Mathematical Statistics
Stockholm University

The real-time growth rate of stochastic epidemics on random intersection graphs

Carolina Fransson

Research Report 2022:8

ISSN 1650-0377

Postal address:

Mathematical Statistics
Dept. of Mathematics
Stockholm University
SE-106 91 Stockholm
Sweden

Internet:

http://www.math.su.se

Mathematical Statistics
Stockholm University
Research Report **2022:8**,
http://www.math.su.se

The real-time growth rate of stochastic epidemics on random intersection graphs

Carolina Fransson*

June 2022

Abstract

This paper is concerned with the growth rate of SIR (Susceptible-Infectious-Recovered) epidemics with general infectious period distribution on random intersection graphs. This type of graph is characterized 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.

1 Introduction

In the earliest epidemic models, it is assumed that the disease spreads in a population consisting of homogeneous 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 be socially active than others, or the population may exhibit a more complex social structure than ho-

*Stockholm University, Sweden; carolina.fransson@math.su.se

mogeneous 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 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 (Heesterbeek et al. 2013; Watson 1972).

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 amount of triangles and other short circuits). Some examples include the great circle model (Ball et al. 1997; Ball and Neal 2003; Neal 2008) and the closely related small-world network model (Watts and Strogatz 1998), 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 has been introduced (Ball et al. 1997; Becker and Dietz 1995). A clique may, for instance, represent a household, workplace or school. Models with this feature have been investigated in various forms, see for instance Ball and Neal (2002), Ball et al. (2016), Ball and Sirl (2012), Ball et al. (2009, 2010), and Pellis et al. (2012), 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 (Dye et al. 2015; Nishiura and Chowell 2014). 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 concerns 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 by Ludwig (1975) but was, however, implicitly present in earlier literature

on percolation (Broadbent and Hammersley 1957; Frisch and Hammersley 1963). For this reason, many models lend themselves more readily to analysis of the final outcome than of the real-time-growth rate. Pellis et al. (2011) 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, Ball and Shaw (2015) provided methods to estimate the within-household infection rate for an 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 (Pellis et al. 2011), 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 to point 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 (Pellis et al. 2011). 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 Iksanov and Meiners (2015), 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 Sagitov (2017) calls regenerative Galton-Watson processes and to the concepts of local infectious clumps and global contacts in Ball and Neal (2002), see also Olofsson (1996) 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 (Karoński et al. 1999). 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 amount 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 (Karoński et al. 1999) in which the degrees distributions are mixed Poisson. Epidemics on graphs of this type have previously been studied

in Ball et al. (2014), 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 (Britton et al. 2008), 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 and in section 2.3 we define the epidemic model. The main results are presented in section 3. Section 4.1 and 5 contains 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 notation conventions and abbreviations that will be frequently used 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 $n \in \mathbb{Z}_{\geq 1}$, $[n] = \{1, \ldots, n\}$.

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

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

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

We will make frequent use of the abbreviations MP (mixed Poisson) and SIR (Susceptible $\to$ Infectious $\to$ 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 probability (w.h.p.) if the probability of the event tends to 1 as $n \to \infty$, 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 by Ball et al. (2014), who 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 $n \to \infty$, 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 will give a somewhat brief description of this graph model and refer the reader to Ball et al. (2014) 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 $E(A) = \mu_A$ and $E(B) = \mu_B$. We make the following assumption on A and B.

Assumption 1.

i) $P(A \geq 0) = P(B \geq 0) = 1$

ii) $P(A = 0) < 1$ and $P(B = 0) < 1$

iii) $E(A^2 \log_+ A) < \infty$ and $E(B^2 \log_+ B) < \infty$

We will refer to the condition of assumption 1 iii) as the $x^2 \log x$-condition.

Remark 1. This version of the random intersection graph can be constructed under less strict assumptions on than the $x^2 \log x$-condition (see Ball et al. (2014)), it is however needed here for the approximating branching process to satisfy the classical $x \log x$-condition.

Let $\{A_k\}_k$ and $\{B_k\}_k$ be two sequences of independent copies of A and B, respectively. Let further $V_n = \{v_1, \ldots, v_n\}$ be the vertex set of G_n, and assign the weight A_i to the vertex v_i, $i = 1, \ldots, n$. As an intermediate step, we construct an auxiliary graph G_n^{aux} with vertex set $V_n \cup V_n'$, where $V_n' = \{v_1', \ldots, v_m'\}$ and

$$m = m(n) := \lfloor n\mu_A/\mu_B \rfloor .$$

Assign the weight B_j to the vertex v_j', $j = 1, \ldots, m$. 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

$$Po\left(\frac{A_i B_j}{n\mu_A}\right),$$

independently for pairs v_i, v'_j. Thus in G_n^{aux} the degree of $v_i \in V_n$ has distribution

$$Po\left(A_i \frac{\mu_B^{(n)} \lfloor n\mu_A/\mu_B \rfloor}{n\mu_A}\right) \tag{1}$$

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

$$Po\left(B_j \frac{\mu_A^{(n)}}{\mu_A}\right) \tag{2}$$

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 point out that each clique $\mathcal{C} = (V_\mathcal{C}, E_\mathcal{C})$ of G_n may be viewed as a (directed) subgraph of G_n. Here the vertex set $V_\mathcal{C} \subset V_n$ consists of the neighbours (in G_n^{aux}) of v' where $v' \in V'_n$ is the vertex that corresponds to $\mathcal{C}$, and $E_\mathcal{C}$ is the edge set of the simple directed complete graph $\mathcal{C}$ (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

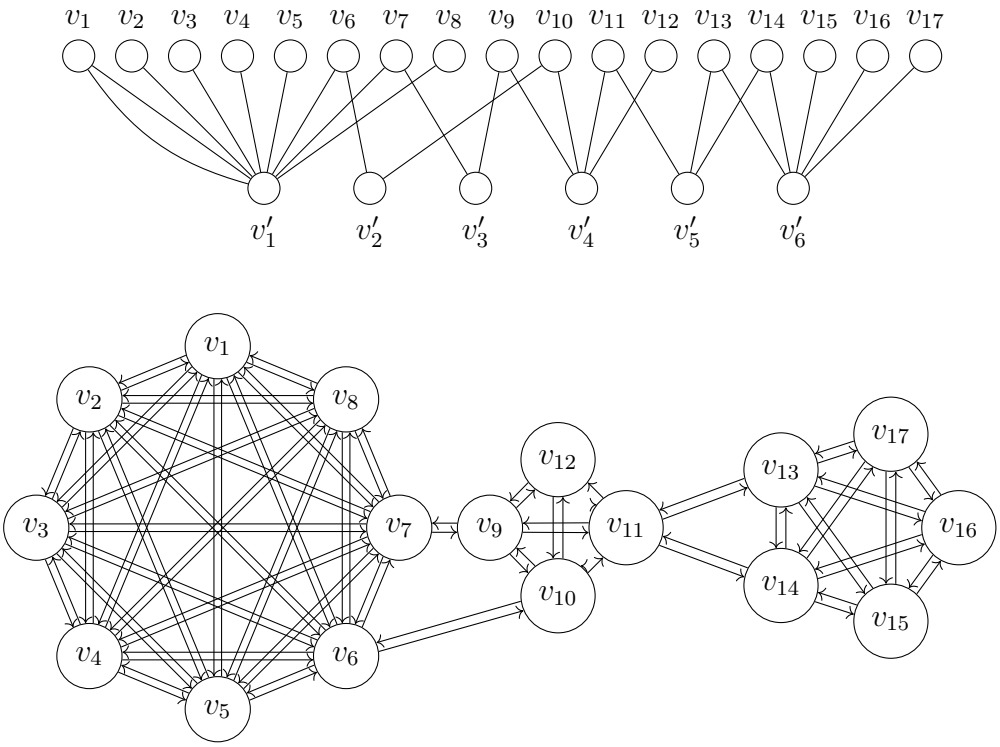

Figure 1: Construction of G_n for $n = 17$ with $|V'_n| = |\{v'_1, \ldots, v'_6\}| = 6$ cliques. Top: the auxiliary graph G_n^{aux}. Bottom: the resulting directed graph G_n.

