

# An epidemic model on a network having two group structures with tunable overlap

Frank Ball\*, Tom Britton† and Peter Neal‡

October 10, 2024

## Abstract

A network epidemic model is studied. The underlying social network has two different types of group structures, households and workplaces, such that each individual belongs to exactly one household and one workplace. The random network is constructed such that a parameter  $\theta$  controls the degree of overlap between the two group structures:  $\theta = 0$  corresponding to all household members belonging to the same workplace and  $\theta = 1$  to all household members belonging to distinct workplaces. On the network a stochastic SIR epidemic is defined, having an arbitrary but specified infectious period distribution, with global (community), household and workplace infectious contacts. The stochastic epidemic model is analysed as the population size  $n \rightarrow \infty$  with the (asymptotic) probability, and size, of a major outbreak obtained. These results are proved in greater generality than existing results in the literature by allowing for any fixed  $0 \leq \theta \leq 1$ , a non-constant infectious period distribution, the presence or absence of global infection and potentially (asymptotically) infinite local outbreaks.

*MSC2020 subject classifications:* Primary 92D30, 60K35; secondary 60J80, 05C80, 91D30.

*Keywords:* SIR epidemic, Final size, Branching process, Coupling, Structured population.

## 1 Introduction

Epidemic models (as well as epidemic outbreaks) are highly affected by heterogeneities in the community in which the epidemic spreads. There has been considerable research into incorporating population structure into epidemic models. One of the simplest and most commonly studied population structures is where the population is partitioned into small mixing groups, often referred to as households (Becker and Dietz [14]). Infectious individuals make both local, within their household, and global, within the community at large, infectious contacts. The probability an infective infects a given housemate is typically an

---

\*School of Mathematical Sciences, University of Nottingham, frank.ball@nottingham.ac.uk

†Department of Mathematics, Stockholm University, tom.britton@math.su.se

‡School of Mathematical Sciences, University of Nottingham, peter.neal@nottingham.ac.uk

order of magnitude higher than the probability they infect a randomly chosen member of the population who resides in a different household. The (asymptotic) probability of a major outbreak and the size of a major outbreak as the size of the population (number of households) tends to infinity can be obtained, see Ball *et al.* [7].

The households model provides a starting point for incorporating realistic population structure into epidemic models for human populations. Human populations contain more intricate social structure which presents challenges in modelling (Ball *et al.* [4]). Households have a geographical location and the spatial location or neighbourhood to which a household belongs can be important in defining the mixing between individuals in different households, and hence the spread of a disease. Increased realism can be obtained by using a hierarchical, three-level-mixing model (Britton *et al.* [16]), in which the population is partitioned into neighbourhoods, which are partitioned into households; with different rates of transmission between an infective and those individuals they share a household with, a neighbourhood with or are a general member of the population. The focus in Britton *et al.* [16] is on inference; Ouboter *et al.* [21] analyse the asymptotic probability and size of a major outbreak for such a three-level-mixing model, with neighbourhoods corresponding to schools, when the infectious period is constant.

As well as close ties with people they live with, individuals have regular contact with individuals through work or school, and hence, increased opportunities to spread a disease with such people. This has led to models with multiple partitions of the population, for example, Andersson [1] and Ball and Neal [8]. In the case of two partitions of the population the model is often referred to as the households-workplaces model. A common assumption, to aid mathematical analysis, is that the two partitions are constructed to have no overlap (see, for example, Ball and Neal [8], Pellis *et al.* [23, 24], Kubasch [18] and Bansaye *et al.* [13]). That is, no two individuals belong to the same household and the same workplace. This contrasts with the hierarchical model where all individuals in the same household belong to the same neighbourhood. Indeed, Ouboter *et al.* [21] prove that if the sizes of households and workplaces are both constant, and the infectious period is also constant, then the asymptotic probability and size of a major outbreak are both greater in the model with no overlap than in the model with complete overlap.

We take workplace as a generic term to cover a mixing group structure outside the household, and to include schools and kindergartens. Therefore, the workplace will typically be a larger unit than the household. There are clearly situations in which substantial overlap occurs between the household and workplace. For example, it is often the case that siblings go to the same kindergarden or school, thus creating overlap between the two group structures. A second example, which is the main inspiration to the current work, is in Patwardhan *et al.* [22] where university students from a group of (major) classes are allocated into dorms. Here dorms play the role of household and classes correspond to workplaces. Patwardhan *et al.* [22] numerically analyse a closely related model and one of the questions they address is whether students should be allocated dorms according to which class they follow (large overlap - in the limit the hierarchical model) or the dorm allocation should be done more randomly (small overlap - in the limit the standard households-workplaces model) in order to minimize effects of epidemic outbreaks. A key finding is that allocation according to classes can not only significantly reduce the severity of an epidemic but it also enhances the effectiveness of vaccinating individuals, chosen uniformly at random, in reducing disease spread.

In the current paper we consider the households-workplaces model with a tunable parameter,  $\theta$ . The model is constructed by first allocating households to workplaces, so that the population structure resembles the neighbourhood model with neighbourhoods labelled by workplace. Then each individual decides independently to be a *mover* with probability  $\theta$ . The movers change their workplace membership but not the household to which they belong. The movers are then randomly allocated to workplace spaces vacated by movers. The non-movers, who we term *remainers*, remain in the workplace and household to which they were originally assigned. In this way the number of individuals in each workplace does not change. A full description of the model is provided in Section 2 with an example of a construction of a population structure given in Figure 4 in Section 5.3.1. The special cases  $\theta = 0$  and  $\theta = 1$  yield complete overlap and no overlap, respectively.

We analyse the spread of an SIR (susceptible  $\rightarrow$  infective  $\rightarrow$  recovered) epidemic, having an arbitrary but specified infectious period distribution, on a network of households and workplaces constructed as above. We assume that the sizes of households and workplaces are both constant and consider the asymptotic regime in which the total population size  $n \rightarrow \infty$ , keeping all other parameters fixed. We consider two scenarios: (a) in which there is no global infection (i.e. with a randomly chosen member of the population), so infection is spread entirely through household and workplace contacts; and (b) where there is global infection, in addition to household and workplace contacts. The papers cited above are concerned with scenario (b), though Andersson [1] also considers scenario (a). We consider epidemics initiated by a single infective, chosen uniformly from the population, and prove that under each scenario that, as  $n \rightarrow \infty$ , the probability of a major outbreak (i.e. one that infects at least  $\log n$  individuals) converges to a constant  $\rho$  and conditional upon the occurrence of a major outbreak, the fraction of the population that gets infected converges in probability to a strictly positive constant  $z$ . For each scenario, we provide recipes for calculating the constants  $\rho$ ,  $z$  and a threshold parameter that determines whether or not  $\rho > 0$ .

Whilst the special cases  $\theta = 0$  and  $\theta = 1$  have been considered in the literature, there has been no general analysis of the case  $0 < \theta < 1$ , where there is partial overlap between the two group structures. Moreover, even for the cases  $\theta = 0$  and  $\theta = 1$ , the only rigorous proofs of a law of large numbers for the size of a major outbreak assume that the infectious period is constant, making the numbers of global, household and workplace infectious contacts made by an infective independent thus simplifying the analysis. The fact that households and workplaces (may) overlap disallows analysing small independent components, so instead we study *local susceptibility sets* and *local infectious clumps* in the spirit of Ball and Neal [8]. In Ball and Neal [8] it was noted that the local infectious clumps (ignoring global infectious contacts) for the households-workplaces model, with  $\theta = 1$ , can be infinite, but attention was focussed on the simpler case where the local infectious clumps are almost surely finite. The infectious clumps have certain *complexes* as their base unit, where a complex is defined as individuals related to a given workplace: the remainers in the original construction, the movers in the original construction (who hence now belong to a different workplace) and the movers joining the workplace (and hence had a different workplace at the start). For both scenarios (a) and (b), the proof of the limiting major outbreak probability uses branching process approximations. For scenario (a), the proof of the law of large numbers for the size of a major outbreak adapts ideas from the proof of a corresponding result for epidemics on random intersection graphs

in Ball et al. [12], while for scenario (b) it uses an extension of the embedding technique of Scalia-Tomba [26, 27]. In both cases, the above-mentioned complexes are fundamental to the extensions, which are far from straightforward.

Beside different notation, our model is similar to the model in Patwardhan et al. [22], although global infectious contacts in the latter are achieved via a stochastic block model. In Patwardhan et al. [22] the list of individuals is gone through sequentially, each time the individual decides to be a mover with probability  $r$  and if so switches workplace with a randomly selected other individual. As a consequence, a given individual remains in their original household if it is not a mover *and* it is not selected to switch workplace with another mover, an event which has asymptotic probability  $(1 - r)e^{-r}$  as  $n \rightarrow \infty$ , whereas individuals remain in their original workplace with asymptotic probability  $1 - \theta$  in our model. (In Patwardhan et al. [22], a given individual may return to their workplace after making one or more switches but that happens with probability zero under the above asymptotic regime.) Patwardhan et al. [22] consider the effect of  $r$  on the time course and final size of the epidemic with particular focus on the cases  $r = 0$  ( $\theta = 0$ ) and  $r = 1$  ( $\theta = 1$ ). Two mean-field approximations for the epidemic are provided in the supplementary material of Patwardhan et al. [22]. However, both approximations perform rather poorly both for the disease dynamics and the final size, unless  $r = 1$  and the epidemic is highly infectious, see Patwardhan et al. [22] supplementary material, Figures S8 and S9. Moreover, Patwardhan et al. [22] note that a very important remark is that their findings indicated above concerning allocation of students to dorms are apparent only from their numerical simulations and do not emerge from numerical integration of their mean-field approximations. By contrast using the framework of Ball and Neal [8] and dealing with the additional mathematical technicalities of allowing (asymptotically) infinite local infectious clumps and overlapping groups, we are able to obtain exact results for the (asymptotic) final size and outbreak probability, which demonstrate clearly (see Figure 3 in Section 4) that disease spread can be increasing in  $\theta$ , i.e. decreasing in overlap.

The remainder of the paper is structured as follows. In the first two short sections (2 and 3) we define the model and state our main results. In Section 4 we illustrate our results and investigate them numerically. In Sections 5 and 6 we prove the main results for scenarios (b), global infection present, and (a), no global infection, respectively. We end with a short discussion in Section 7.

## 2 The epidemic model with two group structures and tunable overlap

### 2.1 The network model

Consider a population of size  $n$  (assumed to be large) having a network structure in which each individual belongs to exactly one household and exactly one workplace. To simplify things we assume all households have size  $h \geq 2$  and all workplaces have size  $w$ , where we make the additional assumption that  $w = dh$  for some integer  $d \geq 1$  (so workplaces are at least as large as and a multiple the size of households). For convenience we consider a population partitioned into households and also into workplaces but the

two groups structures could be other possible partitions, such as households and schools, as in Ouboter et al. [21], or dorms and classes, as in Patwardhan et al. [22].

The parameter  $\theta$  determining the amount of overlap (or rather the lack of overlap) is defined as follows. First we let each workplace consist of  $d$  households, so all individuals of the same household are also in the same workplace implying a complete overlap. Then each individual independently decides whether to be a *mover* with probability  $\theta$  or a *remainder* with the remaining probability  $1 - \theta$ . The workplace spots of movers are made vacant, and all movers are distributed uniformly at random among the vacant spots in the different workplaces, thus creating new workplace constellations but all workplaces still have size  $w = dh$ . As a consequence, the larger  $\theta$ , the fewer pairs of household members who are also work colleagues.

In summary, the parameters of the network (beside the overall population size  $n$  which must be a multiple of  $w$ ) are:  $h$ , the household size,  $w = dh$ , the workplace size, and  $\theta$  the probability of being a mover (the tunable overlap can hence be defined as  $1 - \theta$ ).

## 2.2 The epidemic model on the network

The epidemic is a stochastic continuous-time SIR model in which each infected individual is infectious for an independent random duration  $I$ : the length of the infectious period (possibly preceded by a latent period having an arbitrary distribution). During the infectious period an individual makes three types of infectious contacts at times of independent Poisson processes: at rate  $\beta_H$  the individuals make infectious contact with other members of their household, each time choosing which individual independently and uniformly at random, implying the rate to a specific individual equals  $\beta_H/(h - 1)$ . Additionally, the infectious individual makes contact with workplace individuals at rate  $\beta_W$ , each time with an independent, uniformly selected individual of the same workplace. Finally, the infectious individual makes global contacts at rate  $\beta_G$ , each time with an individual chosen independently and uniformly from the entire population. An infectious contact with a susceptible individual results in the latter getting infected (such individuals starting their infectious contact processes), whereas contacts with already infected people have no effect. After their infectious period is over, the individual recovers and becomes immune thus playing no further role in the epidemic.

The epidemic is initiated by one randomly selected individual being externally infected and the remaining individuals being susceptible. The epidemic runs its course until there are no infectious individuals present in the community when it stops. Each individual is then either recovered or susceptible, and the random distribution of how the recovered and susceptibles are distributed in the community defines the final outcome. The final *number* infected is denoted  $Z$  and referred to as the final size.

For ease of analysis, we assume that each global contact is with an individual chosen uniformly from the entire population, including also the infector itself. Thus a global contact may be with an individual in the infector's own household or workplace. This assumption represents no loss of generality and has no impact on our asymptotic results.

## 2.3 Parametrization

Without loss of generality we assume  $E[I] = 1$  (meaning that time is measured in units of the mean infectious period). Our focus lies on  $h$  and  $w$  being fixed and letting  $n \rightarrow \infty$ . If  $w$  also becomes large with  $n$  we essentially have a households epidemic model, and if both  $h$  and  $w$  grow with  $n$ , the model is very close to being homogeneously mixing.

The epidemic parameters are the three infectious contact rates  $\beta_H$ ,  $\beta_W$  and  $\beta_G$ , and the infectious period distribution  $I$  (assumed to have mean  $E[I] = 1$ ). Later we reparameterize the three infection rates to  $\beta = \beta_H + \beta_W + \beta_G$ ,  $\pi_G = \beta_G/(\beta_H + \beta_W + \beta_G)$  and  $\pi_{H|G^c} = \beta_H/(\beta_H + \beta_W)$ , thus being the overall contact rate, what fraction of the overall contacts rate are global contacts, and what fraction of the remaining (local) contacts are in households, respectively.

## 2.4 Notation

We use  $\xrightarrow{D}$  and  $\xrightarrow{P}$  to denote convergence in distribution and convergence in probability, respectively, and  $\stackrel{D}{=}$  to denote equal in distribution. For a random variable,  $X$  say, taking values in the positive integers,  $f_X$  denotes its probability-generating function (PGF), i.e.  $f_X(s) = E[s^X]$  ( $0 \leq s \leq 1$ ). For any real-valued random variable  $Y$ ,  $\mu_Y$  denotes its mean and  $\phi_Y$  denotes its Laplace transform, so  $\phi_Y(\nu) = E[\exp(-\nu Y)]$  ( $\nu \geq 0$ ). For a positive integer  $n$  and  $p \in [0, 1]$ ,  $\text{Bin}(n, p)$  denotes a binomial random variable with  $n$  trials and success probability  $p$ . For  $\lambda > 0$ ,  $\text{Exp}(\lambda)$  denotes an exponential random variable with rate  $\lambda$  and hence mean  $\lambda^{-1}$ . For  $x \in \mathbb{R}$ ,  $\lfloor x \rfloor$  denotes the usual floor function, so  $\lfloor x \rfloor$  is the greatest integer  $\leq x$ . For a non-negative random variable  $X$ , we write  $Y \sim \text{Po}(X)$  when  $Y$  has a Mixed-Poisson distribution with parameter distributed as  $X$ , i.e. when  $P(Y = k) = E[X^k e^{-X}/k!]$  for  $k = 0, 1, 2, \dots$ .

## 3 Main Results

Let  $Z^{(n)}$  denote the final size of the epidemic in a community of size  $n$ , i.e. the total number getting infected throughout the course of the epidemic. Our first result concerns the probability of a major outbreak. Here we define a major outbreak as  $Z^{(n)} > \log n$ , but any sequence  $h_n$  such that  $h_n \rightarrow \infty$  and  $h_n/n \rightarrow 0$  would do equally well. We also define reproduction numbers  $R_*$  and  $R_L$ , which determine if the asymptotic probability of a major outbreak is strictly positive for the cases  $\pi_G > 0$  and  $\pi_G = 0$ , respectively.

**Theorem 3.1.** *For the epidemic model with two group structures and tunable overlap defined in Section 2,*

$$P(Z^{(n)} > \log n) \rightarrow \rho \quad \text{as } n \rightarrow \infty. \quad (1)$$

*When  $\pi_G > 0$ , the constant  $\rho$  (the major outbreak probability) is given by  $\rho = 1 - \xi$ , where  $\xi$  is the smallest solution in  $[0, 1]$  of  $\phi_A(\beta_G(1 - s)) = s$  and the Laplace transform  $\phi_A(\nu)$  is given by (15) in Section 5.3.3. The major outbreak probability  $\rho$  is strictly positive if and only if  $R_* > 1$ , where  $R_*$  is given by (21) in Section 5.3.4.*

When  $\pi_G = 0$ , the major outbreak probability  $\rho$  is given by (57) in Section 6.4 and  $\rho$  is strictly positive if and only if  $R_L > 1$ , where  $R_L$  is given by (20) in Section 5.3.4.

Our next result concerns the limiting final size in case there is a major outbreak.

**Theorem 3.2.** *Consider the epidemic model with two group structures and tunable overlap defined in Section 2. Suppose that  $R_L > 1$  if  $\pi_G = 0$ , and that  $R_* > 1$  and  $\text{var}(I) < \infty$  if  $\pi_G > 0$ . We then have*

$$\frac{Z^{(n)}}{n} | Z^{(n)} > \log n \xrightarrow{P} z \quad \text{as } n \rightarrow \infty, \quad (2)$$

where the constant  $z$  (the limiting final size) is strictly positive.

If  $\pi_G > 0$ , the limiting final size  $z$  is given by the unique strictly positive root of the equation (7) in Section 5.2, with  $f_S(s)$  being given by (18) in Section 5.3.4.

If  $\pi_G = 0$ , the limiting final size  $z$  is given by (44) in Section 6.2.

**Remark 3.3.** *It is worth noting that if the infectious period is constant, i.e.  $I \equiv 1$ , then the major outbreak probability  $\rho$  in (1) is identical to the final size quantity  $z$  in (2), but this does not hold when the infectious period is random.*

In Section 4 we compute  $z$  (and hence also  $\rho$  if  $I \equiv 1$ ) numerically for various parameter settings and study how it depends on the different model parameters.

## 4 Numerical illustrations

In this section we present numerical results which show the usefulness of the asymptotic results in Theorems 3.1 and 3.2 for finite population size  $n$ , and illustrate the dependence of the limiting final size  $z$  on the probability an individual is a mover,  $\theta$ , the number of households that initially comprise a workplace,  $d = w/h$ , and the distribution of the infectious period,  $I$ .

