

THE REAL-TIME GROWTH RATE OF STOCHASTIC EPIDEMICS ON RANDOM INTERSECTION GRAPHS

CAROLINA FRANSSON,* *Stockholm University*
MADDALENA DONÀ,** *University of Groningen*

Abstract

This paper is concerned with the growth rate of susceptible–infectious–recovered epidemics with general infectious period distribution on random intersection graphs. This type of graph is characterised by the presence of cliques (fully connected subgraphs). We study epidemics on random intersection graphs with a mixed Poisson degree distribution and show that in the limit of large population sizes the number of infected individuals grows exponentially during the early phase of the epidemic, as is generally the case for epidemics on asymptotically unclustered networks. The Malthusian parameter is shown to satisfy a variant of the classical Euler–Lotka equation. To obtain these results we construct a coupling of the epidemic process and a continuous-time multitype branching process, where the type of an individual is (essentially) given by the length of its infectious period. Asymptotic results are then obtained via an embedded single-type Crump–Mode–Jagers branching process.

Keywords: Stochastic SIR epidemic; random intersection graph; cliques; branching process approximation; Malthusian parameter; regenerative branching processes

2010 Mathematics Subject Classification: Primary 60J85; 92D30; 05C80
Secondary 60J80

1. Introduction

In the earliest epidemic models, it is assumed that the disease spreads in a population consisting of indistinguishable individuals exhibiting homogeneous mixing. Since the advent of those early models, there has been considerable interest in incorporating realistic elements from real-world social structures that depart from the simplistic assumption of homogeneity. Such realistic features may take the form both of heterogeneity in social behaviour (some individuals may have a higher proclivity to being socially active than others, or the population may exhibit a more complex social structure than homogeneous mixing) and of biological differences in the ‘susceptibility’ and ‘infectivity’ of individuals.

To give some examples, for deterministic epidemic models this has been manifested through models where the population is stratified into a relatively small number of classes and individuals interact with each other at a rate that is determined by their classes. Individuals may, for

Received 6 September 2022; accepted 3 December 2024.

* Postal address: Stockholms universitet, 106 91 Stockholm, Sweden. Email: c.e.m.fransson@gmail.com

** Postal address: University of Groningen, PO Box 72, 9700 AB Groningen, The Netherlands. Email: m.dona@rug.nl

© The Author(s), 2025. Published by Cambridge University Press on behalf of Applied Probability Trust. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

instance, be spatially separated or stratified by age or sex. This typically gives rise to a system of differential equations, which governs the dynamics of the epidemic [18, 33].

A similar development of increasingly complex social structures has taken place in the field of stochastic epidemic modelling on networks. In particular, a large body of epidemic models that aim to capture the tendency of individuals who know each other to have mutual acquaintances has appeared in the literature. In the context of models where the social network of the population is fully specified by a graph, this means that the graph is clustered (i.e. it contains a considerable number of triangles and other short circuits). Some examples include the great circle model [1, 3, 25] and the closely related small-world network model [34], where individuals typically have both local contacts in a local environment, which exhibits clustering, and global contacts.

In a similar vein, several models that include the presence of small, closely connected groups, or *cliques*, with intense within-clique interactions have been introduced [1, 11]. A clique may, for instance, represent a household, workplace, or school. Models with this feature have been investigated in various forms; see, for instance, [2, 4, 6–8, 30], to name a few.

In the present paper, we study the real-time growth rate of an epidemic that spreads on a random graph whose structure, like that of the above-mentioned models, is characterised by the presence of small (possibly overlapping) highly connected cliques. During the early phase of an epidemic, the number of infectious individuals typically grows exponentially; this is the case for many theoretical models and has also been observed in empirical data [15, 27]. The growth rate is one of the most readily available attributes of an emerging epidemic and it is arguably one of the most natural parameters by which to describe the seriousness of the epidemic. For many models of epidemics on random graphs with clustering, obtaining results that concern the real-time growth rate is, however, more challenging than analysing the final outcome of the epidemic. The reason for this is that results on the final outcome of an epidemic may be obtained without taking the actual chain of transmission into account. This idea was first mentioned in a paper devoted to epidemic modelling [24] but was, however, implicitly present in earlier literature on percolation [14, 17]. For this reason, many models lend themselves more readily to analysis of the final outcome than of the real-time growth rate. In [31] the authors proposed approximate methods for estimating the so-called household reproduction number based on observations of the real-time growth rate in a population structured into small (possibly overlapping) communities, both in the Markovian case and under the arguably strong assumption that the total “infectivity” of an infectious individual and the time points at which the individual transmits the disease are independent. On a related note, [5] provided methods to estimate the within-household infection rate for a susceptible–infectious–recovered (SIR) epidemic among a population of households from the observed real-time growth rate.

As mentioned before, the real-time growth rate of an epidemic in a population with households, schools, and workplaces has previously been studied in [31], where (among other things) heuristic results similar to those presented here were obtained. In this paper we provide rigorous proofs of these results. It is worth pointing out that the methods employed here can be applied to a more general class of random graphs with cliques than the model considered in this paper, and also a more general class of household/school/workplace models than that studied in [31]. In particular, previous results concern only the case where the clique or household sizes are bounded and do not trivially extend to the setting with unbounded clique sizes, whereas the current paper deals with the unbounded case.

The key tool of this paper is a single-type branching process, which we embed in the epidemic process. Our approach is inspired by [19], where a similar embedding was used to

obtain the polynomial rate of convergence of multi-type branching processes. The techniques employed here are also related to what [32] calls regenerative Galton–Watson processes and to the concepts of local infectious clumps and global contacts in [2]; see also [29], which treats multitype branching processes with local dependencies.

To be more specific about the graph model, here we consider the real-time growth rate of epidemics on a random intersection graph [23]. Simply put, a random intersection graph is constructed by dividing the nodes of the graph into groups (a node may belong to zero, one, or several groups) and then connecting nodes that belong to the same group, so that the groups form fully connected (possibly overlapping) subgraphs. Thus, a random intersection graph does, in general, contain a non-negligible number of short circuits, which makes the widely used branching process approximation of the early phase of the epidemic somewhat delicate. Here we consider the real-time growth rate of epidemics on a random intersection graph [23] in which the degree distributions are mixed Poisson. Epidemics on graphs of this type have previously been studied in [9], where expressions for the asymptotic probability of a major outbreak, the final size of a major outbreak, and a threshold parameter were derived. Epidemics on random intersection graphs were also studied in [13], where the clustering of the underlying network is tunable.

This paper is structured as follows. In Section 2.1 we present the notation conventions and abbreviations. Section 2.2 contains an introduction to the underlying graph model, in Section 2.3 we define the epidemic model, and in Section 2.4 we define the approximating branching process. The main results are presented in Section 3. Sections 3.1 and 4 contain some background theory and proofs of the main results.

2. Epidemics on random intersection graphs

2.1. Notation and abbreviations

This section contains a summary of the notation conventions and abbreviations that will be used frequently in this paper.

For any $B \subset \mathbb{R}$ and $x \in \mathbb{R}$ we use the notation $B_{\geq x} = B \cap [x, \infty)$, and $B_{>x}$, $B_{\leq x}$, and $B_{<x}$ are defined analogously.

For $x \in \mathbb{R}$, $\lfloor x \rfloor = \sup \mathbb{Z}_{\leq x}$. For real numbers x and y , $x \vee y = \max(x, y)$ and $\log_+(x) = \log(1 \vee x)$. For any $n \in \mathbb{Z}_{\geq 1}$, $[n] = \{1, \dots, n\}$.

Let $f: \mathbb{R} \rightarrow \mathbb{R}$ and $g: \mathbb{R} \rightarrow \mathbb{R}_{>0}$. We write $f(x) = \mathcal{O}(g(x))$ as $x \rightarrow \infty$ to indicate that $\limsup_{x \rightarrow \infty} |f(x)|/g(x) < \infty$ and $f(x) = o(g(x))$ as $x \rightarrow \infty$ to indicate that $\limsup_{x \rightarrow \infty} |f(x)|/g(x) = 0$. Similarly, $f(x) = \Theta(g(x))$ as $x \rightarrow \infty$ if $f(x) = \mathcal{O}(g(x))$ and $\liminf_{x \rightarrow \infty} |f(x)|/g(x) > 0$.

For a random variable X and an event A , we use the notation $E(X; A) = E(X\mathbf{1}(A))$ where $\mathbf{1}(A)$ is the indicator of A . For a non-negative random variable X , we say that the random variable Y has mixed Poisson distribution with intensity given by (the law of) X if $(Y | X = x) \sim \text{Po}(x)$, and we denote it by $\text{MP}(X)$. For any non-negative integrable random variable X with $\mathbb{E}(X) > 0$, we denote the size-biased version of X by $\bar{X}$, i.e. for any Borel set $B \subset \mathbb{R}$,

$$\mathbb{P}(\bar{X} \in B) = \frac{\mathbb{E}(X; X \in B)}{\mathbb{E}(X)}. \quad (2.1)$$

We will make frequent use of the abbreviations MP (mixed Poisson) and SIR (susceptible $\rightarrow$ infectious $\rightarrow$ recovered). Throughout this paper, G_n denotes a random graph on n vertices generated via the random intersection graph model. We say that an event occurs with high

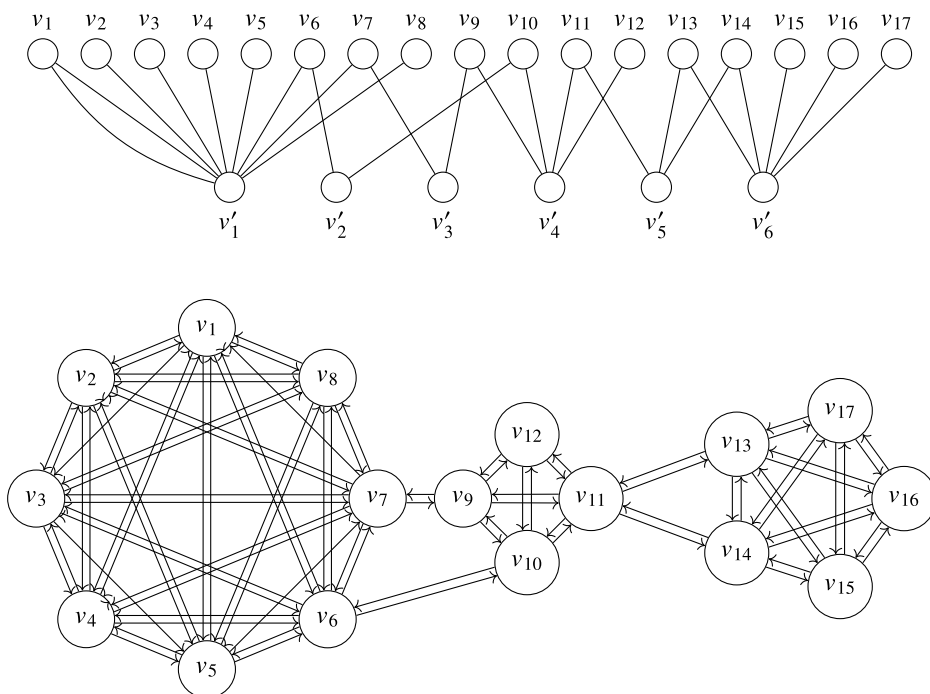


FIGURE 1. Construction of G_n for $n = 17$ with $|V'_n| = |\{v'_1, \dots, v'_6\}| = 6$ cliques. Top: the auxiliary graph G_n^{aux} . Bottom: the resulting directed graph G_n .

probability (w.h.p.) if the probability of the event tends to 1 as $n \rightarrow \infty$, where n is the number of vertices of the graph G_n under consideration.

2.2. The random intersection graph with mixed Poisson degrees

We consider a random intersection graph model where the degrees of the nodes follow a mixed Poisson distribution. Epidemics on this particular type of graph have previously been investigated in [9], which used a branching process coupling to derive expressions for the asymptotic probability of a major outbreak (i.e. that a fraction $\Theta(1)$ of the population contracts the disease in the limit as $n \rightarrow \infty$, where n is the population size), the final size of a major outbreak, and a threshold parameter. In the present paper, we focus on the (exponential) real-time growth rate of an epidemic on a random intersection graph in the early phase of a major outbreak. We give a somewhat brief description of this graph model and refer the reader to [9] for a more detailed account.