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 identically distributed random variables with support in $[0, \infty)$ such that $T_{i_1 j}$ and $T_{i_2 j}$ are independent if $i_1 \neq i_2$. Similarly, we assume that $T_{i_1 j}$ and I_{i_2} are independent if $i_1 \neq i_2$. 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 exponetially 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 the 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. $P(T'_{ij} = 0) < 1/(\mu_{\bar{A}}\mu_{\bar{B}})$ and the distribution of T'_{ij} is non-lattice (i.e. $P(T'_{ij} \in \{\infty, 0, s, 2s, \ldots\}) < 1$ for any $s > 0$).

The first part of Assumption 2, $P(T'_{ij} = 0) < 1/(\mu_{\bar{A}}\mu_{\bar{B}})$, 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 $\varsigma = (v_{i_1}, v_{i_2}, ..., v_{i_k})$ is any finite sequence of vertices of G_n such that $(v_{i_r}, v_{i_{r+1}})$ is an edge of G_n, $r = 1, \ldots k-1$. We define the length $\ell(\varsigma)$ of a path $\varsigma = (v_{i_1}, v_{i_2}, ..., v_{i_k})$ as

$$\ell(\varsigma) = \sum_{r=1}^{k-1} T'_{i_r i_{r+1}}.$$

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

$$d(u, v) := \min_{\varsigma} \ell(\varsigma)$$

where the minimum is taken over all paths $\varsigma \in \Sigma_{uv}$. 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. 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, $i = 1, \ldots, n$, 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 $\Sigma^{\mathcal{C}}_{uv}$ be the collection of paths from u to v restricted

to $\mathcal{C}$. That is, a path ς from u to v belongs $\Sigma_{uv}^{\mathcal{C}}$ whenever every edge of ς 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

$$d_{\mathcal{C}}(u,v) := \min_{\varsigma} \ell(\varsigma)$$

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

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

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 (Nerman 1981) via a coupling of an epidemic process on G_n and a single-type 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, however, 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. To deal with this, we will employ the branching process embedding presented below.

In the early phase of an outbreak, the epidemic process on G_n can be coupled with a branching process Z with type space

$$\mathcal{T}_\theta := \mathcal{T} \cup \{\theta\},$$

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 Nummelin (2004, Def. 4.3). 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 asymptotic results (see Section 4.1 for an overview) that are available for single-type branching processes.

In Z, the type of an individual that was infected via a clique $\mathcal{C}$ of size $|V_{\mathcal{C}}| = 2$ is taken to be θ, and the type of any other individual (except

the individual that corresponds to the initial infectious case) is taken to be the length of its infectious period. 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 is the number of cliques that has to be traversed to reach v_* (including the cliques of v and v_*). If the extinction probability of the approximating branching process Z is strictly smaller than 1 then Z is said to be *supercritical*, and we say that we are in the supercritical regime.

Let $\mathcal{S}$ be the set of all individuals of Z. We assume that there is one individual a_0 of generation 0 from which every other individual of $\mathcal{S}$ stems. We will refer to the common ancestor $a_0 \in \mathcal{S}$ as *the first ancestor* of Z. 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 $MP(A)$ cliques while the number of cliques that a non-initial case is a member of is distributed as $\bar{D}$ cliques, where $D \sim MP(A)$. The following claim is easily checked (see also Ball et al. (2014)).

Claim 1. If $D \sim MP(A)$ then $(\bar{D} - 1) \sim MP(\bar{A})$.

Thus, a non-initial case is a member of approximately $MP(\bar{A})$ cliques that are not yet affected by the disease. By a similar reasoning the size of a clique (excluding the primary case of the clique) reached during the early phase of the epidemic has approximately distribution $MP(\bar{B})$.

Now, 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 $p_k^{\bar{B}} := P(MP(\bar{B}) = k)$ and define the vector $\Gamma^T = \mu_{\bar{A}}(1 \cdot p_1^{\bar{B}}, 2 \cdot p_2^{\bar{B}}, \ldots)$ where $\mu_{\bar{A}} = E(\bar{A}) = E(A^2)/\mu_A$. For a clique $\mathcal{C}$ of size $|V_{\mathcal{C}}| = k$ and any $\lambda \geq 0$ let

$$\mathcal{L}_k^{(\lambda)} := \int_{\mathbb{R}_{\geq 0}} e^{-t\lambda} F_k(dt) \tag{3}$$

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

$$L^{(\lambda)} = (\mathcal{L}_2^{(\lambda)}, \mathcal{L}_3^{(\lambda)}, \ldots)^T. \tag{4}$$

As we will prove in Section 5, the Malthusian parameter can be found by solving the Euler-Lotka equation:

Definition 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 1. Under Assumptions 1 i)-iii) and 2 and if the Malthusian parameter $\alpha > 0$ exists then there exists a non-negative integrable random variable W such that

$$\frac{|Z_t|}{e^{\alpha t}} \xrightarrow{a.s} W$$

where W satisfies $P(\{|Z_t| \not\to 0\} \Delta \{W > 0\}) = 0$.

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 $\{|Z_t| \not\to 0\}$ it is implicit that the limit is taken as t tends to ∞, i.e. $\{|Z_t| \not\to 0\}$ is the event that the branching process population of Z ultimately avoids extinction.

Theorem 2. Let $(G_n)_n$ be a sequence of graphs generated via the random intersection graph model and assume that the assumptions of Theorem 1 hold and let $(\varepsilon_n)_{n\geq 1}$ be a sequence in $(0, \infty)$ that satisfies $\varepsilon_n \log(n) \to \infty$ as $n \to 0$. Then for any $q \geq 2$, $q \neq 3$, that satisfies $E(A^q) < \infty$ and $E(B^q) < \infty$ 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 $\gamma = \min\left(\frac{1}{2}, \frac{q-3/2}{q}\right)$.

4 Branching process theory

To present the main idea that underpins the proof of Theorem 1, we need a slightly more formal description of Z (see e.g. Jagers (1989) for a full formal framework). We assume that each individual of the branching process tree Z has a countable infinite number of children, which makes $\mathcal{S}$ countable. The basic probability space on which Z is defined is given by the product probability space

$$\prod_{v \in \mathcal{S}} (\Omega_v, \mathcal{A}_v).$$

On each $(\Omega_v, \mathcal{A}_v)$ there is defined a point process ξ_v on $[0, \infty] \times \mathcal{T}_\theta$, which we refer to as the reproduction point process of v; a point (t, r) of ξ_v corresponds to a type-r individual produced by v at age t, and there is a one-to-one correspondence between the children of v and the points of ξ_v. Note that $t = \infty$ is allowed; this has the interpretation that v and its descendants are never born. For each $u \in \mathcal{S}$, let $\tau_u \in [0, \infty]$ be the time point of u's birth. We say that $u \in \mathcal{S}$ is *realized* if $\tau_u < \infty$, and write $|v| = n$ to indicate that an individual $v \in \mathcal{S}$ of Z belongs to generation n, $n \geq 0$, and for any $v \in \mathcal{S}$ we denote the type of v by $\sigma(v) \in \mathcal{T}_\theta$.