In Figure 1, histograms are shown of the fraction of the population infected by epidemics with  $d = 1$ ,  $h = w = 4$ ,  $\beta = 3$ ,  $\pi_G = 0.025$ ,  $\pi_{H|G^c} = 0.5$  and  $I \equiv 1$ . Each plot is based on 100,000 simulations of epidemics with one initial infective chosen uniformly at random from the population, with the remainder of the population being susceptible. In the top-left plot,  $\theta = 0.075$  (large overlap) and  $n = 1,000$ , so the population is partitioned into 250 households of size 4, and also into 250 workplaces of size 4. With these parameter values,  $R_* = 0.6541$ , so the epidemic is subcritical as is clear from the shape of the histogram. In the other three plots,  $\theta = 0.4$  (smaller overlap), for which  $R_* = \infty$  as the corresponding epidemic with only local infection (i.e.  $\beta_G = 0$  and  $\beta_H$  and  $\beta_W$  are unchanged) is supercritical. In each of these plots the final size distribution is bimodal, reflecting minor and major outbreaks. When  $n = 1,000$ , there is a sharp distinction between minor and major epidemics, with none of the 100,000 simulated epidemics having

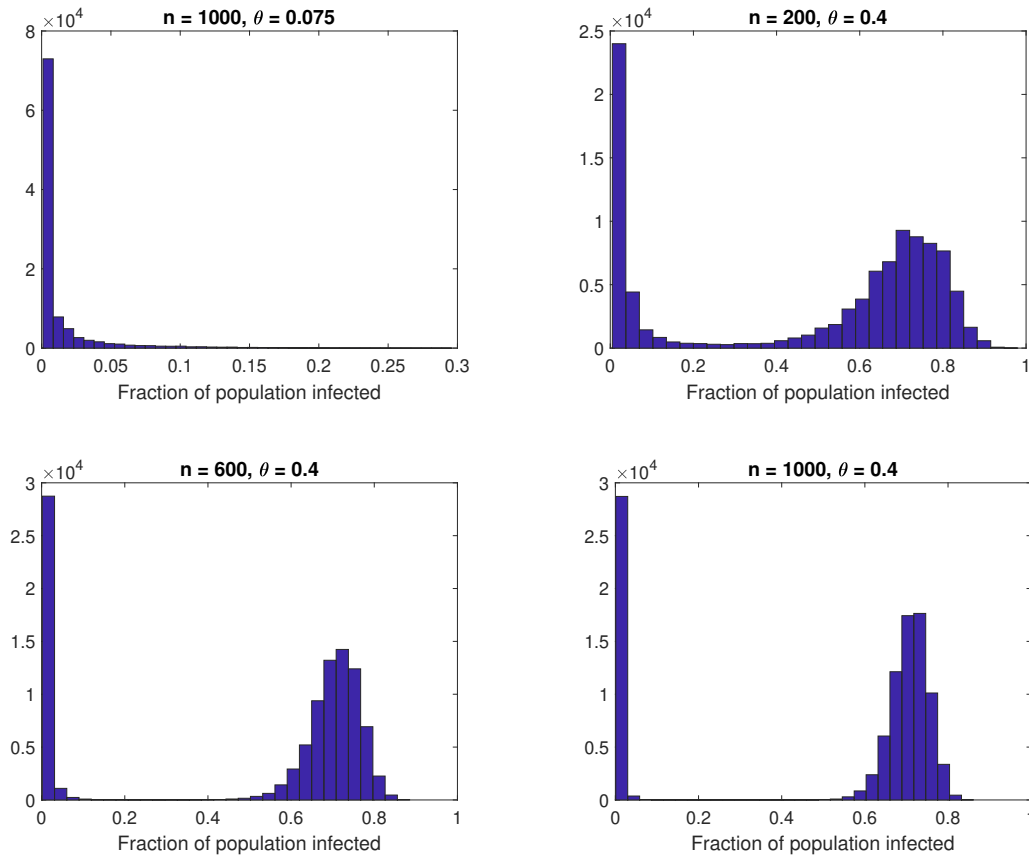


Figure 1: Histograms of simulated fraction of the population infected for epidemics with  $d = 1$ ,  $h = w = 4$ ,  $\beta = 3$ ,  $\pi_G = 0.025$ ,  $\pi_{H|G^c} = 0.5$  and  $I \equiv 1$ . Each plot is based on 100,000 simulations. See text for further details.



a final size in (172, 358). The distinction is less clear when  $n = 600$  and when  $n = 200$  the cut-off between minor and major epidemics is far from clear.

In Figure 2, we explore the accuracy of the limiting (as  $n \rightarrow \infty$ ) final size  $z$  and probability of a major outbreak  $\rho$  for finite population sizes. We assume  $\theta = 0.4$ ,  $h = 4$ ,  $\beta = 3$ ,  $\pi_G = 0.025$ ,  $\pi_{H|G^c} = 0.5$  and  $I \equiv 1$ , and consider epidemics with  $d = 1, 2$  or  $3$  (so  $w = 4, 8$  or  $12$ ) for various population sizes  $n$ . For  $d = 1$  or  $2$ , we consider epidemics in populations of size  $n = 120, 240, 480, 960, 1,440, 1,920, 2,560, 3,200$  and  $3,840$ . The population size  $n$  needs to be a multiple of  $w = hd$ , so when  $d = 3$ , we replaced  $2,560$  and  $3,200$  by  $2,556$  and  $3,204$ , respectively. The cut-offs for distinguishing between minor and major epidemics were determined by inspection of histograms of simulated final sizes. In all cases with  $n > 480$  a cut-off of  $200$  was used. For  $d = 1$ , the cut-offs for  $n = 120, 240, 480$  were  $36, 75, 150$ ; for  $d = 2$ , they were  $50, 100, 150$ ; and for  $d = 3$ , they were  $25, 50, 100$ . For each fixed  $(n, d)$ ,  $n_{\text{sim}} = 10,000$  epidemics were simulated and an estimate  $\hat{\rho}$  of the probability of a major outbreak obtained by the fraction of simulations whose final size was  $\geq$  the cut-off, together with an approximate 95% confidence interval for  $\rho$  given by  $\hat{\rho} \pm 1.96 \sqrt{\hat{\rho}(1 - \hat{\rho})/n_{\text{sim}}}$ . Independently, 10,000 major outbreaks were simulated, using the given cut-off. Let  $\hat{z}$  and  $\hat{\sigma}^2$  be respectively the sample mean and variance of the fraction of the population infected in these simulated epidemics. Then  $z$  is estimated by  $\hat{z}$ , with an approximate 95% confidence interval given by  $\hat{z} \pm 1.96 \hat{\sigma} / \sqrt{n_{\text{sim}}}$ . Note that  $\rho = z$  since the infectious period  $I$  is non-random.

It is clear from Figure 2 that the limiting  $\rho$  and  $z$  provide good approximations even for moderately sized  $n$ . The approximation for the final size  $z$  is better than that for the major outbreak probability  $\rho$  (note the different scales on the  $y$ -axes). In all cases, the limiting  $z$  overestimates the final size for finite  $n$ . That is not the case for major outbreak probability  $\rho$  but note the wider confidence intervals.

In Figure 3, we investigate the dependence of the limiting final size,  $z$ , on the probability an individual is a mover,  $\theta$ , the workplace to household size ratio,  $d$ , and whether the infectious period is non-random or exponential. We assume  $h = 3$ ,  $\beta = 3$ ,  $\pi_G = 0.025$  and  $\pi_{H|G^c} = 0.5$ . In the left panel  $I \equiv 1$  and in the right panel  $I \sim \text{Exp}(1)$ . First note that  $z$  is increasing in both  $\theta$  and  $d$ . Secondly, for fixed  $d$ ,  $z$  is fairly constant with  $\theta$  when  $\theta$  is close to one. This is particularly the case when the infectious period  $I$  is non-random and  $d \geq 2$ , when  $z$  is close to being constant for  $\theta \in [0.5, 1]$ . Recall that  $\theta = 1$  yields the standard households-workplaces model analysed for example in Ball and Neal [8] and Pellis et al. [23], in which the population is partitioned independently into households and into workplaces. Thus, under these circumstances, the standard model in which there is no overlap between the two structures provides a good approximation to the more realistic model having overlap. Thirdly, note that for the parameter values in Figure 3,  $z = 0$  when  $\theta = 0$ . If  $\theta = 0$ , every workplace consists of  $d$  complete households, so it is not possible for the disease to spread outside the workplace of the initial case via local infection. As  $\theta$  increases from 0, with all other parameters held constant, there is a critical value of  $\theta$ ,  $\hat{\theta}_d$  say, when  $R_* = 1$  and the epidemic is critical. Observe that  $\hat{\theta}_d$  decreases with  $d$  and in an obvious notation,  $z'(\hat{\theta}_d)$  is large unless  $d$  is small. Fourthly, given a fixed mean, the choice of the infectious period distribution can have an appreciable impact on epidemic properties. Note that  $z$  is greater when  $I \equiv 1$  than in the corresponding model with  $I \sim \text{Exp}(1)$ . More strikingly,  $\hat{\theta}_d$  can be a lot smaller for the model with  $I \equiv 1$ , the difference being particularly noticeable when  $d = 1$ .

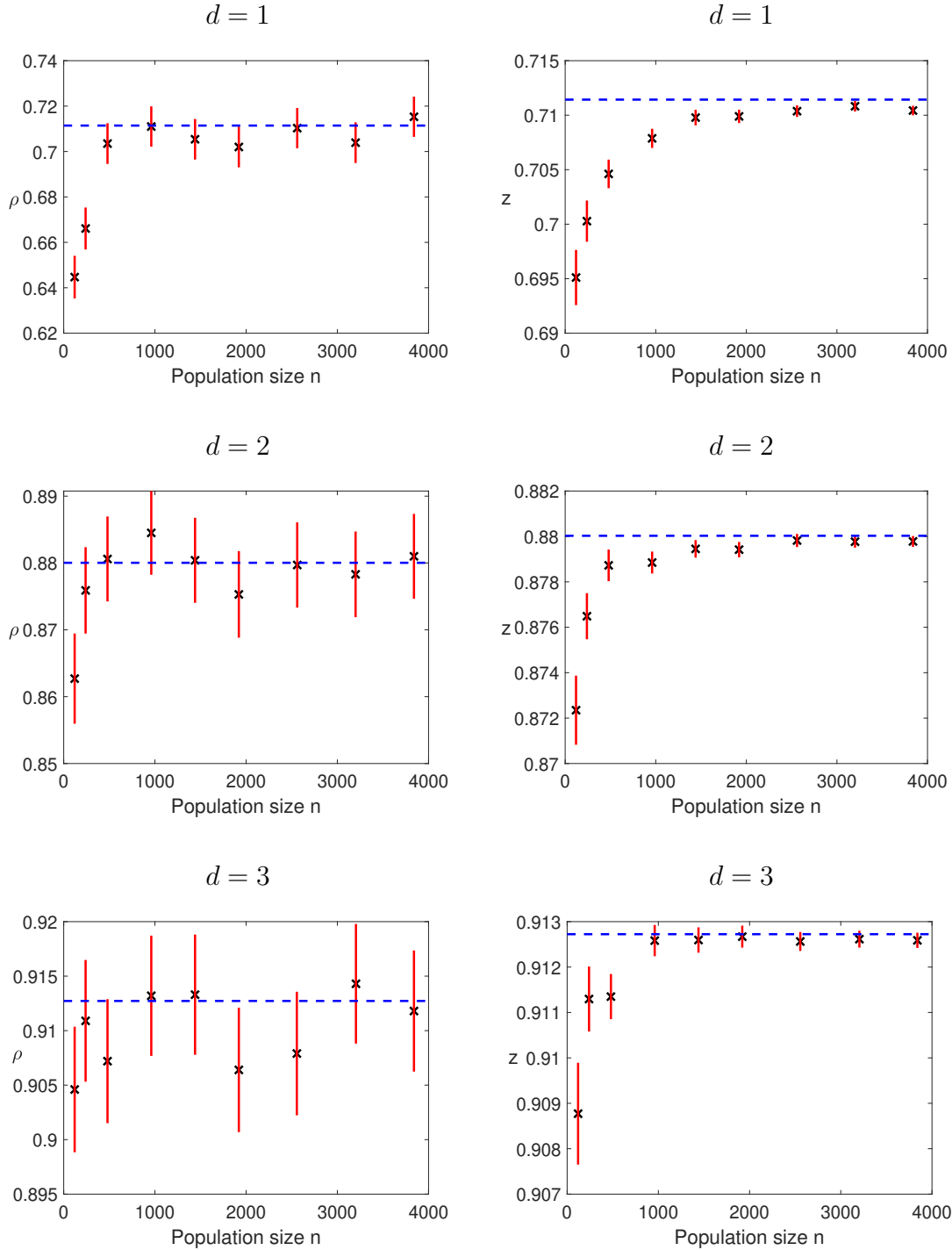


Figure 2: Comparison of limiting and simulated estimates of the probability of a major outbreak  $\rho$  (left column) and the fraction of the population infected by a major outbreak  $z$  (right column) based on 10,000 simulations for each choice of  $(n, d)$ . The blue dashed horizontal lines depict the asymptotic values. The black crosses show the estimated values based on simulations, with the red vertical lines giving approximate 95% confidence intervals. See text for further details.

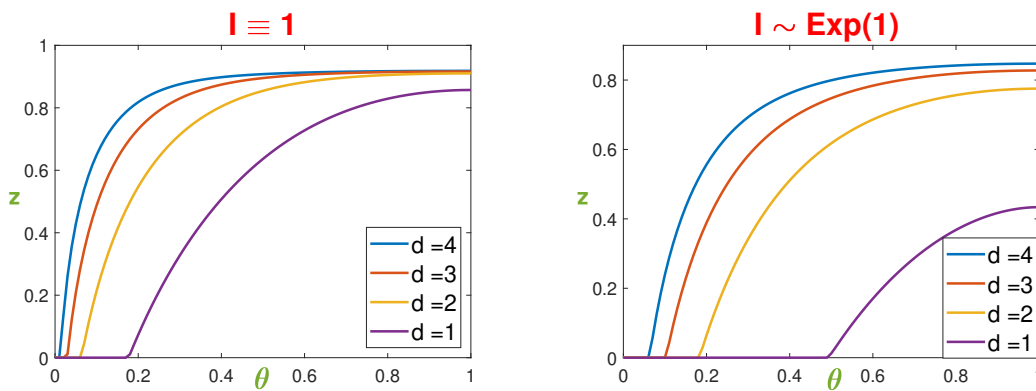


Figure 3: Dependence of the limiting final size,  $z$ , on the probability an individual is a mover,  $\theta$ , for fixed  $d = 1, 2, 3, 4$  when  $h = 3$ ,  $\beta = 3$ ,  $\pi_G = 0.025$  and  $\pi_{H|G^c} = 0.5$ ; in the left panel  $I \equiv 1$  and in the right panel  $I \sim \text{Exp}(1)$ .

## 5 Derivations and proofs when $\pi_G > 0$

### 5.1 Introduction

Let  $\mathcal{E}^{(n)}$  denote the epidemic model defined in Section 2. When  $\pi_G > 0$ , the epidemic model  $\mathcal{E}^{(n)}$  is a special case of the general framework of Ball and Neal [8] for SIR epidemics in populations with two levels of mixing. We first outline that general framework in Section 5.2, introducing key concepts of *local infectious clumps*, *local susceptibility sets* and the *severity* of a local infectious clump, and explain heuristically how the limiting severity of a local infectious clump as  $n \rightarrow \infty$  is related to the event of a major outbreak and the limiting size of a local susceptibility set is related to the final outcome of a major outbreak. In the present model, the structure of local infectious clumps and local susceptibility sets is far more complicated than in the examples considered in Ball and Neal [8], and consequently proofs and calculations of limiting major outbreak probability and final size are considerably more difficult. Local infectious clumps and local susceptibility sets for  $\mathcal{E}^{(n)}$  are analysed in Section 5.3. Crucial to this analysis are *complexes*, defined in Section 5.3.1, which are associated with workplaces but contain also households of movers who moved out of that workplace. The analysis of the limiting local infectious clump as  $n \rightarrow \infty$  is simpler when  $I \equiv 1$ . This is treated in Section 5.3.2, with the extension to the case of general  $I$  and calculation of the limiting major outbreak probability  $\rho$  being given in Section 5.3.3. Local susceptibility sets are analysed in Section 5.3.4, where calculations of the threshold parameter  $R_*$  and the limiting final size of a major outbreak  $z$  are described. Numerical implementation of the calculation of  $R_*$ ,  $z$  and  $\rho$  is considered briefly in Section 5.4. Finally, proofs of Theorems 3.1 and 3.2 for  $\pi_G > 0$  are given in Sections 5.5 and 5.6, respectively.

### 5.2 General framework

We derive the properties of the model by using the general framework of Ball and Neal [8] for SIR epidemics in populations with two levels of mixing. We first outline the key

points of that framework in the current setting. Label the individuals in the population  $1, 2, \dots, n$  and let  $\mathcal{N} = \{1, 2, \dots, n\}$ . (The precise labelling is not important at present.) We construct on  $\mathcal{N}$  a random directed graph  $\mathcal{G}^{(n)}$  of potential *local* (i.e. household or workplace) infectious contacts as follows. For each  $(i, j) \in \mathcal{N}^2$ , with  $i \neq j$ , there is a directed edge from  $i$  to  $j$  in  $\mathcal{G}^{(n)}$  if and only if  $i$ , if infected, makes infectious contact with  $j$ . More precisely, let  $I_i$  denote the length of  $i$ 's infectious period. Then, given  $I_i$ , there is a directed edge from  $i$  to each individual in  $i$ 's household independently with probability  $1 - \exp(-\beta'_H I_i)$ , where  $\beta'_H = \beta_H/(h-1)$ , and a directed edge from  $i$  to each individual in  $i$ 's workplace independently with probability  $1 - \exp(-\beta'_W I_i)$ , where  $\beta'_W = \beta_W/(w-1)$ . If an individual belongs to both  $i$ 's household and workplace, then the directed edge is present with probability  $1 - \exp(-(\beta'_H + \beta'_W)I_i)$ .

We use  $\mathcal{G}^{(n)}$  to define for each  $i \in \mathcal{N}$  their *local infectious clump* and *local susceptibility set* as follows. For  $(i, j) \in \mathcal{N}^2$ , write  $i \rightsquigarrow j$  if and only if there is a chain of directed edges from  $i$  to  $j$ , with the convention that  $i \rightsquigarrow i$ . For  $i \in \mathcal{N}$ , let  $\mathcal{C}_i^{(n)} = \{j \in \mathcal{N} : i \rightsquigarrow j\}$  and  $\mathcal{S}_i^{(n)} = \{j \in \mathcal{N} : j \rightsquigarrow i\}$  be  $i$ 's local infectious clump and local susceptibility set, respectively. Let  $\mathcal{E}^{(n)}$  denote the epidemic model described in Section 2.2. A realisation of the final outcome of  $\mathcal{E}^{(n)}$  can be constructed using  $I_1, I_2, \dots, I_n$  and  $\mathcal{G}^{(n)}$  by attaching to each  $i \in \mathcal{N}$  an independent Poisson process having rate  $\beta_G$  giving the times individuals make global contacts if they are infectious. (As in Section 2.2, each global contact is with an individual chosen independently and uniformly from  $\mathcal{N}$ .) Observe that if  $i$  is infected by a global contact in  $\mathcal{E}^{(n)}$ , then each member of  $\mathcal{C}_i^{(n)}$  is necessarily infected in  $\mathcal{E}^{(n)}$ . Similarly, if  $i$  is an initial susceptible in  $\mathcal{E}^{(n)}$ , then  $i$  is infected in  $\mathcal{E}^{(n)}$  if and only if a member of  $\mathcal{S}_i^{(n)}$  is infected by a global contact or  $\mathcal{S}_i^{(n)}$  contains the initial infective.

For  $i \in \mathcal{N}$ , let  $C_i^{(n)} = |\mathcal{C}_i^{(n)}|$  and  $S_i^{(n)} = |\mathcal{S}_i^{(n)}|$  be the sizes of  $i$ 's local infectious clump and local susceptibility set, respectively, and let  $A_i^{(n)} = \sum_{j \in \mathcal{C}_i^{(n)}} I_j$  be the severity of  $i$ 's local infectious clump. Note that  $C_1^{(n)}, C_2^{(n)}, \dots, C_n^{(n)}$  are identically distributed, with distribution that depends on  $n$  (unless  $\theta = 0$ ), but they are not independent. A similar comment holds for  $S_1^{(n)}, S_2^{(n)}, \dots, S_n^{(n)}$  and  $A_1^{(n)}, A_2^{(n)}, \dots, A_n^{(n)}$ . Let  $C^{(n)}, S^{(n)}$  and  $A^{(n)}$  be random variables distributed as  $C_1^{(n)}, S_1^{(n)}$  and  $A_1^{(n)}$ , respectively. We show below that there exists random variables  $C, S$  and  $A$  such that

$$C^{(n)} \xrightarrow{D} C, \quad S^{(n)} \xrightarrow{D} S \quad \text{and} \quad A^{(n)} \xrightarrow{D} A \quad \text{as } n \rightarrow \infty. \quad (3)$$

The distributions of  $C, S$  and  $A$  are defective, with a mass at infinity, if the epidemic  $\mathcal{E}^{(n)}$  in the absence of global infection (i.e. with  $\beta_G = 0$ ) is supercritical in the limit  $n \rightarrow \infty$ .

Suppose that  $P(C < \infty) = 1$  and recall that the epidemic  $\mathcal{E}^{(n)}$  is initiated by a single individual being externally infected. Then, for large  $n$ , the early stages of  $\mathcal{E}^{(n)}$  can be approximated by a branching process of local infectious clumps, since the probability that clumps arising from distinct global contacts have a non-empty intersection is small. Let  $R^{(n)}$  be the number of global contacts that emanate from a typical local infectious clump. Then  $R^{(n)} \sim \text{Po}(\beta_G A^{(n)})$ , i.e.  $R^{(n)}$  is distributed as a mixed Poisson distribution with random mean  $\beta_G A^{(n)}$ . Thus,  $E[R^{(n)}] = \beta_G E[A^{(n)}]$ . Moreover, since  $I_j$  is independent of the event  $\{j \in \mathcal{C}_i^{(n)}\}$  and  $E[I_j] = 1$ ,

$$E[A_i^{(n)}] = E \left[ \sum_{j=1}^n I_j 1_{\{j \in \mathcal{C}_i^{(n)}\}} \right] = \sum_{i=1}^n E[I_j] P(j \in \mathcal{C}_i^{(n)}) = \sum_{j=1}^n E \left[ 1_{\{j \in \mathcal{C}_i^{(n)}\}} \right] = E[C_i^{(n)}].$$

Hence,  $E[R^{(n)}] = \beta_G E[C^{(n)}]$ . Now  $R^{(n)} \xrightarrow{D} R$  as  $n \rightarrow \infty$ , where  $R \sim \text{Po}(\beta_G A)$ , so the process of local infectious clumps in  $\mathcal{E}^{(n)}$  can be approximated by a Galton-Watson branching process  $\mathcal{B}$ , having a single ancestor and offspring distribution  $R$ .