A graph G_n on n vertices can be constructed via the mixed Poisson random intersection graph model as follows (see Figure 1 for an illustration of this construction). Let A and B be two random variables with expected values $\mathbb{E}(A) = \mu_A$ and $\mathbb{E}(B) = \mu_B$. We make the following assumption on A and B .

Assumption 2.1.

- (i). $\mathbb{P}(A \geq 0) = \mathbb{P}(B \geq 0) = 1$.
- (ii). $\mathbb{P}(A = 0) < 1$ and $\mathbb{P}(B = 0) < 1$.

(iii). $\mathbb{E}(A^2 \log_+ A) < \infty$ and $\mathbb{E}(B^2 \log_+ B) < \infty$.

We will refer to the condition of Assumption 2.1(iii) as the $x^2 \log x$ -condition.

Remark 2.1. This version of the random intersection graph can be constructed under less strict assumptions than the $x^2 \log x$ -condition (see [9]); it is, however, needed here for the approximating branching process to satisfy the classical $x \log x$ -condition (see [20, p. 10]).

Let $\{A_k\}_k$ and $\{B_k\}_k$ be two sequences of independent copies of A and B , respectively. Further, let $V_n = \{v_1, \dots, v_n\}$ be the vertex set of G_n , and assign the weight A_i to the vertex v_i , $i = 1, \dots, n$. As an intermediate step, we construct an auxiliary graph G_n^{aux} with vertex set $V_n \cup V'_n$, where $V'_n = \{v'_1, \dots, v'_m\}$ and $m = m(n) := \lfloor n\mu_A/\mu_B \rfloor$. Assign the weight B_j to the vertex v'_j , $j = 1, \dots, m$. Given the weights of the vertices of V_n and V'_n , for each pair v_i, v'_j of vertices of G_n^{aux} such that $v_i \in V_n$ and $v'_j \in V'_n$, let the number of edges of G_n^{aux} shared by v_i and v'_j have distribution $\text{Po}(A_i B_j / n\mu_A)$, independently for pairs v_i, v'_j . Thus, in G_n^{aux} the degree of $v_i \in V_n$ has distribution

$$\text{Po}\left(A_i \frac{\mu_B^{(n)} \lfloor n\mu_A/\mu_B \rfloor}{n\mu_A}\right), \quad (2.2)$$

where $\mu_B^{(n)} := \sum_{j=1}^m B_j/m$. Similarly, in G_n^{aux} the degree of $v'_j \in V'_n$ has distribution

$$\text{Po}\left(B_j \frac{\mu_A^{(n)}}{\mu_A}\right), \quad (2.3)$$

where $\mu_A^{(n)} := \sum_{i=1}^n A_i/n$. There are no edges of G_n^{aux} between pairs $v_{i_1}, v_{i_2} \in V_n$. Similarly, there are no edges of G_n^{aux} between pairs of vertices of V'_n .

We now obtain the graph G_n from G_n^{aux} by letting two distinct vertices $v_{i_1}, v_{i_2} \in V_n$ of G_n share an edge if and only if v_{i_1} and v_{i_2} of G_n^{aux} have at least one common neighbour in V'_n . Next, we replace each edge of the undirected graph G_n by two directed edges pointing in the opposite direction. The reason for this modification is that in the epidemic model considered in this paper (see Section 2.3) infectious contacts are directed.

For later reference, we note that every vertex $v' \in V'_n$ corresponds to a clique $\mathcal{C}$, with vertex set $V_{\mathcal{C}}$ consisting of neighbours of v' in G_n^{aux} , and with edge set $E_{\mathcal{C}}$ that makes $\mathcal{C}$ a simple directed complete graph (that is, for any pair $u, v \in V_{\mathcal{C}}$ of distinct vertices there are precisely two edges (u, v) , $(v, u) \in E_{\mathcal{C}}$, and $E_{\mathcal{C}}$ contains no self-loops).

2.3. The epidemic model

We consider a stochastic SIR epidemic on G_n . In the SIR model, individuals are classified as susceptible (S), infectious (I), or recovered (R) depending on their current health status. An individual who is classified as infectious can transmit the disease to other individuals in the population; if an infectious individual contacts a susceptible individual then transmission occurs and the susceptible individual immediately becomes infectious. An infectious individual will eventually recover from the disease after some period of time, which we refer to as the *infectious period* of the individual in question (in our model we allow for the infectious period of an individual to be ∞ , which means that once the individual in question has contracted the disease it will remain infectious forever). Recovered individuals are fully immune to the disease; once recovered, an individual plays no further role in the spread of the disease.

For simplicity, we assume that the epidemic starts with one initial infectious case and that the rest of the population is initially fully susceptible to the disease. We assume that the disease spreads in a population of size n , where the underlying social network of the population is represented by G_n . In our model, a ‘close contact’ (a contact that results in transmission if a susceptible individual is contacted by an infectious individual) can only occur between individuals who are neighbours in G_n . Throughout the paper, we will use the terms individual and vertex interchangeably.

To be precise, the epidemic process can be defined as follows. Let I be a random variable with support in $\mathbb{R}_{\geq 0} \cup \{\infty\}$. Each vertex v_i of G_n is equipped with an infectious period I_i , where $\{I_i\}_i$ is a sequence of independent copies of I . Let $\{T_{ij}\}_{ij}$ be a sequence of independent and identically distributed random variables with support in $[0, \infty)$, and assume that they are independent of the infectious periods $(I_i)_{i \geq 1}$. Here, T_{ij} represents the time elapsed from the event that v_i contracts the disease (which might or might not occur) to the event that v_i contacts v_j . In many standard models, T_{ij} are exponentially distributed. For each (directed) edge (v_i, v_j) , we equip (v_i, v_j) with the transmission weight

$$T'_{ij} := \begin{cases} T_{ij} & \text{if } T_{ij} \leq I_i, \\ \infty & \text{if } T_{ij} > I_i. \end{cases}$$

The transmission weight T'_{ij} represents the time elapsed from the event that v_i contracts the disease to the event that v_i makes an infectious contact with v_j , which results in transmission of the disease to v_j if v_j is still susceptible. We make the following assumption on the distribution of T'_{ij} .

Assumption 2.2. $\mathbb{P}(T'_{ij} = 0) < 1/(\mu_{\bar{A}}\mu_{\bar{B}})$ and the distribution of T'_{ij} is non-lattice (i.e. $\mathbb{P}(T'_{ij} \in \{\infty, 0, s, 2s, \dots\}) < 1$ for any $s > 0$).

The first part of Assumption 2.2, $\mathbb{P}(T'_{ij} = 0) < 1/(\mu_{\bar{A}}\mu_{\bar{B}})$, ensures that the approximating branching process does not explode (i.e. that the branching process population does not grow infinitely large in finite time).

A path $\zeta = (v_{i_1}, v_{i_2}, \dots, v_{i_k})$ is any finite sequence of vertices of G_n such that $(v_{i_r}, v_{i_{r+1}})$ is an edge of G_n , $r = 1, \dots, k-1$. We define the length $\ell(\zeta)$ of a path $\zeta = (v_{i_1}, v_{i_2}, \dots, v_{i_k})$ as $\ell(\zeta) = \sum_{r=1}^{k-1} T'_{i_r i_{r+1}}$. Denote the collection of all paths from a vertex u to a vertex v by Σ_{uv} . The distance (transmission time) from u to v is given by $d(u, v) := \min_{\zeta} \ell(\zeta)$, where the minimum is taken over all paths $\zeta \in \Sigma_{uv}$. We make the conventions $d(u, u) = 0$ and $d(u, v) = \infty$ if Σ_{uv} is empty for two distinct vertices u and v .

Remark 2.2. Strictly speaking, d is a quasi-distance rather than a distance since it is not symmetric. We do, however, abuse terminology for convenience.

The initial infected case u_* is then selected according to some rule; a common choice, which we will adhere to here, is to select the initial case uniformly at random. We assume that the initial case u_* contracts the disease at time 0. The time evolution of an outbreak can now be fully specified as follows. An individual v_i , $i = 1, \dots, n$, has contracted the disease at time $t \geq 0$ if and only if $d(u_*, v_i) \leq t$, and v_i has recovered from the disease at t if and only if $d(u_*, v_i) + I_i \leq t$.

We will also need the *within-clique distance*. Let $\mathcal{C}$ be a clique of G_n . For two vertices u and v let $\Sigma_{uv}^{\mathcal{C}}$ be the collection of paths from u to v restricted to $\mathcal{C}$. That is, a path ζ from u to v belongs in $\Sigma_{uv}^{\mathcal{C}}$ whenever every edge of ζ is also an edge of $E_{\mathcal{C}}$. Note that $\Sigma_{uv}^{\mathcal{C}}$ is empty if u and v are not both members of $\mathcal{C}$. The distance from u to v restricted to $\mathcal{C}$ is given by $d_{\mathcal{C}}(u, v) := \min_{\zeta} \ell(\zeta)$,

where the minimum is taken over all paths $\zeta \in \Sigma_{uv}^C$. As before, $d_C(u, u) = 0$ holds whenever $u \in V_C$, and $d_C(u, v) = \infty$ if Σ_{uv}^C is empty.

For any clique C , we refer to the first individuals of C to contract the disease as the *primary cases* of C . That is, given that the disease reaches C (i.e. $\min_{w \in V_C} d(u_*, w) < \infty$), a vertex $u \in V_C$ is a primary case of C if $d(u_*, u) = \min_{w \in V_C} d(u_*, w)$. If $v \in V_C$ contracts the disease but is not a primary case of C , we say that v is a *secondary case* of C , regardless of whether v is infected directly by a primary case or via some other path (which may or may not go via u). Notice that, if we assume that the transmission weight distribution has no atoms, then the primary case is almost surely unique.

2.4. Branching process approximation

At the beginning of an epidemic outbreak, with high probability, infectious individuals contact individuals that are still susceptible. For this reason, in those early stages the epidemic graph is structured as a tree and can therefore be approximated by a branching process. We point out that a salient feature of branching processes is that the lives of individuals that belong to different branches of the branching process tree are independent (conditioned on their types in the multitype case). In the epidemic process, even in the large-population limit, the infectious individuals in a clique ‘compete’ in transmitting the disease to the remaining susceptible individuals in the clique. Therefore, naive attempts to couple a finite-type branching process with the epidemic process will in general give rise to non-local dependencies between the individuals of the branching process tree.

For our scopes, we need a more formal description of a general branching process Z (see, e.g., [21] for a full formal framework).

Let $\mathbb{N}$ be the set of positive integers, $\mathbb{N}^0 := \emptyset$, and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$. The set of all (potential) individuals $\mathcal{S}$ corresponds to the set of all finite tuples of positive integers, i.e. $\mathcal{S} := \bigcup_{n \geq 0} \mathbb{N}^n$. An individual x is identified with a tuple $(x_1, \dots, x_n) \in \mathcal{S}$, indicating that x is the x_n th child of the $\dots$ of the x_2 nd child of the x_1 st child of the ancestor, the latter being $a_0 := \emptyset$. We will refer to the common ancestor $a_0 \in \mathcal{S}$ as the *first ancestor* of Z . We write $|x| = n$ to indicate that x belongs to the n th generation of Z , which is $x \in \mathbb{N}^n$, and we say that a_0 is in generation 0. If $x = (x_1, \dots, x_n)$ and $y = (y_1, \dots, y_m)$, we denote by xy the concatenated individual $(x_1, \dots, x_n, y_1, \dots, y_m)$. If $x = (x_1, \dots, x_{n-1}, x_n)$, $mx = (x_1, \dots, x_{n-1})$ is the mother of x . The individuals of Z can be partially ordered by descent. If $x = m^n y$ for some $n \in \mathbb{N}_0$, then we write $x \preceq y$ (or equivalently $y \succeq x$) to indicate that x is an ancestor of y (we make the convention that an individual is an ancestor of itself and that $m^0 x := x$) and $x < y$ (or $y > x$) to indicate that $x \preceq y$ and $x \neq y$. Similarly, for $\mathcal{J} \subset \mathcal{S}$ and $x \in \mathcal{S}$ we write $\mathcal{J} < x$ to indicate that $y < x$ for some $y \in \mathcal{J}$, and $\mathcal{J} \preceq x$ to indicate that $y \preceq x$ for some $y \in \mathcal{J}$. Note that if $x \in \mathcal{J}$ and there exists $y \in \mathcal{J}$ such that $y < x$, then we still write $\mathcal{J} < x$.