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 $v \in V_{\mathcal{C}}$ for which the time elapsed since the corresponding primary case u contracted the disease is $d_{\mathcal{C}}(u, v) = t$. Each individual $v \in \mathcal{S}$ is assigned an infectious period I_v; given the type $\sigma(v)$ of v, I_v is an independent copy of the generic infectious period if either $\sigma(v) = \theta$ or $v = a_0$ (i.e. v is the first ancestor) , and $I_v = \sigma(v)$ otherwise. For a_0, the points of ξ_{a_0} corresponds to the secondary cases where the corresponding primary case has infectious period I_{a_0} and is a member of $MP(A)$ cliques of independent sizes with distribution $MP(\bar{B})$. Similarly, if $v \neq a_0$ the points of ξ_v corresponds to the secondary cases where the corresponding primary case has infectious period I_v and is a member of $MP(\bar{A})$ cliques, each of size $MP(\bar{B})$. The number $|Z_t|$ of individuals of Z that are alive at time $t \geq 0$, correponds to the number of infectious individuals at t and is given by is the cardinality of the set

$$\{u \in \mathcal{S} : \tau_u \leq t < \tau_u + I_u\}. \tag{5}$$

Before proceeding, we introduce some additional notation. The individuals of Z can be partially ordered by descent; 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 $x \prec y$ (or $y \succ x$) to indicate that $x \preceq y$ and $x \neq y$. Similarly, for $\mathcal{J} \subset \mathcal{S}$ and $x \in \mathcal{S}$ we write $\mathcal{J} \prec x$ to indicate that $y \prec x$ for some $y \in \mathcal{J}$, and $\mathcal{J} \preceq x$ to indicate that $y \preceq x$ for some $y \in \mathcal{J}$.

To arrive at Theorem 1, 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 $\mathcal{S}_\theta \subset \mathcal{S}$ be the set of the type-θ individuals of Z. For $x \in \mathcal{S}_\theta$, define the block $\mathscr{B}_x$ as follows:

$$\mathscr{B}_x := \{y \in \mathcal{S} : x \preceq y, \text{ and whenever } x \prec z \text{ for some } z \in \mathcal{S}_\theta \text{ then } z \not\preceq y\}. \tag{6}$$

In words, for any $x \in \mathcal{S}_\theta$ the block $\mathscr{B}_x$ 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 $\mathcal{S}_\theta$ and the children of $x \in \mathcal{S}_\theta$ (seen as an individual of Y) the type-θ children of individuals of $\mathscr{B}_x$ (in Z). That is, if we define $\mathscr{J}_n$, $n \geq 1$, recursively as follows

$$\mathscr{J}_0 = \{a_0\}$$

and

$$\mathscr{J}_n = \{x \in \mathcal{S}_\theta : x \succ \mathscr{J}_{n-1} \text{ and whenever } \mathscr{J}_{n-1} \prec z \prec x \text{ it holds that } z \notin \mathcal{S}_\theta\},$$

then $\mathscr{J}_n$ 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. Nerman (1981)) 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

$$\phi_x(t) = |\{y \in \mathscr{B}_x : \tau_y \leq t < \tau_y + I_y\}| \tag{7}$$

for each $x \in \mathscr{J} = \cup_{n \geq 0} \mathscr{J}_n$ and we say that

$$Y_t^{\phi} := \sum_{x \in \mathscr{J}} \phi_x(t - \tau_x)$$

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 belongs to the block $\mathscr{B}_x$ 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

$$|Z_t| = Y_t^{\phi} \tag{8}$$

where $Y_t^{\phi} = \sum_{x \in \mathscr{J}} \phi_x(t - \tau_x)$ is the branching process population of Y counted with the characteristic ϕ defined in (7) at t.

Remark 3. It is worth to point out that the embedding technique employed here does not require the presence of cliques of size 2 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. 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 dth 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 dth infected case of $\mathcal{C}$, the infectious period of v is independent of the actual paths of transmission within $\mathcal{C}$.

4.1 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 Nerman (1981) and also in the more recent paper by Iksanov and Meiners (2015).

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 3, except that the first ancestor of $\tilde{Z}$ is of type θ. Let further $\tilde{Y}$ be the corresponding embedded single-type branching process. In what remains, we will recycle the notation from section 3 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 $\mathscr{B}_x$ and the random characteristic ϕ are analogous to the definitions in (7) and (6), 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 $\xi^{(\alpha)} = \int_{\mathbb{R}_{\geq 0}} e^{-\alpha t}\xi(dt) = \sum_{x\in\mathscr{J}} e^{-\alpha\tau_x}$ where $\mathscr{J} = \cup_{n\geq 0}\mathscr{J}_n$ is the space of all individuals of $\tilde{Y}$ and α is the Malthusian parameter, i.e. $E(\xi^{(\alpha)}) = 1$. We define the measure ν on $\mathbb{R}_{\geq 0}$ by $\nu(t) = \nu[0,t] := E(\xi(t))$. Theorem 3 stated below is a special case of Nerman (1981, Theorem 5.4) and will lead us to the a.s. convergence of Theorem 1. In order to state Theorem 3 we need the following conditions.

Condition 1 (Finite mean age at childbearing)**.** The mean age at childbearing β defined by

$$\beta := E\left(\int_{\mathbb{R}_{\geq 0}} te^{-\alpha t}\xi(dt)\right)$$

is finite.

Condition 2 $(x\log x)$**.** The random variable

$$\xi^{(\alpha)}\log_+\left(\xi^{(\alpha)}\right)$$

has finite expectation.

Condition 3. There exists some non-negative real-valued non-increasing integrable function g such that

$$\int_{\mathbb{R}_{\geq 0}} \frac{1}{g(t)} e^{-\alpha t}\nu(dt) < \infty.$$

Condition 4. There exists some non-negative real-valued non-increasing integrable function g such that the expectation of

$$\sup_{t\geq 0} \frac{1}{g(t)\wedge 1} e^{-\alpha t}\phi(t)$$

is finite.

Theorem 3 (Theorem 5.4, Nerman (1981))**.** Under conditions 1 to 4 above it holds almost surely that

$$\frac{\left|\tilde{Y}_t^{\phi}\right|}{e^{\alpha t}} \to \widehat{W} m_{\infty}$$

as $t \to \infty$, where the random variable $\widehat{W}$ has mean $E(\widehat{W}) = 1$ and $P(\{\widehat{W} = 0\}) = P(|Z_t| \to 0)$, and $m_{\infty} \in (0, \infty)$ is a constant that depends on ϕ.

Remark 4. Under the conditions of Theorem 3, applying Theorem 3 to each of the children of the first ancestor of Z gives that

$$\frac{|Z_t|}{e^{\alpha t}} = \frac{\left|Y_t^{\phi}\right|}{e^{\alpha t}} \to W := (\widehat{W}^{(1)} e^{-\alpha \tau_1} + \ldots + \widehat{W}^{(J)} e^{-\alpha \tau_J}) m_{\infty} \tag{9}$$

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

5 Proofs

In Section 5.1, we prove Theorem 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 1. In Section 5.2, we prove Theorem 2. The main step in the proof is to establish upper bounds on the total variation distance of the degree distribution in (1) and $Po(\bar{A})$ and of the distribution in (2) and $Po(\bar{B})$.

5.1 Proof of Theorem 1

Recall that the random measure ξ (defined in section 4.1) is the point process of reproduction on $\mathbb{R}_{\geq 0}$ of a generic individual of $\tilde{Y}$ and that the measure ν on $\mathbb{R}_{\geq 0}$ is defined as $\nu(t) = \nu[0, t] := E(\xi(t))$. 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 $p_k^{\bar{B}} = P(MP(\bar{B}) = k)$ for $k \in \mathbb{Z}_{\geq 1}$ and that $L^{(\alpha)}$ is the vector displayed in (4).