Let  $\xi$  be the extinction probability of  $\mathcal{B}$ . Then  $\xi$  is given by the smallest solution in  $[0, 1]$  of  $f_R(s) = s$ . Conditioning on  $A$ ,

$$f_R(s) = E[E[s^R|A]] = E[\exp(-\beta_G A(1-s))] = \phi_A(\beta_G(1-s)). \quad (4)$$

Further,  $\xi < 1$  if and only if  $R_* > 1$ , where

$$R_* = E[R] = \beta_G E[A] = \beta_G E[C]. \quad (5)$$

Thus,  $R_*$  serves as a threshold parameter for the epidemic  $\mathcal{E}^{(n)}$ , in that as  $n \rightarrow \infty$  a major outbreak occurs with non-zero probability if and only if  $R_* > 1$ . Note that  $\rho$  in Theorem 3.1 is given by  $1 - \xi$ .

The above assumes that  $P(C < \infty) = 1$ . If  $P(C < \infty) < 1$ , then the branching process  $\mathcal{B}$  is defined as above but now  $R_* = \infty$  and an individual may have infinitely many offspring. The branching process  $\mathcal{B}$  can still be used to approximate the early stages of  $\mathcal{E}^{(n)}$  for large  $n$ , since an individual having infinitely many offspring in  $\mathcal{B}$  necessarily corresponds to a major outbreak in  $\mathcal{E}^{(n)}$ . The above heuristic argument is made rigorous in the proof of Theorem 3.1 in Section 5.5.

Suppose that  $n$  is large and a major outbreak occurs. Let  $z$  be the fraction of the population that are infected. Since  $E[I] = 1$ , the total severity of the epidemic is approximately  $nz$ , so the probability a given individual avoids global contact throughout the epidemic is approximately

$$\pi = \exp\left(-\frac{\beta_G}{n}nz\right) = \exp(-\beta_G z). \quad (6)$$

The probability an individual chosen uniformly at random from the population is infected by the epidemic is given by  $z$ . Recall that such an individual, if initially susceptible, avoids infection from the epidemic if and only if (i) all members of its susceptibility set avoid global infection and (ii) its susceptibility set does not contain the initial infective. The former event has asymptotic probability 0 if  $S = \infty$  and the latter event has asymptotic probability 1 if  $S < \infty$ . For large  $n$ , individuals avoid global infection approximately independently, so  $1 - z \approx f_S(\pi) = E[\pi^S]$ . In Section 5.6, we show that these approximations are exact in the limit as  $n \rightarrow \infty$  so, using (6),  $z$  satisfies

$$1 - z = f_S(e^{-\beta_G z}). \quad (7)$$

Note that  $s = 1 - z$  satisfies the equation governing the extinction probability of a Galton-Watson process having offspring distribution  $\text{Po}(\beta_G S)$ . Further, by exchangeability

$$E[C^{(n)}] = \frac{1}{n} E\left[\sum_{i=1}^n \sum_{j=1}^n 1_{\{i \rightsquigarrow j\}}\right] = \frac{1}{n} E\left[\sum_{j=1}^n \sum_{i=1}^n 1_{\{j \rightsquigarrow i\}}\right] = E[S^{(n)}], \quad (8)$$

so letting  $n \rightarrow \infty$  yields  $\beta_G E[S] = \beta_G E[C] = R_*$ . It follows from standard branching process theory that if  $R_* \leq 1$ , then  $z = 0$  is the only solution of (7), while if  $R_* > 1$ ,

there is a unique second solution in  $(0, 1]$ . When  $R_* > 1$ , this second solution gives the fraction of the population that is infected by a major outbreak in the limit  $n \rightarrow \infty$ .

In Section 5.3.4, we show that  $S$ , the size of a typical local susceptibility set in the limit  $n \rightarrow \infty$ , is distributed as the total progeny of a 3-type branching process. The PGF of  $S$  is given in (18) and  $R_*$  is given in (21). When  $I \equiv 1$ , the possible directed edges in  $\mathcal{G}^{(n)}$  are present independently, with probability  $1 - \exp(-\beta'_H)$  if the two individuals are in the same household but not the same workplace,  $1 - \exp(-\beta'_W)$  if they are in the same workplace but not the same household, and  $1 - \exp(-(\beta'_H + \beta'_W))$  if they are in both the same household and the same workplace. It follows that  $C^{(n)} \stackrel{D}{=} S^{(n)}$ , so  $C \stackrel{D}{=} S$ . Note that in this case  $A^{(n)} = C^{(n)}$ , whence  $A \stackrel{D}{=} C \stackrel{D}{=} S$ , and (4) and (7) imply that  $\rho = z$ . When  $I$  is non-constant,  $C$  is distributed as the total progeny of a  $3(h + w - 2)$ -type branching process and  $A$  as the sum of the lifetimes of all individuals in that branching process. The derivation of the Laplace transform  $\phi_A(\nu)$  in that case is outlined in Section 5.3.3.

Although this section is concerned with the case  $\pi_G > 0$ , we note here that when  $\pi_G = 0$ , we have  $\rho = P(C = \infty)$  and  $z = P(S = \infty)$ .

## 5.3 Local infectious clumps and susceptibility sets

### 5.3.1 Complexes

In order to apply the theory outlined in Section 5.2, we need to obtain the distribution of the random variables  $C$ ,  $A$  and  $S$ , describing the size and severity of a typical local infectious clump and the size of a typical local susceptibility set, respectively, in the limit as  $n \rightarrow \infty$ . For this purpose, we first define a *complex*. There are  $m = n/w$  complexes in the network, where  $m$  is the number of workplaces. Each complex is derived from a workplace containing  $d$  households with individuals belonging to 1 or 2 complexes depending on whether they are a remainder (belong to a single complex) or a mover (belongs to two complexes). We label the workplaces  $1, 2, \dots, m$  and associate the  $i^{\text{th}}$  complex with the  $i^{\text{th}}$  workplace. There are 3 types of individuals in the  $i^{\text{th}}$  complex:

- Movers who start in workplace  $i$  (their household belongs to workplace  $i$ ) but move to a different workplace.
- Remainders who start in the workplace  $i$  (their household belongs to workplace  $i$ ) and remain in that workplace.
- Movers who start in a different workplace (their household is part of a different workplace) and move into workplace  $i$ .

We consider complex  $i$ . We label the households in complex  $i$ ,  $(i, 1), (i, 2), \dots, (i, d)$ . Then complex  $i$  is comprised of  $2d + 1$  groups of individuals. These are:

- Group  $2j - 1$  ( $j = 1, 2, \dots, d$ ). These are the individuals in household  $(i, j)$  who remain in workplace  $i$ .
- Group  $2j$  ( $j = 1, 2, \dots, d$ ). These are the individuals in household  $(i, j)$  who move to a different workplace.

- Group  $2d + 1$ . The individuals who move into workplace  $i$ .

Therefore in complex  $i$ , household-based infections take place in household  $(i, j)$  amongst the individuals in groups  $2j - 1$  and  $2j$  and workplace-based infections involve groups  $j = 1, 3, \dots, 2d + 1$  (the odd numbered groups). We note that group  $2j - 1$  ( $j = 1, 2, \dots, d$ ) can be empty if all members of household  $(i, j)$  move to different workplaces. We still associate the household  $(i, j)$  with workplace  $i$ .

In Figure 4 we illustrate how complexes and the population are constructed using an example where  $h = d = 2$  and  $m = 4$  giving  $w = 4$  and  $n = 16$ .

When  $\theta = 0$ , everyone is a remainder, the complexes do not intersect and each complex consists of a workplace that is partitioned into  $d$  households. The distributions of  $C^{(n)}$ ,  $A^{(n)}$  and  $S^{(n)}$  are each independent of  $n$ , so  $C^{(n)} \stackrel{D}{=} C$ ,  $A^{(n)} \stackrel{D}{=} A$  and  $S^{(n)} \stackrel{D}{=} S$  for all  $n$ . The threshold parameter  $R_*$  and the Laplace transform  $\phi_A(\nu)$  can be computed, at least in principle, using exact results for the final outcome of stochastic multitype SIR epidemics given in Ball [2], Section 3, or Picard and Lefèvre [25], Section 4. The probability mass function of  $S$  can be computed in principle using Ball [3], Lemma 4.1. In practice, these computations are possible only for suitably small  $d$ . In the following, we assume that  $\theta > 0$ .

### 5.3.2 Local infectious clumps

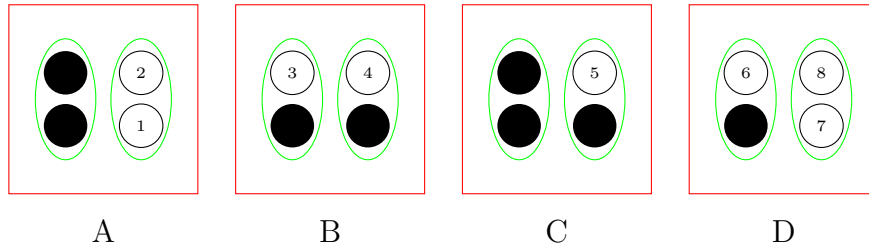
Suppose first that the infectious period  $I$  is non-random, so  $P(I = 1) = 1$ . We construct the local infectious clump of a typical individual,  $i^*$  say, on a generation basis by considering local epidemic spread within successive complexes. Note that apart from the initial generation, a local epidemic *enters* a complex through one of two ways: either a mover in one of groups  $2, 4, \dots, 2d$  is infected in their workplace or a mover in group  $2d + 1$  is infected in their household. The initial individual  $i^*$  can be any member of a complex.

We observe that the groups  $1, 3, \dots, 2d - 1$  are exchangeable, as are the groups  $2, 4, \dots, 2d$ . Therefore we only need to consider the behaviour of the local spread in a complex in the cases where the initial infective in a complex is in group 1, 2 or  $2d + 1$ . Let  $R$  denote that the initial infective is in group 1 (or  $3, 5, \dots, 2d - 1$ ), that is, the individual is a remainder. Let  $H$  denote that the initial infective is in group 2 (or  $4, 6, \dots, 2d$ ), the individual belongs to complex  $i$  through their household. Let  $W$  denote that the initial infective is in group  $2d + 1$ , the individual belongs to complex  $i$  through the workplace.

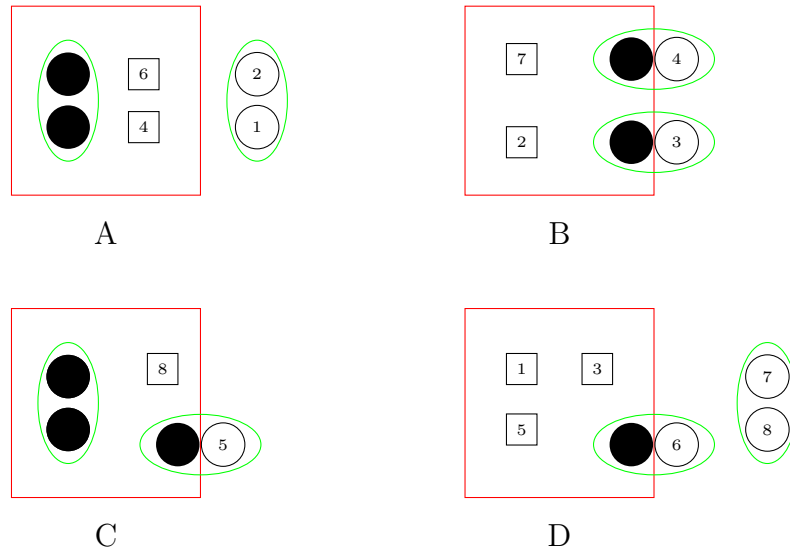
The initial infective in a complex being of type  $X$  is informative about the structure of the complex. For example, if  $X = W$ , then we know that there is at least one mover in the workplace of the complex. Without loss of generality, if the initial individual is of type  $R$  or  $H$  we assign them to group 1 or 2, respectively. Let  $M_1 \sim \text{Bin}(n - 1, \theta)$  and for  $j = 2, 3, \dots, d$ , let  $M_j \sim \text{Bin}(n, \theta)$ . The number of individuals in groups  $2j - 1$  and  $2j$  ( $j = 2, 3, \dots, d$ ) are  $n - M_j$  and  $M_j$ , respectively, regardless of the type of the initial infective. If  $X = R$ , then the number of individuals in groups 1, 2 and  $2d + 1$  are  $n - M_1$ ,  $M_1$  and  $\sum_{l=1}^d M_l$ , respectively, otherwise there are  $n - 1 - M_1$ ,  $1 + M_1$  and  $1 + \sum_{l=1}^d M_l$  individuals in groups 1, 2 and  $2d + 1$ , respectively.

If  $n$  is large, the probability that, in the early stages of the construction of an infectious clump, there is a complex that contains movers who either moved into the complex from

(a)



(b)



(c)

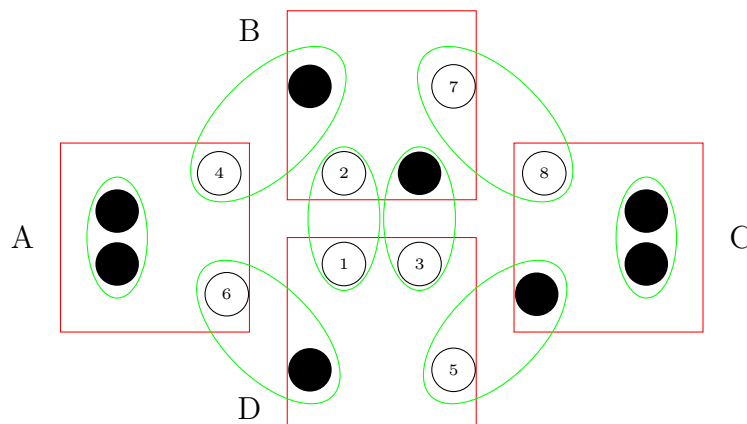


Figure 4: Illustration of the construction of a population of  $n = 16$  individuals divided into 4 workplaces, labelled  $A, B, C, D$ , of size  $4 (= w)$  with each workplace consisting of  $2 (= d)$  households of size  $2 (= h)$ . Red squares denote workplaces and green ellipses denote households.

(a) The initial workplaces with solid circles representing remainders and open circles representing movers with movers numbered. There are 8 movers and 8 remainders.

(b) The construction of complexes. Open squares represent the places created in a workplace by movers and the number in the square denotes the mover who fills the place. Note that in complexes A and D there are households consisting completely of movers with the household residing outside the workplace.

(c) The final population structure created from the complexes.



the same workplace or moved out of their original workplace to the same workplace is small. Thus, we approximate the size  $C^{(n)}$  of the clump by assuming that for each complex, all the members from groups  $2, 4, \dots, 2d$  are in different workplaces and all members of group  $2d + 1$  are in different households. Further, we assume that apart from the initial infective in a complex, all movers in a complex have not been involved previously in the construction of the clump. Under these assumptions,  $C^{(n)}$  is approximated by the total progeny of a 3-type branching process, denoted by  $\mathcal{B}_C$ , which we now describe.

Consider the spread of a local epidemic within a typical complex defined as above. For  $X, Y = R, H, W$ , let  $Z_{XY}$  be defined as follows. Let  $Z_{XR}$  denote the number of remainder individuals (excluding the initial infective if  $X = R$ ) infected in a complex when there is a single initial infective in group type  $X$ . Let  $Z_{XH}$  denote the number of individuals in group  $2d + 1$  (excluding the initial infective if  $X = W$ ) infected in the complex when there is a single infective in group type  $X$ . Individuals infected in group  $2d + 1$  will be infected in the workplace and will start an epidemic in a new complex through their household. In other words, in that other complex they will belong to group  $2, 4, \dots, 2d$ . Similarly, let  $Z_{XW}$  denote the number of individuals in groups  $2, 4, \dots, 2d$  (excluding the initial infective if  $X = H$ ) infected in the complex when there is a single infective in group type  $X$ . Individuals infected in group  $2, 4, \dots, 2d$  will be infected in the household and will start an epidemic in a new complex through their workplace. In other words, in that other complex they will belong to group  $2d + 1$ .

If  $i^*$ , the initial member of the local infectious clump is a remainder, then the initial individual in  $\mathcal{B}_C$  has type  $R$  and its offspring is distributed as  $(Z_{RR}, Z_{RH}, Z_{RW})$ . If  $i^*$  is a mover, then  $i^*$  belongs to two complexes and its offspring is distributed as  $(Z_{HR}, Z_{HH}, Z_{HW}) + (Z_{WR}, Z_{WH}, Z_{WW})$ , where the two random vectors are independent since  $I$  is non-random. All subsequent individuals in the branching process will be of type  $H$  or  $W$ , and their offspring are distributed as  $(Z_{HR}, Z_{HH}, Z_{HW})$  or  $(Z_{WR}, Z_{WH}, Z_{WW})$ , respectively. Note that individuals in the branching process  $\mathcal{B}_C$  correspond to complexes in the local infectious clump.

Consider the branching process  $\hat{\mathcal{B}}_C$ , in which there is a single ancestor, who is either of type  $H$  or  $W$ , and the offspring distributions of type- $H$  and type- $W$  individuals are  $(Z_{HR}, Z_{HH}, Z_{HW})$  and  $(Z_{WR}, Z_{WH}, Z_{WW})$ , respectively. Thus, type- $R$  individuals may exist in  $\hat{\mathcal{B}}_C$  but they have no offspring. Let  $\hat{Z}_H$  be the total progeny of  $\hat{\mathcal{B}}_C$  (i.e the total number of individuals infected and not the total number of complexes infected), counting all types and including the ancestor, given that the ancestor is of type  $H$ , and define  $\hat{Z}_W$  similarly. Also, for  $X = R, H, W$  and  $0 \leq s_1, s_2, s_3 \leq 1$ , let

$$g_X(s_1, s_2, s_3) = \mathbb{E} [s_1^{Z_{XR}} s_2^{Z_{XH}} s_3^{Z_{XW}}].$$

By taking expectations with respect to the offspring of the ancestor,

$$f_{\hat{Z}_H}(s) = sg_H(s, f_{\hat{Z}_H}(s), f_{\hat{Z}_W}(s)) \quad \text{and} \quad f_{\hat{Z}_W}(s) = sg_W(s, f_{\hat{Z}_H}(s), f_{\hat{Z}_W}(s)). \quad (9)$$

For each  $s \in [0, 1]$ , the equations (9) determine  $f_{\hat{Z}_H}(s)$  and  $f_{\hat{Z}_W}(s)$ .

Let  $C$  be the total progeny of  $\mathcal{B}_C$ , counting all types and including the ancestor, given that the ancestor is a mover with probability  $\theta$ , otherwise it is a remainder. Then, conditioning on the type of the ancestor and the spread within its complex(es),

$$f_C(s) = (1 - \theta)sg_R(s, f_{\hat{Z}_H}(s), f_{\hat{Z}_W}(s)) + \theta g_H(s, f_{\hat{Z}_H}(s), f_{\hat{Z}_W}(s))f_{\hat{Z}_W}(s). \quad (10)$$

(Note, using (9), that for  $s > 0$ , the second term on the right-hand side of (10) may be expressed as  $\theta s^{-1} f_{\hat{Z}_H}(s) f_{\hat{Z}_W}(s)$ ; we use the form in (10) as it is also valid when  $s = 0$ .)

**Lemma 5.1.** *For the model with a constant infectious period,*

$$C^{(n)} \xrightarrow{D} C \quad \text{as } n \rightarrow \infty.$$

*Proof.* We prove the lemma by coupling realisations of the local infectious clump  $\mathcal{C}_{i^*}^{(n)}$  of an individual  $i^*$  in  $\mathcal{E}^{(n)}$  and the branching process  $\mathcal{B}_C$ , both of which are constructed on a generation basis, taking complexes in the same order. We start with a realisation of  $\mathcal{C}_{i^*}^{(n)}$  and use the local spread within its complexes to give a realisation of  $\mathcal{B}_C$  until in  $\mathcal{C}_{i^*}^{(n)}$  we encounter a complex which contains a mover whose original workplace is the same as that of another mover used previously in the construction of  $\mathcal{C}_{i^*}^{(n)}$ . Let  $D^{(n)}$  be the number of complexes used in  $\mathcal{C}_{i^*}^{(n)}$  before we encounter such a complex, where  $D^{(n)} = \infty$  if no such complex is encountered. As soon as such a complex is encountered, the offsprings of all subsequent individuals in  $\mathcal{B}_C$  are sampled independently of  $\mathcal{C}_{i^*}^{(n)}$  from their appropriate distributions.

