all have the same response to vaccination. We assume that, for each $i \in \mathcal{J}$, a vaccinated type-i individual has relative susceptibility a_i (compared to a non-vaccinated type-i individual) and, if it becomes infected, relative infectivity b_i. The rate of each Poisson process associated with infection of such a vaccinated individual is multiplied by a_i (if $a_i \leq 1$ this can be interpreted as the vaccinee 'repelling' each incoming infectious contact independently with probability $1 - a_i$) and, if infected, the rates at which it makes infectious contacts are all multiplied by b_i. Typically a_i and b_i are both in the interval $[0, 1]$, but this is not necessary (they may take values larger than 1), and for convenience we write $\boldsymbol{a} = (a_j, \ j \in \mathcal{J})$ and $\boldsymbol{b} = (b_j, \ j \in \mathcal{J})$. (If $\boldsymbol{a} = \boldsymbol{0}$ the vaccine is called *perfect*, whilst if $\boldsymbol{b} = \boldsymbol{1}$, so vaccination affects only susceptibility, it is called *leaky*.)

We allow for the vaccine allocation to be any *household category based* regime, which can be defined by $(x_{\boldsymbol{nv}}, \ \boldsymbol{n} \in \mathcal{N}, \ \boldsymbol{0} \leq \boldsymbol{v} \leq \boldsymbol{n})$, where $x_{\boldsymbol{nv}}$ is the proportion of households of category $\boldsymbol{n}$ in which $\boldsymbol{v} = (v_1, v_2, \ldots, v_J)$ individuals of the various types are vaccinated. This class of vaccine allocation regimes includes (i) choosing households uniformly at random and vaccinating all residents of a particular type or types (perhaps all types), (ii) choosing individuals of each type uniformly at random and vaccinating either just them, or possibly everyone of some or all types in their household. The issue of the existence and determination of possible optimal allocation strategies naturally arises, but we do not address this issue here. Note also that our vaccine allocation model assumes that households are chosen for vaccination independently of the degrees of their constituent individuals. In partucular, this does not allow for so-called 'acquaintance vaccination' addressed by Cohen *et al.* (2003) and Britton *et al.* (2007).

An important characteristic of any vaccine allocation regime is the *coverage* of each type of individual that it yields, that is the proportion of such individuals in the population that are vaccinated. In our framework this

coverage is given by

$$c_i = \sum_{\boldsymbol{n} \in \mathcal{N}} \tilde{\rho}_{\boldsymbol{n}}^{(i)} \sum_{\boldsymbol{v}=\boldsymbol{0}}^{\boldsymbol{n}} x_{\boldsymbol{n}\boldsymbol{v}} \frac{v_i}{n_i} \qquad (i \in \mathcal{J})$$

and the overall coverage is given by $c = \sum_{j \in \mathcal{J}} \nu_j c_j$. Note that c_i is also the probability that a type-i individual chosen uniformly at random from the population is vaccinated, which is how the above formula may be derived.

We show now that the effect of vaccination in a multitype NHM with a non-random vaccine allocated according to a strategy that depends only on household categories has the same final outcome properties as a $2J$-type NHM, where the extra types correspond to vaccinated versions of the original J types of individual. It is immediate that this approach works for, say, the standard (single- or multitype) households model in which global contacts arise from homogeneous mixing; but in our situation the random network might present a problem. In the 'true' model we construct the population structure with J types of individual and then vaccinate, yielding $2J$ types of individuals; whereas in our $2J$-type NHM we construct the population (in particular the global network) having already determined which individuals are vaccinated. It is not immediately clear that the joint distribution of the number of unvaccinated and vaccinated neighbours of a typical type-i individual is the same under both models.

The parameters of the $2J$-type NHM with the same outcomes as the J-type NHM with vaccination are as follows. Suppose that the parameters of the original J-type NHM are $\rho = (\rho_{\boldsymbol{n}})$, $(f_{\boldsymbol{D}^{(i)}}(\boldsymbol{s}), i \in \mathcal{J})$, $(\phi^{(i)}(\theta), i \in \mathcal{J})$, $\Lambda^{(L)} = (\lambda_{ij}^{(L)})$ and $\Lambda^{(G)} = (\lambda_{ij}^{(G)})$. Suppose also that the vaccine action is non-random with parameters $\boldsymbol{a} = (a_j)$ and $\boldsymbol{b} = (b_j)$ and the vaccine allocation regime is given by $(x_{\boldsymbol{n}\boldsymbol{v}})$. In the $2J$-type model, the types $1, 2, \ldots, J$ and $J+1, J+2, \ldots, 2J$ correspond, respectively, to unvaccinated and vaccinated individuals of types $1, 2, \ldots, J$ in the original model. It is clear that the household category distribution of the $2J$-type model should be

$$^V\!\rho_{(\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)})} = \rho_{\boldsymbol{n}^{(1)} + \boldsymbol{n}^{(2)}} \, x_{\boldsymbol{n}^{(1)} + \boldsymbol{n}^{(2)}, \boldsymbol{n}^{(2)}},$$

for all $\boldsymbol{n}^{(1)}, \boldsymbol{n}^{(2)} \in \mathcal{N}$. It is also clear that the infectious period distributions should be defined by $^V\!\phi^{(i)}(\theta) = {}^V\!\phi^{(i+J)}(\theta) = \phi^{(i)}(\theta)$ for all $i \in \mathcal{J}$. A little thought makes it clear that the infection rate matrices $^V\!\Lambda_L$ and $^V\!\Lambda_G$ are of

the form

$$
{}^V\Lambda = \begin{pmatrix}
\lambda_{11} & \cdots & \lambda_{1J} & \lambda_{11}a_1 & \cdots & \lambda_{1J}a_J \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
\lambda_{J1} & \cdots & \lambda_{JJ} & \lambda_{J1}a_1 & \cdots & \lambda_{JJ}a_J \\
b_1\lambda_{11} & \cdots & b_1\lambda_{1J} & b_1\lambda_{11}a_1 & \cdots & b_1\lambda_{1J}a_J \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
b_J\lambda_{J1} & \cdots & b_J\lambda_{JJ} & b_J\lambda_{J1}a_1 & \cdots & b_J\lambda_{JJ}a_J
\end{pmatrix}.
$$

A little thought suggests that the degree distributions ${}^V\boldsymbol{D}^{(i)} \overset{\mathcal{D}}{=} {}^V\boldsymbol{D}^{(i+J)}$ ($i \in \mathcal{J}$), where ${}^V\boldsymbol{D}^{(i)} = ({}^VD_1^{(i)}, {}^VD_2^{(i)}, \ldots, {}^VD_{2J}^{(i)})$, should be given by the PGF (defined for $\boldsymbol{s} \in [0,1]^{2J}$)

$$
f_{{}^V\boldsymbol{D}^{(i)}}(\boldsymbol{s}) = \mathrm{E}\left[\boldsymbol{s}^{{}^V\boldsymbol{D}^{(i)}}\right] = f_{\boldsymbol{D}^{(i)}}((1-c_1)s_1 + c_1 s_{1+J}, \cdots, (1-c_J)s_J + c_J s_{2J}).
$$

This follows because, in a large population, in the true model half-edges emanating from a given type-i individual are paired with a vaccinated individual independently with probability c_j for a type $i \to j$ half-edge.