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

Proof of Lemma 1. Below $*$ denotes convolution, i.e. for two cumulative distribution functions F and G

$$F * G(t) = \int_{-\infty}^{t} G(t - s) F(ds).$$

We have

$$\nu(t) = \gamma_1 F_2(t) + \sum_r \sum_{(m_1,\ldots,m_r)} \gamma_{m_1}\gamma_{m_2}\cdots\gamma_{m_r} F_{m_1+1} * \ldots * F_{m_r+1} * F_2(t)\gamma_1 \tag{10}$$

where the sums run over $\mathbb{Z}_{\geq 1}$ and $\mathbb{Z}^r_{\geq 1}$. Taking the Laplace transform of the right hand side in (10) and writing this in vector form gives that the Malthusian parameter α is the solution of

$$\int_{\mathbb{R}_{\geq 0}} e^{-\alpha t}\nu(dt) = \gamma_1 \mathcal{L}_2(\alpha) \sum_{n=0}^{\infty} \left(\Gamma_{\geq 2} \cdot L^{(\alpha)}_{\geq 3}\right)^n = 1 \tag{11}$$

where $\Gamma^T_{\geq 2} = (\gamma_2, \gamma_3, \ldots)$ and the elements of $L^{(\alpha)}_{\geq 3} = (\mathcal{L}^{(\alpha)}_3, \mathcal{L}^{(\alpha)}_4, \ldots)^T$ are defined in (3). Since $\Gamma_{\geq 2} \cdot L^{(\alpha)}_{\geq 3} < 1$ (shown below), (11) reduces to

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

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

It remains to show that $\Gamma_{\geq 2} \cdot L^{(\alpha)}_{\geq 3} < 1$. By the $x^2 \log x$ assumption,

$$\Gamma \cdot L^{(\lambda)} \leq \Gamma \cdot (1, 1, \ldots, 1)^T < \infty$$

for all $\lambda \geq 0$, 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

$$\Gamma \cdot L^{(\lambda)} \to P(T'_{ij} = 0)\mu_{\bar{A}}\mu_{\bar{B}} < 1$$

as $\lambda \to \infty$, the Malthusian parameter exists and is unique. □

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_1K_2 as

$$K_1K_2(r, A) := \int_E K_1(r, ds)K_2(s, A),\ A \in \mathcal{E}, r \in E.$$

For any $m \geq 1$, $K^m_1 := K_1K^{m-1}_1$ and $K^0_1 := I$ where I is the identity kernel $I(r, A) := \mathbb{1}(r \in A)$ for any $A \in \mathcal{E}$. If f is a ($\mathcal{E}$-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 $\mu(r, A \times B) = E(\xi_r(A \times B))$ 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 $K_{(\lambda)}(r, ds \times dt) := e^{-\lambda t}\mu(r, ds \times dt)$. Let further

$$\widehat{K}(r, ds) := \int_{\mathbb{R}_{\geq 0}} K_{(\alpha)}(r, ds \times dt), \tag{12}$$

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

$$I_{\{\theta\}^c}(r, B) = I(r, B \cap \mathcal{T}_\theta \setminus \{\theta\}). \tag{13}$$

Remark 5. 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

$$h(x) = G_\theta(x, \{\theta\}) \tag{14}$$

and the measure π by

$$\pi(A) = \widehat{K}G_\theta(\theta, A). \tag{15}$$

Then h is harmonic for $\widehat{K}$ (see Nummelin (2004, Proposition 4.6)), i.e. $\widehat{K}h = h$. 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 $\beta = \int te^{-\alpha t}\nu(dt) < \infty$ where ν is the measure in (10).

Lemma 2. $0 < \beta < \infty$.

Proof of Lemma 2. Let $\varepsilon > 0$ be small so that $\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varepsilon)} < 1$, and let the constant C_ε be such that $C_\varepsilon e^{-(\alpha-\varepsilon)t} \geq te^{-\alpha t}$ for all $t \geq 0$. Then

$$\beta = \int te^{-\alpha t}\nu(dt) \leq C_\varepsilon \int e^{-(\alpha-\varepsilon)t}\nu(dt)$$

$$= C_\varepsilon \gamma_1 \mathcal{L}_2(\alpha - \varepsilon) \sum_{n=0}^{\infty} \left(\Gamma_{\geq 2} \cdot L_{\geq 3}^{(\alpha-\varepsilon)}\right)^n < \infty.$$

□

5.1.1 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 $\mathscr{J}_n$ 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 Jagers (1989) or Biggins and Kyprianou (2004)).

Following Jagers (1989), we say that a set of individuals $L \subset \mathcal{S}$ is a *stopping line* if for any pair $y, x \in L$ it holds $x \not\prec y$. 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 $\mathcal{G}_\ell := \sigma(\cup_{x \not\succeq \ell} \mathcal{G}_x)$ 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 $\{L \preceq \ell\}$ belongs to $\mathcal{G}_\ell$. Here $L \preceq \ell$ means that for any $x \in L$ there is $y \in \ell$ such that $x \preceq y$.

Note that $\mathscr{J}_n$ 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

$$\widehat{W}_n := \sum_{x \in \mathscr{J}_n} e^{-\alpha \tau_x}. \tag{16}$$

and

$$\widetilde{W}_n = \frac{1}{h(\theta)} \sum_{|u|=n} e^{-\alpha \tau_u} h(\sigma_u) \tag{17}$$

where the sums run over all individuals of generation n of $\tilde{Y}$ and $\tilde{Z}$, respectively. It is well known (Biggins and Kyprianou 2004) 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 $\mathcal{G}_\varnothing$ is the trivial σ-algebra). Similarly, $\{\widehat{W}_n\}_{n \in \mathbb{Z}_{\geq 0}}$ is a martingale with respect to $\{\mathcal{G}_{\mathscr{J}_n}\}_n$. For later reference, we now state a special case of results presented in Biggins and Kyprianou (2004).

Proposition 1 (c.f. Biggins and Kyprianou (2004), Theorem 6.1 and Lemma 6.2)**.** Let $\{\widehat{W}_n\}_n$ and $\{\widetilde{W}_n\}_n$ be as in (16) and (17). Then, with probability 1, the limits $\lim_n \widetilde{W}_n$ and $\lim_n \widehat{W}_n$ exist and

$$\lim_n \widetilde{W}_n = \lim_n \widehat{W}_n.$$

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

$$E(\widehat{W}_1 \log_+ \widehat{W}_1) < \infty. \tag{18}$$

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

Lemma 3. Let $J \sim MP(X)$, where X is a non-negative integrable random variable with $P(X = 0) < 1$. Then $E(X^2 \log_+ X) < \infty$ is necessary and sufficient for $E(\bar{J} \log_+ \bar{J}) < \infty$ to hold.

Proof of Lemma 3. First note that

$$E(\bar{J} \log_+ \bar{J}) = E(J^2 \log_+ J)/E(J) = E(J^2 \log_+ J)/E(X).$$

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

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

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

Lemma 4. Under assumption 1 $\{\widehat{W}_n\}_n$, satisfies the $x \log x$ condition in (18).