Note that for  $k = 1, 2, \dots$ , if  $C_{i^*}^{(n)} = k$  and  $D^{(n)} > k + 1$ , then  $C_{i^*}^{(n)} = C$ , since in the construction of  $\mathcal{C}_{i^*}^{(n)}$ , each new complex necessarily contains at least one new individual. Hence, for  $k = 1, 2, \dots$ ,

$$\begin{aligned} \mathrm{P}(C^{(n)} \leq k) &= \mathrm{P}(C^{(n)} \leq k, D^{(n)} > k + 1) + \mathrm{P}(C^{(n)} \leq k, D^{(n)} \leq k + 1) \\ &= \mathrm{P}(C \leq k, D^{(n)} > k + 1) + \mathrm{P}(C^{(n)} \leq k, D^{(n)} \leq k + 1). \end{aligned}$$

Also,

$$\mathrm{P}(C \leq k) = \mathrm{P}(C \leq k, D^{(n)} > k + 1) + \mathrm{P}(C \leq k, D^{(n)} \leq k + 1),$$

It follows that

$$|\mathrm{P}(C^{(n)} \leq k) - \mathrm{P}(C \leq k)| \leq \mathrm{P}(D^{(n)} \leq k + 1). \quad (11)$$

Recall that  $\theta > 0$ . Let  $T^{(n)}$  be the total number of movers in  $\mathcal{E}^{(n)}$ . A simple application of Chebyshev's inequality yields  $\mathrm{P}(T^{(n)} < \theta n/2) \rightarrow 0$  as  $n \rightarrow \infty$ . The maximum size of a complex is  $2w = 2hd$ , so the number of movers in the first  $k + 1$  complexes is at most  $2w(k + 1)$ . Given,  $T^{(n)} \geq \theta n/2$  the probability that any two given distinct movers come from the same original workplace is bounded above by  $(w - 1)/(\frac{1}{2}\theta n - 1)$ . Thus,

$$\mathrm{P}(D^{(n)} \leq k + 1) \leq \mathrm{P}(T^{(n)} < \theta n/2) + \binom{2w(k + 1)}{2} \frac{w - 1}{\frac{1}{2}\theta n - 1} \mathrm{P}(T^{(n)} > \theta n/2) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Hence, for  $k = 1, 2, \dots$ ,

$$\lim_{n \rightarrow \infty} \mathrm{P}(C^{(n)} \leq k) = \mathrm{P}(C \leq k),$$

and the lemma follows.  $\square$

When the infectious period  $I$  is random, the local infectious clump  $\mathcal{C}_{i^*}^{(n)}$  can still be constructed on a generation basis and coupled to a limiting process as in Lemma 5.1. However, the limiting process is not a branching process, since the infectious periods of movers infected in a within-complex epidemic (i.e. individuals who subsequently start epidemics in

new complexes) are not independent of the size of that within-complex epidemic. The limiting process can be made a branching process by further typing, for example, by also typing infected movers by the number they infect in their household/workplace in the next complex, as we describe in Section 5.3.3.

### 5.3.3 Major outbreak probability

Recall from Section 5.2 that the major outbreak probability  $\rho = 1 - \xi$ , where  $\xi$  is the smallest root in  $[0, 1]$  of  $f_R(s) = s$ , with  $f_R(s)$  being given by (4). Suppose first that  $I \equiv 1$ . Then, as noted in Section 5.2, we have that  $A \stackrel{D}{=} C$  and  $\rho = z$ . We describe the computation of  $z$  (for the case of general  $I$ ) in Section 5.3.4 below, thus enabling  $\rho$  to be calculated. Alternatively, substituting  $A \stackrel{D}{=} C$  into (4) yields  $f_R(s) = f_C(e^{-\beta_G(1-s)})$ , so  $f_R(s)$  can be obtained using (10). We outline below how the distribution of the limiting severity  $A$  can be obtained for general  $I$  in a constructive manner, from which  $\rho$  can be obtained using (4). The key building block is the epidemic within a complex and in particular the joint distribution of the severity of the within-complex epidemic and the number of infectious contacts outside the complex made by movers infected within the complex.

For  $X = W, H$  and  $l = 1, 2, \dots, w - 1$ , let  $A^{X,l}$  be the severity, excluding the initial infective, in a complex, where the initial infective in the complex is a mover who belongs to the complex only through its workplace if  $X = W$  (household if  $X = H$ ) and the initial infective infects  $l$  individuals within the complex (either in the workplace or household depending on whether  $X = W$  or  $X = H$ ). Note that if we know the number of infectious contacts the initial infective makes within a complex then the size of the epidemic in the complex is conditionally independent of the initial infective's infectious period. In addition, for  $Y = W, H$  and  $k = 1, 2, \dots, w - 1$ , let  $Z_{Y,k}^{X,l}$  denote the number of type  $(Y, k)$  movers infected in the complex with an initial infective of type- $(X, l)$ . A type- $(Y, k)$  individual is a mover who belongs to another complex through their workplace if  $Y = W$  (household if  $Y = H$ ) and infects  $k$  individuals within their other complex. We are interested in the joint distribution of  $(A^{X,l}, Z^{X,l})$ , where  $\mathbf{Z}^{X,l} = (Z_{H,1}^{X,l}, \dots, Z_{H,h-1}^{X,l}, Z_{W,1}^{X,l}, \dots, Z_{W,w-1}^{X,l})$  is a random vector of length  $w + h - 2$ .

For  $X = W, H$  and  $l = 1, 2, \dots, w - 1$ , let  $\hat{A}^{X,l}$  be the severity (sum of infectious periods), excluding the initial infective, in a *branch* of the infectious clump, where the initial infective in a branch is a mover who belongs to the initial complex in a branch only through its workplace if  $X = W$  (household if  $X = H$ ) and the initial infective infects  $l$  individuals within the complex (either in the workplace or household depending on whether  $X = W$  or  $X = H$ ). A branch of an infectious clump starts with a mover individual who begins a complex epidemic and consists of all subsequent complex epidemics emanating from the initial infectious complex in the branch. By the (asymptotic) branching process of the growth of an infectious clump it follows that

$$\hat{A}^{X,l} \stackrel{D}{=} A^{X,l} + \sum_{(Y,k)} \sum_{i=1}^{Z_{Y,k}^{X,l}} \hat{A}^{Y,k,i}, \quad (12)$$

where  $\hat{A}^{Y,k,1}, \hat{A}^{Y,k,2}, \dots$  are independent and identically distributed copies of  $\hat{A}^{Y,k}$  and sums are equal to 0 if vacuous.

For a given complex structure, i.e. sizes of groups  $1, 2, \dots, 2d + 1$ ,  $\mathbf{Z}^{X,l}$  is a vector of final state random variables defined on a multitype SIR epidemic and the joint Laplace transform generating function of  $(A^{X,l}, \mathbf{Z}^{X,l})$  can be computed using Ball and O'Neill [9], Theorem 5.1, or Ball [3], Theorem 4.2. The unconditional joint Laplace transform generating function of  $(A^{X,l}, \mathbf{Z}^{X,l})$  can then be obtained by taking expectations with respect to the complex structure. For fixed  $\nu \geq 0$ , a set of  $w + h - 2$  non-linear equations which determine  $(\phi_{\hat{A}^{H,1}}(\nu), \dots, \phi_{\hat{A}^{H,h-1}}(\nu), \phi_{\hat{A}^{W,1}}(\nu), \dots, \phi_{\hat{A}^{W,w-1}}(\nu))$  can then be derived using (12) enabling, at least in principle, the  $\phi_{\hat{A}^{X,l}}(\nu)$  to be computed.

Consider the initial individual in the infectious clump. Let  $I$  denote its infectious period, and  $Q_H$  and  $Q_W$  denote the number of individuals the initial infective infects within their household and workplace, respectively. Then  $Q_H|I \sim \text{Bin}(h - 1, 1 - \exp(-\beta_H I))$  and  $Q_W|I \sim \text{Bin}(w - 1, 1 - \exp(-\beta_W I))$ , where  $Q_H$  and  $Q_W$  are conditionally independent given  $I$ . Let  $\tilde{A}_M$  denote the severity of an infectious clump where the initial infective is a mover. Conditional on the triple  $(I, Q_H, Q_W)$ ,

$$\tilde{A}_M \stackrel{D}{=} I + \hat{A}^{H,Q_H} + \hat{A}^{W,Q_W}, \quad (13)$$

with  $I$ ,  $\hat{A}^{H,Q_H}$  and  $\hat{A}^{W,Q_W}$  conditionally independent given  $Q_H$  and  $Q_W$ . (The initial infective is responsible for instigating two branches, originating from each of the complexes to which it belongs.) By conditioning first on  $I$  and then on  $(Q_H, Q_W)$ , the Laplace transform  $\phi_{\tilde{A}_M}(\nu)$  can be computed using (13).

Consider now the case when the initial infective in an infectious clump is a remainder. Then the initial infective belongs to only one complex; in all subsequent complexes in the clump, the initial infective in the complex is a mover, but in this first complex they are a remainder. Let  $A^{R,(j,l)}$  denote the severity, excluding the initial infective, in a complex, where the initial infective in the complex is a remainder who makes contact with  $j$  of its housemates and  $l$  of its workplace colleagues. Note that some individuals can belong to both the same household and workplace as the initial infective so that the number of individuals infected by the initial infective can be less than  $j + l$ . For  $Y = W, H$  and  $k = 1, 2, \dots, w - 1$ , let  $Z_{Y,k}^{R,(j,l)}$  denote the number of type  $(Y, k)$  movers infected in the complex with an initial infective of type  $(R, j, l)$ . Let  $\tilde{A}_R$  denote the severity of an infectious clump where the initial infective is a remainder. Conditional on the triple  $(I, Q_H, Q_W)$ ,

$$\tilde{A}_R \stackrel{D}{=} I + A^{R,(Q_H,Q_W)} + \sum_{(Y,k)} Z_{Y,k}^{R,(Q_H,Q_W)} \sum_{i=1} \hat{A}^{Y,k,i}, \quad (14)$$

where  $\hat{A}^{Y,k,1}, \hat{A}^{Y,k,2}, \dots$  are independent and identically distributed copies of  $\hat{A}^{Y,k}$  and are independent of  $(I, A^{R,(Q_H,Q_W)})$  given  $Z^{R,(Q_H,Q_W)}$ . For each  $(j, l)$ , the joint Laplace transform generating function of  $(A^{R,(j,l)}, \mathbf{Z}^{R,(j,l)})$  can be computed similarly to that of  $(A^{X,l}, \mathbf{Z}^{X,l})$  above, whence  $\phi_{\tilde{A}_R}(\nu)$  can be computed using (14).

Finally we note that  $A$  is a mixture of  $\tilde{A}_M$  and  $\tilde{A}_R$  with mixing probabilities  $\theta$  and  $1 - \theta$ , so

$$\phi_A(\nu) = \theta \phi_{\tilde{A}_M}(\nu) + (1 - \theta) \phi_{\tilde{A}_R}(\nu), \quad (15)$$

and using (4),  $\rho = 1 - \xi$ , where  $\xi$  is the smallest root of  $\phi_A(\beta_G(1 - s)) = s$  in  $[0, 1]$ .

### 5.3.4 Local susceptibility sets

The local susceptibility set  $\mathcal{S}_{i^*}^{(n)}$  of an individual  $i^*$  in  $\mathcal{E}^{(n)}$  can be constructed on a generation basis in an analogous fashion to the construction of  $\mathcal{C}_{i^*}^{(n)}$  in Section 5.3.2. We first construct the susceptibility set of  $i^*$  within the one or two complexes to which  $i^*$  belongs. We then construct the susceptibility sets of all movers (apart from  $i^*$  if it is a mover) in that susceptibility set in their other complex, and so on. An equivalent result to Lemma 5.1 holds. Moreover, the limiting process is necessarily a branching process, even if  $I$  is random, since when considering the susceptibility  $\mathcal{S}_{C,j^*}$  say, of a single individual  $j^*$  within a complex  $C$ , the distribution of  $\mathcal{S}_{C,j^*}$  does not depend on  $I_{j^*}$ .

A similar argument to the proof of Lemma 5.1 shows that  $S^{(n)} \xrightarrow{D} S$  as  $n \rightarrow \infty$ , where  $S$  is the total progeny of a 3-type branching process, denoted by  $\mathcal{B}_S$ , which is defined analogously to  $\mathcal{B}_C$  in Section 5.3.2. For  $X, Y = R, H, W$ , let  $Z_{XY}^S$  be defined similarly to  $Z_{XY}$  in Section 5.3.2 but for the susceptibility set of an individual within a typical complex. Thus, in the above notation,  $Z_{XY}^S$  is the number of type- $Y$  individuals in  $\mathcal{S}_{C,j^*} \setminus \{j^*\}$ , given  $j^*$  is of type  $X$ . For  $X = H, W$ , let  $\hat{Z}_X^S$  be the total number of individuals, including the ancestor, in the branching process  $\hat{\mathcal{B}}_S$  that is defined analogously to  $\hat{\mathcal{B}}_C$  in the obvious fashion. For  $X = R, H, W$  and  $0 \leq s_1, s_2, s_3 \leq 1$ , let

$$g_X^S(s_1, s_2, s_3) = \mathbb{E} \left[ s_1^{Z_{XR}^S} s_2^{Z_{XH}^S} s_3^{Z_{XW}^S} \right]. \quad (16)$$

Then, cf. (9), for each  $s \in [0, 1]$ , the PGFs  $f_{\hat{Z}_H^S}(s)$  and  $f_{\hat{Z}_W^S}(s)$  of  $\hat{Z}_H^S$  and  $\hat{Z}_W^S$ , respectively, are determined by

$$f_{\hat{Z}_H^S}(s) = s g_H^S(s, f_{\hat{Z}_H^S}(s), f_{\hat{Z}_W^S}(s)) \quad \text{and} \quad f_{\hat{Z}_W^S}(s) = s g_W^S(s, f_{\hat{Z}_H^S}(s), f_{\hat{Z}_W^S}(s)). \quad (17)$$

Let  $S$  be the total progeny of  $\mathcal{B}_S$ . Then, cf. (10),

$$f_S(s) = (1 - \theta) s g_R^S(s, f_{\hat{Z}_H^S}(s), f_{\hat{Z}_W^S}(s)) + \theta g_H^S(s, f_{\hat{Z}_H^S}(s), f_{\hat{Z}_W^S}(s)) f_{\hat{Z}_W^S}(s). \quad (18)$$

Recall from (5) that  $R_* = \beta_G \mu_S$ , since (8) implies  $\mu_C = \mu_S$ . Conditioning on the first generation of  $\hat{\mathcal{B}}_S$  or suitable differentiation of (17) yields

$$\begin{pmatrix} \mu_{\hat{Z}_H^S} \\ \mu_{\hat{Z}_W^S} \end{pmatrix} = \begin{pmatrix} 1 + \mu_{Z_{HR}^S} \\ 1 + \mu_{Z_{WR}^S} \end{pmatrix} + \mathbf{M}^S \begin{pmatrix} \mu_{\hat{Z}_H^S} \\ \mu_{\hat{Z}_W^S} \end{pmatrix},$$

where

$$\mathbf{M}^S = \begin{pmatrix} \mu_{Z_{HH}^S} & \mu_{Z_{HW}^S} \\ \mu_{Z_{WH}^S} & \mu_{Z_{WW}^S} \end{pmatrix}. \quad (19)$$

Let  $\zeta^S$  be the maximum eigenvalue (Perron-Frobenius root) of  $\mathbf{M}^S$ . In Section 6.4 we show that  $\zeta^S$  serves as a threshold parameter for the epidemic with  $\pi_G = 0$ , i.e. for the epidemic with only local infection, so we also denote  $\zeta^S$  by  $R_L$ . Thus,

$$R_L (= \zeta^S) = \frac{\mu_{Z_{HH}^S} + \mu_{Z_{WW}^S} + \sqrt{(\mu_{Z_{HH}^S} - \mu_{Z_{WW}^S})^2 + 4\mu_{Z_{HW}^S}\mu_{Z_{WH}^S}}}{2}. \quad (20)$$

Observe that  $\mu_{\hat{Z}_H^S} = \mu_{\hat{Z}_W^S} = \infty$  if  $R_L \geq 1$ , otherwise

$$\begin{pmatrix} \mu_{\hat{Z}_H^S} \\ \mu_{\hat{Z}_W^S} \end{pmatrix} = (I - \mathbf{M}^S)^{-1} \begin{pmatrix} 1 + \mu_{Z_{HR}^S} \\ 1 + \mu_{Z_{WR}^S} \end{pmatrix}.$$

Differentiation of (18) then yields

$$R_* = \begin{cases} \beta_G \left[ 1 - 2\theta + (1 - \theta)(\mu_{Z_{RR}^S} + \mu_{Z_{RH}^S} \mu_{\hat{Z}_H^S} + \mu_{Z_{RW}^S} \mu_{\hat{Z}_W^S}) + \theta(\mu_{\hat{Z}_H^S} + \mu_{\hat{Z}_W^S}) \right] & \text{if } R_L < 1, \\ \infty & \text{if } R_L \geq 1. \end{cases} \quad (21)$$

## 5.4 Numerical computation

We outline computation of  $R_*$  and  $z$  (and hence also  $\rho$  when  $I \equiv 1$ ). Computation of  $\rho$  when  $I$  is non-constant is indicated in Section 5.3.3 above but its implementation is computationally more intensive and intricate, owing to the appreciably larger dimension of the approximation branching process  $\mathcal{B}_C$ .

For  $X = R, H, W$ , let  $\mathbf{Z}_X^S = (Z_{XR}^S, Z_{XH}^S, Z_{XW}^S)$ . Note that  $\mathbf{Z}_X^S$  has finite support, so its joint PGF  $g_X^S(s_1, s_2, s_3)$  and mean vector are easily computed from its joint PMF (probability mass function). Thus, given the joint PMFs of  $\mathbf{Z}_X^S$  ( $X = R, H, W$ ), the PGF  $f_S(s)$  is easily computed using (17) and (18), enabling the limiting final size  $z$  to be computed using (7), and  $R_*$  is easily computed using (21). We outline how the joint PMF of  $\mathbf{Z}_R^S$  may be computed; the joint PMFs of  $\mathbf{Z}_H^S$  and  $\mathbf{Z}_W^S$  may be computed similarly.

Recall the binomially distributed independent random variables  $M_1, M_2, \dots, M_d$  defined in Section 5.3.2. Consider the susceptibility set,  $\mathcal{S}_{C,j^*}$  say, of a single individual  $j^*$  within a typical complex. Given  $j^*$  is a remainder, the number of individuals in the complex in groups  $2j - 1$  and  $2j$  ( $j = 1, 2, \dots, d$ ) are  $n - M_j$  and  $M_j$ , respectively, and the number of individuals in group  $2d + 1$  is  $\sum_{i=1}^d M_i$ . Let  $\mathbf{S}_R^C = (S_{R,1}^C, S_{R,2}^C, \dots, S_{R,2d+1}^C)$ , where  $S_{R,i}^C$  is the number of group- $i$  individuals in the complex susceptibility set  $\mathcal{S}_{C,j^*}$ , excluding  $j^*$  if  $i = 1$ . Let  $\mathbf{M} = (M_1, M_2, \dots, M_d)$  and  $\mathbf{m}$  denote a realisation of  $\mathbf{M}$ . For each possible  $\mathbf{m}$ , the joint PMF of  $\mathbf{S}_R^C | \mathbf{M} = \mathbf{m}$  may be computed using Ball [3], Lemma 4.1. The joint PMF of  $\mathbf{S}_R^C$  may then be computed using the law of total probability and the joint PMF of  $\mathbf{Z}_R^S$  follows easily.

Let  $\boldsymbol{\ell} = (\ell_1, \ell_2, \dots, \ell_{2d+1})$  be a possible realisation of  $\mathbf{S}_R^C$ . Using Ball [3], Lemma 4.1,  $P(\mathbf{S}_R^C = \boldsymbol{\ell} | \mathbf{M} = \mathbf{m})$  takes the form

$$P(\mathbf{S}_R^C = \boldsymbol{\ell} | \mathbf{M} = \mathbf{m}) = \alpha_R(\mathbf{m}, \boldsymbol{\ell}) \beta_R(\boldsymbol{\ell}),$$

where  $\beta_R(\boldsymbol{\ell})$  is defined in terms of multivariate Gontcharoff polynomials (Lefèvre and Picard [19]). The  $\alpha_R(\mathbf{m}, \boldsymbol{\ell})$  are readily computed but the  $\beta_R(\boldsymbol{\ell})$  need to be obtained by solving a triangular system of linear equations, which is numerically prohibitive unless  $d$  and  $n$  are suitably small. However, note that the  $\beta_R(\boldsymbol{\ell})$  are independent of  $\mathbf{m}$  and hence need only computing once.