As mentioned above, the main thing to prove is that the joint distribution of the number of unvaccinated and vaccinated neighbours of a typical type-i individual is the same under (i) the 'true' model where the network is constructed based on J types of individual and then vaccination takes place and (ii) the computationally convenient model where individuals are vaccinated and then the network is constructed based on the resulting $2J$ types of individual. In model (ii) this distribution is part of the specification of the model and is given by $f_{{}^V\boldsymbol{D}^{(i)}}(\boldsymbol{s})$. Therefore, in model (i), for each $i \in {}^V\mathcal{J}$, write $\boldsymbol{K}^{(i)} = (K_j^{(i)}, j \in {}^V\mathcal{J})$, where $K_j^{(i)}$ is the number of type-j neighbours of a typical type-i individual and ${}^V\mathcal{J} = \{1, 2, \ldots, 2J\}$. From the construction it is clear that $\boldsymbol{K}^{(i)} \overset{\mathcal{D}}{=} \boldsymbol{K}^{(i+J)}$ for all $i \in \mathcal{J}$. First condition on the degree $\boldsymbol{D}^{(i)} = (D_j^{(i)}, j \in \mathcal{J})$ of the individual in the construction of the network and then on $\boldsymbol{N}^{(i)} = (N_j^{(i)}, j \in \mathcal{J})$, where $N_j^{(i)}$ is the number of type-j (in terms of the original network) individuals who are vaccinated. Using the facts that $(N_j^{(i)}|D_j^{(i)}) \sim \mathrm{Bin}(D_j^{(i)}, c_j)$ and that these are all independent we find that, for $\boldsymbol{s} \in [0,1]^{2J}$,

$$
f_{\boldsymbol{K}^{(i)}}(\boldsymbol{s}) = \mathrm{E}\left[\boldsymbol{s}^{\boldsymbol{K}^{(i)}}\right] = \mathrm{E}_{\boldsymbol{D}^{(i)}, \boldsymbol{N}^{(i)}}\left[\prod_{j=1}^{J} s_j^{D_j^{(i)} - N_j^{(i)}} s_{j+J}^{N_j^{(i)}}\right]
$$

$$
= \mathrm{E}_{\boldsymbol{D}^{(i)}}\left[\prod_{j=1}^{J}((1-c_j)s_j + c_j s_{j+J})^{D_j^{(i)}}\right] = f_{{}^V\boldsymbol{D}^{(i)}}(\boldsymbol{s}).
$$

We have thus shown that the necessary joint distributions are the same. The key point here is that whether or not an individual is vaccinated is independent of the degree of every individual in the population.

It is then trivial (though somewhat tedious) to verify that the $(2J)^2$-type branching process approximations of both the epidemic and the size of an individual's susceptibility set obtained using constructions (i) and (ii) have the same offspring distributions and thus that the two models for vaccination discussed above have the same final size outcomes. Checking these details proves the following result which, whilst interesting from a conceptual viewpoint, also has the significant implication that computer code that can find the properties of a general J-type NHM can simply be given different input to calculate the effects of vaccination on such a model.

Theorem 2. *The J-type NHM with vaccination and the $2J$-type NHM defined above have the same final outcomes, in the sense that the forward processes associated with the two models are identically distributed, as are the backward processes.*

Remark. Note that in the $2J$-type NHM defined referred to here there may be some types which do not occur and thus can be left out of any analysis. For example, if no type i individual (in the J-type model) is vaccinated then there is no type $i + J$ individual in the $2J$-type model. Similarly, if all type i individuals (in the J-type model) are vaccinated then there is no type i individual in the $2J$-type model. From a numerical perspective, reducing the number of types of individuals also reduces computation times appreciably.

Note that Theorem 2 can be extended to deal with a more general vaccine action model. Becker and Starczak (1998) introduced (in the context of a single-type epidemic model) a vaccine action model in which the vaccine response of an individual of type i who is vaccinated is described by a random vector $(A^{(i)}, B^{(i)})$ with a finite set of possible outcomes. The results given in this section consider the special case where there is a single outcome, i.e. $P((A^{(i)}, B^{(i)}) = (a_i, b_i)) = 1$ for all $i \in \mathcal{J}$. Another important special case that can be dealt with under this more general framework is the all-or-nothing vaccine action model, where $P((A^{(i)}, B^{(i)}) = (0, 0)) = 1 - P((A^{(i)}, B^{(i)}) = (1, 1)) = \varepsilon_i$ for each $i \in \mathcal{J}$. Thus, under this vaccine action model, a vaccinated type-i individual is rendered completely immune with probability ε_i, otherwise the vaccine has no effect.

5 Numerical examples

In this section we look briefly at some numerical results, both comparing our analytical asymptotic quantities of interest to empirical estimates of these quantities from simulations of finite populations and indicating the sort of information that can be gleaned from our vaccination results. Throughout this section we use for our calculations a 2-type NHM, the types representing say children (type 1) and adults (type 2), with the following parameters. The household category distribution is given in Table 5, it results in the proportions of individuals being $\boldsymbol{\nu} = (\frac{3}{5}, \frac{2}{5})$. We assume that the degrees of

Table 1: Distribution of household categories.

		# children			
		0	1	2	3
# adults	0	-	0	0	0
	1	0.205	0.04	0.04	0.02
	2	0.195	0.15	0.25	0.1

the individuals are $\boldsymbol{D}^{(1)} \equiv (15, 4)$ and $\boldsymbol{D}^{(2)} \equiv (6, 10)$ and that the infectious periods for both adults and children are fixed and equal to 1. Lastly, we assume that the contact rate matrices are

$$\Lambda_L = \begin{pmatrix} \frac{4}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{2}{15} \end{pmatrix} \qquad \text{and} \qquad \Lambda_G = \begin{pmatrix} \frac{2}{15} & \frac{1}{15} \\ \frac{1}{15} & \frac{1}{30} \end{pmatrix}.$$

This model has $R_* \approx 1.54$, $p_{\mathrm{maj}} \approx 0.47$ and $z \approx 0.51$.