Let $(\Omega_\emptyset, \mathcal{A}_\emptyset)$ be a measurable space. Using the terminology in [21], we call $\omega \in \Omega_\emptyset$ a possible *life career* of an individual; any property of an individual can be seen as a measurable function on $(\Omega_\emptyset, \mathcal{A}_\emptyset)$. For any individual $x \in \mathcal{S}$, we denote by $\tau_x \in [0, \infty]$ the time at which it is born; we say that x is realised if $\tau_x < \infty$, while $\tau_x = \infty$ means that x and its (possible) descendants are never born. At birth, any child gets a type in the type space E , which is equipped with a countably generated σ -algebra $\mathcal{E}$. We denote by $\sigma(x) \in E$ the type of x . For each $x \in \mathcal{S}$ we create an independent copy $(\Omega_x, \mathcal{A}_x)$ of the space $(\Omega_\emptyset, \mathcal{A}_\emptyset)$; each space $(\Omega_x, \mathcal{A}_x)$ carries a point process ξ_x on $\mathbb{R}_+ \times E$, defined as

$$\xi_x(A \times B) := \#\{k \in \mathbb{N} : \tau_{xk} - \tau_x \in A, \sigma(xk) \in B\}, \quad A \in \mathcal{B}, B \in \mathcal{E},$$

with $\mathcal{B}$ denoting the Borel σ -algebra of $\mathbb{R}_+$, and with xk being the concatenated individual $(x_1, \dots, x_n, k)$. A point (t, r) of ξ_x indicates a type- r individual produced by x at time t , and there's a one-to-one correspondence between the children of x and the points of ξ_x .

The population space of Z is then $(\Omega, \mathcal{A}) := (E \times \prod_{x \in \mathcal{S}} \Omega_x, \mathcal{E} \times \prod_{x \in \mathcal{S}} \mathcal{A}_x)$. An element of $(\Omega, \mathcal{A})$ consists of a type for the ancestor and a life career for all individuals of $\mathcal{S}$.

The epidemic process on G_n can be approximated, in its early phase, by a multi-type branching process whose individuals correspond to infectious individuals in the epidemic process. Following the description above, we will now define a branching process Z , similar to the one presented in [9], that describes the initial course of the epidemic. From now on, we will talk about individuals of Z and infectious individuals of the epidemic model interchangeably, keeping in mind their correspondence.

The type space of Z is $\mathcal{T}_\theta := \mathcal{T} \cup \{\theta\}$, where $\mathcal{T}$ is the support of the generic infectious period I , and the extra point θ is an atom for the reproduction kernel of Z , in the sense of [28, Definition 4.3]. If an infectious individual is, in the epidemic process, the secondary case of a clique of size two, then the corresponding element in the branching process is assigned type θ . The type of any other individual, apart from the ancestor (i.e. the first infectious individual), is taken to be the length of its infectious period. We notice that the infectious period I_v of $v \in \mathcal{S}$ is an independent copy of the general infectious period if $\sigma(v) = \theta$ or if $v = a_0$; $I_v = \sigma(v)$ otherwise. For this reason, θ -individuals generate independent and identically distributed branches of the branching process. This feature is crucial for the embedded process defined in Section 3.1. The individuals of Z are divided into generations by attributing all secondary cases in a clique to the primary case of the clique in question, even though it may well be that a secondary case does not get infected directly by the primary case. In other words, if we follow the epidemic trail from v back to the initial case v_* then the generation of v in Z corresponds, in the epidemic process, to the number of cliques that have to be traversed to reach v_* (including the cliques of v and v_*).

Each space $(\Omega_v, \mathcal{A}_v)$, $v \in \mathcal{S}$, is equipped with a point process ξ_v . The law of ξ_v can be described as follows. Given the type of an individual v , its point process of reproduction is independent of the lives of the individuals that do not stem from v . A point (t, r) of ξ_v corresponds to a secondary type- r case $u \in V_C$ for which the time elapsed since the corresponding primary case v contracted the disease is $d_C(v, u) = t$.

Observe that the law of the life of a_0 is usually different from the laws of the other branching process individuals since the initial case is assumed to be selected uniformly at random from the population, whereas individuals of subsequent generations represent infectious cases whose degree distribution is size biased. For this reason, the initial case is a member of $\text{MP}(A)$ cliques while the number of cliques that a non-initial case is a member of is distributed as $\bar{D}$ cliques, where D is distributed as $\text{MP}(A)$. It is readily verified (see also [9]) that, if D has $\text{MP}(A)$ distribution, then $(\bar{D} - 1)$ has $\text{MP}(\bar{A})$ distribution. Thus, asymptotically, a non-initial case is a member of $\text{MP}(\bar{A})$ cliques that are not yet affected by the disease. Note that the size-biased distribution appears here because the probability that a vertex belongs to a clique is proportional to its number of edges. Similarly, the size of a clique (excluding the primary case of the clique) reached during the early phase of the epidemic has asymptotic distribution given by $\text{MP}(\bar{B})$. Here, the size biasing is present because the size of a clique is proportional to the number of edges of its correspondent vertex in the auxiliary graph; see the construction in Section 2.2. Thus, the points of ξ_{a_0} correspond to the secondary cases where the corresponding primary case has infectious period I_{a_0} and is a member of $\text{MP}(A)$ cliques of independent sizes with distribution $\text{MP}(\bar{B})$. Similarly, if $v \neq a_0$, the points of ξ_v correspond to the secondary cases where

the corresponding primary case has infectious period I_v and is a member of $\text{MP}(\bar{A})$ cliques, each of size $\text{MP}(\bar{B})$.

The number $|Z_t|$ of individuals of Z that are alive at time $t \geq 0$ corresponds to the number of infectious individuals at t and is given by the cardinality of the set $\{u \in \mathcal{S} : \tau_u \leq t < \tau_u + I_u\}$.

If the extinction probability of Z is strictly smaller than 1 then Z is said to be *supercritical*, and we say that we are in the supercritical regime.

3. Main results

In this section we present the main results and give a rough outline of the ideas behind the proofs. These proofs rely on asymptotic results on finite-type branching processes (see, e.g., [26]) via a coupling of an epidemic process on G_n and a single-type branching process.

The main idea of this section is to embed a single-type branching process Y in Z by letting the type- θ individuals of Z be the individuals of Y . This allows us to employ the almost sure (a.s.) asymptotic results (see Section 3.1 for an overview) that are available for single-type branching processes.

We now need to introduce some extra notation. Let F_k , $k \in \mathbb{N}_{\geq 2}$, be the cumulative distribution function of the transmission time from the primary case in a clique of size k to another (specific) member of the clique. That is to say, for a clique $\mathcal{C}$ with $|V_{\mathcal{C}}| = k$ and two fixed individuals $u, v \in V_{\mathcal{C}}$, F_k is the cumulative distribution function of the within-clique distance $d_{\mathcal{C}}(u, v)$. Let $p_k^{\bar{B}} := \mathbb{P}(\text{MP}(\bar{B}) = k)$ and define the vector $\Gamma^{\top} = \mu_{\bar{A}}(1 \cdot p_1^{\bar{B}}, 2 \cdot p_2^{\bar{B}}, \dots)$, where $\mu_{\bar{A}} = \mathbb{E}(\bar{A}) = \mathbb{E}(A^2)/\mu_A$. For a clique $\mathcal{C}$ of size $|V_{\mathcal{C}}| = k$ and any $\lambda \geq 0$, let

$$\mathcal{L}_k^{(\lambda)} := \int_{\mathbb{R}_{\geq 0}} e^{-t\lambda} F_k(dt) \quad (3.1)$$

be the Laplace transform of the transmission time within $\mathcal{C}$, and define the vector

$$L^{(\lambda)} = (\mathcal{L}_2^{(\lambda)}, \mathcal{L}_3^{(\lambda)}, \dots)^{\top}. \quad (3.2)$$

Note that, for $\lambda = 0$, the mass at infinity is typically not included in the integral defined in (3.1).

As we will prove in Section 4, the Malthusian parameter can be found by solving the Euler–Lotka equation.

Definition 3.1. *The Malthusian parameter is the unique solution $\alpha > 0$ of $\Gamma \cdot L^{(\alpha)} = 1$.*

We are now ready to state our main results.

Theorem 3.1. *Under Assumptions 2.1(i)–(iii) and 2.2, and if the Malthusian parameter $\alpha > 0$ exists, then there exists a non-negative integrable random variable W such that $|Z_t|/e^{\alpha t} \xrightarrow{\text{a.s.}} W$, where W satisfies $\mathbb{P}(\{|Z_t| \not\rightarrow 0\} \Delta \{W > 0\}) = 0$.*

Here, Δ denotes the symmetric difference, i.e. $A \Delta B = (A \setminus B) \cup (B \setminus A)$, and $|Z_t|$ denotes the number of individuals of Z that are alive at time t . In the notation $\{|Z_t| \not\rightarrow 0\}$ it is implicit that the limit is taken as t tends to ∞ . So, $\{|Z_t| \not\rightarrow 0\}$ is the event that the branching process population of Z ultimately avoids extinction.

Theorem 3.2. *Let $(G_n)_n$ be a sequence of graphs generated via the random intersection graph model and assume that the assumptions of Theorem 3.1 hold. Let $(\varepsilon_n)_{n \geq 1}$ be a sequence in $(0, \infty)$ that satisfies $\varepsilon_n \log(n) \rightarrow \infty$ as $n \rightarrow \infty$. Then, for any $q \geq 2$, $q \neq 3$, that satisfies $\mathbb{E}(A^q) < \infty$ and $\mathbb{E}(B^q) < \infty$ there exist couplings of the epidemic process on the G_n and Z such that the*

two processes agree w.h.p. until at least $n^{\gamma-\varepsilon_n}$ individuals have contracted the disease. Here, $\gamma = \min\left(\frac{1}{2}, \left(q - \frac{3}{2}\right)/q\right)$.

3.1. Branching processes counted with random characteristics

This section contains a brief overview of some preliminaries from the theory of branching processes, which we include for completeness. More detailed accounts of the branching process theory can be found in [26], and also in the more recent paper [19].

Before proceeding, we introduce some additional notation to Section 2.4. To arrive at Theorem 3.1, as anticipated before, we embed a single-type branching process Y into the above-described branching process Z . To this end, we partition the individuals of Z into *blocks*. Let $S_\theta \subset S$ be the set of the type- θ individuals of Z . For $x \in S_\theta$, define the block $\mathcal{B}_x$ as follows:

$$\mathcal{B}_x := \{y \in S : x \preceq y, \text{ and whenever } x \prec z \text{ for some } z \in S_\theta \text{ then } z \not\preceq y\}. \quad (3.3)$$

In words, for any $x \in S_\theta$ the block $\mathcal{B}_x$ is the set of descendants of x for which the line of descent back to x does not contain an individual of type θ . The embedded branching process Y is then obtained by letting the individuals of Y be the individuals of S_θ and the children of $x \in S_\theta$ (seen as an individual of Y) the type- θ children of individuals of $\mathcal{B}_x$ (in Z). That is, if we define $\mathcal{J}_n$, $n \geq 1$, recursively as $\mathcal{J}_0 = \{a_0\}$ and

$$\mathcal{J}_n = \{x \in S_\theta : x \succ \mathcal{J}_{n-1} \text{ and whenever } \mathcal{J}_{n-1} \prec z \prec x \text{ then } z \notin S_\theta\},$$

then $\mathcal{J}_n$ consists of the individuals of generation n of Y , $n \geq 0$. Since we are interested in the number of infected individuals at each time point $t \in \mathbb{R}_{\geq 0}$ we count the population of the embedded single-type branching process Y with a certain random characteristic (see, e.g., [26]) which provides the link between the size $|Z_t|$ of the branching process population of Z at t and the embedded branching process Y . Here we consider the special case where the random characteristic ϕ is defined as

$$\phi_x(t) = |\{y \in \mathcal{B}_x : \tau_y \leq t < \tau_y + I_y\}| \quad (3.4)$$

for each $x \in \mathcal{J} = \bigcup_{n \geq 0} \mathcal{J}_n$, and we say that $Y_t^\phi := \sum_{x \in \mathcal{J}} \phi_x(t - \tau_x)$ is the branching process population of Y counted with the characteristic ϕ . In words, $\phi_x(t)$ can be thought of as the number of infectious individuals which belong to the block $\mathcal{B}_x$ at $\tau_x + t$, where τ_x is the time point when x contracts the disease. Thus, the total population size $|Z_t|$ of the approximating branching process Z at the time point t can be recovered from the embedded single-type branching process Y via the relation $|Z_t| = Y_t^\phi$, where $Y_t^\phi = \sum_{x \in \mathcal{J}} \phi_x(t - \tau_x)$ is the branching process population of Y counted with the characteristic ϕ defined in (3.4) at t .