## 5.5 Proof of Theorem 3.1 when $\pi_G > 0$

We prove here a version of Theorem 3.1 in which  $\log n$  is replaced by  $\kappa_n = \log \log n$ . The result for  $\log n$  follows from the proof of Theorem 3.2 in Section 5.6. We use the construction of the final size of  $\mathcal{E}^{(n)}$  via local infectious clumps outlined in Section 5.2. Let  $Z^{(n)}$  denote the final size of  $\mathcal{E}^{(n)}$  and  $\check{Z}^{(n)}$  be the final size of a modified epidemic,  $\check{\mathcal{E}}^{(n)}$ , in which the size of local infectious clumps is capped at  $\kappa_n + 1$ . The latter can be achieved by growing the local infectious clumps in real time and stopping the growth of a clump if and when its size reaches  $\kappa_n + 1$ . Thus, in  $\check{\mathcal{E}}^{(n)}$ , the local infectious clump of individual  $i$  is  $\check{\mathcal{C}}_i^{(n)}$ , where  $\check{\mathcal{C}}_i^{(n)} \subseteq \mathcal{C}_i^{(n)}$  and  $\check{\mathcal{C}}_i^{(n)}$  has size

$$\check{C}_i^{(n)} = \begin{cases} C_i^{(n)} & \text{if } C_i^{(n)} \leq \kappa_n + 1, \\ \lfloor \kappa_n \rfloor & \text{if } C_i^{(n)} > \kappa_n + 1. \end{cases}$$

Let  $\mathcal{B}^{(n)}$  be the branching process of local infected clumps defined analogously to the branching process  $\mathcal{B}$  in Section 5.2, except  $(C, A, R)$  is distributed as  $(C^{(n)}, A^{(n)}, R^{(n)})$ . Let  $Z_I^{(n)}$  be the total number of infected individuals in  $\mathcal{B}^{(n)}$ , i.e. the sum of the sizes of the local infectious clumps of all individuals in  $\mathcal{B}^{(n)}$ . Let  $\check{\mathcal{B}}^{(n)}$  be a modified version of  $\mathcal{B}^{(n)}$  in which the size of local infectious clumps is capped at  $\kappa_n$  and let  $\check{Z}_I^{(n)}$  be the total number of infected individuals in  $\check{\mathcal{B}}^{(n)}$ . Observe that

$$\mathbb{P}(Z^{(n)} \leq \kappa_n) = \mathbb{P}(\check{Z}^{(n)} \leq \kappa_n) \quad \text{and} \quad \mathbb{P}(Z_I^{(n)} \leq \kappa_n) = \mathbb{P}(\check{Z}_I^{(n)} \leq \kappa_n). \quad (22)$$

We construct coupled realisations of  $\check{\mathcal{E}}^{(n)}$  and  $\check{\mathcal{B}}^{(n)}$  as follows. Let  $\check{\mathcal{G}}^{(n)}$  be the random graph with vertex set  $\mathcal{N}$  in which for distinct  $i, j \in \mathcal{N}$  there is an edge between  $i$  and  $j$  if and only if  $i$  and  $j$  belong to the same complex. For  $(i, j) \in \mathcal{N}^2$ , let  $\rho(i, j)$  be the length of the shortest path between  $i$  and  $j$  in  $\check{\mathcal{G}}^{(n)}$ , with  $\rho(i, i) = 0$ . Let  $J_1^{(n)}, J_2^{(n)}, \dots$  be i.i.d., each uniformly distributed on  $\mathcal{N}$ . The initial infective in  $\check{\mathcal{E}}^{(n)}$  is individual  $J_1^{(n)}$  and for  $k = 1, 2, \dots$ , the  $k^{\text{th}}$  global contact in  $\check{\mathcal{E}}^{(n)}$  is with individual  $J_k^{(n)}$ . Let

$$F_n = \bigcap_{i=1}^{\lfloor \kappa_n \rfloor - 1} \bigcap_{j=i+1}^{\lfloor \kappa_n \rfloor} \{\rho(J_i^{(n)}, J_j^{(n)}) > 2(\kappa_n + 1)\}.$$

If  $F_n$  occurs, then  $\check{\mathcal{C}}_{J_k^{(n)}}^{(n)}$  ( $k = 1, 2, \dots, \lfloor \kappa_n \rfloor$ ) are independent, so  $(\check{C}_{J_k^{(n)}}^{(n)}, \check{R}_{J_k^{(n)}}^{(n)})$  ( $k = 1, 2, \dots, \lfloor \kappa_n \rfloor$ ) can be used to give the clump sizes and offsprings of the first  $\lfloor \kappa_n \rfloor$  individuals in  $\check{\mathcal{B}}^{(n)}$ . (Here,  $\check{R}_i^{(n)}$  is the number of global contacts that emanate from  $\check{\mathcal{C}}_i^{(n)}$ .) If  $F_n^C$  occurs, then we construct a realisation of  $\check{\mathcal{B}}^{(n)}$  independently of  $\check{\mathcal{E}}^{(n)}$ . Now,

$$\mathbb{P}(\check{Z}^{(n)} \leq \kappa_n) = \mathbb{P}(\check{Z}^{(n)} \leq \kappa_n, F_n) + \mathbb{P}(\check{Z}^{(n)} \leq \kappa_n, F_n^C)$$

and

$$\mathbb{P}(\check{Z}_I^{(n)} \leq \kappa_n) = \mathbb{P}(\check{Z}_I^{(n)} \leq \kappa_n, F_n) + \mathbb{P}(\check{Z}_I^{(n)} \leq \kappa_n, F_n^C).$$

By construction, if  $F_n$  occurs, then  $\check{Z}^{(n)} \leq \kappa_n$  if and only if  $\check{Z}_I^{(n)} \leq \kappa_n$ , so

$$\left| \mathbb{P}(\check{Z}^{(n)} \leq \kappa_n) - \mathbb{P}(\check{Z}_I^{(n)} \leq \kappa_n) \right| \leq \mathbb{P}(F_n^C).$$

Note that, since each individual belongs to 1 or 2 complexes and each complex contains at most  $2w$  individuals, for  $K = 1, 2, \dots$ , there are at most  $H(K) = 2 \sum_{i=1}^K (2w)^i \leq 4(2w)^K$  individuals within  $\rho$ -distance  $K$  of a given individual, so

$$P(F_n^C) \leq \frac{\kappa_n(\kappa_n - 1)4(2w)^{2(\kappa_n+1)}}{n} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Thus, recalling (22),

$$\lim_{n \rightarrow \infty} \left| P(Z^{(n)} \leq \kappa_n) - P(Z_I^{(n)} \leq \kappa_n) \right| = 0. \quad (23)$$

Recall that  $\xi$  is the extinction probability of  $\mathcal{B}$  and let  $\xi_n$  be the extinction probability of  $\mathcal{B}^{(n)}$ . Now  $R^{(n)} \xrightarrow{D} R$  as  $n \rightarrow \infty$ , so  $\lim_{n \rightarrow \infty} \xi_n = \xi$  (cf. Britton et al. [15], Lemma 4.1). Hence,

$$\limsup_{n \rightarrow \infty} P(Z_I^{(n)} \leq \kappa_n) \leq \limsup_{n \rightarrow \infty} P(Z_I^{(n)} < \infty) = \limsup_{n \rightarrow \infty} \xi_n = \xi. \quad (24)$$

Also, for fixed  $k = 1, 2, \dots$ ,

$$\liminf_{n \rightarrow \infty} P(Z_I^{(n)} \leq \kappa_n) \geq \liminf_{n \rightarrow \infty} P(Z_I^{(n)} \leq k) = P(Z_I \leq k), \quad (25)$$

where  $Z_I$  is the total number of individuals infected in  $\mathcal{B}$ . Now,  $P(Z_I \leq k) \rightarrow \xi$  as  $k \rightarrow \infty$ , so letting  $k \uparrow \infty$  in (25) yields

$$\liminf_{n \rightarrow \infty} P(Z_I^{(n)} \leq \kappa_n) \geq \xi,$$

which together with (24) implies  $\lim_{n \rightarrow \infty} P(Z_I^{(n)} \leq \kappa_n) = \xi$ . Thus, using (23),

$$\lim_{n \rightarrow \infty} P(Z^{(n)} \leq \kappa_n) = \xi,$$

whence  $\lim_{n \rightarrow \infty} P(Z^{(n)} < \kappa_n) = \rho = 1 - \xi$ , which completes the proof of Theorem 3.1, with  $\log n$  replaced by  $\kappa_n$ .

Note that a similar argument to the above shows that

$$\lim_{n \rightarrow \infty} P(Z^{(n)} \leq k) = P(Z_I \leq k) \quad (k = 1, 2, \dots),$$

so  $Z^{(n)} \xrightarrow{D} Z_I$  as  $n \rightarrow \infty$ . The random variable  $Z_I$  is defective, with a mass at infinity, if  $R_* > 1$ .

## 5.6 Proof of Theorem 3.2 when $\pi_G > 0$

To prove Theorem 3.2, we use the adaptation of the embedding technique of Scalia-Tomba [26, 27] to the general two-levels of mixing setting; see Ball and Neal [8], Section 6.2 and Ball and Sirl [11], Section 3.4.2.

Let  $I_1, I_2, \dots, I_n$  and the random directed graph  $\mathcal{G}^{(n)}$  be as in Section 5.2 and let  $L_1, L_2, \dots, L_n$  be an independent sequence of i.i.d. exponential random variables, each having rate  $\beta_G$ . Given,  $I_1, I_2, \dots, I_n$ ,  $\mathcal{G}^{(n)}$  and  $L_1, L_2, \dots, L_n$ , we construct for each  $t \geq 0$  a realisation of



the final outcome of an epidemic  $\tilde{\mathcal{E}}^{(n)}(t)$  as follows. Note that in this construction  $t$  refers to global infectious pressure and not time. We determine first who is infected globally; specifically, for  $i \in \mathcal{N}$ , individual  $i$  is infected globally if and only if  $L_i \leq t$ . We then use the random directed graph  $\mathcal{G}^{(n)}$  to determine who is subsequently infected locally. Specifically, for  $i \in \mathcal{N}$ , individual  $i$  is infected locally if and only if at least one member of  $i$ 's local susceptibility set  $\mathcal{S}_i^{(n)}$  is infected globally. For  $i \in \mathcal{N}$ , let

$$\begin{aligned}\chi_i^{(n)}(t) &= 1_{\{i \text{ is infected in } \tilde{\mathcal{E}}^{(n)}(t)\}} \\ &= \begin{cases} 1 & \text{if } \min_{j \in \mathcal{S}_i^{(n)}} L_j \leq t, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

Let

$$R_{\bullet}^{(n)}(t) = \sum_{i=1}^n \chi_i^{(n)}(t) \quad \text{and} \quad A_{\bullet}^{(n)}(t) = \sum_{i=1}^n I_i \chi_i^{(n)}(t)$$

be respectively the total size and severity of the epidemic  $\tilde{\mathcal{E}}^{(n)}(t)$ .

To connect with the epidemic  $\mathcal{E}^{(n)}$ , consider an epidemic,  $\hat{\mathcal{E}}^{(n)}$  say, that is initiated by exposing the population to  $T_0^{(n)}$  units of global infectious pressure, so each individual is exposed to  $\bar{T}_0^{(n)} = n^{-1}T_0^{(n)}$  units of global infectious pressure. These  $T_0^{(n)}$  units of global infectious pressure may infect some individuals globally and the consequent local spread will lead to  $A_{\bullet}^{(n)}(\bar{T}_0^{(n)})$  further units of global infectious pressure. Thus, the population is now exposed to  $T_0^{(n)} + A_{\bullet}^{(n)}(\bar{T}_0^{(n)})$  units of global infectious pressure and the process can be continued in the obvious fashion. For  $k = 0, 1, \dots$ , let

$$\bar{T}_{k+1}^{(n)} = \bar{T}_0^{(n)} + A_{\bullet}^{(n)}(\bar{T}_k^{(n)}).$$

Then  $k_*^{(n)} = \min\{k : \bar{T}_{k+1}^{(n)} = \bar{T}_k^{(n)}\}$  is well-defined as the population is finite. Let  $T_{\infty}^{(n)} = T_{k_*}^{(n)}$  and  $\bar{T}_{\infty}^{(n)} = n^{-1}T_{\infty}^{(n)}$ . Note that

$$\bar{T}_{\infty}^{(n)} = \min\{t \geq 0 : t = \bar{T}_0^{(n)} + \bar{A}_{\bullet}^{(n)}(t)\}, \quad (26)$$

where  $\bar{A}_{\bullet}^{(n)}(t) = n^{-1}A_{\bullet}^{(n)}(t)$ . The total size and severity of the epidemic  $\hat{\mathcal{E}}^{(n)}$  are given by  $R_{\bullet}^{(n)}(\bar{T}_{\infty}^{(n)})$  and  $A_{\bullet}^{(n)}(\bar{T}_{\infty}^{(n)})$ , respectively.

The epidemics  $\mathcal{E}^{(n)}$  and  $\hat{\mathcal{E}}^{(n)}$  have different laws, since  $\mathcal{E}^{(n)}$  is initiated by a single individual chosen uniformly at random from the population being externally infected and the number of initial infectives in  $\hat{\mathcal{E}}^{(n)}$ ,  $\hat{Z}_{\text{init}}^{(n)}$  say, is distributed as  $R_{\bullet}^{(n)}(\bar{T}_0^{(n)})$ . However, owing to the lack-of-memory property of the exponential distribution,  $\mathcal{E}^{(n)}$  has the same law as  $\hat{\mathcal{E}}^{(n)}$  conditioned upon  $\hat{Z}_{\text{init}}^{(n)} = 1$ . Suppose that  $\mathbb{P}(T_0^{(n)} = 1) = 1$ . Then,  $\hat{Z}_{\text{init}}^{(n)} \sim \text{Bin}(n, 1 - \exp(-\beta_G/n))$ , so

$$\mathbb{P}(\hat{Z}_{\text{init}}^{(n)} = 1) = n \exp(-(n-1)\beta_G/n)(1 - \exp(-\beta_G/n)) \rightarrow \beta_G \exp(-\beta_G) \quad \text{as } n \rightarrow \infty.$$

Thus,  $\liminf_{n \rightarrow \infty} \mathbb{P}(\hat{Z}_{\text{init}}^{(n)} = 1) > 0$  and (cf. Jansen [17]) it follows that convergence in probability results for  $\hat{\mathcal{E}}^{(n)}$  hold also for  $\hat{\mathcal{E}}^{(n)} | \hat{Z}_{\text{init}}^{(n)} = 1$ , and hence for  $\mathcal{E}^{(n)}$ .

Recall the random variable  $S$  defined at (3) in Section 5.2, describing the size of a typical local susceptibility set in the limit  $n \rightarrow \infty$ . For  $t \geq 0$ , let  $r(t) = 1 - f_S(e^{-\beta_G t})$  and  $\bar{R}_{\bullet}^{(n)}(t) = n^{-1}R_{\bullet}^{(n)}(t)$ .

**Lemma 5.2.** *Suppose that  $\text{var}(I) < \infty$ . Then, for any  $t_0 > 0$ ,*

$$\sup_{t_0 \leq t < \infty} |\bar{R}_{\bullet}^{(n)}(t) - r(t)| \xrightarrow{P} 0 \quad \text{and} \quad \sup_{t_0 \leq t < \infty} |\bar{A}_{\bullet}^{(n)}(t) - r(t)| \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty. \quad (27)$$

*Proof.* We prove the result for  $\bar{A}_{\bullet}^{(n)}(t)$ . The result for  $\bar{R}_{\bullet}^{(n)}(t)$  is proved similarly. Recall that we assume  $E[I] = 1$ .

For  $t \geq 0$ , let

$$\begin{aligned} a^{(n)}(t) &= E[\bar{A}_{\bullet}^{(n)}(t)] = E[I_1 \chi_1^{(n)}(t)] \\ &= E[I_1(1 - \exp(-\beta_G t S_1^{(n)}))] \\ &= 1 - E[\exp(-\beta_G t S_1^{(n)})], \end{aligned}$$

since  $I_1$  and  $S_1^{(n)}$  are independent. We show below that, for each  $t \in (0, \infty)$ ,  $\bar{A}_{\bullet}^{(n)}(t) - a^{(n)}(t) \xrightarrow{P} 0$  as  $n \rightarrow \infty$ . Now  $a^{(n)}(t) \rightarrow r(t)$  as  $n \rightarrow \infty$ , since  $S_1^{(n)} \xrightarrow{D} S$  as  $n \rightarrow \infty$  and  $0 \leq \exp(-\beta_G t S_1^{(n)}) \leq 1$  for all  $n$ . Thus,  $\bar{A}_{\bullet}^{(n)}(t) \xrightarrow{P} r(t)$  as  $n \rightarrow \infty$ . This convergence holds also for  $t = \infty$ , since  $\bar{A}_{\bullet}^{(n)}(\infty) = n^{-1} \sum_{i=1}^n I_i \xrightarrow{P} 1 = r(\infty)$  as  $n \rightarrow \infty$  by the weak law of large numbers. The random function  $\bar{A}_{\bullet}^{(n)}(t)$  is non-decreasing in  $t$  on  $[t_0, \infty]$  and  $r(t)$  is continuous and increasing on  $[t_0, \infty]$ . The second result in (27) then follows by a similar argument to the proof of Ball and Lyne [6], Lemma 5.1, see Ball and Sirl [11], page 187.

To complete the proof, we show that for each  $t \in (0, \infty)$ ,  $\bar{A}_{\bullet}^{(n)}(t) - a^{(n)}(t) \xrightarrow{P} 0$  as  $n \rightarrow \infty$ . By Chebyshev's inequality

$$P(|\bar{A}_{\bullet}^{(n)}(t) - a^{(n)}(t)| > \epsilon) \leq \frac{\text{var}(\bar{A}_{\bullet}^{(n)}(t))}{\epsilon^2} = \frac{\text{var}(A_{\bullet}^{(n)}(t))}{n^2 \epsilon^2},$$

so a sufficient condition for  $\bar{A}_{\bullet}^{(n)}(t) - a^{(n)}(t) \xrightarrow{P} 0$  as  $n \rightarrow \infty$  is

$$\lim_{n \rightarrow \infty} n^{-2} \text{var}(A_{\bullet}^{(n)}(t)) = 0. \quad (28)$$

For  $t \geq 0$ , let  $\tilde{\chi}_i^{(n)}(t) = 1 - \chi_i^{(n)}(t)$  and  $\tilde{A}_{\bullet}^{(n)}(t) = \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t)$ . Let  $\sigma_I^2 = \text{var}(I)$ . Noting that  $A_{\bullet}^{(n)}(t) + \tilde{A}_{\bullet}^{(n)}(t) = \sum_{i=1}^n I_i$ , we have that

$$n\sigma_I^2 = \text{var}(A_{\bullet}^{(n)}(t)) + \text{var}(\tilde{A}_{\bullet}^{(n)}(t)) + 2\text{cov}(A_{\bullet}^{(n)}(t), \tilde{A}_{\bullet}^{(n)}(t)).$$

Further,

$$\begin{aligned} \text{cov}(A_{\bullet}^{(n)}(t), \tilde{A}_{\bullet}^{(n)}(t)) &= \text{cov}\left(\sum_{i=1}^n I_i(1 - \tilde{\chi}_i^{(n)}(t)), \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t)\right) \\ &= \text{cov}\left(\sum_{i=1}^n I_i, \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t)\right) - \text{var}(\tilde{A}_{\bullet}^{(n)}(t)), \end{aligned}$$

so (28) holds if

$$\lim_{n \rightarrow \infty} n^{-2} \text{var}(\tilde{A}_{\bullet}^{(n)}(t)) = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} n^{-2} \text{cov}\left(\sum_{i=1}^n I_i, \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t)\right) = 0. \quad (29)$$

For  $t \geq 0$ , let  $\hat{A}_\bullet^{(n)}(t) = \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t) 1_{\{S_i^{(n)} \leq \kappa_n\}}$ , where  $\kappa_n = \log \log n$  as in Section 5.5. Then, using the independence of  $I_i$  and  $S_i^{(n)}$  and recalling that  $E[I] = 1$ ,

$$\begin{aligned} \left| E[\tilde{A}_\bullet^{(n)}(t)] - E[\hat{A}_\bullet^{(n)}(t)] \right| &= E \left[ \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t) 1_{\{S_i^{(n)} > \kappa_n\}} \right] \\ &= \sum_{i=1}^n E \left[ I_i \exp(-\beta_G t S_i^{(n)}) 1_{\{S_i^{(n)} > \kappa_n\}} \right] \\ &\leq n g_n(t), \end{aligned} \tag{30}$$

where  $g_n(t) = \exp(-\beta_G t \kappa_n)$ . Further,

$$\begin{aligned} E \left[ \tilde{A}_\bullet^{(n)}(t)^2 \right] &= E \left[ \sum_{i=1}^n I_i \tilde{\chi}_i^{(n)}(t) \sum_{j=1}^n I_j \tilde{\chi}_j^{(n)}(t) \right] \\ &= \sum_{i=1}^n \sum_{j=1}^n E \left[ I_i I_j \exp(-\beta_G t |S_i^{(n)} \cup S_j^{(n)}|) \right]. \end{aligned}$$

Thus,

$$\begin{aligned} \left| E[\tilde{A}_\bullet^{(n)}(t)^2] - E[\hat{A}_\bullet^{(n)}(t)^2] \right| &= \sum_{i=1}^n \sum_{j=1}^n E \left[ I_i I_j \exp(-\beta_G t |S_i^{(n)} \cup S_j^{(n)}|) \left( 1 - 1_{\{S_i^{(n)} \leq \kappa_n\}} 1_{\{S_j^{(n)} \leq \kappa_n\}} \right) \right] \\ &\leq \sum_{i=1}^n \sum_{j=1}^n E \left[ I_i I_j \exp(-\beta_G t |S_i^{(n)} \cup S_j^{(n)}|) \left( 1_{\{S_i^{(n)} > \kappa_n\}} + 1_{\{S_j^{(n)} > \kappa_n\}} \right) \right] \\ &\leq 2g_n(t) \sum_{i=1}^n \sum_{j=1}^n E[I_i I_j] \\ &= 2g_n(t) [n\sigma_I^2 + n^2]. \end{aligned} \tag{31}$$

It follows from (30) and (31) that, for  $t > 0$ ,

$$\lim_{n \rightarrow \infty} n^{-2} \text{var}(\tilde{A}_\bullet^{(n)}(t)) = 0 \iff \lim_{n \rightarrow \infty} n^{-2} \text{var}(\hat{A}_\bullet^{(n)}(t)) = 0, \tag{32}$$

since  $g_n(t) \rightarrow 0$  as  $n \rightarrow \infty$ .