Figure 1 shows the apparent convergence of empirical estimates of both the probability of and expected proportion of individuals infected by a major outbreak in finite populations to the asymptotic values $p_{\mathrm{maj}} = \sum_{j \in \mathcal{J}} \nu_j p_{\mathrm{maj}}^{(j)}$ and $z = \sum_{j \in \mathcal{J}} \nu_j z^{(j)}$. Even for m as small as around 500 (under 1500 individuals) the asymptotic values appear to be within 0.01 of the empirical estimates. The empirical estimates are (each) based on 10,000 simulations of the relevant finite-population model. Each simulation of our model involves generating a random population with the desired structure (both the household categories, which are independent and identically distributed, and the network of possible global contacts) and running one epidemic on it with the initial infective chosen uniformly at random from the whole population; we do *not* simulate all of the epidemic processes on a single randomly generated network.
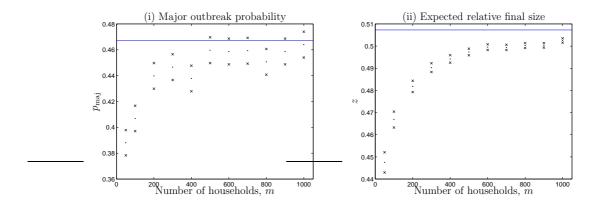
Figure 1: Simulation-based estimates of (i) the probability and (ii) the expected relative final size of a major outbreak (the dots and crosses represent point estimates ± 2 standard errors) against number of households, together with asymptotic value (solid line), for the MTNHM. Each estimate is based on 10,000 simulations.

In small populations the determination of a cut-off for whether a particular final size constitutes a major or minor outbreak can be difficult – the population has to be moderately large for the distinction to be clear. We determine this cut-off by inspecting histograms (not shown) of the relative final size for our simulations and we find that for m larger than about 200 a cut-off of 0.2 of the population size is appropriate for the parameter values we use. If the parameters are chosen so that major outbreaks are smaller then a lower cut-off is necessary and a larger number of households is needed for the distinction between minor and major outbreaks to be clear.

Note that the standard errors of the empirical estimates are appreciably smaller for z than p_{maj}. Loosely speaking, this is because in each simulation we have a single realisation of whether or not a major outbreak occurs, whereas in each simulation that gives a major outbreak we have many realisations of whether an individual is ultimately infected by a major outbreak (though they are highly correlated). For p_{maj}, estimated as $\hat{p}_{\mathrm{maj}}$, the standard error is $(\hat{p}_{\mathrm{maj}}(1 - \hat{p}_{\mathrm{maj}})/n_0)^{1/2}$, where $n_0 = 10,000$ is the number of simulations, and the standard error of z is $\hat{\sigma} n_1^{-1/2}$, where $\hat{\sigma}^2$ is the sample variance of the relative final sizes of major outbreaks and n_1 is the number of simulations that resulted in major outbreaks.

Reassuringly, these comments and observations regarding the approximation of quantities of interest associated with finite populations using our asymptotic results are much the same as those in Section 5 of Ball *et al.*

22

(2009) and Section 4.1 of Ball *et al.* (2010) concerning the single type version of our model.

To demonstrate the potential importance of the multiple type framework of this paper we now consider the affect of vaccination on this 2-type NHM. We assume that the vaccine action is non-random and given by $\boldsymbol{a} = (0.5, 0.8)$ and $\boldsymbol{b} = (0.8, 1)$; i.e. that the response in children is greater than that in adults. We consider three different vaccine allocation strategies: (i) vaccinate individuals chosen uniformly at random, (ii) vaccinate type-1 individuals uniformly at random and then, if there is sufficient vaccine to vaccinate all type-1 individuals, also vaccinate individuals of type 2 chosen uniformly at random from the population, (iii) first vaccinate type-2 individuals chosen uniformly at random and then, if there is sufficient vaccine, also vaccinate individuals of type 1. Figure 2 shows the effect that vaccinating with these different strategies can have. For the parameters we have used individuals of type 1 are both more likely to spread infection and are better protected by the vaccine, thus vaccination strategies that target these type-1 individuals are more effective in reducing the spread of infection. Note especially the very different critical vaccination coverages for the different allocation methods; the most effective strategy requires only about 45% as much vaccine as the least effective strategy to bring the model below threshold and eliminate the possibility of major outbreaks.

Of course there are many other ways to allocate limited vaccine supplies based only on the household category structure, but here we emphasise the effect of having different types of individuals in the population. Behaviour analogous to that in Ball and Lyne (2006) is observed if, for example, we explore vaccine allocation strategies involving vaccinating whole households at a time or which attempt to replicate in some way the 'equalising' strategy of Ball *et al.* (1997, Section 5.2).

6 Proofs

6.1 Network properties

In this section we give some details concerning the properties of the random multigraph that we use to model the network of possible global contacts. In particular, we concern ourselves with the numbers of self-loops and parallel edges in the graph and the number of leftover half-edges that arise if the total numbers of $i \to j$ and $j \to i$ half-edges are not equal.

The number of self-loops is easily counted in exactly the same way as in the standard configuration model. The same arguments as used in Theo-
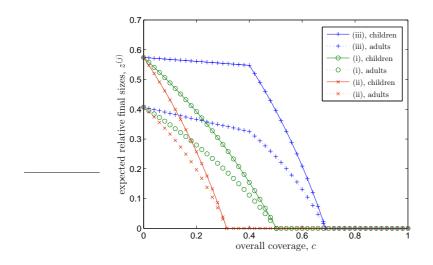
Figure 2: Effect of vaccination on the expected relative final size in the event of a major outbreak of 2-type NHM with differing 'type-dependent' vaccine allocation regimes. Solid lines correspond to the final size amongst children, dotted lines to the final size amongst adults and different markers indicate different vaccine allocation regimes

rem 3.1.2 of Durrett (2007) show that the numbers of self-loops 'on' type-j individuals are asymptotically, as $m \to \infty$, mutually independent Poisson random variables with mean $\mathrm{E}[D_j^{(jj)}(D_j^{(jj)} - 1)]/\mathrm{E}[D_j^{(jj)}] = \mu_j^{(jj)}$, so the total number of self-loops is asymptotically Poisson with mean $\sum_{j \in \mathcal{J}} \mu_j^{(jj)}$. Thus the proportion of half-edges involved in self-loops converges in probability to 0 as $m \to \infty$.

The number of parallel edges is not so easily dealt with as there are correlations between the numbers of, say, $i \to j$ and $i \to k$ half-edges of a typical type-i individual. The same method as in Durrett's book can be employed to show that the expected number of parallel edges between type-i and type-j individuals is asymptotically $\mu_i^{(ji)}\mu_j^{(ij)}/2$ for $i \neq j$ and $(\mu_j^{(jj)}/2)^2$ when $i = j$. Since the total number of half-edges is asymptotically proportional to the number of households, it then follows immediately from Markov's inequality that the proportion of half-edges involved in parallel edges converges in probability to 0 as $m \to \infty$.