Proof. It is known (Iksanov and Meiners 2015, Proposition 4.1) that the inequality in (18) holds if and only if $\{\widehat{W}_n\}_n$ is uniformly integrable, which is also equivalent to $E(\widehat{W}) = 1$, where $\widehat{W}$ is the almost sure limit $\lim_n \widehat{W}_n$. By Proposition 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

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

Combining this with Assumption 1, Lemma 3 and Iksanov and Meiners (2015, Corollary 2.1), the $x \log x$ condition holds for Y if we can show that almost surely

$$\sup_{x>2} \left(\frac{\sum_i \mathbb{1}(H(\varsigma_i) \geq x^{-1})}{\log_+(x)} \right) < \infty \tag{19}$$

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

$$R((r,0), A \times B) := \frac{1}{h(r)} \int_{A \times B} h(s) K(r, ds \times dt)$$

$$R((r,t), A \times B) := R((r,0), A \times (B-t)_{\geq 0})$$

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

Let $p_1 : \mathcal{T}_\theta \times \mathbb{R}_{\geq 0} \to \mathcal{T}_\theta$ and $p_2 : \mathcal{T}_\theta \times \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ 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 $\{\varsigma'_0, \varsigma'_1, \ldots\}$ is a Markov chain on $\mathcal{T}_\theta$ with transition measure R_1:

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

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

Now, it is easily verified that θ is a (positive) recurrent state of $\{\varsigma'_0, \varsigma'_1, \ldots\}$. Indeed, let M_i, $i \geq 1$ be the number of steps until $\{p_1(\varsigma_0), p_1(\varsigma_1), \ldots\}$ revisits θ for the ith 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 13)

$$P(M = m) = R_1(I_{\{\theta^c\}} R_1)^{m-1}(\theta, \{\theta\})$$

$$= \int_{\mathbb{R}_{\geq 0}^{m-1}} K(r_{m-1}, \{\theta\} \times \mathbb{R}_{\geq 0}) K(r_{m-2}, dr_{m-1} \times \mathbb{R}_{\geq 0}) \ldots K(r_1, dr_2 \times \mathbb{R}_{\geq 0}) K(\theta, dr_1 \times \mathbb{R}_{\geq 0})$$

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

Thus $P(M = \infty) = 0$ and M is geometrically distributed. Thus the inequality in (19) holds almost surely by the law of large numbers applied to $\{\varsigma''_{M_k} + \ldots \varsigma''_{M_{k+1}-1}\}_k$. □

We now turn our attention to conditions 3 and 4.

Lemma 5. Conditions 3 and 4 are satisfied under assumption 1.

Proof. If we take g to be the mapping $t \mapsto e^{-\varphi x}$ then conditions 3 and 4 are satisfied 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

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

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

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

for any $t \geq 0$. Thus

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

If the random characteristic ϕ is as in (7) then condition 4 is satisfied for the same choice of g.

□

Proof of Theorem 1. Assume that Assumption 1 i) - iii) hold. By Lemma 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 2 Condition 1 holds, by Lemma 4 Condition 2 holds and by Lemma 5 Conditions 3 and 4 hold. Hence the conditions of Theorem 3 are satisfied, and by Remark 4 the convergence in (9) holds.

□

5.2 Proof of Theorem 2

Below we describe a probabilistically equivalent construction of G_n^{aux}. 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, that is $p([0,x]) = P(A \leq x)$ 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 $A_1, \ldots, A_n$, let $A_{(n)}$ be a random variable which follows the empirical distribution of $A_1, \ldots, A_n$ and let p_n be the corresponding probability measure, i.e.

$$p_n([0,x]) = P(A_{(n)} \leq x | A_1, \ldots, A_n) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}(A_i \leq x) \tag{20}$$

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

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

Similarly, conditioned on the weights $B_1, \ldots, B_m$, let $B_{(n)}$ be a random variable which follows the empirical distribution of $B_1, \ldots, B_m$ 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 $\mathcal{E}_0 = \mathcal{E}_0' = \mathcal{N}_0 = \varnothing$. The component of G_n^{aux} that contains u is now constructed by iterating the following steps for $t = 1, 2, \ldots$.

1. If $t = 1$ let $v = u$ be the vertex that is currently being explored. Generate the downshifted *group degree* D of v from the distribution given in (1). Here D represents the (additional) number of cliques that v is a member of.

2. Draw D elements $B_{(1)}, \ldots, B_{(D)}$ from the multiset $\{B_1, \ldots, B_m\}$ independently with replacement. The probability to select $B_k \in \{B_1, \ldots, B_m\}$ 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 $\bar{B}_{(n)}$.

3. Let $v'_{(j)} \in V_n'$ be the vertex that corresponds to $B_{(j)}$, $j = 1, \ldots, D$. For each $v'_{(j)} \in \{v_{(1)}, \ldots, v_{(D)}\}$, if $v'_{(j)}$ is not a member of the set $\mathcal{E}'_{t-1} \subset V_n'$ of hitherto explored vertices carry out step 3 (a) to 3 (d) below. 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 the steps below.

   (a) Generate the downshifted clique size D'_j of $v'_{(j)}$ from the distribution given in (2) with $B_{(j)}$ in place of B_j.

(b) Select D'_j elements $A^j_{(1)}, \ldots, A^j_{(D'_j)}$ from the multiset $\{A_1, \ldots, A_n\}$ independently with replacement. The probability to select 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 $A_{(n)}$.

(c) Let $v^j_k \in V_n$ be the vertex that corresponds to $A^j_{(k)}$, $k = 1, \ldots, D'_j$.

(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.

4. Update the set $\mathcal{E}'_t$ of explored cliques by putting $\mathcal{E}'_t = \mathcal{E}'_{t-1} \cup \{v'_{(1)}, \ldots, v'_{(D_i)}\}$.

5. Update the set $\mathcal{N}_t$ of neighbours of explored vertices by putting $\mathcal{N}_t = \mathcal{N}_{t-1} \cup \{v^j_k : j = 1, \ldots, D_i \text{ and } k = 1 \ldots, D'_j\}$.

6. Update the set $\mathcal{E}_t$ of explored vertices by putting $\mathcal{E}_t = \mathcal{E}_{t-1} \cup \{v\}$.

7. 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 tth non-initial case).

If the nodes $v'_{(1)}, \ldots, v'_{(D)}$ are not distinct or $\{v'_{(1)}, \ldots, v'_{(D)}\} \cap \mathcal{E}_{t-1} \neq \varnothing$ in step 3 for some iteration $t \in \{0, 1, \ldots\}$ 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 $v^j_k \in E_{t-1}$ for some $j \in \{1, \ldots, D\}$ and $k \in \{1, \ldots, D'_j\}$ then the coupling breaks down.

The following claim, which is a variant of the birthday problem, ensures that with high probability 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.

Claim 2. Let $j_n = o(\sqrt{n})$ and let A have finite second moment. Suppose that we draw elements from the multiset $\{A_1, \ldots, A_n\}$ independently with replacement, and that the probability that $A_i \in \{A_1, \ldots, A_n\}$ is selected in a specific draw is proportional to A_i. It then holds that the first j_n drawn elements are distinct with P-probability tending to 1 as $n \to \infty$.