Remark 3.1. It is worth pointing out that the embedding technique employed here does not require the presence of cliques of size two in the underlying graph model. Indeed, in a more general setting, an embedding of a single-type branching process may be obtained by letting the individuals of the single-type branching process be represented by the vertices that are the last to be infected in their clique, if all clique members get infected. This embedding relies on the observation that if a clique $\mathcal{C}$ (of size $|V_{\mathcal{C}}| = d \geq 2$, say) is fully infected then the d th individual of $\mathcal{C}$ to be infected does not compete with the other infected cases of $\mathcal{C}$ in transmitting the disease to the remaining susceptible individuals of $\mathcal{C}$. Thus, given that v is the d th infected case of $\mathcal{C}$, the infectious period of v is independent of the actual paths of transmission within $\mathcal{C}$.

We begin by stating an asymptotic result for single-type branching processes where the type of the ancestor is the same as the type of the other individuals. To this end, let $\tilde{Z}$ be a branching process that behaves like a copy of Z , where Z is the branching process in Section 2.4, except that the first ancestor of $\tilde{Z}$ is of type θ . Further, let $\tilde{Y}$ be the corresponding embedded single-type branching process. In what follows, we will recycle the notation from Section 2.4 for ease of notation. That is, we denote the type space of $\tilde{Z}$ by $\mathcal{T}_\theta$, $\mathcal{S}$ denotes the space of individuals of $\tilde{Z}$, the block $\mathcal{B}_x$ and the random characteristic ϕ are analogous to the definitions in (3.4) and (3.3), respectively, and so forth.

Let the random measure ξ be the point process of reproduction on $\mathbb{R}_{\geq 0}$ of a generic individual of the single-type branching process $\tilde{Y}$, and let $\xi^{(\alpha)} = \int_{\mathbb{R}_{\geq 0}} e^{-\alpha t} \xi(dt) = \sum_{x \in \mathcal{J}} e^{-\alpha \tau_x}$, where $\mathcal{J} = \bigcup_{n \geq 0} \mathcal{J}_n$ is the space of all individuals of $\tilde{Y}$ and α is the Malthusian parameter, i.e. $\mathbb{E}(\xi^{(\alpha)}) = 1$. We define the measure η on $\mathbb{R}_{\geq 0}$ by $\eta(t) = \eta[0, t] := \mathbb{E}(\xi(t))$. Theorem 3.3 is a special case of [26, Theorem 5.4] and will lead us to the a.s. convergence of Theorem 3.1. In order to state Theorem 3.3 we need the following conditions.

Condition 3.1. (*Finite mean age at childbearing.*) The mean age at childbearing β defined by $\beta := \mathbb{E}(\int_{\mathbb{R}_{\geq 0}} t e^{-\alpha t} \xi(dt))$ is finite.

Condition 3.2. ($(x \log x)$.) The random variable $\xi^{(\alpha)} \log_+ (\xi^{(\alpha)})$ has finite expectation.

Condition 3.3. There exists some non-negative, real-valued, non-increasing, integrable function g_1 such that $\int_{\mathbb{R}_{\geq 0}} e^{-\alpha t} / g_1(t) \eta(dt) < \infty$.

Condition 3.4. There exists some non-negative, real-valued, non-increasing, integrable function g_2 such that the expectation of $\sup_{t \geq 0} e^{-\alpha t} \phi(t) / (g_2(t) \wedge 1)$ is finite. Recall that ϕ is the random characteristic defined in (3.4).

Theorem 3.3. ([26, Theorem 5.4].) Under Conditions 3.1–3.4, almost surely, $|\tilde{Y}_t^\phi| / e^{\alpha t} \rightarrow \widehat{W} m_\infty$ as $t \rightarrow \infty$, where the random variable $\widehat{W}$ has mean $\mathbb{E}(\widehat{W}) = 1$ and $\mathbb{P}(\{\widehat{W} = 0\}) = \mathbb{P}(|Z_t| \rightarrow 0)$, and $m_\infty \in (0, \infty)$ is a constant that depends on ϕ .

Remark 3.2. Under the conditions of Theorem 3.3, applying Theorem 3.3 to each of the children of the first ancestor of Z gives

$$\frac{|Z_t|}{e^{\alpha t}} = \frac{|Y_t^\phi|}{e^{\alpha t}} \rightarrow W := (\widehat{W}^{(1)} e^{-\alpha \tau_1} + \dots + \widehat{W}^{(J)} e^{-\alpha \tau_J}) m_\infty \quad (3.5)$$

almost surely as $t \rightarrow \infty$, where J is the number of children of the first ancestor of Z that are born in $[0, \infty)$, the time points $\tau_1, \dots, \tau_J$ are the birth times of those children, and $\widehat{W}^{(1)}, \dots, \widehat{W}^{(J)}$ are J copies of $\widehat{W}$ (which are not independent in general).

4. Proofs

In Section 4.1 we prove Theorem 3.1 by showing that there is a coupling of the branching process Z and a single-type branching process whose Malthusian parameter is given in Definition 3.1. In Section 4.2 we prove Theorem 3.2. The main step in the proof is to establish upper bounds on the total variation distance of the degree distribution in (2.2) and $\text{Po}(\bar{A})$, and of the distribution in (2.3) and $\text{Po}(\bar{B})$.

4.1. Proof of Theorem 3.1

Recall that the random measure ξ (defined in Section 3.1) is the point process of reproduction on $\mathbb{R}_{\geq 0}$ of a generic individual v of $\tilde{Y}$, and that the measure η on $\mathbb{R}_{\geq 0}$ is defined as $\eta(t) = \eta[0, t] := \mathbb{E}(\xi(t))$. Also recall that $\Gamma = (\gamma_k)_k$ is the vector with elements of the form $\gamma_k = \mu_{\bar{A}} k p_k^{\bar{B}}$, where $p_k^{\bar{B}} = \mathbb{P}(\text{MP}(\bar{B}) = k)$ for $k \in \mathbb{Z}_{\geq 1}$, and that $L^{(\alpha)}$ is the vector displayed in (3.2).

Lemma 4.1. *In the supercritical regime, the Malthusian parameter $\alpha > 0$ of Y exists if and only if $\mathbb{P}(T'_{ij} = 0) < 1/(\mu_{\bar{A}}\mu_{\bar{B}})$ and is then the unique solution of $\Gamma \cdot L^{(\alpha)} = 1$.*

Proof. Here, $*$ denotes convolution, i.e. for two cumulative distribution functions F and G , $F * G(t) = \int_{-\infty}^{\infty} G(t-s)F(ds)$. Note that the convolution function is symmetric. The expected number of children of v that are θ -individuals born up to time t is given by

$$\eta(t) = \gamma_1 F_2(t) + \sum_r \sum_{(m_1, \dots, m_r)} \gamma_{m_1} \gamma_{m_2} \cdots \gamma_{m_r} F_{m_1+1} * \cdots * F_{m_r+1} * F_2(t) \gamma_1, \quad (4.1)$$

where the sums run over $\mathbb{Z}_{\geq 1}$ and $\mathbb{Z}_{\geq 1}^r$. To see this, we look at v 's offspring in Z . If v belongs to a clique of size two, then we can have a θ -newborn in the first generation, and this is taken into account by the first addend of (4.1). After this, we consider the second generation (corresponding to $r = 1$ in the formula) and we look for all possible paths that pass via a clique of generic size $m_1 + 1$ and end on a clique of size two, and so on. Taking the Laplace transform of the right-hand side in (4.1) and writing this in vector form gives that the Malthusian parameter α of Y is the solution of

$$\int_{\mathbb{R}_{\geq 0}} e^{-\alpha t} \eta(dt) = \gamma_1 \mathcal{L}_2^{(\alpha)} \sum_{n=0}^{\infty} (\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)})^n = 1, \quad (4.2)$$

where $\Gamma_{\geq 2}^{\top} = (\gamma_2, \gamma_3, \dots)$ and the elements of $L_{\geq 3}^{(\alpha)} = (\mathcal{L}_3^{(\alpha)}, \mathcal{L}_4^{(\alpha)}, \dots)^{\top}$ are defined in (3.1). Since $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)} < 1$ (shown below), (4.2) reduces to

$$\frac{\gamma_1 \mathcal{L}_2^{(\alpha)}}{1 - \Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)}} = 1.$$

That is, $\Gamma \cdot L^{(\alpha)} = 1$.

It remains to show that $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)} < 1$. By the $x^2 \log x$ assumption, for all $\lambda \geq 0$, $\Gamma \cdot L^{(\lambda)} \leq \Gamma \cdot (1, 1, \dots, 1)^{\top} < \infty$ and, since the approximating branching process is supercritical, $\Gamma \cdot L^{(0)} > 1$. Since $\Gamma \cdot L^{(\lambda)}$ is continuous and strictly decreasing in λ with $\Gamma \cdot L^{(\lambda)} \rightarrow \mathbb{P}(T'_{ij} = 0) \mu_{\bar{A}} \mu_{\bar{B}} < 1$ as $\lambda \rightarrow \infty$, there exists an α such that $\gamma_1 \mathcal{L}_2^{(\alpha)} + \Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)} = 1$; given that $\gamma_1 > 0$, for that α we have $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)} < 1$. We conclude that the Malthusian parameter exists and is unique. $\square$

To proceed, we need some additional notation and terminology. For two kernels K_1 and K_2 (defined on the same measurable space $(E, \mathcal{E})$), we define the convolution kernel $K_1 K_2$ as

$$K_1 K_2(r, A) := \int_E K_1(r, ds) K_2(s, A), \quad A \in \mathcal{E}, \quad r \in E.$$

For any $m \geq 1$, $K_1^m := K_1 K_1^{m-1}$ and $K_1^0 := I$, where I is the identity kernel $I(r, A) := \mathbf{1}(r \in A)$ for any $A \in \mathcal{E}$. If f is an $\mathcal{E}$ -measurable function on E then we define the function $K_1 f$ as $K_1 f(\cdot) := \int_E f(s) K_1(\cdot, ds)$, and similarly, for a measure η on $(E, \mathcal{E})$ we define $\eta K(\cdot) = \int \eta(ds) K(s, \cdot)$. For any $A \in \mathcal{E}$, let I_A be the kernel $I_A(r, B) = I(r, A \cap B)$.

Recall that we denote (a generic copy of) the point process of reproduction on $\mathcal{T}_\theta \times \mathbb{R}_{\geq 0}$ of a type- r individual of $\tilde{Z}$ by ξ_r , and let $\mu(r, A \times B) = \mathbb{E}(\xi_r(A \times B))$ be the expected number of offspring of a type in $A \subset \mathcal{T}_\theta$ produced by a type- r individual (born at time 0) in $B \subset \mathbb{R}_{\geq 0}$. For $\lambda \in \mathbb{R}$, define the kernel $K_{(\lambda)}(r, ds \times dt) := e^{-\lambda t} \mu(r, ds \times dt)$. Also let $\widehat{K}(r, ds) := \int_{\mathbb{R}_{\geq 0}} K_{(\alpha)}(r, ds \times dt)$, and let $G_\theta = \sum_{n=0}^{\infty} (I_{\{\theta\}^c} \widehat{K})^n$ be the potential kernel of $I_{\{\theta\}^c} \widehat{K}$. Here, $\{\theta\}^c = \mathcal{T}_\theta \setminus \{\theta\}$ and

$$I_{\{\theta\}^c}(r, B) = I(r, B \cap \mathcal{T}_\theta \setminus \{\theta\}). \quad (4.3)$$

Remark 4.1. Note that, for any $A \subset \mathcal{T}$ and $s \in \mathcal{T}_\theta$, $G_\theta(s, A)$ is the expected number of descendants of an individual u of $\tilde{Z}$, $\sigma(u) = s$, that are members of the same block as u and whose type belongs to A discounted by their time of birth. Similarly, $G_\theta(s, \{\theta\})$ is the expected number of type- θ individuals stemming from u whose mother belongs to the same block as u discounted by their time of birth.

Define the function h by $h(x) = G_\theta(x, \{\theta\})$ and the measure π by $\pi(A) = \widehat{K} G_\theta(\theta, A)$. The kernel $\widehat{K}$ is 1-recurrent in the sense of [28, Definition 3.2]; see [21, 22]. Then, from [28, Proposition 4.6], it follows that h is harmonic for $\widehat{K}$, i.e. $h = \widehat{K}h$. Similarly, π is invariant for $\widehat{K}$, i.e. $\pi = \pi \widehat{K}$.