The number of leftover half-edges when constructing the sub-networks of individuals of the same type is at most one, and theory from the standard configuration model implies that this leftover edge, if it exists, can be safely ignored. Results concerning the number of leftover half-edges in

the construction of the sub-networks involving individuals of different types are more complicated. The number of half-edges leftover from the pairing of type-i and type-j individuals is the absolute value of the random variable $X_{ij} = \sum_{k=1}^{N_i} D_j^{(i)}(k) - \sum_{k=1}^{N_j} D_i^{(j)}(k)$, where $D_j^{(i)}(k)$ is the j-th entry of $\boldsymbol{D}^{(i)}(k)$, itself the k-th independent copy of the random variable $\boldsymbol{D}^{(i)}$. It follows, because of the balance condition (1), that each X_{ij} has zero mean. (Of course this is the reason that (1) is imposed.) The variance of X_{ij} is easily shown to be proportional to N_i (or N_j, the ratio of N_i and N_j must be the same as that of ν_i and ν_j) and thus to m. Thus each X_{ij} has mean 0 and variance proportional to m.

Now recall that the main reason for our interest in these quantities is to get information about the probability of picking an '$i \to j$ half-edge' that is unpaired when sampling uniformly at random from all $i \to j$ half-edges. Note also that the number of $i \to j$ half-edges is, asymptotically, proportional to m. We are thus interested in a quantity which is 0 if $X_{ij} \le 0$ and proportional to X_{ij}/m if $X_{ij} > 0$. It is trivial to show that X_{ij}/m has asymptotic mean 0 and variance c_{ij}/m, where c_{ij} depends on ν_i, ν_j and the variances of $D_j^{(i)}$ and $D_i^{(j)}$. It then follows that for each pair $(i, j) \in \mathcal{J}^2$, the proportion of '$i \to j$ half-edges' that are unpaired tends to 0 in probability as $m \to \infty$. In fact it is not hard to establish that the X_{ij}'s satisfy a multivariate central limit theorem with zero means and variances proportional to $\sqrt{m}$.

6.2 Strong approximation

In Section 3.1.1, we described heuristically a branching process approximation to the proliferation of infected households in the early stages of the epidemic. Here we provide a rigorous basis for this, culminating in a proof of the strong approximation theorem, Theorem 1.

6.2.1 Notation, assumptions and preliminary results

Recall, from Section 2, the definition of $\mathcal{N}$ and a household *category* $\boldsymbol{n} = (n_1, n_2, \ldots, n_J) \in \mathcal{N}$. We also define a household *configuration*

$$(\boldsymbol{n}; \boldsymbol{d}_1^{(1)}, \boldsymbol{d}_2^{(1)}, \ldots, \boldsymbol{d}_{n_1}^{(1)}; \boldsymbol{d}_1^{(2)}, \boldsymbol{d}_2^{(2)}, \ldots, \boldsymbol{d}_{n_2}^{(2)}; \ldots; \boldsymbol{d}_1^{(J)}, \boldsymbol{d}_2^{(J)}, \ldots, \boldsymbol{d}_{n_J}^{(J)}),$$

where $\boldsymbol{d}_j^{(i)} = (d_{j1}^{(i)}, d_{j2}^{(i)}, \ldots, d_{jJ}^{(i)})$ and $d_{jk}^{(i)}$ is the number of type-k half-edges emanating from the j-th type-i individual in the household. We assume that the individuals in a category $\boldsymbol{n}$ household have local labels (i, j), where $i = 1, 2, \ldots, J$ and $j = 1, 2, \ldots, n_i$, and that $\boldsymbol{d}_j^{(i)}$ gives the degrees of the

individual with local label (i, j). Thus the ordering of the $\boldsymbol{d}_1^{(i)}, \boldsymbol{d}_2^{(i)}, \ldots, \boldsymbol{d}_{n_i}^{(i)}$ in a configuration is important.

Let $\mathcal{A}$ be the set of all household configurations. Note that $\mathcal{A}$ is countable and let $\alpha_1, \alpha_2, \ldots$ be an enumeration of $\mathcal{A}$. We consider a sequence of epidemics, $E^{(m)}$ say, where for $m = 1, 2, \ldots$, the epidemic is among a population comprising, for every $\alpha \in \mathcal{A}$, $n_\alpha^{(m)}$ households of configuration α. We assume that the total number of households in $E^{(m)}$ is m, i.e. that for all $m = 1, 2, \ldots$,

$$\sum_{\alpha \in \mathcal{A}} n_\alpha^{(m)} = m.$$

For $\alpha = (\boldsymbol{n}; \boldsymbol{d}_1^{(1)}, \boldsymbol{d}_2^{(1)}, \ldots, \boldsymbol{d}_{n_1}^{(1)}; \ldots; \boldsymbol{d}_1^{(J)}, \boldsymbol{d}_2^{(J)}, \ldots, \boldsymbol{d}_{n_J}^{(J)})$ and $(i, k) \in \mathcal{J}^2$, let

$$d_\alpha(i, k) = \sum_{j=1}^{n_i} d_{jk}^{(i)}$$

be the total type $i \to k$ degree of a household in configuration α. We assume that

(A1) for all $\alpha \in \mathcal{A}$, $m^{-1} n_\alpha^{(m)} \to p_\alpha$ as $m \to \infty$, where $\sum_{\alpha \in \mathcal{A}} p_\alpha = 1$ and $\mu_{\mathrm{HD}}(i, k) = \sum_{\alpha \in \mathcal{A}} d_\alpha(i, k) p_\alpha < \infty$ for all $(i, k) \in \mathcal{J}^2$; and

(A2) for all $(i, k) \in \mathcal{J}^2$, $m^{-1} \sum_{\alpha \in \mathcal{A}} d_\alpha(i, k) n_\alpha^{(m)} \to \mu_{\mathrm{HD}}(i, k)$ as $m \to \infty$.

If $n_\alpha^{(m)}$ ($m = 1, 2, \ldots$, $\alpha \in \mathcal{A}$) are prescribed then the convergence in (A1) and (A2) is in the usual real analysis sense. If there is an element of randomness in the household categories and/or degree configurations then the convergence in (A1) and (A2) is assumed to hold almost surely. In most models where the latter is relevant, these assumptions are a consequence of the strong law of large numbers, as is the case for the model described in Section 2, provided the means $\mathrm{E}[\|\boldsymbol{H}\|]$ and $\mathrm{E}[\|\boldsymbol{D}^{(j)}\|]$ ($j \in \mathcal{J}$) are all finite. For ease of argument and simplicity of presentation, we assume now that $n_\alpha^{(m)}$ ($m = 1, 2, \ldots$, $\alpha \in \mathcal{A}$) are prescribed.