Proof of Claim 2. For $k = 1, 2, \ldots$, let $A^{(k)}$ be the kth element that is drawn from $\{A_1, \ldots, A_n\}$ and let $E_n(\cdot)$ be the conditional expectation operator given $A_1, \ldots, A_n$. Conditioned on $A_1, \ldots, A_n$, for $k \geq 2$ the probability that $A^{(k)}$ is not distinct from $A^{(j)}$ for some $j \in \{1, \ldots, k-1\}$ is smaller than or equal to

$$E_n\left(\frac{A^{(1)} + \ldots + A^{(k-1)}}{A_1 + \ldots + A_n}\right) = (k-1)\frac{A_1^2 + \ldots + A_n^2}{(A_1 + \ldots + A_n)^2}.$$

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

$$\sum_{k=1}^{j_n}(k-1)\frac{A_1^2+\ldots+A_n^2}{(A_1+\ldots+A_n)^2}$$

$$=\frac{(j_n-1)j_n\,(A_1^2+\ldots+A_n^2)}{2\,(A_1+\ldots+A_n)^2}$$

$$=\left(\frac{(j_n-1)j_n}{n}\right)\left(\frac{(A_1^2+\ldots+A_n^2)\,/n}{2(\mu_A^{(n)})^2}\right)$$

Since A has finite second moments

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

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

$$P(A^{(1)},\ldots,A^{(j_n)} \text{ are distinct})$$

$$\geq 1-E\left(1\wedge\left(\frac{(j_n-1)j_n}{n}\right)\frac{(A_1^2+\ldots+A_n^2)\,/n}{2(\mu_A^{(n)})^2}\right)$$

where the right side tends to 1 as $n\to\infty$. □

In the construction algorithm described above, 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 2 we find upper bounds on the coupling error between $MP(\bar{A}_{(n)})$ and $MP(\bar{A})$ and between $MP(\bar{B}_{(n)})$ and $MP(\bar{B})$ which we state in Proposition 2 and Lemma 6.

Given the weights $A_1,\ldots,A_n$, for a coupling $\mathscr{C}$ of A and $A_{(n)}$ we denote the corresponding probability measure and expectation by $P_{\mathscr{C}}$ and $E_{\mathscr{C}}$, respectively. A coupling of two random variables with distributions $MP(\bar{A})$ and $MP(\bar{A}_{(n)})$ can be constructed by first constructing a coupling $\mathscr{C}$ of their respective intensities $\bar{A}$ and $A_{(n)}$, then generating a joint realization $(\bar{A}', \bar{A}'_{(n)})$ of these intensities according to $\mathscr{C}$ 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 $\frac{1}{2}d_{TV}(Po(\bar{A}'), Po(\bar{A}'_{(n)}))$. Here d_{TV} denotes the total variation distance, i.e. for $a, b \in \mathbb{R}_{\geq 0}$

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

Thus, given the distribution of $A_{(n)}$ the probability of a miscoupling is given by
$E_{\mathscr{C}}(d_{TV}(Po(\bar{A}), Po(\bar{A}_{(n)})))/2$.

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

$$\mathscr{C}_n := \underset{\mathscr{C}}{\operatorname{argmin}} E_{\mathscr{C}} \left| \sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}} \right| \tag{21}$$

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

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

Proposition 2. Assume that $E(A^q) < \infty$ for $q > \frac{3}{2}$. Let $(\varepsilon_n)_{n\geq 1}$ be a sequence in $(0, \infty)$ such that if $q \neq 3$ then $\varepsilon_n \log(n) \to \infty$ as $n \to 0$ and if $q = 3$ then $\varepsilon_n \log(n) - \log(\log(n)) \to \infty$ as $n \to 0$. Then

$$P\left(E_{\mathscr{C}_n}(d_{TV}(Po(\bar{A}), Po(\bar{A}_{(n)}))) \geq n^{-\gamma+\varepsilon_n}\right) \to 0 \quad \text{as } n \to \infty$$

where $\gamma = \frac{1}{2} \wedge \frac{q-3/2}{q}$.

Remark 6. Proposition 2 holds also if A, $\bar{A}$ and $\bar{A}_{(n)}$ are replaced by B, $\bar{B}$ and $\bar{B}_{(n)}$.

Proof of Proposition 2. We have that (the first inequality follows from Barbour et al. (1992, Theorem 1.C))

$$E_{\mathscr{C}_n}\left(d_{TV}(Po(\bar{A}), Po(\bar{A}_{(n)}))\right)$$

$$\leq E_{\mathscr{C}_n}\left(\frac{1}{\sqrt{\bar{A}_{(n)} \vee \bar{A}}} \left|\bar{A} - \bar{A}_{(n)}\right|\right)$$

$$\leq E_{\mathscr{C}_n}\left(\frac{1}{\frac{1}{2}\left(\sqrt{\bar{A}} + \sqrt{\bar{A}_{(n)}}\right)} \left|\sqrt{\bar{A}} + \sqrt{\bar{A}_{(n)}}\right| \left|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}\right|\right)$$

$$= 2E_{\mathscr{C}_n}\left(\left|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}\right|\right).$$

Hence, by Proposition 3 below,

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

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 $P(D_n^c) = p(0)^n$, $p(0) < 1$. □

By the degree distributions of G_n^{aux} given in (1) and (2), in order to arrive at Theorem 2 we will also need a bound on the coupling error of $MP\left(\bar{B}_{(n)}\mu_A^{(n)}/\mu_A\right)$ and $MP\left(\bar{B}_{(n)}\right)$ and of $MP\left(\bar{A}_{(n)}\mu_B^{(n)}\lfloor n\mu_A/\mu_B\rfloor/n\mu_A\right)$ and $MP\left(\bar{A}_{(n)}\right)$.

Proposition 3 (c.f. Fournier and Guillin (2015), Theorem 1)**.** If $E(A^q) < \infty$ for $q > 3/2$ then

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

where $\mathscr{C}_n$ is the coupling in (21) and D_n is the event that $\max_{i\in[n]}(A_i) > 0$.

Proof of Proposition 3. This proof is, in part, analogous to the proof of Theorem 1 in Fournier and Guillin (2015) and is presented here in full for completeness. The differences between the present proof and the proof of Theorem 1 in Fournier and Guillin (2015) arise mainly due to the size-biasing of the weights.

Throughout, $C_1, C_2, \ldots$ 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

$$E_{\mathscr{C}_n}\left|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}\right|$$

is the 1-Wasserstein distance between the distributions of $\sqrt{\bar{A}}$ and $\sqrt{\bar{A}_{(n)}}$. Hence, with the notation $2^m F = \{2^m u : u \in F\}$ for any event F, we have by Lemma 5 and 6 in Fournier and Guillin (2015)

$$E_{\mathscr{C}_n}\left|\sqrt{\bar{A}}-\sqrt{\bar{A}_{(n)}}\right|$$

$$\leq C_1\sum_{m\geq 0}2^m\sum_{\ell\geq 0}2^{-\ell}\sum_{F\in\mathcal{P}_\ell}|\nu_n(2^mF\cap U_m)-\nu(2^mF\cap U_m)| \tag{22}$$

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

$$|\nu_n(U)-\nu(U)|$$

$$=\left|\bar{p}_n(U^2)-\bar{p}(U^2)\right|$$

$$=\left|\frac{\sum_{i=1}^n A_i\mathbb{1}(A_i\in U^2)}{n\mu_A^{(n)}}-\frac{E(A\mathbb{1}(A\in U^2))}{\mu_A}\right|$$

$$\leq\frac{1}{n\mu_A^{(n)}}\sum_{i=1}^n A_i\mathbb{1}(A_i\in U^2)\left|1-\frac{\mu_A^{(n)}}{\mu_A}\right|+\frac{1}{\mu_A}\left|\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(A_i\in U^2)-E(A\mathbb{1}(A\in U^2))\right|. \tag{23}$$

The second term in the right hand side in (23) satisfies (with the first inequality following from Jensen's inequality)

$$E\left(\left|\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(A_i\in U^2)-E(A\mathbb{1}(A\in U^2))\right|\right)$$

$$\leq\sqrt{Var\left(\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(A_i\in U^2)\right)} \tag{24}$$

$$\leq(\sup U)^2\sqrt{\frac{1}{n}p(U^2)}$$

and, again by the triangle inequality,

$$E\left(\left|\frac{1}{n}\sum_{i=1}^n A_i\mathbb{1}(A_i\in U^2)-E(A\mathbb{1}(A\in U^2))\right|\right)\leq 2(\sup U)^2p(U^2). \tag{25}$$