Note that, by exchangeability,

$$\begin{aligned} n^{-2} \text{var}(\hat{A}_\bullet^{(n)}(t)) &= n^{-2} \sum_{i=1}^n \sum_{j=1}^n \text{cov} \left( I_i \tilde{\chi}_i^{(n)}(t) 1_{\{S_i^{(n)} \leq \kappa_n\}}, I_j \tilde{\chi}_j^{(n)}(t) 1_{\{S_j^{(n)} \leq \kappa_n\}} \right) \\ &= n^{-1} \sum_{j=1}^n \text{cov} \left( I_1 \tilde{\chi}_1^{(n)}(t) 1_{\{S_1^{(n)} \leq \kappa_n\}}, I_j \tilde{\chi}_j^{(n)}(t) 1_{\{S_j^{(n)} \leq \kappa_n\}} \right) \\ &= \text{cov} \left( I_1 \tilde{\chi}_1^{(n)}(t) 1_{\{S_1^{(n)} \leq \kappa_n\}}, I_J \tilde{\chi}_J^{(n)}(t) 1_{\{S_J^{(n)} \leq \kappa_n\}} \right), \end{aligned} \tag{33}$$

where  $J$  is chosen uniformly at random from  $\mathcal{N}$ .

Recall from Section 5.5 the graph  $\tilde{\mathcal{G}}^{(n)}$  and the associated distance function  $\rho(i, j)$ . Let  $X_i^{(n)}(t) = I_i \tilde{X}_i^{(n)}(t) 1_{\{S_i^{(n)} \leq \kappa_n\}}$  ( $i \in \mathcal{N}$ ) and  $U^{(n)} = 1_{\{\rho(1, J) \leq 3\kappa_n\}}$ . Then,

$$\begin{aligned} \text{cov} \left( X_1^{(n)}(t), X_J^{(n)}(t) \right) &= \mathbb{E} \left[ \text{cov} \left( X_1^{(n)}(t), X_J^{(n)}(t) \mid U^{(n)} \right) \right] \\ &\quad + \text{cov} \left( \mathbb{E} \left[ X_1^{(n)}(t) \mid U^{(n)} \right], \mathbb{E} \left[ X_J^{(n)}(t) \mid U^{(n)} \right] \right). \end{aligned}$$

Let  $p_n = \mathbb{P}(U^{(n)} = 1)$  and note that  $p_n \leq \frac{H(3\kappa_n)}{n}$ , where  $H(K) \leq 4(2w)^K$ ; see just before (23) in Section 5.5. Hence,  $p_n \rightarrow 0$  as  $n \rightarrow \infty$ . Further,  $X_1^{(n)}(t)$  and  $X_J^{(n)}(t)$  are independent if  $U^{(n)} = 0$ , since  $\{i \in \mathcal{N} : \rho(1, i) \leq \kappa_n + 1\}$  and  $\{j \in \mathcal{N} : \rho(j, J) \leq \kappa_n + 1\}$  are disjoint. Thus,

$$\begin{aligned} \left| \mathbb{E} \left[ \text{cov} \left( X_1^{(n)}(t), X_J^{(n)}(t) \mid U^{(n)} \right) \right] \right| &\leq p_n \left| \text{cov} \left( X_1^{(n)}(t), X_J^{(n)}(t) \mid U^{(n)} = 1 \right) \right| \\ &\leq p_n \mathbb{E}[I^2] \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Also,

$$\begin{aligned} \left| \text{cov} \left( \mathbb{E} \left[ X_1^{(n)}(t) \mid U^{(n)} \right], \mathbb{E} \left[ X_J^{(n)}(t) \mid U^{(n)} \right] \right) \right| \\ &= p_n(1 - p_n) \left( \mathbb{E} \left[ X_J^{(n)}(t) \mid U^{(n)} = 1 \right] - \mathbb{E} \left[ X_J^{(n)}(t) \mid U^{(n)} = 0 \right] \right)^2 \\ &\leq p_n(1 - p_n) \mathbb{E}[I]^2 \rightarrow 0 \quad \text{as } n \rightarrow \infty. \end{aligned}$$

Thus,  $\text{cov} \left( X_1^{(n)}(t), X_J^{(n)}(t) \right) \rightarrow 0$  as  $n \rightarrow \infty$ , so the first condition in (29) is satisfied using (32) and (33). The second condition in (29) is proved similarly, so the details are left to the reader.  $\square$

**Remark 5.3.** Note that  $\bar{A}_\bullet^{(n)}(0) = a^{(n)}(0) = 0$ , so (27) holds with  $t_0 = 0$  if  $\lim_{t \downarrow 0} r(t) = 0$ , a necessary and sufficient condition for which is  $R_L \leq 1$ .

Suppose that  $\mathbb{P}(T_0^{(n)} = 1) = 1$  and that  $R_* > 1$ . Note that the equation  $r(t) = t$  is the same equation satisfied by  $z$  at (7) in Section 5.2. Now,  $r(0) = 0$ ,  $r'(0) = R_* > 1$  and  $r''(t) < 0$  for  $t \geq 0$ . Thus,  $r(t)$  is concave on  $[0, \infty)$ . Further,  $r(\infty) = 1$ , so there exists a unique  $z \in (0, \infty)$  satisfying  $r(z) = z$  and  $r'(z) < 1$ . (If  $r'(z) = 1$  then  $r(z) = \int_0^z r'(x) dx > \int_0^z 1 dx = z$ , contradicting  $r(z) = z$ .) It follows from (26) and the second result in (27) that  $\min(\bar{T}_\infty^{(n)}, |\bar{T}_\infty^{(n)} - z|) \xrightarrow{\mathbb{P}} 0$  as  $n \rightarrow \infty$ , and hence from the first result in (27) that

$$\min(\bar{R}_\bullet^{(n)}(\bar{T}_\infty^{(n)}), |\bar{R}_\bullet^{(n)}(\bar{T}_\infty^{(n)}) - z|) \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty. \quad (34)$$

Recall that  $\bar{R}_\bullet^{(n)}(\bar{T}_\infty^{(n)}) = n^{-1} R_\bullet^{(n)}(\bar{T}_\infty^{(n)})$ , where  $R_\bullet^{(n)}(\bar{T}_\infty^{(n)})$  is the total size of the epidemic  $\hat{\mathcal{E}}^{(n)}$ , and that convergence in probability results that hold for  $\hat{\mathcal{E}}^{(n)}$ , hold also for  $\mathcal{E}^{(n)}$ . Let  $\bar{Z}^{(n)} = n^{-1} Z^{(n)}$ , where  $Z^{(n)}$  is the total size of the epidemic  $\mathcal{E}^{(n)}$ . It follows from the above

that  $\bar{Z}^{(n)} \xrightarrow{P} \bar{Z}$  as  $n \rightarrow \infty$ , where the random variable  $\bar{Z}$  has support  $\{0, z\}$ . All that remains is to determine  $P(\bar{Z} = 0)$ . Using Theorem 3.1, with  $\log n$  replaced by  $\kappa_n$ ,

$$\liminf_{n \rightarrow \infty} P(\bar{Z}^{(n)} \leq \frac{z}{2}) \geq \liminf_{n \rightarrow \infty} P(Z^{(n)} \leq \kappa_n) = \xi. \quad (35)$$

To obtain an upper bound, first note using (34) that, for any  $\epsilon \in (0, z/2)$ ,

$$\limsup_{n \rightarrow \infty} P(\bar{Z}^{(n)} \leq \frac{z}{2}) = \limsup_{n \rightarrow \infty} P(\bar{Z}^{(n)} \leq \epsilon). \quad (36)$$

For  $K = 1, 2, \dots$ , let  $\check{\mathcal{E}}_K^{(n)}$  be a modification of the epidemic  $\mathcal{E}^{(n)}$  in which the size of all local infectious clumps is capped at  $K$  and let  $\check{Z}_K^{(n)}$  be the total size of  $\check{\mathcal{E}}_K^{(n)}$ . Clearly,

$$P(\bar{Z}^{(n)} \leq \epsilon) \leq P(\check{Z}_K^{(n)} \leq n\epsilon). \quad (37)$$

To derive an upper bound for  $P(\check{Z}_K^{(n)} \leq n\epsilon)$ , we extend a technique which dates back to Whittle [28]. A realisation of  $\check{\mathcal{E}}_K^{(n)}$  can be constructed in an analogous fashion to that of  $\check{\mathcal{E}}^{(n)}$  in Section 5.5, by using a sequence  $J_1^{(n)}, J_2^{(n)}, \dots$  of i.i.d. random variables that are uniformly distributed on  $\mathcal{N}$  to determine the individuals contacted by successive global contacts. A lower bound process may be obtained by deleting any global contact (and ignoring the consequent local infectious clump) that is with an individual whose  $\rho$ -distance from any previously globally contacted individual (including the initial infective) is  $\leq 2K$ . While the total size of  $\check{\mathcal{E}}_K^{(n)}$  is  $\leq \epsilon n$ , the probability that a global contact is deleted is bounded above by  $\frac{\epsilon n H(2K)}{n} \leq 4\epsilon(2w)^{2K}$ . Thus,

$$P(\check{Z}_K^{(n)} \leq n\epsilon) \leq P(\check{Z}_{I,K,\epsilon}^{(n)} \leq n\epsilon), \quad (38)$$

where  $\check{Z}_{I,K,\epsilon}^{(n)}$  is the total number of individuals infected in the branching process  $\check{\mathcal{B}}_{K,\epsilon}^{(n)}$ , whose law is similar to that of  $\mathcal{B}^{(n)}$ , except the size of local infectious clumps is capped at  $K$  and each individual (i.e. local infectious clump) is deleted at birth independently with probability  $\min(4\epsilon(2w)^{2K}, 1)$ . Let  $\mathcal{B}_{K,\epsilon}$  be an analogous branching process derived from  $\mathcal{B}$ . Let  $\xi_n(K, \epsilon)$  and  $\xi(K, \epsilon)$  be the extinction probabilities of  $\check{\mathcal{B}}_{K,\epsilon}^{(n)}$  and  $\mathcal{B}_{K,\epsilon}$ , respectively, and note that  $\lim_{n \rightarrow \infty} \xi_n(K, \epsilon) = \xi(K, \epsilon)$ . Using (36), (37) and (38),

$$\limsup_{n \rightarrow \infty} P(\bar{Z}^{(n)} \leq \frac{z}{2}) \leq \limsup_{n \rightarrow \infty} P(\check{Z}_{I,K,\epsilon}^{(n)} \leq n\epsilon) \leq \limsup_{n \rightarrow \infty} \xi_n(K, \epsilon) = \xi(K, \epsilon). \quad (39)$$

Now  $\xi(K, \epsilon) \downarrow \xi(K)$  as  $\epsilon \downarrow 0$ , where  $\xi(K)$  is the extinction probability for the branching process derived from  $\mathcal{B}$  in which clumps are capped at size  $K$ , and  $\xi(K) \downarrow \xi$  as  $K \uparrow \infty$ . Thus, (35) and (39) yield

$$\lim_{n \rightarrow \infty} P(\bar{Z}^{(n)} \leq \frac{z}{2}) = \xi,$$

whence  $P(\bar{Z} = 0) = \xi = 1 - P(\bar{Z} = z)$ .

Recall from the end of Section 5.5 that  $Z^{(n)} \xrightarrow{D} Z_I$  as  $n \rightarrow \infty$ . It follows easily from the above that, if  $(v_n)$  is any sequence of nonnegative integers satisfying  $v_n \rightarrow \infty$  and  $n^{-1}v_n \rightarrow 0$  as  $n \rightarrow \infty$ , then

$$\lim_{n \rightarrow \infty} P(Z^{(n)} \geq v_n) = 1 - \xi$$

and, if  $R_* > 1$ ,

$$\bar{Z}^{(n)} | Z^{(n)} \geq v_n \xrightarrow{P} z \quad \text{as } n \rightarrow \infty.$$

## 6 Derivations and proofs when $\pi_G = 0$

### 6.1 Introduction and overview

In this section we consider the special case where  $\pi_G = 0$ , so that all infection is local through the households and workplaces. Throughout this section we assume that  $\theta > 0$ , since otherwise the complexes and workplaces coincide and in the absence of global infection the epidemic is restricted to its original workplace. We prove Theorems 3.1 and 3.2 concerning the probability ( $\rho$ ) and final size ( $z$ ) of a major outbreak hold in this setting. Note that  $\rho = 0$  unless the local infectious clumps and local susceptibility sets are supercritical, i.e. unless  $R_L > 1$ , in which case  $z$  and  $\rho$  are given by (44) and (57), respectively.

Let  $i^*$  denote the initial infective. We couple the construction of the local infections clump for individual  $i^*$ ,  $\mathcal{C}_{i^*}^{(n)}$ , to a *forward* branching process  $\mathcal{B}_C$ . Then we show that

$$P(C^{(n)} > \log n) \rightarrow \rho \quad \text{as } n \rightarrow \infty,$$

where  $C^{(n)} = |\mathcal{C}_{i^*}^{(n)}|$  and  $\rho$  is the non-extinction probability of  $\mathcal{B}_C$  given by (57). Furthermore, we show that there exists  $\delta' > 0$  such that

$$P(C^{(n)} > \delta'n | C^{(n)} > \log n) \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (40)$$

In the absence of global infection,  $Z^{(n)} = C^{(n)}$ .

Similarly, we couple the construction of the local infections susceptibility set,  $\mathcal{S}^{(n)}$ , to a *backward* branching process  $\mathcal{B}_S$ . We show that

$$P(S^{(n)} > \log n) \rightarrow z \quad \text{as } n \rightarrow \infty, \quad (41)$$

where  $z$  is the non-extinction probability of  $\mathcal{B}_S$  given by (44). Consider a typical initially susceptible individual  $j^*$ , with local susceptibility set  $\mathcal{S}_{j^*}^{(n)}$  having size  $S_{j^*}^{(n)} \sim S^{(n)}$ . If  $S_{j^*}^{(n)} < \log n$ , then with high probability (i.e. with probability tending to one as  $n \rightarrow \infty$ ),  $\mathcal{S}_{j^*}^{(n)}$  does not contain  $i^*$ , so  $j^*$  is not infected by the epidemic. Suppose instead that  $S_{j^*}^{(n)} \geq \log n$ . Then, in the event of a major outbreak, it follows from (40) that  $C^{(n)} > \delta'n$  with high probability, and consequently  $\mathcal{S}_{j^*}^{(n)} \cap \mathcal{C}^{(n)}$  is non-empty with high probability and  $j^*$  is infected by the epidemic. The events  $S_{j_1^*}^{(n)} < \log n$  and  $S_{j_2^*}^{(n)} < \log n$ , where  $j_1^*$  and  $j_2^*$  are chosen uniformly at random from the initial susceptibles, are asymptotically independent and consequently  $n^{-1}Z^{(n)} | Z^{(n)} > \log n \xrightarrow{P} z$  as  $n \rightarrow \infty$ . This is the intuition underlying the proof. Rather than considering whether  $S_{j^*}^{(n)} < \log n$ , it is convenient to consider whether  $\mathcal{S}_{j^*}^{(n)}$  goes extinct before generation  $\ell_n$ , where  $\ell_n = \lceil a \log \log n \rceil$  for suitable choice of  $a > 0$ , cf. Ball et al. [12].

The remainder of this section is organised as follows. In Section 6.2, we derive some asymptotic properties of local infectious clumps and susceptibility sets, focusing mainly on the latter. In Section 6.3, we consider the final outcome of the epidemic. We show first in Lemma 6.3 that the fraction of the population whose local susceptibility set reaches generation  $\ell_n$  converges in probability to  $z$ . Then, in Lemma 6.4, we consider the special

case when  $I \equiv 1$  and show that, conditional upon the occurrence of a major outbreak, the difference between that fraction and  $n^{-1}Z^{(n)}$  converges in probability to 0 as  $n \rightarrow \infty$ , leading to

$$\frac{Z^{(n)}}{n} \Big| Z^{(n)} > \log n \xrightarrow{P} z \quad \text{as } n \rightarrow \infty,$$

i.e. Theorem 3.2 in the case  $\pi_G = 0$  and  $I \equiv 1$ . Finally, in Section 6.4, we extend the proof to the case of general  $I$ .

## 6.2 Asymptotic properties of local infectious clumps and susceptibility sets

For some  $a > 0$ , let  $\ell_n = \lceil a \log \log n \rceil$ , where further requirements on  $a$  will be defined as required. We start by coupling the construction of  $\mathcal{C}^{(n)}$  to  $\mathcal{B}_C$  and  $\mathcal{S}^{(n)}$  to  $\mathcal{B}_S$  over the first  $\ell_n$  generations, where the initial individual is generation 0. Since the maximum size of a complex is  $2w$  (all individuals in the workplace are movers), we have that the number of individuals whose shortest path from a randomly selected individual has length at most  $\ell_n$ , is bounded above by

$$\sum_{i=0}^{\ell_n} 2(2w)^i = 2 \frac{[2w]^{\ell_n+1} - 1}{2w - 1}, \quad (42)$$

where the factor 2 comes from mover individuals belonging to two complexes. For any  $\epsilon > 0$ , the right hand side of (42) is less than  $n^\epsilon$  for all sufficiently large  $n$ . The total number of movers in the population is  $\text{Bin}(n, \theta)$ , so the proportion of the population who are movers is approximately  $\theta$ . Given any single complex contains at most  $2w$  movers it is straightforward using the birthday problem, see, for example, Ball and Donnelly [5], p.4, to show that with probability tending to 1, distinct complexes are used in the first  $\ell_n$  generations of  $\mathcal{S}^{(n)}$  or  $\mathcal{C}^{(n)}$ . In other words, the growth of the population from a randomly selected individual is, with high probability, a tree-like configuration of complexes over the first  $\ell_n$  generations *c.f.* the proof of Lemma 5.1. This can be taken further. Let  $\mathcal{K}^{(n)} = \mathcal{K}_S^{(n)} \cup \mathcal{K}_C^{(n)}$  be a set of  $K = K_S + K_C \in \mathbb{N}$  randomly selected individuals in a population of size  $n$ , where for the  $K_S$  ( $K_C$ ) individuals in  $\mathcal{K}_S^{(n)}$  ( $\mathcal{K}_C^{(n)}$ ) we construct the susceptibility set (clump) to generation  $\ell_n$ . Then with probability tending to 1, no complex is used more than once in the construction of any of the susceptibility sets or clumps for the first  $\ell_n$  generations.

In order to obtain  $z$  we employ a construction of a local susceptibility set similar to the local infectious clump given in Section 5.3.2. In Section 5.3.2 and Lemma 5.1, the focus is on the probability mass function of  $C^{(n)}$  and its limit  $C$ . In the absence of global infection, we simply need to consider whether or not the limiting branching process,  $\mathcal{B}_S$ , goes extinct.