For $\alpha \in \mathcal{A}$ and $(i, k) \in \mathcal{J}^2$, let $\tilde{p}_\alpha(i, k) = d_\alpha(i, k) p_\alpha / \mu_{\mathrm{HD}}(i, k)$ and $\tilde{p}_\alpha^{(m)}(i, k) = d_\alpha(i, k) n_\alpha^{(m)} / \sum_{\beta \in \mathcal{A}} d_\beta(i, k) n_\beta^{(m)}$. Note that, for $(i, k) \in \mathcal{J}^2$, $(\tilde{p}_\alpha(i, k), \alpha \in \mathcal{A})$ and $(\tilde{p}_\alpha^{(m)}(i, k), \alpha \in \mathcal{A})$ ($m = 1, 2, \ldots$) are proper probability distributions and (A1) and (A2) imply that $\tilde{p}_\alpha^{(m)}(i, k) \to \tilde{p}_\alpha(i, k)$ as $m \to \infty$ (for every $\alpha \in \mathcal{A}$ and $(i, k) \in \mathcal{J}^2$). Recall that $\alpha_1, \alpha_2, \ldots$ is an enumeration of $\mathcal{A}$. For $(i, k) \in \mathcal{J}^2$, let $\tilde{r}_0(i, k) = 0$ and, for $l = 1, 2, \ldots$, let $\tilde{r}_l(i, k) = \sum_{p=1}^{l} \tilde{p}_{\alpha_p}(i, k)$. For $m = 1, 2, \ldots$, $l = 0, 1, \ldots$ and $(i, k) \in \mathcal{J}^2$, define $\tilde{r}_l^{(m)}(i, k)$ similarly.

For $m = 1, 2, \ldots$, label the m households in $E^{(m)}$ as $(m, 1), (m, 2), \ldots, (m, m)$, in such a way that if $\alpha(m, j)$ denotes the configuration of household (m, j) then, for any m,

$$j_1 < j_2 \implies \alpha(m, j_1) \prec \alpha(m, j_2),$$

where, for $\alpha_k, \alpha_l \in \mathcal{A}$, $\alpha_k \prec \alpha_l$ if and only if $k \leq l$. Thus, the labelling of the households in $E^{(m)}$ preserves the ordering induced by the enumeration $\alpha_1, \alpha_2, \ldots$ of $\mathcal{A}$. For $m = 1, 2, \ldots$, let $r_0^{(m)} = 0$ and, for $l = 1, 2, \ldots$, let $r_l^{(m)} = \sum_{p=1}^{\infty} p_{\alpha_p}^{(m)}$, where, for $\alpha \in \mathcal{A}$, $p_\alpha^{(m)} = n_\alpha^{(m)}/m$. For $(i, k) \in \mathcal{J}^2$, define the piecewise-linear function $g_{(i,k)}^{(m)} : [0, 1) \to [0, 1)$ by

$$g_{(i,k)}^{(m)}(x) = r_l^{(m)} + \frac{r_{l+1}^{(m)} - r_l^{(m)}}{\tilde{r}_{l+1}^{(m)}(i, k) - \tilde{r}_l^{(m)}(i, k)} \, (x - \tilde{r}_l^{(m)}(i, k)) \tag{9}$$

$$\text{if } x \in [\tilde{r}_l^{(m)}(i, k), \tilde{r}_{l+1}^{(m)}(i, k)) \quad (l = 0, 1, \ldots).$$

Also, for $(i, k) \in \mathcal{J}^2$, define the piecewise-linear function $g_{(i,k)} : [0, 1) \to [0, 1)$ analagously to $g_{(i,k)}^{(m)}$ but with $r_l^{(m)}$ and $\tilde{r}_l^{(m)}(i, k)$ replaced by r_l and $\tilde{r}_l(i, k)$, where $r_0 = 0$ and, for $l = 1, 2, \ldots$, $r_l = \sum_{p=1}^{l} p_{\alpha_p}$.

Observe that $r_l^{(m)} \to r_l$ as $m \to \infty$ ($l \in \mathbb{Z}_+$) and, for $(i, k) \in \mathcal{J}^2$, $\tilde{r}_l^{(m)}(i, k) \to \tilde{r}_l(i, k)$ as $m \to \infty$ ($l \in \mathbb{Z}_+$). For each $(i, k) \in \mathcal{J}^2$, define the set $\tilde{R}(i, k) = \{\tilde{r}_0(i, k), \tilde{r}_1(i, k), \ldots\}$.

Lemma 2. *For $(i, k) \in \mathcal{J}^2$ and $x \in [0, 1) \setminus \tilde{R}(i, k)$, $g_{(i,k)}^{(m)}(x) \to g_{(i,k)}(x)$ as $m \to \infty$.*

Proof. Fix $(i, k) \in \mathcal{J}^2$ and $x \in [0, 1) \setminus \tilde{R}(i, k)$. Then $x \in (\tilde{r}_l(i, k), \tilde{r}_{l+1}(i, k))$ for some $l \in \mathbb{Z}_+$. Now $\tilde{r}_l^{(m)}(i, k) \to \tilde{r}_l(i, k)$ and $\tilde{r}_{l+1}^{(m)}(i, k) \to \tilde{r}_{l+1}(i, k)$ as $m \to \infty$, so $x \in (\tilde{r}_l^{(m)}(i, k), \tilde{r}_{l+1}^{(m)}(i, k))$ for all sufficiently large m. The result then follows by letting $m \to \infty$ in (9). $\qquad\square$

6.2.2 Construction of epidemic and branching processes

Let $(\Omega, \mathcal{F}, \mathrm{P})$ be a probability space on which are defined the following mutually independent random variables:

(i) for $(i, k) \in \mathcal{J}^2$, $U_0^{(1)}(i, k), U_1^{(1)}(i, k), \ldots$ and $U_1^{(2)}(i, k), U_2^{(2)}(i, k), \ldots$, all IID $U(0, 1)$ random variables;

(ii) for $(i, k) \in \mathcal{J}^2$ and $p = 1, 2, \ldots$, a sequence of random vectors $\boldsymbol{Y}_{p1}(i, k), \boldsymbol{Y}_{p2}(i, k), \ldots$ which are IID copies of the random vector $\boldsymbol{Y}_p(i, k)$ defined below;

(iii) for $p = 1, 2, \ldots$, a random vector $\boldsymbol{Y}_p^{(0)}$ defined below.