In order to use the finite-type branching process toolbox, we need to verify that the mean age at childbearing β of Y is finite, i.e. that $\beta = \int t e^{-\alpha t} \eta(dt) < \infty$, where η is the measure in (4.1).

Lemma 4.2. $0 < \beta < \infty$.

Proof. Let $\varepsilon > 0$ be small so that $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varepsilon)} < 1$; note that such an ε exists: $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\lambda)}$ is continuous and $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)} < 1$, as shown in the proof of Lemma 4.1. Let the constant C_ε be such that $C_\varepsilon e^{-(\alpha-\varepsilon)t} \geq t e^{-\alpha t}$ for all $t \geq 0$. Then

$$\beta = \int t e^{-\alpha t} \eta(dt) \leq C_\varepsilon \int e^{-(\alpha-\varepsilon)t} \eta(dt) = C_\varepsilon \gamma_1 \mathcal{L}_2^{(\alpha-\varepsilon)} \sum_{n=0}^{\infty} (\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varepsilon)})^n < \infty. \quad \square$$

4.1.1. Optimal lines and the $x \log x$ condition. In this paper, we consider two ways of dividing the individuals of the approximating branching process $\tilde{Z}$ into generations: either generation n consists of the individuals of $\mathcal{S}_n := \{x \in \mathcal{S} : |x| = n\}$ (i.e. of the individuals separated from the first ancestor by a line of descent of length n), or generation n consists of the individuals of $\mathcal{J}_n$, which leads us to the embedded branching process $\tilde{Y}$. There is a close connection between these two ways of viewing generations and the concept of *stopping lines* (see [21] or [12]).

Following [21], we say that a set of individuals $L \subset \mathcal{S}$ is a *stopping line* if, for any pair $y, x \in L$, $x \not\prec y$. In other words, a stopping line L cuts across the branching process tree Z in the sense that if $x \in L$ then no descendants or ancestors of x (apart from the individual x itself) are members of L . For any $x \in \mathcal{S}$, let $\mathcal{G}_x$ be the σ -algebra generated by the lives (infectious periods and reproduction processes) of the ancestors of x (including x), and for any non-random stopping line ℓ , let $\mathcal{G}_\ell := \sigma(\cup_{x \not\prec \ell} \mathcal{G}_x)$ be the σ -algebra generated by the lives of the individuals that do not have an ancestor in ℓ . Mirroring the concept of optimal stopping times, we say that a line L is *optimal* if, for any non-random stopping line ℓ , the event $\{L \leq \ell\}$ belongs to $\mathcal{G}_\ell$.

Here, $L \preceq \ell$ means that for any $x \in \ell$ there is $y \in L$ such that $y \preceq x$. Note that $\mathcal{J}_n$ is an optimal line for each $n \geq 0$. Note also that $\mathcal{S}_n$ is an optimal line, $n \geq 0$.

For each $n \in \mathbb{Z}_{\geq 0}$, define

$$\widehat{W}_n := \sum_{x \in \mathcal{J}_n} e^{-\alpha \tau_x} \quad (4.4)$$

and

$$\widetilde{W}_n = \frac{1}{h(\theta)} \sum_{|u|=n} e^{-\alpha \tau_u} h(\sigma_u), \quad (4.5)$$

where the sums run over all individuals of generation n of $\tilde{Y}$ and $\tilde{Z}$, respectively. It is well known (see, e.g., [12]) that $\{\widetilde{W}_n\}_{n \in \mathbb{Z}_{\geq 0}}$ is a martingale with respect to $\mathcal{F} := \{\mathcal{F}_n\}_n$, where $\mathcal{F}_n := \mathcal{G}_{\mathcal{S}_n}$ is generated by the lives of the individuals up to generation n (of $\tilde{Z}$) for $n \geq 1$ (we make the convention that $\mathcal{G}_{\emptyset}$ is the trivial σ -algebra). Similarly, $\{\widehat{W}_n\}_{n \in \mathbb{Z}_{\geq 0}}$ is a martingale with respect to $\{\mathcal{G}_{\mathcal{J}_n}\}_n$. For later reference, we now state a special case of results presented in [12].

Proposition 4.1. (c.f. [12, Theorem 6.1 and Lemma 6.2].) *Let $\{\widehat{W}_n\}_n$ and $\{\widetilde{W}_n\}_n$ be as in (4.4) and (4.5). Then, with probability 1, the limits $\lim_n \widehat{W}_n$ and $\lim_n \widetilde{W}_n$ exist and $\lim_n \widehat{W}_n = \lim_n \widetilde{W}_n$.*

Now, the $x \log x$ condition for the single-type branching process $\tilde{Y}$ takes the form

$$\mathbb{E}(\widehat{W}_1 \log_+ \widehat{W}_1) < \infty. \quad (4.6)$$

The following two lemmas assert that $\tilde{Y}$ satisfies the $x \log x$ condition under the $x^2 \log x$ condition.

Lemma 4.3. *Let J have $MP(X)$ distribution, where X is a non-negative integrable random variable with $\mathbb{P}(X=0) < 1$. Then $\mathbb{E}(X^2 \log_+ X) < \infty$ is necessary and sufficient for $\mathbb{E}(\bar{J} \log_+ \bar{J}) < \infty$ to hold.*

Proof. First, note that $\mathbb{E}(\bar{J} \log_+ \bar{J}) = \mathbb{E}(J^2 \log_+ J) / \mathbb{E}(\bar{J}) = \mathbb{E}(J^2 \log_+ J) / \mathbb{E}(X)$.

Since $x \mapsto x^2 \log_+ x$ is convex, necessity now follows from Jensen's inequality for conditional expectations. Sufficiency follows from

$$\begin{aligned} \mathbb{E}(J^2 \log_+ J) &= \mathbb{E}\left(\sum_{k \geq 0} k^2 \log_+ k \frac{X^k e^{-X}}{k!}\right) \\ &= \mathbb{E}\left(X \sum_{k \geq 0} (k+1) \log_+ (k+1) \frac{X^k e^{-X}}{k!}\right) \\ &= \mathbb{E}\left(X \sum_{k \geq 0} \log_+ (k+1) \frac{X^k e^{-X}}{k!}\right) + \mathbb{E}\left(X \sum_{k \geq 0} k \log_+ (k+1) \frac{X^k e^{-X}}{k!}\right) \\ &= \mathbb{E}\left(X \sum_{k \geq 0} \log_+ (k+1) \frac{X^k e^{-X}}{k!}\right) + \mathbb{E}\left(X^2 \sum_{k \geq 0} \log_+ (k+2) \frac{X^k e^{-X}}{k!}\right) \\ &\leq \mathbb{E}(X \log_+ (X+1)) + \mathbb{E}(X^2 \log_+ (X+2)), \end{aligned}$$

where the inequality follows from Jensen's inequality applied to the concave (on $\mathbb{R}_{\geq 0}$) functions $x \mapsto \log_+(x+1)$ and $x \mapsto \log_+(x+2)$. $\square$

Lemma 4.4. *Under Assumption 2.1, $\{\widehat{W}_n\}_n$ satisfies the $x \log x$ condition in (4.6).*

Proof. It is known (see [19, Proposition 4.1]) that the inequality in (4.6) holds if and only if $\{\widehat{W}_n\}_n$ is uniformly integrable, which is also equivalent to $\mathbb{E}(\widehat{W}) = 1$, where $\widehat{W}$ is the almost sure limit $\lim_n \widehat{W}_n$. By Proposition 4.1, $\lim_n \widehat{W}_n = \lim_n \widetilde{W}_n$ almost surely. Thus, in order to verify that $\{\widehat{W}_n\}_n$ is uniformly integrable it is sufficient to verify that the almost sure limit of $\widetilde{W}_n$ has mean 1.

Now note that h is bounded: for any $x \in \mathcal{T}_\theta$ we have

$$h(x) = G_\theta(x, \{\theta\}) \leq 1 + \mathbb{E}(\bar{A})\mathbb{E}(\bar{B})\pi(\{\theta\}) = 1 + \mathbb{E}(\bar{A})\mathbb{E}(\bar{B}).$$

Combining this with Assumption 2.1, Lemma 4.3 and [12, Corollary 2.1], the $x \log x$ condition holds for Y if we can show that, almost surely,

$$\sup_{x>2} \left(\frac{\sum_i \mathbf{1}(H(\varsigma_i) \geq x^{-1})}{\log_+(x)} \right) < \infty, \quad (4.7)$$

where $\varsigma = (\varsigma_0, \varsigma_1, \dots)$, $\varsigma_0 = (\theta, 0)$, is a Markov chain on $\mathcal{T}_\theta \times \mathbb{R}_{\geq 0}$ with transmission measure given by

$$\begin{aligned} R((r, 0), A \times B) &:= \frac{1}{h(r)} \int_{A \times B} h(s) K_{(\alpha)}(r, ds \times dt), \\ R((r, t), A \times B) &:= R((r, 0), A \times (B - t)_{\geq 0}), \end{aligned}$$

where $B - t = \{b - t : b \in B\}$ and $H((r, t)) = e^{-\alpha t} h(r)$.

Let $p_1 : \mathcal{T}_\theta \times \mathbb{R}_{\geq 0} \rightarrow \mathcal{T}_\theta$ and $p_2 : \mathcal{T}_\theta \times \mathbb{R}_{\geq 0} \rightarrow \mathbb{R}_{\geq 0}$ be the projections onto the first and second coordinate, respectively, and put $\varsigma'_j = p_1(\varsigma_j)$ and $\varsigma''_j = p_2(\varsigma_j)$ for $j \geq 0$. Then $\{\varsigma'_0, \varsigma'_1, \dots\}$ is a Markov chain on $\mathcal{T}_\theta$ with transition measure R_1 :

$$R_1(r, B) = \frac{1}{h(r)} \int_B h(s) K_{(\alpha)}(r, ds \times \mathbb{R}_{\geq 0}) = \frac{1}{h(r)} \int_B h(s) \widehat{K}(r, ds)$$

for (measurable) $B \subset \mathcal{T}_\theta$.

Now, it is easily verified that θ is a (positive) recurrent state of $\{\varsigma'_0, \varsigma'_1, \dots\}$. Indeed, let M_i , $i \geq 1$, be the number of steps until $\{p_1(\varsigma_0), p_1(\varsigma_1), \dots\}$ revisits θ for the i th time, given that $p_1(\varsigma_0) = \theta$: $M_0 := 1$ and, for $i \geq 1$, $M_i = \inf\{m > M_{i-1} : p_1(\varsigma_m) = \theta\}$. Then, for $m \geq 1$ (recall that $I_{\{\theta\}^c}$ is the operator in 4.3),

$$\begin{aligned} \mathbb{P}(M = m) &= R_1(I_{\{\theta\}^c} R_1)^{m-1}(\theta, \{\theta\}) \\ &= \int_{\mathbb{R}_{\geq 0}^{m-1}} K_{(\alpha)}(r_{m-1}, \{\theta\} \times \mathbb{R}_{\geq 0}) K_{(\alpha)}(r_{m-2}, dr_{m-1} \times \mathbb{R}_{\geq 0}) \cdots \\ &\quad \times K_{(\alpha)}(r_1, dr_2 \times \mathbb{R}_{\geq 0}) K_{(\alpha)}(\theta, dr_1 \times \mathbb{R}_{\geq 0}) \\ &= \gamma_1 \mathcal{L}_2^{(\alpha)} (\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)})^{m-1} = (1 - \Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)}) (\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha)})^{m-1}. \end{aligned}$$

Thus, $\mathbb{P}(M = \infty) = 0$ and M is geometrically distributed, and hence the inequality in (4.7) holds almost surely by the law of large numbers applied to $\{\varsigma''_{M_k} + \dots + \varsigma''_{M_{k+1}-1}\}_k$. $\square$

We now turn our attention to Conditions 3.3 and 3.4.