Combining (24) and (25) gives that the second term in the right hand side in (23) satisfies

$$\begin{aligned}&\frac{1}{\mu_A}\left|\frac{1}{n}\sum_{i=1}^{n}A_i\mathbb{1}(A_i\in U^2)-E(A\mathbb{1}(A\in U^2))\right|\\&\leq C_2(\sup U)^2\left(\sqrt{\frac{1}{n}p(U^2)}\wedge p(U^2)\right).\end{aligned}\tag{26}$$

In order to find a similar upper bound on the first term in the right hand side in (23) we define the event S_n as

$$S_n:=\{\mu_A^{(n)}\leq\mu_A-\kappa\}\tag{27}$$

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

$$\begin{aligned}&E\left(\frac{1}{n\mu_A^{(n)}}\sum_{i=1}^{n}A_i\mathbb{1}(A_i\in U^2)\left|1-\frac{\mu_A^{(n)}}{\mu_A}\right|;S_n^c\cap D_n\right)\\&\leq\frac{(\sup U)^2}{(\mu_A-\kappa)}\left(\frac{1}{n}E\left(\sum_{i=1}^{n}\mathbb{1}(A_i\in U^2)\left|1-\frac{\mu_A^{(n)}}{\mu_A}\right|\right)\wedge(2p(U^2))\right)\\&\leq\frac{(\sup U)^2}{(\mu_A-\kappa)}\left(\sqrt{\frac{1}{n^3}E\left(\left(\sum_{i=1}^{n}\mathbb{1}(A_i\in U^2)\right)^2\right)Var\left(\frac{A}{\mu_A}\right)}\wedge(2p(U^2))\right)\\&\leq\frac{(\sup U)^2}{(\mu_A-\kappa)}\left(\sqrt{\frac{1}{n^3}np(U^2)(np(U^2)+1)Var\left(\frac{A}{\mu_A}\right)}\wedge(2p(U^2))\right)\\&\leq C_3(\sup U)^2\left(\sqrt{\frac{1}{n}p(U^2)}\wedge p(U^2))\right).\end{aligned}\tag{28}$$

By the inequalities in (26) and (28) together with the fact that for any $G\subset[0,\infty)$ it holds that $\sup(G\cap U_m^2)\leq\sup(U_m^2)=2^{2m}$, we have

$$\sum_{F\in\mathcal{P}_\ell} E\left(|\nu_n(2^m F\cap U_m)-\nu(2^m F\cap U_m)|\,;D_n\cap S_n^c\right)$$

$$\leq C_4\left(2^{2m}\sum_{F\in\mathcal{P}_\ell}\left(\sqrt{\frac{1}{n}p(2^{2m}F^2\cap U_m^2)}\wedge p(2^{2m}F^2\cap U_m^2)\right)\right) \tag{29}$$

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

$$\sum_{F\in\mathcal{P}_\ell}\left(\sqrt{\frac{1}{n}p(2^{2m}F^2\cap U_m^2)}\wedge p(2^{2m}F^2\cap U_m^2)\right)$$

$$\leq\left(\frac{1}{\sqrt{n}}\sum_{F\in\mathcal{P}_\ell}\sqrt{p(2^{2m}F^2\cap U_m^2)}\right)\wedge p(U_m^2)$$

$$\leq C_5\left(\left(\frac{1}{\sqrt{n}}\sqrt{|P_\ell|\,p(U_m^2)}\right)\wedge 2^{-2qm}\right) \tag{30}$$

$$\leq C_5\left(\left(\frac{1}{\sqrt{n}}\sqrt{2^\ell E(A^q)2^{-2qm}}\right)\wedge 2^{-2qm}\right)$$

$$\leq C_6\left(\frac{1}{\sqrt{n}}2^{\ell/2-qm}\wedge 2^{-2qm}\right).$$

Hence, by (22), (29) and (30),

$$E\left(E_{\mathscr{C}_n}\left(\left|\sqrt{\bar{A}}-\sqrt{\bar{A}_{(n)}}\right|\right);S_n^c\cap D_n\right)$$

$$\leq C_7\sum_{m\geq 0}2^m\sum_{\ell\geq 0}2^{-\ell+2m}\left(\frac{1}{\sqrt{n}}2^{\ell/2-qm}\wedge 2^{-2qm}\right)$$

$$\leq C_7\sum_{m\geq 0}2^{3m}\sum_{\ell\geq 0}2^{-\ell/2}\left(\frac{1}{\sqrt{n}}2^{-qm}\wedge 2^{-2qm}\right)$$

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

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

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

$$\leq C_8 \sum_{m=0}^{m_n} \frac{1}{\sqrt{n}}2^{-m(q-3)} + C_8 \sum_{m>m_n}^{\infty} 2^{-m(2q-3)}$$

$$= O\left(n^{-\left(1-\frac{3}{2q}\right)}\right).$$

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

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

$$\leq C_8 \frac{a_n}{\sqrt{n}} + C_8 \sum_{m\geq a_n}^{\infty} 2^{-m(2q-3)}$$

$$= O\left(n^{-1/2}\log(n)\right).$$

Thus it only remains to bound the expectation of $E_{\mathscr{C}_n}\left(\left|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}\right|\right)$ on $S_n \cap D_n$, where S_n is the event in (27). 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 $i = 1, \ldots, n$,

$$E\left(E_{\mathscr{C}_n}\left(\left|\sqrt{\bar{A}} - \sqrt{\bar{A}_{(n)}}\right|\right); S_n \cap D_n\right)$$

$$\leq P(S_n \cap D_n)E(\sqrt{\bar{A}}) + E\left(\sqrt{\bar{A}_{(n)}}; S_n \cap D_n\right)$$

$$= P(S_n \cap D_n)E(\sqrt{\bar{A}}) + E\left(\frac{\sum_{i=1}^n A_i^{\frac{3}{2}}}{\sum_{i=1}^n A_i}; S_n \cap D_n\right)$$

$$\leq P(S_n \cap D_n)E(\sqrt{\bar{A}}) + E\left(\sum_{i=1}^n \sqrt{A_i}; S_n \cap D_n\right)$$

$$\leq P(S_n \cap D_n)E(\sqrt{\bar{A}}) + E\left(n^{3/2}\sqrt{\mu_A - \kappa}; S_n \cap D_n\right)$$

$$= \mathcal{O}(n^{3/2} P(S_n))$$

$$= \mathcal{O}(n^{3/2} e^{n(\mu_a - \kappa)} E(e^{-A})^n)$$

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

□

Lemma 6. Let ε_n be as in Proposition 2 and $E(A^2), E(B^2) < \infty$. Then, w.h.p.,

$$d_{TV}\left(Po\left(\bar{B}_{(n)}\mu_A^{(n)}/\mu_A\right), Po\left(\bar{B}_{(n)}\right)\right) \leq n^{-\frac{1}{2}+\varepsilon_n} \tag{31}$$

and

$$d_{TV}\left(Po\left(\bar{A}\mu_B^{(n)} \lfloor n\mu_A/\mu_B \rfloor / n\mu_A\right), Po\left(\bar{A}_{(n)}\right)\right) \leq n^{-\frac{1}{2}+\varepsilon_n}. \tag{32}$$

Proof of Lemma 6. Define

$$H_n := \left\{\left|1 - \mu_A^{(n)}/\mu_A\right| \leq n^{-\frac{1}{2}+\frac{1}{2}\varepsilon_n}\right\}.$$

By Chebyshev's inequality $P\left(H_n^c\right) = \mathcal{O}\left(n^{-\varepsilon_n}\right)$. Now (again by Barbour et al. (1992, Theorem 1.C))

$$E\left(d_{TV}\left(Po\left(\bar{B}_{(n)}\mu_A^{(n)}/\mu_A\right), Po\left(\bar{B}_{(n)}\right)\right); H_n\right)$$

$$\leq E\left(\left|\bar{B}_{(n)}\mu_A^{(n)}/\mu_A - \bar{B}_{(n)}\right|; H_n\right)$$

$$\leq E\left(\bar{B}_{(n)} n^{-\frac{1}{2}+\frac{1}{2}\varepsilon_n}\right) = \mathcal{O}(n^{-\frac{1}{2}+\frac{1}{2}\varepsilon_n}).$$

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

$$P\left(d_{TV}\left(Po\left(\bar{B}_{(n)}\mu_A^{(n)}/\mu_A\right), Po\left(\bar{B}_{(n)}\right)\right) \geq n^{-\frac{1}{2}+\varepsilon_n}\right)$$

$$\leq P(H_n^c) + \mathcal{O}(n^{-\frac{1}{2}\varepsilon_n}).$$

This proves the inequality in (31), and the proof of the inequality in (32) is completely analogous.