The construction of  $\mathcal{B}_S$  given in Section 5.3.4 can again be used with  $\mathbf{Z}_X^S = (Z_{XH}^S, Z_{XW}^S)$  being the number of new complex susceptibility sets that are added to a local susceptibility set (offspring in the branching process) from a type  $X$  complex susceptibility set. Let  $j^*$  denote the initial individual for whom we are constructing a local susceptibility set. If  $j^*$

is a mover then they belong to two complexes and its offspring distribution is  $\mathbf{Z}_H^S + \mathbf{Z}_W^S$ , where the two random vectors are independent. If  $j^*$  is a remainder, then without loss of generality they belong to group 1 and  $\mathbf{Z}_R^S = (Z_{RH}^S, Z_{RW}^S)$  gives the number of movers in the complex who belong to the susceptibility set of  $j^*$ . Note that, beyond the identity (mover or remainder) of the initial individual, we are not interested in any remainders who belong to the complex susceptibility set as they play no role in the extinction probability. For  $X = R, H, W$  and  $\mathbf{s} \in [0, 1]^2$ , let

$$\bar{g}_X^S(\mathbf{s}) = \mathbb{E} \left[ s_1^{Z_{XH}^S} s_2^{Z_{XW}^S} \right].$$

**Remark 6.1.** Note that  $\bar{g}_X^S(\mathbf{s}) = g_X^S(1, s_1, s_2)$  where  $g_X^S$  is defined in (16). Thus,  $\bar{g}_X^S$  can be computed using methods described in Section 5.4.

Let  $\boldsymbol{\eta}^S = (\eta_H^S, \eta_W^S)$  denote the smallest solution in  $[0, 1]^2$  of

$$\boldsymbol{\eta}^S = (\bar{g}_H^S(\boldsymbol{\eta}^S), \bar{g}_W^S(\boldsymbol{\eta}^S)), \quad (43)$$

where  $\eta_H^S$  ( $\eta_W^S$ ) is the probability that the branching process  $\mathcal{B}_S$  goes extinct from a single type  $H$  ( $W$ ) individual. Then  $z$ , the non-extinction of  $\mathcal{B}_S$ , satisfies

$$1 - z = (1 - \theta)\bar{g}_R^S(\boldsymbol{\eta}^S) + \theta\eta_H^S\eta_W^S. \quad (44)$$

For  $i = 1, 2, \dots$ , let  $\mathbf{V}_i^S = (V_{H,i}^S, V_{W,i}^S)$  denote the number of individuals in the  $i^{\text{th}}$  generation of  $\mathcal{B}_S$ . Similarly, let  $\mathbf{V}_i^{S(n)} = (V_{H,i}^{S(n)}, V_{W,i}^{S(n)})$  denote the number of individuals in the  $i^{\text{th}}$  generation of  $\mathcal{S}^{(n)}$ . Since with probability tending to 1,  $\mathcal{S}^{(n)}$  can be coupled to  $\mathcal{B}_S$  to coincide over the first  $\ell_n$  generations, we have that

$$\mathbb{P} \left( \mathbf{V}_{\ell_n}^{S(n)} \neq \mathbf{V}_{\ell_n}^S \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

We use properties of branching processes to explore  $\mathbf{V}_{\ell_n}^S$ .

The mean offspring matrix,  $\mathbf{M}^S$ , given in (19), with Perron-Frobenius eigenvalue  $\zeta^S (= R_L)$ , determines the behaviour of the branching process  $\mathcal{B}_S$ . Note that  $\boldsymbol{\eta}^S = (1, 1)$  if  $\zeta^S \leq 1$  and  $\boldsymbol{\eta}^S$  is the unique solution of (43) if  $\zeta^S > 1$ . Hence,  $z > 0$  if and only if  $\zeta^S > 1$ . Also, for  $0 < \theta < 1$ , all elements of  $\mathbf{M}^S$  are non-zero and the branching process  $\mathcal{B}_S$  is aperiodic. Therefore by Mode [20], p.19, (8.2), there exists a random variable  $W^S$  with  $\mathbb{P}(W^S = 0) = 1 - z$  and a continuous distribution on  $(0, \infty)$ , such that

$$\lim_{k \rightarrow \infty} \frac{1}{[\zeta^S]^k} \mathbf{V}_k^S = W^S \mathbf{v}^S \quad \text{a.s.}, \quad (45)$$

where  $\mathbf{v}^S > \mathbf{0}$  is the normalised left eigenvector of  $\mathbf{M}^S$  corresponding to the Perron-Frobenius root  $\zeta^S$ . For  $\theta = 1$ , the branching process  $\mathcal{B}_S$  has period 2 and the arguments are easily modified by letting  $\ell_n = \lceil a \log \log n \rceil + 1$  if  $\lceil a \log \log n \rceil$  is odd and using single-type branching process theory.



**Lemma 6.2.** *Suppose that  $\zeta^S > 1$ . Let  $c^S = \log \zeta^S > 0$  and  $a = 5/(2c^S)$ , and recall that  $\ell_n = \lceil a \log \log n \rceil$ . Then*

$$\mathbb{P} \left( [\log n]^2 < \mathbf{V}_{\ell_n}^{S^{(n)}} < [\log n]^3 \right) \rightarrow z \quad \text{as } n \rightarrow \infty, \quad (46)$$

where  $z$  satisfies (44).

*Proof.* Note that

$$[\zeta^S]^{\ell_n} = \exp(c^S \lceil a \log \log n \rceil) \approx \exp \left( \log [\log n]^{c^S a} \right) = [\log n]^{c^S a}.$$

Therefore

$$\mathbf{V}_{\ell_n}^S \approx [\log n]^{c^S a} W^S \mathbf{v}^S,$$

where  $W^S = 0$  if and only if the branching process goes extinct. Therefore, for any  $0 < d_0 < c^S a < d_1$ , conditional on non-extinction of the branching process,  $\mathcal{B}_S$ , we have that for all sufficiently large  $n$ , the number of individuals in generation  $\ell_n$  lies between  $[\log n]^{d_0}$  and  $[\log n]^{d_1}$ . Since  $ac^S = 2.5$ , (46), follows immediately.  $\square$

It is straightforward to adapt the arguments used in the proof of Lemma 6.2 to prove (41). Let  $0 < p_0 \leq 1$  then using (45), we have that the  $\mathbb{P}(\mathbf{V}_{\lceil p_0 \ell_n \rceil}^S \neq \mathbf{0}) \rightarrow z$  as  $n \rightarrow \infty$ . For sufficiently small  $p_0$ , the number of individuals in the first  $p_0 \ell_n$  generations is less than  $[\log n]$  with probability tending to 1 as  $n \rightarrow \infty$ . However, if the susceptibility set reaches generation  $\lceil p_0 \ell_n \rceil$  without going extinct, it will, with probability tending to 1, reach generation  $\ell_n$  without going extinct, as the number of individuals in generation  $\lceil p_0 \ell_n \rceil$  tends to infinity as  $n \rightarrow \infty$ .

### 6.3 Size of major outbreaks

For  $n = 1, 2, \dots$  and  $j = 1, 2, \dots, n$ , let  $\check{S}_j^{(n)}$  denote the number of individuals in generation  $\ell_n$  of the susceptibility set of individual  $j$ . In particular, we are interested in whether or not  $\check{S}_j^{(n)} = 0$ , i.e. whether or not the susceptibility set goes extinct by generation  $\ell_n$ . Let

$$\check{Z}^{(n)} = \sum_{j=1}^n 1_{\{\check{S}_j^{(n)} \neq 0\}}$$

denote the number of susceptibility sets which do not go extinct by generation  $\ell_n$ .

**Lemma 6.3.** *Let  $z$  satisfy (44), then*

$$\frac{\check{Z}^{(n)}}{n} \xrightarrow{\mathbb{P}} z \quad \text{as } n \rightarrow \infty. \quad (47)$$

*Proof.* Using Markov's inequality, (47) holds if for  $m = 1, 2$ ,

$$\mathbb{E} \left[ \left( \frac{\check{Z}^{(n)}}{n} \right)^m \right] \rightarrow z^m \quad \text{as } n \rightarrow \infty. \quad (48)$$

Since the individuals in the population are exchangeable, we have that

$$\mathbb{E} \left[ \frac{\check{Z}^{(n)}}{n} \right] = \mathbb{P} \left( \check{S}_1^{(n)} \neq 0 \right) \rightarrow z \quad \text{as } n \rightarrow \infty.$$

Similarly, we have that

$$\mathbb{E} \left[ \left( \frac{\check{Z}^{(n)}}{n} \right)^2 \right] = \frac{1}{n} \mathbb{P} \left( \check{S}_1^{(n)} \neq 0 \right) + \frac{n-1}{n} \mathbb{P} \left( \check{S}_1^{(n)} \neq 0, \check{S}_2^{(n)} \neq 0 \right), \quad (49)$$

where individuals 1 and 2 denote two individuals selected uniformly at random from the population.

Let  $E_n$  be the event that two randomly selected individuals 1 and 2 are within a distance  $2\ell_n$  of each other, under the metric  $\rho$  defined just after (22). Using (42) it is straightforward to show that, for any  $\epsilon > 0$ , the number of individuals within a distance  $2\ell_n$  of individual 1 is at most  $n^\epsilon$ , for all sufficiently large  $n$ . Hence,  $\mathbb{P}(E_n) \rightarrow 0$  as  $n \rightarrow \infty$ . Conditional upon  $E_n^C$ , the growth of the susceptibility sets of individuals 1 and 2 over the first  $\ell_n$  generations are independent. Therefore

$$\begin{aligned} \mathbb{P} \left( \check{S}_1^{(n)} \neq 0, \check{S}_2^{(n)} \neq 0 \right) &= \mathbb{P} \left( \check{S}_1^{(n)} \neq 0, \check{S}_2^{(n)} \neq 0, E_n \right) + \mathbb{P} \left( E_n^C \right) \prod_{j=1}^2 \mathbb{P} \left( \check{S}_j^{(n)} \neq 0 | E_n^C \right) \\ &\rightarrow z^2 \quad \text{as } n \rightarrow \infty. \end{aligned} \quad (50)$$

The lemma follows since (50) and (49) imply (48) holds for  $m = 2$ .  $\square$

In Lemma 6.4 we show, conditional on  $Z^{(n)} (= C^{(n)}) > \log n$ , *i.e.* a major epidemic outbreak has occurred, that those infected in the epidemic consist mainly of those individuals whose susceptibility set does not go extinct in  $\ell_n$  generations. We prove the lemma under the assumption that  $I \equiv 1$ , a constant infectious period. In this case the construction of infectious clumps and susceptibility sets are identical, and  $\mathbf{V}_{\ell_n}^{C^{(n)}} \stackrel{D}{=} \mathbf{V}_{\ell_n}^{S^{(n)}}$ , where  $\mathbf{V}_{\ell_n}^{C^{(n)}}$  is the number of individuals in the  $\ell_n^{\text{th}}$  generation of the infectious clump. Observe that Lemma 6.4 yields Theorem 3.2 with  $\pi_G = 0$  (recall  $R_L = \zeta^S$ ). In Section 6.4, we show how the arguments are adjusted to general  $I$ .

**Lemma 6.4.** *Suppose that  $I \equiv 1$  and  $\zeta^S > 1$ , and let  $z$  satisfy (44). Then*

$$\frac{C^{(n)}}{n} \mid C^{(n)} > \log n = \frac{Z^{(n)}}{n} \mid Z^{(n)} > \log n \xrightarrow{\mathbb{P}} z, \quad \text{as } n \rightarrow \infty. \quad (51)$$

*Proof.* Note that  $C^{(n)} \stackrel{D}{=} S^{(n)}$ , since  $I \equiv 1$ , so (40) implies  $\lim_{n \rightarrow \infty} \mathbb{P}(C^{(n)} > \log n) = z$ , where  $z > 0$  as  $\zeta^S > 1$ . Hence, Lemma 6.3 implies

$$\frac{\check{Z}^{(n)}}{n} \mid C^{(n)} > \log n \xrightarrow{\mathbb{P}} z \quad \text{as } n \rightarrow \infty. \quad (52)$$

Thus, to prove (51), we show that

$$\frac{\check{Z}^{(n)}}{n} - \frac{C^{(n)}}{n} \mid C^{(n)} > \log n \xrightarrow{\mathbb{P}} 0 \quad \text{as } n \rightarrow \infty,$$

with the lemma then following immediately from  $C^{(n)} = Z^{(n)}$  and (52).

By Markov's inequality, for any  $\epsilon > 0$ ,

$$\mathbb{P} \left( \left| \frac{\check{Z}^{(n)}}{n} - \frac{C^{(n)}}{n} \right| > \epsilon \mid C^{(n)} > \log n \right) \leq \frac{1}{\epsilon} \mathbb{E} \left[ \left| \frac{\check{Z}^{(n)}}{n} - \frac{C^{(n)}}{n} \right| \mid C^{(n)} > \log n \right]. \quad (53)$$

We can bound the expectation on the right-hand side of (53) by

$$\begin{aligned} & \mathbb{E} \left[ \left| \frac{\check{Z}^{(n)}}{n} - \frac{C^{(n)}}{n} \right| \mid C^{(n)} > \log n \right] \\ &= \mathbb{E} \left[ \frac{1}{n} \left| \sum_{j=1}^n \left( 1_{\{\check{S}_j^{(n)} \neq 0\}} - 1_{\{i^* \in \mathcal{S}_j^{(n)}\}} \right) \right| \mid C^{(n)} > \log n \right] \\ &\leq \frac{1}{n} \sum_{j=1}^n \mathbb{E} \left[ \left| 1_{\{\check{S}_j^{(n)} \neq 0\}} - 1_{\{i^* \in \mathcal{S}_j^{(n)}\}} \right| \mid C^{(n)} > \log n \right] \\ &= \mathbb{P} \left( 1_{\{\check{S}_1^{(n)} \neq 0\}} \neq 1_{\{i^* \in \mathcal{S}_1^{(n)}\}} \mid C^{(n)} > \log n \right), \end{aligned} \quad (54)$$

where the final equality in (54) follows from the exchangeability of individuals.

Let  $\check{E}_n$  be the event that the length of the shortest path between individuals 1 and  $i^*$  is less than or equal to  $\ell_n$ . Then  $\mathbb{P}(\check{E}_n) \rightarrow 0$  as  $n \rightarrow \infty$  and

$$\mathbb{P} \left( \check{S}_1^{(n)} = 0, i^* \in \mathcal{S}_1^{(n)} \mid \check{E}_n^C \right) = 0.$$

Hence,  $\mathbb{P} \left( \check{S}_1^{(n)} = 0, i^* \in \mathcal{S}_1^{(n)} \mid C^{(n)} > \log n \right) \rightarrow 0$  as  $n \rightarrow \infty$ .

We complete the proof of Lemma 6.4 by showing that

$$\mathbb{P} \left( \check{S}_1^{(n)} \neq 0, i^* \notin \mathcal{S}_1^{(n)} \mid C^{(n)} > \log n \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty. \quad (55)$$

Following Lemma 6.2 and the subsequent arguments, for  $a = 5/(2c^S)$  and  $\ell_n = \lceil a \log \log n \rceil$ , we have that

$$\mathbb{P} \left( \mathbf{V}_{\ell_n}^{C^{(n)}} > [\log n]^2 \mid C^{(n)} > \log n \right) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

From generation  $\ell_n$  we couple the growth of the infectious clump to a lower bound branching process. As  $n \rightarrow \infty$ , the probability that the total number of movers who belong to complexes only through a household (workplace) exceeds  $\theta n/2$  tends to 1. The number of movers who belong to a given complex only through a household (workplace) is at most  $w$ . Therefore, conditional on the number of movers exceeding  $\theta n/2$ , the probability of selecting a given complex, when selecting a mover who belongs to a complex through a household (workplace) at random, is at most  $2w/(\theta n)$ . It follows that if an infectious clump is comprised of  $K$  complexes then the probability of attempting to add an existing complex from within the clump to the clump is at most  $K \times 2w/(\theta n)$ .

Fix  $0 < \delta < [\zeta^S - 1]/\zeta^S$  and let  $\delta' = \delta\theta/(2w)$ . Construct  $\mathcal{C}_{i^*}^{(n)}$  from generation  $\ell_n$  onwards until  $\delta'n$  complexes have been added to  $\mathcal{C}_{i^*}^{(n)}$  as follows. Consider the growth of the infectious clump from each of the  $\mathbf{V}_{\ell_n}^{C^{(n)}} > [\log n]^2$  individuals in generation  $\ell_n$  in turn.

We choose complexes to add to the infectious clump according to the number of movers in the complex. Whilst fewer than  $\delta'n$  complexes have been added to  $\mathcal{C}_{i^*}^{(n)}$ , the probability we attempt to add an existing complex to the clump is less than  $\delta$ . We construct a lower bound branching process,  $\mathcal{B}_\delta^L$ , where the addition of each complex is culled independently with probability  $\delta$ . Since  $[1 - \delta]\zeta^S > 1$ , the extinction probability,  $z_{L,\delta}$ , of  $\mathcal{B}_\delta^L$  is less than 1. It follows that

$$\mathrm{P}\left(C^{(n)} > \delta'n \mid C^{(n)} > \log n\right) \rightarrow 1 \quad \text{as } n \rightarrow \infty.$$

Given that  $\mathrm{P}(\check{S}_1^{(n)} > [\log n]^2 \mid \check{S}_1^{(n)} \neq 0) \rightarrow 1$  as  $n \rightarrow \infty$ , we have, using the theorem of total probability, that

$$\begin{aligned} \mathrm{P}\left(i^* \notin \mathcal{S}_1^{(n)} \mid \check{S}_1^{(n)} \neq 0, C^{(n)} > \log n\right) &\leq \mathrm{P}\left(i^* \notin \mathcal{S}_1^{(n)} \mid \check{S}_1^{(n)} > [\log n]^2, C^{(n)} > \delta'n\right) \\ &\quad + \mathrm{P}\left(\check{S}_1^{(n)} \leq [\log n]^2 \mid \check{S}_1^{(n)} \neq 0, C^{(n)} > \log n\right) \\ &\quad + \mathrm{P}\left(C^{(n)} \leq \delta'n \mid \check{S}_1^{(n)} \neq 0, C^{(n)} > \log n\right). \end{aligned} \quad (56)$$

Using a similar argument to (50), we have that  $\mathrm{P}(\check{S}_1^{(n)} \neq 0, C^{(n)} > \log n) \rightarrow z^2$  as  $n \rightarrow \infty$ . It is then straightforward to show that the latter two terms on the right hand side of (56) converge to 0 as  $n \rightarrow \infty$ . Finally

$$\mathrm{P}\left(i^* \notin \mathcal{S}_1^{(n)} \mid \check{S}_1^{(n)} > [\log n]^2, C^{(n)} > \delta'n\right) \leq \left[1 - \frac{[\log n]^2}{n}\right]^{\delta'n} \rightarrow 0 \quad \text{as } n \rightarrow \infty,$$

with (55) following immediately, completing the proof of the lemma.  $\square$

## 6.4 Non-constant infectious periods

In the proof of Lemma 6.4 we have assumed  $I \equiv 1$ . We now discuss how the arguments need to be changed for general, random  $I$  using the construction of complex infectious clumps given in Section 5.3.3. (Note that the constructions of susceptibility sets do not change and the coupling of  $\mathcal{S}^{(n)}$  to  $\mathcal{B}^S$  carries over.) In Section 5.3.3, the focus is on the distribution of the severity of an infectious clump whereas here we focus on the probability of extinction of the infectious clump.

Let  $\mathbf{Z}^{X,l}$  be defined, as in Section 5.3.3, as a random vector of length  $w + h - 2$  for the number of movers of each type infected in a complex infectious clump where the initial infective is of type  $(X, l)$ . Let  $\boldsymbol{\eta}^C = (\eta_{H,1}^C, \dots, \eta_{H,h-1}^C, \eta_{W,1}^C, \dots, \eta_{W,w-1}^C)$ , where  $\eta_{X,l}^C$  is the probability of extinction of the branching process starting from a single complex infectious clump of type  $(X, l)$ . Then  $\boldsymbol{\eta}^C$  is the smallest solution in  $[0, 1]^{w+h-2}$  of

$$\boldsymbol{\eta}^C = \mathbf{g}^C(\boldsymbol{\eta}^C) = (g_{H,1}^C(\boldsymbol{\eta}^C), \dots, g_{H,h-1}^C(\boldsymbol{\eta}^C), g_{W,1}^C(\boldsymbol{\eta}^C), \dots, g_{W,w-1}^C(\boldsymbol{\eta}^C))$$

where for  $\mathbf{x} \in [0, 1]^{w+h-2}$ ,

$$g_{X,l}^C(\mathbf{x}) = \mathrm{E} \left[ \prod_{i=1}^{h-1} x_i^{Z_{H,i}^{X,l}} \times \prod_{j=1}^{w-1} x_{j+h-1}^{Z_{W,j}^{X,l}} \right].$$

Then  $g_{X,l}^C(\mathbf{x})$ , the PGF of  $\mathbf{Z}^{(X,l)}$ , can be obtained in a similar manner to the PGF of  $\mathbf{Z}_X^S$  in Section 5.4. By conditioning on the complex structure given by  $M_1, M_2, \dots, M_d$ , defined in Section 5.3.2, and who the initial infective in the complex infects,  $g_{X,l}^C(\mathbf{x})$  can be derived using Ball and O'Neill [9], Theorem 5.1 or Ball [3], Theorem 4.2.