The random vector $\boldsymbol{Y}_p(i, k) = (Y_p^{(i,k)}(j, l), \ (j, l) \in \mathcal{J}^2)$ is defined as follows. Take a household of configuration α_p. If $d_{\alpha_p}(i, k) = 0$ then $\boldsymbol{Y}_p(i, k) = \boldsymbol{0}$. If $d_{\alpha_p}(i, k) > 0$ then suppose, for ease of exposition, that $\alpha_p = (\boldsymbol{n}; \boldsymbol{d}_1^{(1)}, \boldsymbol{d}_2^{(1)}, \ldots, \boldsymbol{d}_{n_1}^{(1)}; \boldsymbol{d}_1^{(2)}, \boldsymbol{d}_2^{(2)}, \ldots, \boldsymbol{d}_{n_2}^{(2)}; \ldots; \boldsymbol{d}_1^{(J)}, \boldsymbol{d}_2^{(J)}, \ldots, \boldsymbol{d}_{n_J}^{(J)})$. Choose one of the $d_{\alpha_p}(i, k)$ type $i \to k$ half-edges emanating from the household uniformly at random. Suppose that the household is infected (globally) along this half-edge, consider the ensuing local epidemic and, for $(j, l) \in \mathcal{J}^2$, let $Y_p^{(i,k)}(j, l)$ be the number of $j \to l$ half-edges emanating from this household along which the infection is subsequently transmitted. (Note that $Y_p^{(i,k)}(i, k) = 0$ if $d_{\alpha_p}(i, k) = 1$.) The random vector $\boldsymbol{Y}_p^{(0)}$ is defined similarly to $\boldsymbol{Y}_p(1, 1)$ except, as it corresponds to the initial infective in $E^{(m)}$, all $d_{\alpha_p}(1, 1)$ type $1 \to 1$ half-edges are available to make global infectious contacts.

For ease of exposition, we assume that $E^{(m)}$ is initiated by a single type-1 infective chosen by sampling uniformly from all of the $T_D^{(m)}(1, 1) = \sum_{j=1}^m d_{\alpha(m,j)}(1, 1)$ type $1 \to 1$ half-edges present in $E^{(m)}$. Thus we assume implicitly that $T_D^{(m)}(1, 1) \geq 1$. Our argument is easily extended to other assumptions about initial infectives, though the details become cumbersome to write out.

The limiting 'two-type' branching process (in fact a $2J^2$-type branching process, however it is helpful to separate the types into two groups – a full description is given later in the paragraph), $Z = (Z^{(1)}, Z^{(2)})$ (cf. discussion immediately preceding Theorem 1), is defined on $(\Omega, \mathcal{F}, \mathrm{P})$ as follows. The household configuration of the household containing the ancestor (who is of type 1) is given as follows. For $(i, k) \in \mathcal{J}^2$, let $f_{(i,k)} : [0, 1) \to \mathcal{A}$ be defined by

$$f_{(i,k)}(x) = \alpha_l \qquad \text{if } x \in [\tilde{r}_{l-1}(i, k), \tilde{r}_l(i, k)) \quad (l = 1, 2, \ldots).$$

Then the household configuration of the ancestor is $f_{(1,1)}(U_0^{(1)}(1, 1))$. Suppose that this configuration is α_{i_0}. Then the ancestor has $\boldsymbol{Y}_{i_0}^{(0)}$ type-1 offspring and $\boldsymbol{d}_{\alpha_{i_0}} - \boldsymbol{Y}_{i_0}^{(0)}$ type-2 offspring, where $\boldsymbol{d}_\alpha = (d_\alpha(i, k), \ (i, k) \in \mathcal{J}^2)$. Type-1 individuals correspond to infected households and type-2 individuals to neighbours of infected households which are not themselves infected. Type-2 individuals therefore do not have offspring. Individuals are also typed according to the nature of the global connection through which they arise from their parent individual. A type $i \to k$ individual corresponds to a household whose connection to its parent household is through a global link between a type-i individual in the parent household and a type k individual in the household of interest. We thus refer to type $(1, i \to k)$ and type

$(2, i \to k)$ individuals in the branching process, where $(i, k) \in \mathcal{J}^2$. For $l = 1, 2, \dots$ and $(i, k) \in \mathcal{J}^2$, the l-th type $(1, i \to k)$ individual born in Z has household configuration $f_{(k,i)}(U_l^{(1)}(k, i))$. Suppose that this household configuration is α_j and that it is the p-th type $(1, i \to k)$ individual having household configuration α_j born in Z. Then it has $\boldsymbol{Y}_{pl}(k, i)$ type-1 offspring and $\boldsymbol{d}_{\alpha_j} - \boldsymbol{Y}_{pl}(k, i) - \boldsymbol{\delta}(k, i)$ type-2 offspring, where $\boldsymbol{\delta}(k, i) = (\delta^{(k,i)}(j, l), (j, l) \in \mathcal{J}^2)$ and $\delta^{(k,i)}(j, l)$ is 1 if $(k, i) = (j, l)$ and 0 otherwise.

For $m = 1, 2, \dots$, a realisation of $E^{(m)}$ is constructed on $(\Omega, \mathcal{F}, \mathrm{P})$ as follows. The initial infective resides in the household labelled $(m, [mg_{(1,1)}^{(m)}(U_0^{(1)}(1, 1))] + 1)$. (For $x \in \mathbb{R}$, $[x]$ is the greatest integer $\leq x$.) Suppose that this household has configuration $\alpha_{i_0(m)}$. This household thus has $\boldsymbol{d}_{\alpha_{i_0(m)}}$ half-edges emanating from it and $\boldsymbol{Y}_{\alpha_{i_0(m)}}^{(0)}$ is used to determine, in the obvious fashion, which of these half-edges are 'live' and propogate the infection and which are 'dead' and do not. Now take these $\boldsymbol{Y}_{\alpha_{i_0(m)}}^{(0)}$ live half-edges in turn. For $(i, k) \in \mathcal{J}^2$, the l-th such $i \to k$ half-edge is paired with a $k \to i$ half-edge from the household labelled $(m, [mg_{(k,i)}(U_l^{(1)}(k, i))] + 1)$. Similarly, take the dead half-edges in turn, with the l-th such $i \to k$ half-edge being paired with a $k \to i$ half-edge from the household labelled $(m, [mg_{(k,i)}(U_l^{(2)}(k, i))] + 1)$. The epidemic $E^{(m)}$ continues in an obvious fashion until a half-edge is paired with one from a household used previously in the construction of $E^{(m)}$. If that happens then the construction needs modifying, but the detail of this modification is not important for our purposes. The key observation is that $E^{(m)}$ and Z coincide until the time such a pairing is first attempted in $E^{(m)}$.