Lemma 4.5. *Conditions 3.3 and 3.4 are satisfied under Assumption 2.1.*

Proof. The mapping $t \mapsto e^{-\varphi t}$ satisfies Conditions 3.3 and 3.4 if $\varphi > 0$ is small enough. Indeed, let $\varphi \in (0, \alpha)$ be small, so that $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varphi)} < 1$, and put $g(t) = e^{-\varphi t}$. Then

$$\int_0^\infty \frac{1}{g(t)} e^{-\alpha t} \eta(dt) = \int_0^\infty e^{-(\alpha-\varphi)t} \eta(dt) = \gamma_1 \mathcal{L}_2^{(\alpha-\varphi)} \sum_{n=0}^\infty (\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varphi)})^n < \infty.$$

In the following, x is a generic type- θ individual of Z with $\phi = \phi_x$, $\mathcal{B} = \mathcal{B}_x$, and $\tau_x = 0$. Clearly, since $e^{-\alpha t}/g(t)$ is non-increasing in t we have

$$\frac{e^{-\alpha t} \phi(t)}{g(t)} \leq \sum_{y \in \mathcal{B}} \frac{e^{-\alpha \tau_y}}{g(\tau_y)}$$

for any $t \geq 0$. Thus

$$\mathbb{E} \left(\sup_{t \geq 0} \frac{e^{-\alpha t} \phi(t)}{g(t)} \right) \leq \mathbb{E} \left(\sum_{y \in \mathcal{B}} e^{-(\alpha-\varphi)\tau_y} \right) = \sum_{n=1}^\infty (\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varphi)})^n < \infty.$$

If the random characteristic ϕ is as in (3.4), then the same function g also satisfies Condition 3.4. $\square$

Proof of Theorem 3.1. Assume that Assumptions 2.1(i)–(iii) hold. By Lemma 4.1, the Malthusian parameter for the single-type branching process $\tilde{Y}$ is the unique solution $\alpha > 0$ of $\Gamma \cdot L^{(\alpha)} = 1$. By Lemma 4.2, Condition 3.1 holds; by Lemma 4.4, Condition 3.2 holds; and by Lemma 4.5, Conditions 3.3 and 3.4 hold. Hence, the conditions of Theorem 3.3 are satisfied, and by Remark 3.2 the convergence in (3.5) holds.

4.2. Proof of Theorem 3.2

Here we describe a probabilistically equivalent construction of G_n^{aux} , that was defined in Section 2.2. This alternative way of constructing G_n^{aux} is useful in the branching process approximation of the epidemic process since it allows us to run the epidemic process and construct G_n in unison.

Throughout this section we denote the probability measure of A by p , i.e. $p([0, x]) = \mathbb{P}(A \leq x)$ for $x \in [0, \infty)$. Similarly, we denote the probability measure of the size-biased version $\bar{A}$ of A by $\bar{p}$. Given the weights $A_1, \dots, A_n$, let $A_{(n)}$ be a random variable that follows the empirical distribution of $A_1, \dots, A_n$, and let p_n be the corresponding probability measure, i.e.

$$p_n([0, x]) = \mathbb{P}(A_{(n)} \leq x \mid A_1, \dots, A_n) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}(A_i \leq x)$$

for $x \in [0, \infty)$. Also let $\bar{A}_{(n)}$ denote the size-biased version of $A_{(n)}$, and let $\bar{p}_n$ be the corresponding probability measure (see the definition in (2.1)), i.e.

$$\bar{p}_n([0, x]) = \frac{1}{n\mu_A^{(n)}} \sum_{i=1}^n A_i \mathbf{1}(A_i \leq x).$$

Similarly, conditioned on the weights $B_1, \dots, B_m$, let $B_{(n)}$ be a random variable that follows the empirical distribution of $B_1, \dots, B_m$, and let $\bar{B}_{(n)}$ be the size-biased version of $B_{(n)}$.

To construct G_n^{aux} , start by picking some vertex u of V_n according to some rule (e.g. uniformly at random). Typically, u represents the initial case of the epidemic. Put $\mathcal{E}_0 = \mathcal{E}'_0 = \mathcal{N}_0 = \emptyset$. The component of G_n^{aux} that contains u is now constructed by iterating the following steps for $t = 1, 2, \dots$:

- (i). If $t = 1$, let $v = u$ be the vertex that is currently being explored. Generate the downshifted (i.e. reduced by one) *group degree* D of v from the distribution given in (2.2). Here, D represents the (additional) number of cliques of which v is a member.
- (ii). Draw D elements, $B_{(1)}, \dots, B_{(D)}$, from the multiset of (realised) weights $\{B_1, \dots, B_m\}$ independently with replacement. The probability of selecting $B_k \in \{B_1, \dots, B_m\}$ in a specific draw is given by $B_k/m\mu_B^{(n)}$, where $\mu_B^{(n)} = \sum_{k=1}^m B_k/m$. In other words, we generate D independent copies of $\bar{B}_{(n)}$.
- (iii). Let $v'_{(j)} \in V'_n$ be the vertex that corresponds to $B_{(j)}$, $j = 1, \dots, D$. For each $v'_{(j)} \in \{v_{(1)}, \dots, v_{(D)}\}$, if $v'_{(j)}$ is not a member of the set $\mathcal{E}'_{t-1} \subset V'_n$ of hitherto explored vertices, carry out the following steps 3a to 3d. If $v'_{(j)} \in \mathcal{E}'_{t-1}$ then the clique that corresponds to $v'_{(j)}$ is already explored, so $v'_{(j)}$ is excluded from these steps.
 - (a) Generate the downshifted clique size D'_j of $v'_{(j)}$ from the distribution given in (2.3) with $B_{(j)}$ in place of B_j .
 - (b) Select D'_j elements $A^j_{(1)}, \dots, A^j_{(D'_j)}$ from the multiset of (realised) weights $\{A_1, \dots, A_n\}$ independently with replacement. The probability of selecting A_k in a specific draw is given by $A_k/n\mu_A^{(n)}$, where $\mu_A^{(n)} = \sum_{k=1}^n A_k/n$. In other words, we generate D'_j independent copies of $\bar{A}_{(n)}$.
 - (c) Let $v^j_k \in V_n$ be the vertex that corresponds to $A^j_{(k)}$, $k = 1, \dots, D'_j$.
 - (d) Add an edge between each pair of distinct vertices in $\{v\} \cup \{v^j_k\}_{k=1}^{D'_j} \setminus \mathcal{E}_{t-1}$ to G_n .
- (iv). Update the set $\mathcal{E}'_t$ of explored cliques by putting $\mathcal{E}'_t = \mathcal{E}'_{t-1} \cup \{v'_{(1)}, \dots, v'_{(D)}\}$.
- (v). Update the set $\mathcal{N}'_t$ of neighbours of explored vertices by putting $\mathcal{N}'_t = \mathcal{N}'_{t-1} \cup \{v^j_k : j = 1, \dots, D; k = 1, \dots, D'_j\}$.
- (vi). Update the set $\mathcal{E}_t$ of explored vertices by putting $\mathcal{E}_t = \mathcal{E}_{t-1} \cup \{v\}$.
- (vii). If $\mathcal{N}'_t = \mathcal{E}_t$ then the construction of the component is complete and we exit the algorithm. Otherwise, update v by picking some new vertex in $\mathcal{N}'_t \setminus \mathcal{E}_t$. If G_n^{aux} is constructed as the epidemic progresses, then the new vertex v is the t th non-initial case.

If the nodes $v'_{(1)}, \dots, v'_{(D)}$ are not distinct or $\{v'_{(1)}, \dots, v'_{(D)}\} \cap \mathcal{E}_{t-1} \neq \emptyset$ in step (iii) for some iteration $t \in \{0, 1, \dots\}$ then the coupling of the approximating branching process and the epidemic process breaks down. Similarly, if in some iteration t the v^j_k are not all distinct, or $v^j_k \in \mathcal{E}_{t-1}$ for some $j \in \{1, \dots, D\}$ and $k \in \{1, \dots, D'_j\}$, then the coupling breaks down.

The following lemma, which is a variant of the birthday problem, ensures that w.h.p. the coupling of the approximating branching process and the epidemic process holds during the first $o(\sqrt{n})$ steps of the construction algorithm. The proof is included here for completeness.

Lemma 4.6. *Let $j_n = o(\sqrt{n})$, and let A have finite second moment. Suppose that we draw elements from the multiset $\{A_1, \dots, A_n\}$ independently with replacement, and that the probability that $A_i \in \{A_1, \dots, A_n\}$ is selected in a specific draw is proportional to A_i . Then the first j_n drawn elements are distinct with P -probability tending to 1 as $n \rightarrow \infty$.*

Proof. For $k = 1, 2, \dots$, let $A^{(k)}$ be the k th element that is drawn from $\{A_1, \dots, A_n\}$, and let $\mathbb{E}_n(\cdot)$ be the conditional expectation operator given $A_1, \dots, A_n$. Conditioned on $A_1, \dots, A_n$, for $k \geq 2$ the probability that $A^{(k)}$ is not distinct from $A^{(j)}$ for some $j \in \{1, \dots, k-1\}$ is less than or equal to

$$\mathbb{E}_n\left(\frac{A^{(1)} + \dots + A^{(k-1)}}{A_1 + \dots + A_n}\right) = (k-1) \frac{A_1^2 + \dots + A_n^2}{(A_1 + \dots + A_n)^2}.$$

Thus, by the union bound, conditioned on $\{A_1, \dots, A_n\}$ the probability that the first j_n drawn elements are not distinct is less than or equal to

$$\begin{aligned} \sum_{k=1}^{j_n} (k-1) \frac{A_1^2 + \dots + A_n^2}{(A_1 + \dots + A_n)^2} &= \frac{(j_n-1)j_n(A_1^2 + \dots + A_n^2)}{2(A_1 + \dots + A_n)^2} \\ &= \left(\frac{(j_n-1)j_n}{n}\right) \left(\frac{(A_1^2 + \dots + A_n^2)/n}{2(\mu_A^{(n)})^2}\right). \end{aligned}$$

Since A has finite second moments,

$$\frac{(A_1^2 + \dots + A_n^2)/n}{2(\mu_A^{(n)})^2}$$

converges to $\mathbb{E}(A^2)/2\mu_A^2$ in $\mathbb{P}$ -probability as $n \rightarrow \infty$. Hence,

$$\mathbb{P}(A^{(1)}, \dots, A^{(j_n)} \text{ are distinct}) \geq 1 - \mathbb{E}\left(1 \wedge \left(\frac{(j_n-1)j_n}{n}\right) \frac{(A_1^2 + \dots + A_n^2)/n}{2(\mu_A^{(n)})^2}\right),$$

where the right-hand side tends to 1 as $n \rightarrow \infty$. $\square$

In the construction algorithm described previously, the weights of explored vertices in V_n and V'_n are generated from the distributions of $\bar{A}_{(n)}$ and $\bar{B}_{(n)}$, whereas the weights in the approximating branching are generated from the distributions of $\bar{A}$ and $\bar{B}$. Therefore, in order to prove Theorem 3.2 we find upper bounds on the coupling error between $\text{MP}(\bar{A}_{(n)})$ and $\text{MP}(\bar{A})$, and between $\text{MP}(\bar{B}_{(n)})$ and $\text{MP}(\bar{B})$, which we state in Proposition 4.2 and Lemma 4.7.

Given the weights $A_1, \dots, A_n$, for a coupling $\mathcal{C}$ of A and $A_{(n)}$ we denote the corresponding probability measure and expectation by $\mathbb{P}_{\mathcal{C}}$ and $\mathbb{E}_{\mathcal{C}}$, respectively. A coupling of two random variables with distributions $\text{MP}(\bar{A})$ and $\text{MP}(\bar{A}_{(n)})$ can be constructed by first constructing a coupling $\mathcal{C}$ of their respective intensities $\bar{A}$ and $A_{(n)}$, then generating a joint realization $(\bar{A}', \bar{A}'_{(n)})$ of these intensities according to $\mathcal{C}$ and in the next step using these intensities to generate two random variables from the Poisson distribution. If in the last step a maximal coupling is used

then the (conditional) probability of a miscoupling is given by $\frac{1}{2}d_{\text{TV}}(\text{Po}(\bar{A}'), \text{Po}(\bar{A}_{(n)}'))$. Here, d_{TV} denotes the total variation distance, i.e. for $a, b \in \mathbb{R}_{\geq 0}$,

$$d_{\text{TV}}(\text{Po}(a), \text{Po}(b)) = \sum_{k \geq 0} \left| \frac{a^k e^{-a}}{k!} - \frac{b^k e^{-b}}{k!} \right|.$$

Thus, given the distribution of $A_{(n)}$, the probability of a miscoupling is given by $\frac{1}{2}\mathbb{E}_{\mathcal{C}}(d_{\text{TV}}(\text{Po}(\bar{A}), \text{Po}(\bar{A}_{(n)})))$.