Returning to the initial infective, let  $\mathbf{Z}^R = (Z_{H,1}^R, \dots, Z_{H,h-1}^R, Z_{W,1}^R, \dots, Z_{W,w-1}^R)$  be the number of offspring of each type where the initial individual is a remainder with PGF  $g^R(\mathbf{x})$ . Then we have that

$$1 - \rho = (1 - \theta)g^R(\boldsymbol{\eta}^C) + \theta E [g^{H,Q_H}(\boldsymbol{\eta}^C)g^{W,Q_W}(\boldsymbol{\eta}^C)], \quad (57)$$

with  $g^{H,0}(\boldsymbol{\eta}^C) = g^{W,0}(\boldsymbol{\eta}^C) = 1$ , since from  $(H, 0)$  and  $(W, 0)$  individuals no infectious clump is created.

By coupling  $\mathcal{C}_{i^*}^{(n)}$  to the branching process  $\mathcal{B}_C$  in analogous fashion to  $\mathcal{S}^{(n)}$  to  $\mathcal{B}_S$ , we have that

$$P(C^{(n)} > \log n) \rightarrow \rho \quad \text{as } n \rightarrow \infty,$$

where  $\rho$  satisfies (57).

Let  $\tilde{\mathbf{M}}^C$  be the mean offspring matrix of the branching process  $\mathcal{B}_C$  (excluding the initial generation) and let  $\tilde{\zeta}^C$  be the maximum eigenvalue of  $\tilde{\mathbf{M}}^C$ . For  $0 < \theta < 1$ , we have that the branching process  $\mathcal{B}_C$  is aperiodic and again by Mode [20], p.19, (8.2), there exists a random variable  $W^C$  with  $P(W^C = 0) = 1 - \rho$  and a continuous distribution on  $(0, \infty)$ , such that

$$\lim_{k \rightarrow \infty} \frac{1}{[\tilde{\zeta}^C]^k} \mathbf{V}_k^C = W^C \mathbf{v}^C \quad \text{a.s.},$$

where  $\mathbf{v}^C > \mathbf{0}$  is the normalised left eigenvector of  $\tilde{\mathbf{M}}^C$  corresponding to  $\tilde{\zeta}^C$ . It follows that

$$P(\mathbf{V}_{\ell_n}^{C^{(n)}} > [\log n]^2 | C^{(n)} > \log n) \rightarrow 1 \quad \text{as } n \rightarrow \infty,$$

and employing a lower bound branching process as before,

$$P(C^{(n)} > \delta' n | C^{(n)} > \log n) \rightarrow 1 \quad \text{as } n \rightarrow \infty. \quad (58)$$

As before, the argument is easily modified if  $\theta = 1$ .

We describe how to combine the forward infectious clump with  $w + h - 2$  types with the backward 2-type susceptibility sets. From the proof of Lemma 6.2, there is, conditional on the susceptibility set surviving to generation  $\ell_n$ , at least  $[\log n]^2$  individuals in generation  $\ell_n$  with probability tending to 1 as  $n \rightarrow \infty$ . We grow the local infectious clump  $\mathcal{C}_{i^*}^{(n)}$  from generation  $\ell_n$  in a similar fashion as in the proof of Lemma 6.4. However, each time a mover is added to the infectious clump, we determine first who that mover is by sampling uniformly from the set of available movers, including those of the right type ( $H$  or  $W$ ) in the  $\ell_n^{\text{th}}$  generation of the susceptibility set  $\mathcal{S}_1^{(n)}$ . If the chosen mover belongs to the susceptibility set then individual 1 is infected by the epidemic and we can stop growing the infectious clump. Otherwise we determine the finer type of the mover (number of individuals it infects in its new complex) and carry on growing the infectious clump. In

view of (58), we can then follow similar arguments to the proof of Lemma 6.4 to show that an infectious clump containing at least  $\delta'n$  individuals will infect a susceptibility set which does not go extinct in the first  $\ell_n$  generations with probability tending to 1 as  $n \rightarrow \infty$ .

Finally, we show that the maximum eigenvalues associated with the forward and backward branching process  $\mathcal{B}_C$  and  $\mathcal{B}_S$  are equal, so as is clear on intuitive grounds,  $\lim_{n \rightarrow \infty} \mathbb{P}(C^{(n)} > \log n) > 0$  if and only if  $\lim_{n \rightarrow \infty} \mathbb{P}(S^{(n)} > \log n) > 0$ .

**Lemma 6.5.** *The maximum eigenvalues of  $\tilde{\mathbf{M}}^C$  and  $\mathbf{M}^S$  are equal, i.e.*

$$\tilde{\zeta}^C = \zeta^S. \quad (59)$$

*Proof.* Consider the spread of a clump within a typical single complex and for  $X, Y = H, W$  define  $Z_{XY}^C$  analogously to  $Z_{XY}^S$ . Recall that in  $\mathcal{B}_C$ , type- $H$  individuals are partitioned into type- $(H, 0), (H, 1), \dots, (H, h-1)$  individuals and type- $W$  individuals are partitioned into type- $(W, 0), (W, 1), \dots, (W, w-1)$  individuals. Note that in the spread of an epidemic within a complex, the infectious period of an individual if it becomes infected is independent of the event that it becomes infected. Thus, the finer type of each type- $H$  individual infected in the complex epidemic is independently distributed, with probability  $p_H(i)$  of being type  $(H, i)$ , where

$$p_H(i) = \mathbb{E} \left[ \binom{h-1}{i} \left(1 - e^{-\beta'_H I}\right)^i e^{-(h-1-i)\beta'_H I} \right] \quad (i = 1, 2, \dots, h-1),$$

with  $\beta'_H = \beta_H/(h-1)$ . Similarly, each type- $W$  individual is independently type  $(W, i)$  with probability  $p_W(i)$ , where

$$p_W(i) = \mathbb{E} \left[ \binom{w-1}{i} \left(1 - e^{-\beta'_W I}\right)^i e^{-(w-1-i)\beta'_W I} \right] \quad (i = 1, 2, \dots, w-1),$$

with  $\beta'_W = \beta_W/(w-1)$ . (Although the types of individuals are independent, the type of an individual is not independent of the size of the clump unless  $I$  is non-random.)

For  $i = 1, 2, \dots, h-1$  and  $X = H, W$ , let  $Z_{(H,i),X}^C$  be the number of type- $X$  offspring of a typical type- $(H, i)$  individual in  $\mathcal{B}^C$  and define  $Z_{(W,i),X}^C$  similarly for  $i = 1, 2, \dots, w-1$ . Exploiting the above-mentioned independence of types, the mean numbers of type- $(H, j_1)$  and type- $(W, j_2)$  individuals in the  $k^{\text{th}}$  generation of  $\mathcal{B}^C$  given that the initial individual has type  $(H, i)$  are given by

$$\left( \mu_{Z_{(H,i),H}^C}, \mu_{Z_{(H,i),W}^C} \right) (\mathbf{M}^C)^{k-1} \begin{pmatrix} p_H(j_1) & 0 \\ 0 & p_W(j_2) \end{pmatrix}, \quad (60)$$

where

$$\mathbf{M}^C = \begin{pmatrix} \mu_{Z_{HH}^C} & \mu_{Z_{HW}^C} \\ \mu_{Z_{WH}^C} & \mu_{Z_{WW}^C} \end{pmatrix}.$$

If the initial individual has type  $(W, i)$  then these means are given by

$$\left( \mu_{Z_{(W,i),H}^C}, \mu_{Z_{(W,i),W}^C} \right) (\mathbf{M}^C)^{k-1} \begin{pmatrix} p_H(j_1) & 0 \\ 0 & p_W(j_2) \end{pmatrix}. \quad (61)$$

Note that  $\tilde{\zeta}_C$  equals the asymptotic geometric rate of growth of  $E[\mathbf{V}_k^C]$ . Hence, it follows from (60) and (61) that  $\tilde{\zeta}_C = \zeta_C$ , where  $\zeta_C$  is the maximum eigenvalue of  $\mathbf{M}^C$ . To complete the proof we show that the maximum eigenvalues of  $\mathbf{M}^C$  and  $\mathbf{M}^S$  are equal.

Consider a complex  $\mathcal{C}$  having structure  $\mathbf{A} = \mathbf{a}$ , where  $a_i$  denotes the number of individuals in group  $i$ , and at least two movers but without the restriction that group 2 contains at least one mover. Let  $\mathcal{H}$  be the set of individuals in  $\mathcal{C}$  that belong to groups  $2, 4, \dots, 2d$  and  $\mathcal{W}$  set of individuals in group  $2d+1$ , so  $|\mathcal{H}| = |\mathcal{W}| = a_{2d+1}$ . Construct on  $\mathcal{C}$  a random directed graph  $\mathcal{G}_C$  of potential local contacts in an analogous fashion to  $\mathcal{G}^{(n)}$  in Section 5.2. For  $i \in \mathcal{H} \cup \mathcal{W}$ , let  $\mathcal{C}_i^C = \{j \in \mathcal{C} : i \rightsquigarrow j\}$  and  $\mathcal{S}_i^C = \{j \in \mathcal{C} : j \rightsquigarrow i\}$  be respectively the local infectious clump and local susceptibility set of  $i$  in  $\mathcal{C}$ . For  $X, Y = \mathcal{H}, \mathcal{W}$ , let  $\hat{Z}_{XY}^C(\mathbf{a}) = |\mathcal{C}_i^C \cap Y|$  and  $\hat{Z}_{XY}^S(\mathbf{a}) = |\mathcal{S}_i^C \cap Y|$ , where  $i$  is chosen uniformly at random from  $X$  and the dependence on the complex structure  $\mathbf{a}$  is indicated. Then, for  $X, Y = \mathcal{H}, \mathcal{W}$ ,

$$\begin{aligned} \mu_{\hat{Z}_{XY}^C(\mathbf{a})} &= \frac{1}{a_{2d+1}} \sum_{i \in X} E[|\mathcal{C}_i^C \cap Y|] = \frac{1}{a_{2d+1}} E \left[ \sum_{i \in X} \sum_{j \in Y} 1_{i \rightsquigarrow j} \right] \\ &= \frac{1}{a_{2d+1}} \sum_{j \in Y} E[|\mathcal{S}_j^C \cap X|] = \mu_{\hat{Z}_{YX}^S(\mathbf{a})}. \end{aligned} \quad (62)$$

Observe that (62) continues to hold if for all  $i \in \mathcal{H} \cup \mathcal{W}$ , both  $\mathcal{C}_i^C$  and  $\mathcal{S}_i^C$  do not include  $i$ . With that modification,

$$\mathbf{M}^C = \begin{pmatrix} E[\mu_{\hat{Z}_{\mathcal{H}\mathcal{W}}^C(\mathbf{A})}] & E[\mu_{\hat{Z}_{\mathcal{H}\mathcal{H}}^C(\mathbf{A})}] \\ E[\mu_{\hat{Z}_{\mathcal{W}\mathcal{W}}^C(\mathbf{A})}] & E[\mu_{\hat{Z}_{\mathcal{W}\mathcal{H}}^C(\mathbf{A})}] \end{pmatrix}$$

and

$$\mathbf{M}^S = \begin{pmatrix} E[\mu_{\hat{Z}_{\mathcal{H}\mathcal{W}}^S(\mathbf{A})}] & E[\mu_{\hat{Z}_{\mathcal{H}\mathcal{H}}^S(\mathbf{A})}] \\ E[\mu_{\hat{Z}_{\mathcal{W}\mathcal{W}}^S(\mathbf{A})}] & E[\mu_{\hat{Z}_{\mathcal{W}\mathcal{H}}^S(\mathbf{A})}] \end{pmatrix},$$

where expectations are with respect to the complex structure  $\mathbf{A}$ . Using (62),

$$\mathbf{M}^S = \begin{pmatrix} \mu_{Z_{HH}^S} & \mu_{Z_{HW}^S} \\ \mu_{Z_{WH}^S} & \mu_{Z_{WW}^S} \end{pmatrix} = \begin{pmatrix} \mu_{Z_{WW}^C} & \mu_{Z_{HW}^C} \\ \mu_{Z_{WH}^C} & \mu_{Z_{HH}^C} \end{pmatrix},$$

so  $\mathbf{M}^C$  and  $\mathbf{M}^S$  have the same eigenvalues, whence  $\zeta^C = \zeta^S$ .  $\square$

## 7 Discussion

In this paper we studied an epidemic model for a community having two different group structures, households and workplaces, allowed to partly overlap, and asymptotic properties of the epidemic were derived. For ease of exposition, we have presented our results within the framework of an SIR model in which the infection rates remain constant throughout an infective's infectious period. Theorems 3.1 and 3.2 can be generalised to the case where infectivity varies stochastically over an individual's infective period, with  $I$  denoting an individual's final cumulative infectivity.

It would of course be interesting to generalise the model to make it more realistic. One obvious generalisation would be to allow households as well as workplaces to have different sizes. Such generalisation is not expected to result in any qualitative new insights.

Perhaps more interesting would be to allow individuals to have more heterogeneities: currently the only difference between individuals is their random infectious period. Allowing for varying overall contact rates will have no effect since this could be contained in the infectious period, but allowing how the contact rate divides between household, workplace and community to be different, and also allowing susceptibility to vary, would be an interesting as well as from an applied point of view relevant, extension of the model. Another interesting extension would be to consider non-pharmaceutical interventions (NPIs), typically acting differently on the three types of contact, to see how this alters an ongoing epidemic.

A natural theoretical extension would be to supplement the law of large numbers for the size of major outbreak in Theorem 3.2 with a central limit theorem. Figure 1 in Section 4 suggests that  $Z^{(n)}|Z^{(n)} > \log n$  is asymptotically normal when  $\pi_G > 0$  and the embedding construction in Section 5.6 offers a possible method of proof (cf. Ball et al. [7] for the households model). However, except in the special case  $\theta = 0$ , proving a functional central limit theorem analogue of Lemma 5.2 is not straightforward and a useful expression for the asymptotic variance is difficult to obtain, owing to overlap of complexes.

The numerical illustrations suggest that the final size  $z$  increases with  $\theta$  (the amount of non-overlap between group structures) and workplace size  $w$ . Other numerical calculations (not shown) suggest that  $z$  also increases with  $\pi_G$  (the fraction of contacts that are global) and household size  $h$ . An obvious open problem is to prove these results in generality, or to find counterexamples.

Another direction for future work is investigation of vaccination strategies under our model. Patwardhan et al. [22] use Monte Carlo simulations to show that when a fraction of the population is vaccinated with a perfect vaccine (i.e. one which necessarily renders the recipient fully immune) by selecting individuals for vaccination uniformly at random, there is a greater reduction in both the peak and final size of an outbreak when there is full overlap ( $\theta = 0$ ) than when there is no overlap ( $\theta = 1$ ), provided  $\beta$  is calibrated suitably for the two models. The methodology of this paper can be extended to allow the effect of such vaccination on final size to be investigated more systematically. Other vaccination-related questions could also be explored. For example, is it better to target vaccination at movers or remainers, and what is the optimal vaccine allocation strategy for a given vaccination coverage?

Although the reproduction numbers  $R_*$  and  $R_L$  serve as threshold parameters for the model with  $\pi_G > 0$  and  $\pi_G = 0$ , respectively, they do have some limitations; see Pellis et al. [23] for a discussion of  $R_*$  for the model with no overlap ( $\theta = 1$ ). For example, the fact that  $R_*$  can be infinite means that it can be completely uninformative about the effort that is required to bring an epidemic below threshold. Even though  $R_L$  is necessarily finite, it can be misleading when comparing different models (for example, models with different values of  $\theta$ ), since it is based on the proliferation of infected complexes, rather than infected individuals, and different models may have different complex size distributions. A similar comment applies to  $R_*$ . It is possible to generalise the individual-based reproduction numbers  $R_0$  and  $R_I$ , given for the case  $\theta = 1$  in Ball et al. [10], Sections 4.2 and 4.5, respectively, to general  $\theta$ , though numerical calculation of the former is generally prohibitive. These will be considered in a subsequent paper.



# Acknowledgements

Tom Britton is grateful to the Swedish Research Council (grant 2020-0474) for financial support.

# References

- [1] Andersson, H. (1999) Epidemic models and social networks. *Math. Scientist* **24**, 128–147.
- [2] Ball, F. (1986) A unified approach to the distribution of total size and total area under the trajectory of infectives in epidemic models. *Adv. Appl. Prob.* **18**, 289–310.
- [3] Ball, F. (2019) Susceptibility sets and the final outcome of collective Reed-Frost epidemics. *Methodol. Comput. Appl. Probab.* **21**, 401–421.
- [4] Ball, F., Britton, T., House, T., Isham, V., Mollison, D., Pellis, L. and Scalia Tomba, G. (2015) Seven challenges for metapopulation models of epidemics, including households models. *Epidemics* **10**, 63–67.
- [5] Ball, F. and Donnelly, P. (1995) Strong approximations for epidemic models. *Stoch. Proc. Appl.* **55**, 1–21.
- [6] Ball, F. and Lyne, O.D. (2001) Stochastic multitype SIR epidemics among a population partitioned into households. *Adv. Appl. Prob.* **33**, 99–123.
- [7] Ball, F., Mollison, D. and Scalia-Tomba, G. (1997) Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7**, 46–89.
- [8] Ball, F. and Neal, P. (2002) A general model for stochastic SIR epidemics with two levels of mixing. *Math. Biosci.* **180**, 73–102.
- [9] Ball, F. and O’Neill, P. (1999) The distribution of general final state random variables for stochastic epidemic models. *J. Appl. Prob.* **36**, 473–491.
- [10] Ball, F., Pellis, L. and Trapman, P. (2016) Reproduction numbers for epidemic models with households and other social structures. II. Comparisons and implications for vaccination. *Math. Biosci.* **274**, 108–139.
- [11] Ball, F. and Sirl, D. (2019) Stochastic SIR epidemics in structured populations. In *Stochastic Epidemic Models with Inference*, 123–237, Springer.
- [12] Ball, F., Sirl, D. and Trapman, P. (2014) Epidemics on random intersection graphs. *Ann. Appl. Prob.* **24**, 1081–1128.
- [13] Bansaye, V., Deslandes, F., Kubasch, M. and Vergu, E. (2024) The epidemiological footprint of contact structures in models with two levels of mixing. *arXiv:2303.05287v2*.

- [14] Becker, N. and Dietz, K. (1995) The effect of the household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127**, 207–219.
- [15] Britton, T., Janson, S. and Martin-Löf, A. (2007) Graphs with specified degree distributions, simple epidemics, and local vaccination strategies. *Adv. Appl. Prob.* **39**, 922–948.
- [16] Britton, T., Kypraios, T. and O’Neill, P. (2011) Inference for epidemics with three levels of mixing: methodology and application to a measles outbreak. *Scan. J. Stat.* **38**, 578–599.
- [17] Janson, S. (2009) The probability that a random multigraph is simple. *Comb. Probab. Comput.* **18**, 205–225.
- [18] Kubasch, M. (2023) Large population limit for a multilayer SIR model including households and workplaces. *arXiv:2305.17064v2*.
- [19] Lefèvre, C. and Picard, P. (1990) A non-standard family of polynomials and the final size distribution of Reed-Frost epidemic processes. *Adv. Appl. Prob.* **22**, 25–48.
- [20] Mode, C.J. (1971) *Multitype Branching Processes: Theory and Applications*. Elsevier, New York.
- [21] Ouboter, T., Meester, R. and Trapman, P. (2016) Stochastic SIR epidemics in a population with households and schools. *J. Math. Biol.* **72**, 1177–1193.
- [22] Patwardhan, S., Rao, V. K., Fortunato, S. and Radicchi, F. (2023) Epidemic spreading in group-structured populations. *Phys. Rev. X* **13**, 041054 (10 pages).
- [23] Pellis, L., Ferguson, N.M. and Fraser, C. (2009) Threshold parameters for a model of epidemic spread among households and workplaces. *J. R. Soc. Interface.* **6**, 979–987.
- [24] Pellis, L., Ferguson, N.M. and Fraser, C. (2011) Epidemic growth rate and household reproduction number in communities of households, schools and workplaces. *J. Math. Biol.* **63**, 691–734.
- [25] Picard, P. and Lefèvre, C. (1990) A unified analysis of the final size and severity distribution in collective Reed-Frost epidemic processes. *Adv. Appl. Prob.* **22**, 269–294.
- [26] Scalia-Tomba, G. (1985) Asymptotic final-size distribution for some chain-binomial processes. *Adv. Appl. Prob.* **17**, 477–495.
- [27] Scalia-Tomba, G. (1990) On the asymptotic final size distribution of epidemics in heterogeneous populations. In *Stochastic Processes in Epidemic Theory*, 189–196, Springer.
- [28] Whittle, P. (1955) The outcome of a stochastic epidemic – a note on Bailey’s paper. *Biometrika* **42** 116–122.