Define the sets

$$F_1 = \{\omega \in \Omega : g_{(i,k)}(U_l^{(p)}(i, k)) \neq g_{(i',k')}(U_{l'}^{(p')}(i', k'))$$
$$\text{for all } (i, k, p, l) \neq (i', k', p', l')\}$$

and

$$F_2 = \bigcap_{(i,k) \in \mathcal{J}^2} \bigcap_{l=0}^{\infty} \bigcap_{p=1}^{2} \{\omega \in \Omega : U_l^{(p)}(i, k) \notin \tilde{R}(i, k)\}.$$

Note that, since F_1 and F_2 can be expressed as countable intersections and the maps $g_{(i,k)}$ are injective, $\mathrm{P}(F_1) = \mathrm{P}(F_2) = 1$. Let $F_{\text{EXT}} = \{\omega \in \Omega : Z(\omega) \text{ goes extinct}\}$, let $\hat{\boldsymbol{Z}} = (\hat{\boldsymbol{Z}}^{(1)}, \hat{\boldsymbol{Z}}^{(2)})$ denote the total progeny of Z, not including the initial ancestor, and let $\hat{\boldsymbol{E}}^{(m)}$ denote the total number of households infected in $E^{(m)}$, not including the initially infected household. (The elements of $\hat{\boldsymbol{Z}}^{(1)}$, $\hat{\boldsymbol{Z}}^{(2)}$ and $\hat{\boldsymbol{E}}^{(m)}$ correspond to the types $(i \to k)$ in an obvious way.) We are now ready to prove Theorem 1, as stated in Section 3.1.1.

6.2.3   Proof of Theorem 1

Let $F = F_1 \cap F_2 \cap F_{\text{EXT}}$, so that $F \subseteq F_{\text{EXT}}$ and $\mathrm{P}(F) = \mathrm{P}(F_{\text{EXT}})$. Suppose that $\omega \in F$. Then all J^2 elements of $\hat{\boldsymbol{Z}}^{(1)}$ and $\hat{\boldsymbol{Z}}^{(2)}$ are finite. Let

$$\mathcal{M} = \{(k, i, p, l) : (k, i) \in \mathcal{J}^2, p \in \{1, 2\},$$
$$l \in \{0, 1, \ldots, \hat{Z}_{ik}^{(1)}(\omega) + \hat{Z}_{ik}^{(2)}(\omega)\}\}$$

and define

$$M(\omega) = \min\{\, |g_{(k,i)}(U_l^{(p)}(\omega)) - g_{(k',i')}(U_{l'}^{(p')}(\omega))|\},$$

where the minimum is over $(k, i, p, l) \in \mathcal{M}$ and $(k', i', p', l') \in \mathcal{M}$ with $(k, i, p, l) \neq (k', i', p', l')$. Note that $M(\omega) > 0$, since $\omega \in F_1$.

Now, since $\omega \in F_2$, in view of Lemma 2, for all sufficiently large m the configurations of all households used in $E^{(m)}$ and Z coincide. The same household is attempted to be used twice in $E^{(m)}$ only if there exists $(k, i, p, l) \neq (k', i', p', l')$ such that $0 \le l \le \hat{Z}_{ik}^{(1)}(\omega) + \hat{Z}_{ik}^{(2)}(\omega)$ and $0 \le l' \le \hat{Z}_{i'k'}^{(1)}(\omega) + \hat{Z}_{i'k'}^{(2)}(\omega)$ and

$$|g_{(k,i)}^{(m)}(U_l^{(p)}(\omega)) - g_{(k',i')}^{(m)}(U_{l'}^{(p')}(\omega))| \le m^{-1},$$

and hence only if $M^{(m)}(\omega) \le m^{-1}$, where $M^{(m)}(\omega)$ is defined similarly to $M(\omega)$ but with $g_{(k,i)}$ replaced by $g_{(k,i)}^{(m)}$. Lemma 2 implies that, for all sufficiently large m, $M^{(m)}(\omega) > \frac{1}{2}M(\omega) > 0$. Thus, for all sufficiently large m, $M^{(m)}(\omega) > m^{-1}$ and $E^{(m)}$ and Z coincide, proving part (i).

To prove part (ii), suppose that $\omega \in F^* = F_1 \cap F_2 \cap (\Omega \setminus F_{\text{EXT}})$. Then $F^* \subseteq \Omega \setminus F_{\text{EXT}}$ and $\mathrm{P}(F^*) = \mathrm{P}(\Omega \setminus F_{\text{EXT}})$. Let $\boldsymbol{z} \in \mathbb{Z}_+^{J^2}$ be arbitrary. Stop the construction of Z as soon as its total 'type-1' progeny is at least $\boldsymbol{z}$ (elementwise). The above argument then shows that $\hat{\boldsymbol{E}}^{(m)}(\omega) \ge \boldsymbol{z}$ for all sufficiently large m. Part (ii) of the theorem follows as $\boldsymbol{z}$ is arbitrary.

7   Concluding remarks

We have analysed the final outcome properties of a stochastic SIR epidemic on a population sturcture featuring significant departures from traditional homogeneous mixing; individuals of several types may make local contacts with individuals in the same household and global contacts with their neighbours in a random graph with specified degree distribution. We have proved a strong approximation theorem for the epidemic process which implies a threshold theorem for the epidemic model and provides a method of calculating the probability of a major outbreak when there are few initial infectives.

We have also given a method for calculating the expected relative final size of such major outbreaks, assuming multitype analogues of the results of Ball *et al.* (2009, Section 6.5). Vaccination has also been considered, we have shown that vaccination can be included in our model by simply indroducing extra types of individual, corresponding to vaccinated versions of the original types. We have also briefly indicated that these analytic results, valid for large populations, capture the behaviour of moderately sized finite populations quite well and shown through an example that the effects of vaccinating different types of individual can be appreciably different.

There are of course many possible extensions of our model. In principle it is simple to allow the distributions of the degrees $(\boldsymbol{D}^{(j)},\ j \in \mathcal{J})$ and the local and global infection rate matrices $\Lambda^{(L)}$ and $\Lambda^{(G)}$ to depend on the household category $\boldsymbol{n}$. Although such generalisations are easily accommodated in the framework of our model, they have the potential to make any computation hugely more expensive. In the context of the final size quantities we consider in this paper it is also true that the infectious period distributions and contact rates affect the model only through the joint distributions of the total local and global infectious pressure exerted on each type of individual. Specifically, suppose $J_{ki}^{(j)}(t)$ $(j, i \in \mathcal{J},\ k \in \{L, G\})$ is a suitable random function giving the infectious pressure exerted by a typical type-j individual on each of its 'k-neighbours' of type j. Then we can specify the model in terms of contacts at the points of inhomogeneous Poisson processes with rate function $J_{ki}^{(j)}(t)$; our current specification being the special case where $J_{ki}^{(j)}(t) = \lambda_{ji}^{(k)} \mathbb{1}_{\{t \leq T^{(j)}\}}$. The final size quantities we consider depend on these random functions only through the joint Laplace-Stieltjes transforms