Let the coupling $\mathcal{C}_n$ of $\bar{A}'$ and $\bar{A}_{(n)}'$ be given by

$$\mathcal{C}_n := \underset{\mathcal{C}}{\text{argmin}} \mathbb{E}_{\mathcal{C}} |\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}| \quad (4.8)$$

for each $n \geq 1$, where the minimum extends over all couplings $\mathcal{C}$ of A and $A_{(n)}$.

In the following proposition, $A_1, \dots, A_n$ are random with respect to $\mathbb{P}$. Thus, under $\mathbb{P}$, $\mathcal{C}_n$ is a coupling of $\bar{p}$ and the random probability measure $\bar{p}_n$.

Proposition 4.2. Assume that $\mathbb{E}(A^q) < \infty$ for $q > \frac{3}{2}$. Let $(\varepsilon_n)_{n \geq 1}$ be a sequence in $(0, \infty)$ such that if $q \neq 3$ then $\varepsilon_n \log(n) \rightarrow \infty$ as $n \rightarrow 0$ and if $q = 3$ then $\varepsilon_n \log(n) - \log(\log(n)) \rightarrow \infty$ as $n \rightarrow 0$. Then $\mathbb{P}(\mathbb{E}_{\mathcal{C}_n}(d_{\text{TV}}(\text{Po}(\bar{A}), \text{Po}(\bar{A}_{(n)}))) \geq n^{-\gamma + \varepsilon_n}) \rightarrow 0$ as $n \rightarrow \infty$, where $\gamma = \frac{1}{2} \wedge (q - \frac{3}{2})/q$.

Remark 4.2. Proposition 4.2 also holds if $A, \bar{A}$, and $\bar{A}_{(n)}$ are replaced by $B, \bar{B}$, and $\bar{B}_{(n)}$.

Proof of Proposition 4.2. We have (the first inequality following from [10, Theorem 1.C])

$$\begin{aligned} \mathbb{E}_{\mathcal{C}_n}(d_{\text{TV}}(\text{Po}(\bar{A}), \text{Po}(\bar{A}_{(n)}))) &\leq \mathbb{E}_{\mathcal{C}_n} \left(\frac{1}{\sqrt{\bar{A}_{(n)}} \vee \bar{A}} |\bar{A} - \bar{A}_{(n)}| \right) \\ &\leq \mathbb{E}_{\mathcal{C}_n} \left(\frac{1}{\frac{1}{2}(\sqrt{\bar{A}} + \sqrt{\bar{A}_{(n)}})} |\sqrt{\bar{A}} + \sqrt{\bar{A}_{(n)}}| \cdot |\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}| \right) \\ &= 2\mathbb{E}_{\mathcal{C}_n} (|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}|). \end{aligned}$$

Hence, by Proposition 4.3,

$$\mathbb{E}(\mathbb{E}_{\mathcal{C}_n}(d_{\text{TV}}(\text{Po}(\bar{A}), \text{Po}(\bar{A}_{(n)}))); D_n) = \begin{cases} \mathcal{O}(n^{-(q-3/2)/q}) & \text{if } \frac{3}{2} < q < 3, \\ \mathcal{O}(n^{-1/2} \log(n)) & \text{if } q = 3, \\ \mathcal{O}(n^{-1/2}) & \text{if } q > 3, \end{cases}$$

where D_n is the event $\max_{i \in [n]} (A_i) > 0$. The assertion of the proposition now follows from Markov's inequality and $\mathbb{P}(D_n^c) = p(0)^n$, $p(0) < 1$. $\square$

By the degree distributions of G_n^{aux} given in (2.2) and (2.3), in order to arrive at Theorem 3.2 we will also need a bound on the coupling error of $\text{MP}(\bar{B}_{(n)}\mu_A^{(n)}/\mu_A)$ and $\text{MP}(\bar{B}_{(n)})$, and of $\text{MP}(\bar{A}_{(n)}\mu_B^{(n)}/n\mu_A/\mu_B)$ and $\text{MP}(\bar{A}_{(n)})$.

Proposition 4.3. (c.f. [16, Theorem 1].) If $\mathbb{E}(A^q) < \infty$ for $q > \frac{3}{2}$ then

$$\mathbb{E}(\mathbb{E}_{\mathcal{C}_n} (|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}|); D_n) = \begin{cases} \mathcal{O}(n^{-(q-3/2)/q}) & \text{if } \frac{3}{2} < q < 3, \\ \mathcal{O}(n^{-1/2} \log(n)) & \text{if } q = 3, \\ \mathcal{O}(n^{-1/2}) & \text{if } q > 3, \end{cases}$$

where $\mathcal{C}_n$ is the coupling in (4.8) and D_n is the event that $\max_{i \in [n]} (A_i) > 0$.

Proof. This proof is, in part, analogous to the proof of [16, Theorem 1] and is presented here in full for completeness. The differences arise mainly due to the size-biasing of the weights.

Throughout, $C_1, C_2, \dots$ are positive constants that depend only on q and the distribution of A , and $U \subset [0, \infty)$ is a generic Borel set.

Note that $\mathbb{E}_{\mathcal{C}_n} |\sqrt{A} - \sqrt{A_{(n)}}|$ is the 1-Wasserstein distance between the distributions of $\sqrt{A}$ and $\sqrt{A_{(n)}}$. Hence, with the notation $2^m F = \{2^m u : u \in F\}$ for any event F , by [16, Lemmas 5 and 6],

$$\mathbb{E}_{\mathcal{C}_n} |\sqrt{A} - \sqrt{A_{(n)}}| \leq C_1 \sum_{m \geq 0} 2^m \sum_{\ell \geq 0} 2^{-\ell} \sum_{F \in \mathcal{P}_\ell} |\eta_n(2^m F \cap U_m) - \eta(2^m F \cap U_m)|, \quad (4.9)$$

where $\mathcal{P}_\ell$ is the partition of $[0, 1]$ that consists of $\{0\}$ and $2^{-\ell+1}k + (0, 2^{-\ell+1}]$ for $k \in \{0, 1, \dots, 2^{\ell-1} - 1\}$, $U_0 = [0, 1]$ and $U_m = [0, 2^m] \setminus [0, 2^{m-1}]$ for $m \geq 1$, and η and η_n are the probability distributions of $\sqrt{A}$ and $\sqrt{A_{(n)}}$, respectively. Now, with the notation $U^2 = \{u^2 : u \in U\}$, by the triangle inequality,

$$\begin{aligned} |\eta_n(U) - \eta(U)| &= |\bar{p}_n(U^2) - \bar{p}(U^2)| \\ &= \left| \frac{\sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2)}{n\mu_A^{(n)}} - \frac{\mathbb{E}(A \mathbf{1}(A \in U^2))}{\mu_A} \right| \\ &\leq \frac{1}{n\mu_A^{(n)}} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) \left| 1 - \frac{\mu_A^{(n)}}{\mu_A} \right| \\ &\quad + \frac{1}{\mu_A} \left| \frac{1}{n} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) - \mathbb{E}(A \mathbf{1}(A \in U^2)) \right|. \end{aligned} \quad (4.10)$$

The second term on the right-hand side in (4.10) satisfies (with the first inequality following from Jensen's inequality)

$$\begin{aligned} \mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) - \mathbb{E}(A \mathbf{1}(A \in U^2)) \right| \right) &\leq \sqrt{\text{Var} \left(\frac{1}{n} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) \right)} \\ &\leq (\sup U)^2 \sqrt{\frac{1}{n} p(U^2)} \end{aligned} \quad (4.11)$$

and, again by the triangle inequality,

$$\mathbb{E} \left(\left| \frac{1}{n} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) - \mathbb{E}(A \mathbf{1}(A \in U^2)) \right| \right) \leq 2(\sup U)^2 p(U^2). \quad (4.12)$$

Combining (4.11) and (4.12) gives that the second term on the right-hand side in (4.10) satisfies

$$\frac{1}{\mu_A} \left| \frac{1}{n} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) - \mathbb{E}(A \mathbf{1}(A \in U^2)) \right| \leq C_2 (\sup U)^2 \left(\sqrt{\frac{1}{n} p(U^2)} \wedge p(U^2) \right). \quad (4.13)$$

In order to find a similar upper bound on the first term on the right-hand side in (4.10) we define the event S_n as

$$S_n := \left\{ \mu_A^{(n)} \leq \mu_A - \kappa \right\}, \quad (4.14)$$

where $\kappa \in (0, \mu_A)$ is a fixed constant such that $e^{(\mu_A - \kappa)} \mathbb{E}(e^{-A}) < 1$. Then, with the second inequality following from Hölder's inequality and the last inequality from $\sum_{i=1}^n \mathbf{1}(A_i \in U^2) \sim \text{Bin}(n, p(U^2))$,

$$\begin{aligned} & \mathbb{E} \left(\left| \frac{1}{n\mu_A^{(n)}} \sum_{i=1}^n A_i \mathbf{1}(A_i \in U^2) \right| \left| 1 - \frac{\mu_A^{(n)}}{\mu_A} \right|; S_n^c \cap D_n \right) \\ & \leq \frac{(\sup U)^2}{(\mu_A - \kappa)} \left(\frac{1}{n} \mathbb{E} \left(\sum_{i=1}^n \mathbf{1}(A_i \in U^2) \left| 1 - \frac{\mu_A^{(n)}}{\mu_A} \right| \right) \wedge (2p(U^2)) \right) \\ & \leq \frac{(\sup U)^2}{(\mu_A - \kappa)} \left(\sqrt{\frac{1}{n^3} \mathbb{E} \left(\left(\sum_{i=1}^n \mathbf{1}(A_i \in U^2) \right)^2 \right) \text{Var} \left(\frac{A}{\mu_A} \right)} \wedge (2p(U^2)) \right) \\ & \leq \frac{(\sup U)^2}{(\mu_A - \kappa)} \left(\sqrt{\frac{1}{n^3} np(U^2)(np(U^2) + 1) \text{Var} \left(\frac{A}{\mu_A} \right)} \wedge (2p(U^2)) \right) \\ & \leq C_3 (\sup U)^2 \left(\sqrt{\frac{1}{n} p(U^2)} \wedge p(U^2) \right). \end{aligned} \quad (4.15)$$

By the inequalities in (4.13) and (4.15), together with the fact that, for any $G \subset [0, \infty)$, $\sup(G \cap U_m^2) \leq \sup(U_m^2) = 2^{2m}$, we have

$$\begin{aligned} & \sum_{F \in \mathcal{P}_\ell} \mathbb{E}(|\eta_n(2^m F \cap U_m) - \eta(2^m F \cap U_m)|; D_n \cap S_n^c) \\ & \leq C_4 \left(2^{2m} \sum_{F \in \mathcal{P}_\ell} \left(\sqrt{\frac{1}{n} p(2^{2m} F^2 \cap U_m^2)} \wedge p(2^{2m} F^2 \cap U_m^2) \right) \right) \end{aligned} \quad (4.16)$$

and (with the second inequality following from the Cauchy–Schwarz inequality and the fact that, since $\mathbb{E}(A^q)$ is finite, $\mathbb{P}(U_m^2) \leq \mathbb{E}(A^q) 2^{-2qm}$)

$$\begin{aligned} & \sum_{F \in \mathcal{P}_\ell} \left(\sqrt{\frac{1}{n} p(2^{2m} F^2 \cap U_m^2)} \wedge p(2^{2m} F^2 \cap U_m^2) \right) \leq \left(\frac{1}{\sqrt{n}} \sum_{F \in \mathcal{P}_\ell} \sqrt{p(2^{2m} F^2 \cap U_m^2)} \right) \wedge p(U_m^2) \\ & \leq C_5 \left(\left(\frac{1}{\sqrt{n}} \sqrt{|\mathcal{P}_\ell| p(U_m^2)} \right) \wedge 2^{-2qm} \right) \\ & \leq C_5 \left(\left(\frac{1}{\sqrt{n}} \sqrt{2^\ell \mathbb{E}(A^q) 2^{-2qm}} \right) \wedge 2^{-2qm} \right) \\ & \leq C_6 \left(\frac{1}{\sqrt{n}} 2^{\ell/2 - qm} \wedge 2^{-2qm} \right). \end{aligned} \quad (4.17)$$