□

Proof of Theorem 2. Let ε_n, q and γ be as in Theorem 2, and let $T \in \mathbb{N}_0$ be the number of iterations in the construction algorithm on page 22 (i.e. $\mathcal{N}_T = \mathcal{E}_T$). By Claim 2, 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 Claim 2 with Proposition 2 gives the assertion of Theorem 2 by the union bound. $\square$

References

Ball, Frank, Denis Mollison, and Gianpaolo Scalia-Tomba (1997). "Epidemics with two levels of mixing". In: *The Annals of Applied Probability*, pp. 46–89.

Ball, Frank and Peter Neal (2002). "A general model for stochastic SIR epidemics with two levels of mixing". In: *Mathematical Biosciences* 180.1–2, pp. 73 –102.

– (2003). "The great circle epidemic model". In: *Stochastic Processes and their Applications* 107.2, pp. 233 –268.

Ball, Frank, Lorenzo Pellis, and Pieter Trapman (2016). "Reproduction numbers for epidemic models with households and other social structures II: Comparisons and implications for vaccination". In: *Mathematical Biosciences* 274, pp. 108 –139.

Ball, Frank and Laurence Shaw (2015). "Estimating the within-household infection rate in emerging SIR epidemics among a community of households". In: *Journal of Mathematical Biology* 71.6, pp. 1705–1735.

Ball, Frank and David Sirl (Mar. 2012). "An SIR epidemic model on a population with random network and household structure, and several types of individuals". In: *Adv. in Appl. Probab.* 44.1, pp. 63–86.

Ball, Frank, David Sirl, and Pieter Trapman (2009). "Threshold behaviour and final outcome of an epidemicon a random network with household structure". In: *Advances in Applied Probability* 41.3, pp. 765–796.

– (2010). "Analysis of a stochastic SIR epidemic on a random network incorporating household structure". In: *Mathematical Biosciences* 224.2, pp. 53 –73.

– (2014). "Epidemics on random intersection graphs". In: *The Annals of Applied Probability* 24.3, pp. 1081–1128.

Barbour, Andrew, Lars Holst, and Svante Janson (1992). *Poisson Approximation*. Oxford science publications. Clarendon Press.

Becker, Niels G. and Klaus Dietz (1995). "The effect of household distribution on transmission and control of highly infectious diseases". In: *Mathematical Biosciences* 127.2, pp. 207 –219.

Biggins, John and Andreas Kyprianou (June 2004). "Measure change in multitype branching". In: *Adv. in Appl. Probab.* 36.2, pp. 544–581.

Britton, Tom, Maria Deijfen, Andreas N. Lagerås, and Mathias Lindholm (Sept. 2008). "Epidemics on random graphs with tunable clustering". In: *J. Appl. Probab.* 45.3, pp. 743–756.

Broadbent, Simon and John Hammersley (1957). "Percolation processes: I. Crystals and mazes". In: *Mathematical Proceedings of the Cambridge Philosophical Society* 53.3, 629–641.

Dye, Christopher, Bruce Aylward, Philippe Barboza, Luke Bawo, Eric Bertherat, Pepe Bilivogui, Isobel Blake, Rick Brennan, Sylvie Briand, Jethro Chakauya, Kennedy Chitala, Roland M. Conteh, Anne Cori, Alice Croisier, Dangou Jean-Marie, Boubacar Diallo, Christl A. Donnelly, Tim Eckmanns, Neil M. Ferguson, and Zabulon Yoti (Jan. 2015). "Ebola virus disease in West

Africa–the first 9 months". In: *The New England journal of medicine* 372, p. 189.

Fournier, Nicolas and Arnaud Guillin (2015). "On the rate of convergence in Wasserstein distance of the empirical measure". In: *Probability Theory and Related Fields* 162.3, pp. 707–738.

Frisch, Harry and John Hammersley (1963). "Percolation Processes and Related Topics". In: *Journal of the Society for Industrial and Applied Mathematics* 11.4, pp. 894–918. eprint: `https://doi.org/10.1137/0111066`.

Heesterbeek, Hans, Tom Britton, and Odo Diekmann (2013). *Mathematical Tools for Understanding Infectious Disease Dynamics*. Princeton University Press.

Iksanov, Alexander and Matthias Meiners (2015). "Rate of convergence in the law of large numbers for supercritical general multi-type branching processes". In: *Stochastic Processes and their Applications* 125.2, pp. 708 –738.

Jagers, Peter (1989). "General branching processes as Markov fields". In: *Stochastic Processes and their Applications* 32.2, pp. 183 –212.

Karoński, Michał, Edward Scheinerman, and Karen Singer-cohen (1999). "On Random Intersection Graphs: The Subgraph Problem". In: *Combinatorics, Probability and Computing* 8.1-2, 131–159.

Ludwig, Donald (1975). "Final size distribution for epidemics". In: *Mathematical Biosciences* 23.1, pp. 33 –46.

Neal, Peter (2008). "The SIS Great Circle Epidemic Model". In: *Journal of Applied Probability* 45.2, pp. 513–530.

Nerman, Olle (1981). "On the convergence of supercritical general (C-M-J) branching processes". In: *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 57.3, pp. 365–395.

Nishiura, Hiroshi and Gerado Chowell (2014). "Early transmission dynamics of Ebola virus disease (EVD), West Africa, March to August 2014". In: *Eurosurveillance* 19.36, 20894.

Nummelin, Esa (2004). *General Irreducible Markov Chains and Non-Negative Operators*. Cambridge Tracts in Mathematics. Cambridge University Press.

Olofsson, Peter (Feb. 1996). "Branching processes with local dependencies". In: *Ann. Appl. Probab.* 6.1, pp. 238–268.

Pellis, Lorenzo, Frank Ball, and Pieter Trapman (2012). "Reproduction numbers for epidemic models with households and other social structures. I. Definition and calculation of R_0". In: *Mathematical Biosciences* 235.1, pp. 85–97.

Pellis, Lorenzo, Neil M Ferguson, and Christophe Fraser (2011). "Epidemic growth rate and household reproduction number in communities of households, schools and workplaces". In: *Journal of mathematical biology* 63.4, pp. 691–734.

Sagitov, Serik (2017). *Regenerative multi-type Galton-Watson processes*. eprint: `arXiv:1708.02636`.

Watson, Raymond Keith (1972). "On an epidemic in a stratified population". In: *Journal of Applied Probability* 9.3, pp. 659–666.
Watts, Duncan and Steven Strogatz (July 1998). "Collective Dynamics of Small World Networks". In: *Nature* 393, pp. 440–2.