$$\varphi^{(j)}(\boldsymbol{\theta}^{(L)}, \boldsymbol{\theta}^{(G)}) = \mathrm{E}\left[\exp\left(-\sum_{\substack{i \in \mathcal{J} \\ k \in \{L, G\}}} \theta_i^{(k)} \int_0^\infty J_{ki}^{(j)}(t)\, \mathrm{d}t\right)\right] \qquad (j \in \mathcal{J}),$$

where, for $k \in \{L, G\}$, $\boldsymbol{\theta}^{(k)} = (\theta_i^{(k)},\ i \in \mathcal{J}) \in \mathbb{R}_+^J$. The results in this paper carry over with, for all $\boldsymbol{i}, \boldsymbol{l} \in \mathbb{Z}_+^J$, occurrences of $\phi^{(j)}(\sum_{k \in \mathcal{J}} \lambda_{jk}^{(L)} i_k + \lambda_{jk}^{(G)} l_k)$ replaced by $\varphi^{(j)}(\boldsymbol{i}, \boldsymbol{l})$; this quantity being the probability an infected type-j individual fails to make infectious contact with any of a given set of $\boldsymbol{i}$ local and $\boldsymbol{l}$ global neighbours. (These quantities can also be household category dependent.)

Another of our assumptions that can be easily modified is that there is just a single initial infective chosen uniformly at random from the population. As long as the number of initial infectives remains constant as $m \to \infty$ the threshold behaviour of forward process remains the same. Simple condi-

tioning arguments give modifications to the PGF $G_i(\boldsymbol{s})$ (equation (5)) which allows the probability of a major outbreak to be calculated if it is assumed that the initial infective resides in a household of some particular category or configuration, has a given degree, or even that there are multiple initial infectives in the same household. If multiple individuals are initially infected but they are chosen uniformly at random from the population then we can exploit the fact that the epidemics they initiate evolve independently (in the early stages).

A further line of enquiry, currently ongoing, concerns the analysis of 'acquaintance' vaccine allocation methods such as those investigated by Cohen *et al.* (2003) and Britton *et al.* (2007). Another interesting research direction is to explore the possibility of indroducing a model for the degrees of individuals in the population which moves away from the IID assumption (but of course still satisfies (A1) and (A2)), with the intention of there being fewer 'leftover' half-edges. A further challenge is the identifucation of flexible classes of multivariate degree distributions (with dependencies) which are realistic but still susceptible to analysis, which for our purposes means that evaluating their joint PGFs is feasible.

Acknowledgements

References

ANDERSSON, H. (1998). Limit theorems for a random graph epidemic model. *Ann. Appl. Probab.* **8,** 1331–1349.

BALL, F. G. (1986). A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Adv. Appl. Probab.* **18,** 289–310.

BALL, F. G. (2000). Susceptibility sets and the final outcome of stochastic SIR epidemic models. Research Report 00-09. Division of Statistics, School of Mathematical Sciences, University of Nottingham.

Ball, F. G. and Lyne, O. D. (2001). Stochastic multitype SIR epidemics among a population partitioned into households. *Adv. Appl. Probab.* **33,** 99–123.

Ball, F. G. and Lyne, O. D. (2006). Optimal vaccination schemes for epidemics among a population of households, with application to variola minor in Brazil. *Stat. Methods Med. Res.* **15,** 481–497.

Ball, F. G., Mollison, D. and Scalia-Tomba, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Probab.* **7,** 46–89.

Ball, F. G. and Neal, P. J. (2002). A general model for stochastic SIR epidemics with two levels of mixing. *Math. Biosci.* **180,** 73–102.

Ball, F. G. and Neal, P. J. (2003). The great circle epidemic model. *Stochastic Process. Appl.* **107,** 233–268.

Ball, F. G. and Neal, P. J. (2008). Network epidemic models with two levels of mixing. *Math. Biosci.* **212,** 69–87.

Ball, F. G. and O'Neill, P. D. (1999). The distribution of general final state random variables for stochastic epidemic models. *J. Appl. Probab.* **36,** 473–491.

Ball, F. G., Sirl, D. J. and Trapman, P. (2009). Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv. Appl. Probab.* **41,** 765–796.

Ball, F. G., Sirl, D. J. and Trapman, P. (2010). Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.* **224,** 53–73.

Bartoszyński, R. (1972). On a certain model of an epidemic. *Zastos. Mat.* **13,** 139–151.

Becker, N. G. and Dietz, K. (1995). The effect of household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127,** 207–219.

Becker, N. G. and Starczak, D. N. (1998). The effect of random vaccine response on the vaccination coverage required to prevent epidemics. *Math. Biosci.* **154,** 117–135.

Britton, T., Deijfen, M., Lagerås, A. N. and Lindholm, M. (2008). Epidemics on random graphs with tunable clustering. *J. Appl. Probab.* **45,** 743–756.

BRITTON, T., JANSON, S. AND MARTIN-LÖF, A. (2007). Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Adv. Appl. Probab.* **39,** 922–948.

COHEN, R., HAVLIN, S. AND BEN AVRAHAM, D. (2003). Efficient immunization strategies for computer networks and populations. *Phys. Rev. Lett.* **91,** 247901.

DURRETT, R. (2007). *Random graph dynamics.* Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge.

GLEESON, J. P. AND MELNIK, S. (2009). Analytical results for bond percolation and k-core sizes on clustered networks. *Phys. Rev. E* **80,** 046121.

KENAH, E. AND ROBINS, J. M. (2007). Second look at the spread of epidemics on networks. *Phys. Rev. E* **76,** 036113.

MILLER, J. C. (2009). Spread of infectious disease through clustered populations. *J. R. Soc. Interface* **6,** 1121–1134.

MODE, C. J. (1971). *Multitype branching processes. Theory and applications.* Modern Analytic and Computational Methods in Science and Mathematics, No. 34. American Elsevier Publishing Co., Inc., New York.

MOLLOY, M. AND REED, B. (1995). A critical point for random graphs with a given degree sequence. *Rand. Struct. Alg.* **6,** 161–179.

NEWMAN, M. E. J. (2002). Spread of epidemic disease on networks. *Phys. Rev. E* **66,** 016128.

NEWMAN, M. E. J., STROGATZ, S. H. AND WATTS, D. J. (2001). Spread of epidemic disease on networks. *Phys. Rev. E* **64,** 026118.

NEWMAN, M. E. J., WATTS, D. J. AND STROGATZ, S. H. (2002). Random graph models of social networks. *Proc. Natl. Acad. Sci. USA* **99,** 2566–2572.

PELLIS, L., FERGUSON, N. M. AND FRASER, C. (2008). The relationship between real-time and discrete-generation models of epidemic spread. *Math. Biosci.* **216,** 63–70.

TRAPMAN, P. (2007). On analytical approaches to epidemics on networks. *Theor. Pop. Biol.* **71,** 160–173.

WATSON, R. K. (1972). On an epidemic in a stratified population. *J. Appl. Probab.* **9,** 659–666.