Hence, by (4.9), (4.16), and (4.17),

$$\begin{aligned} \mathbb{E}(\mathbb{E}_{\mathcal{E}_n}(|\sqrt{A} - \sqrt{A_{(n)}}|); S_n^c \cap D_n) &\leq C_7 \sum_{m \geq 0} 2^m \sum_{\ell \geq 0} 2^{-\ell+2m} \left(\frac{1}{\sqrt{n}} 2^{\ell/2-qm} \wedge 2^{-2qm} \right) \\ &\leq C_7 \sum_{m \geq 0} 2^{3m} \sum_{\ell \geq 0} 2^{-\ell/2} \left(\frac{1}{\sqrt{n}} 2^{-qm} \wedge 2^{-2qm} \right) \\ &\leq C_8 \sum_{m \geq 0} \left(\frac{1}{\sqrt{n}} 2^{m(3-q)} \wedge 2^{-m(2q-3)} \right). \end{aligned}$$

If $q > 3$ then

$$C_8 \sum_{m \geq 0} \left(\frac{1}{\sqrt{n}} 2^{m(3-q)} \wedge 2^{-m(2q-3)} \right) \leq C_9 \frac{1}{\sqrt{n}}.$$

If $q \in (\frac{3}{2}, 3)$ then, with $m_n = \lfloor \log(n)/(2q \log(2)) \rfloor$,

$$C_8 \sum_{m \geq 0} \left(\frac{1}{\sqrt{n}} 2^{m(3-q)} \wedge 2^{-m(2q-3)} \right) \leq C_8 \sum_{m=0}^{m_n} \frac{1}{\sqrt{n}} 2^{-m(q-3)} + C_8 \sum_{m > m_n} 2^{-m(2q-3)} = O(n^{-(1-3/2q)}).$$

If $q = 3$ then, with $a_n = \lfloor \log(n)/\log(2) \rfloor$,

$$C_8 \sum_{m \geq 0} \left(\frac{1}{\sqrt{n}} 2^{m(3-q)} \wedge 2^{-m(2q-3)} \right) \leq C_8 \frac{a_n}{\sqrt{n}} + C_8 \sum_{m \geq a_n} 2^{-m(2q-3)} = O(n^{-1/2} \log(n)).$$

Thus, it only remains to bound the expectation of $\mathbb{E}_{\mathcal{E}_n}(|\sqrt{A} - \sqrt{A_{(n)}}|)$ on $S_n \cap D_n$, where S_n is the event in (4.14). Now, with the third inequality following from the fact that on S_n we have $\sqrt{A_i} \leq \sqrt{n(\mu_A - \kappa)}$ for $i = 1, \dots, n$,

$$\begin{aligned} \mathbb{E}(\mathbb{E}_{\mathcal{E}_n}(|\sqrt{A} - \sqrt{A_{(n)}}|); S_n \cap D_n) &\leq \mathbb{P}(S_n \cap D_n) \mathbb{E}(\sqrt{A}) + \mathbb{E}(\sqrt{A_{(n)}}; S_n \cap D_n) \\ &= \mathbb{P}(S_n \cap D_n) \mathbb{E}(\sqrt{A}) + \mathbb{E}\left(\frac{\sum_{i=1}^n A_i^{3/2}}{\sum_{i=1}^n A_i}; S_n \cap D_n\right) \\ &\leq \mathbb{P}(S_n \cap D_n) \mathbb{E}(\sqrt{A}) + \mathbb{E}\left(\sum_{i=1}^n \sqrt{A_i}; S_n \cap D_n\right) \\ &\leq \mathbb{P}(S_n \cap D_n) \mathbb{E}(\sqrt{A}) + \mathbb{E}(n^{3/2} \sqrt{\mu_A - \kappa}; S_n \cap D_n) \\ &= O(n^{3/2} \mathbb{P}(S_n)) = O(n^{3/2} e^{n(\mu_A - \kappa)} \mathbb{E}(e^{-A})^n), \end{aligned}$$

where the last step follows from the Chernoff bound $\mathbb{P}(S_n) \leq e^{n(\mu_A - \kappa)} \mathbb{E}(e^{-A})^n$. The assertion now follows by recalling that $e^{(\mu_A - \kappa)} \mathbb{E}(e^{-A}) < 1$. $\square$

Lemma 4.7. Let ε_n be as in Proposition 4.2, and $\mathbb{E}(A^2), \mathbb{E}(B^2) < \infty$. Then, w.h.p.,

$$d_{\text{TV}}(\text{Po}(\bar{B}_{(n)} \mu_A^{(n)} / \mu_A), \text{Po}(\bar{B}_{(n)})) \leq n^{-1/2+\varepsilon_n}, \quad (4.18)$$

$$d_{\text{TV}}(\text{Po}(\bar{A}_{(n)} \mu_B^{(n)} \lfloor n\mu_A / \mu_B \rfloor / n\mu_A), \text{Po}(\bar{A}_{(n)})) \leq n^{-1/2+\varepsilon_n}. \quad (4.19)$$

Proof. Define $H_n := \{ |1 - \mu_A^{(n)} / \mu_A| \leq n^{-1/2+(1/2)\varepsilon_n} \}$. By Chebyshev's inequality, $\mathbb{P}(H_n^c) = \mathcal{O}(n^{-\varepsilon_n})$. Now (again by [10, Theorem 1.C]),

$$\begin{aligned} \mathbb{E}(d_{\text{TV}}(\text{Po}(\overline{B}_{(n)}\mu_A^{(n)} / \mu_A), \text{Po}(\overline{B}_{(n)})); H_n) &\leq \mathbb{E}(|\overline{B}_{(n)}\mu_A^{(n)} / \mu_A - \overline{B}_{(n)}|; H_n) \\ &\leq \mathbb{E}(\overline{B}_{(n)}n^{-1/2+(1/2)\varepsilon_n}) = \mathcal{O}(n^{-1/2+(1/2)\varepsilon_n}). \end{aligned}$$

This implies, using the union bound and Markov's inequality,

$$\mathbb{P}(d_{\text{TV}}(\text{Po}(\overline{B}_{(n)}\mu_A^{(n)} / \mu_A), \text{Po}(\overline{B}_{(n)})) \geq n^{-1/2+\varepsilon_n}) \leq \mathbb{P}(H_n^c) + \mathcal{O}(n^{-1/2\varepsilon_n}).$$

This proves the inequality in (4.18), and the proof of the inequality in (4.19) is completely analogous. $\square$

Proof of Theorem 3.2. Let ε_n , q , and γ be as in Theorem 3.2, and let $T \in \mathbb{N}_0$ be the number of iterations in the construction algorithm on page 16 (i.e. $\mathcal{N}_T = \mathcal{E}_T$). By Lemma 4.6, w.h.p. the vertices of V_n and V'_n that are explored in the first $\lfloor n^{\gamma-\varepsilon_n} \rfloor \wedge T$ steps of the construction algorithm are distinct. Combining Lemma 4.6 with Proposition 4.2 gives the assertion of Theorem 3.2 by the union bound. $\square$

Acknowledgements

The authors would like to thank Pieter Trapman for fruitful discussions and constructive criticism of the paper. We would also like to thank an anonymous referee for their many valuable comments, which contributed to a considerable improvement of the paper.

Funding information

This research is supported by the Swedish Research Council (Vetenskapsrådet) grant 2016-04566.

Competing interests

There were no competing interests to declare which arose during the preparation or publication process of this article.

References

- [1] BALL, F., MOLLISON, D. AND SCALIA-TOMBA, G. (1997). Epidemics with two levels of mixing. *Ann. Appl. Prob.* **7**, 46–89.
- [2] BALL, F. AND NEAL, P. (2002). A general model for stochastic SIR epidemics with two levels of mixing. *Math. Biosci.* **180**, 73–102.
- [3] BALL, F. AND NEAL, P. (2003). The great circle epidemic model. *Stoch. Process. Appl.* **107**, 233–268.
- [4] 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.
- [5] BALL, F. AND SHAW, L. (2015). Estimating the within-household infection rate in emerging SIR epidemics among a community of households. *J. Math. Biol.* **71**, 1705–1735.
- [6] BALL, F. AND SIRL, D. (2012). An SIR epidemic model on a population with random network and household structure, and several types of individuals. *Adv. Appl. Prob.* **44**, 63–86.
- [7] BALL, F., SIRL, D. AND TRAPMAN, P. (2009). Threshold behaviour and final outcome of an epidemic on a random network with household structure. *Adv. Appl. Prob.* **41**, 765–796.
- [8] BALL, F., SIRL, D. AND TRAPMAN, P. (2010). Analysis of a stochastic SIR epidemic on a random network incorporating household structure. *Math. Biosci.* **224**, 53–73.
- [9] BALL, F., SIRL, D. AND TRAPMAN, P. (2014). Epidemics on random intersection graphs. *Ann. Appl. Prob.* **24**, 1081–1128.

- [10] BARBOUR, A., HOLST, L. AND JANSON, S. (1992). *Poisson Approximation*. Clarendon Press, Oxford.
- [11] BECKER, N. G. AND DIETZ, K. (1995). The effect of household distribution on transmission and control of highly infectious diseases. *Math. Biosci.* **127**, 207–219.
- [12] BIGGINS, J. D. AND KYPRIANOU, A. E. (2004). Measure change in multitype branching. *Adv. Appl. Prob.* **36**, 544–581.
- [13] BRITTON, T., DEIJFEN, M., LAGERÅS, A. N. AND LINDHOLM, M. (2008). Epidemics on random graphs with tunable clustering. *J. Appl. Prob.* **45**, 743–756.
- [14] BROADBENT, S. R. AND HAMMERSLEY, J. M. (1957). Percolation processes: I. Crystals and mazes. *Math. Proc. Camb. Phil. Soc.* **53**, 629–641.
- [15] DYE, C. *et al.* (2015). Ebola virus disease in West Africa: The first 9 months. *New England J. Medicine* **372**, 189.
- [16] FOURNIER, N. AND GUILLIN, A. (2015). On the rate of convergence in Wasserstein distance of the empirical measure. *Prob. Theory Relat. Fields* **162**, 707–738.
- [17] FRISCH, H. AND HAMMERSLEY, J. (1963). Percolation processes and related topics. *J. SIAM* **11**, 894–918.
- [18] HEESTERBEEK, H., BRITTON, T. AND DIEKMANN, O. (2013). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.
- [19] IKSANOV, A. AND MEINERS, M. (2015). Rate of convergence in the law of large numbers for supercritical general multi-type branching processes. *Stoch. Process. Appl.* **125**, 708–738.
- [20] JAGERS, P. (1975). *Branching Processes with Biological Applications*. John Wiley, Chichester.
- [21] JAGERS, P. (1989). General branching processes as Markov fields. *Stoch. Process. Appl.* **32**, 183–212.
- [22] JAGERS, P. (1993). Stabilities and instabilities in population dynamics. In *Stability Problems for Stochastic Models*, eds V. V. Kalashnikov and V. M. Zolotarev. Springer, Berlin, pp. 58–67.
- [23] KAROŃSKI, M., SCHEINERMAN, E. AND SINGER-COHEN, K. (1999). On random intersection graphs: The subgraph problem. *Combinatorics Prob. Comput.* **8**, 131–159.
- [24] LUDWIG, D. (1975). Final size distribution for epidemics. *Math. Biosci.* **23**, 33–46.
- [25] NEAL, P. (2008). The SIS great circle epidemic model. *J. Appl. Prob.* **45**, 513–530.
- [26] NERMAN, O. (1981). On the convergence of supercritical general (C-M-J) branching processes. *Z. Wahrscheinlichkeitsth.* **57**, 365–395.
- [27] NISHIURA, H. AND CHOWELL, G. (2014). Early transmission dynamics of ebola virus disease (EVD), West Africa, March to August 2014. *Eurosurveillance* **19**, 20894.
- [28] NUMMELIN, E. (2004). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge University Press.
- [29] OLOFSSON, P. (1996). Branching processes with local dependencies. *Ann. Appl. Prob.* **6**, 238–268.
- [30] PELLIS, L., BALL, F. AND TRAPMAN, P. (2012). Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of inline853. *Math. Biosci.* **235**, 85–97.
- [31] PELLIS, L., FERGUSON, N. AND FRASER, C. (2011). Epidemic growth rate and household reproduction number in communities of households, schools and workplaces. *J. Math. Biol.* **63**, 691–734.
- [32] SAGITOV, S. (2017). Regenerative multi-type Galton–Watson processes. Preprint, arXiv:1708.02636.
- [33] WATSON, R. K. (1972). On an epidemic in a stratified population. *J. Appl. Prob.* **9**, 659–666.
- [34] WATTS, D. AND STROGATZ, S. (1998). Collective dynamics of small world networks. *Nature* **393**, 440–442.