

SURVIVAL AND EXTINCTION OF EPIDEMICS ON RANDOM GRAPHS WITH GENERAL DEGREE

BY SHANKAR BHAMIDI¹, DANNY NAM^{2,*}, OANH NGUYEN^{2,†} AND ALLAN SLY^{2,‡}

¹Department of Statistics and Operations Research, University of North Carolina, bhamidi@email.unc.edu

²Department of Mathematics, Princeton University, *dhn@princeton.edu; †onguyen@princeton.edu; ‡asly@princeton.edu

In this paper we establish the necessary and sufficient criterion for the contact process on Galton–Watson trees (resp., random graphs) to exhibit the phase of extinction (resp., short survival). We prove that the survival threshold λ_1 for a Galton–Watson tree is strictly positive if and only if its offspring distribution ξ has an exponential tail, that is, $\mathbb{E}e^{c\xi} < \infty$ for some $c > 0$, settling a conjecture by Huang and Durrett (2018). On the random graph with degree distribution μ , we show that if μ has an exponential tail, then for small enough λ the contact process with the all-infected initial condition survives for $n^{1+o(1)}$ -time whp (short survival), while for large enough λ it runs over $e^{\Theta(n)}$ -time whp (long survival). When μ is subexponential, we prove that the contact process whp displays long survival for any fixed $\lambda > 0$.

1. Introduction. The contact process is a model of epidemics on networks introduced by Harris in 1974 [11]. Its transitions are given as follows:

- Each vertex is either *infected* or *healthy*.
- Each infected vertex infects each of its neighbors independently at rate λ , and it is healed at rate 1 independently of all the infections.
- Infection and recovery events in the process happen independently from vertex to vertex.

The phase diagrams of the contact processes on \mathbb{Z}^d and on \mathbb{T}_d , the infinite d -ary tree, are well understood. In particular, the contact process on an infinite tree has drawn particular interest, as it has two distinct phase transitions. In a series of beautiful works [16, 26, 29], it was shown that the contact process on \mathbb{T}_d for $d \geq 2$, with an initial infection at the root, has two different thresholds $0 < \lambda_1 < \lambda_2$ such that:

- (*Extinction*) For $\lambda < \lambda_1$, the infection becomes extinct almost surely.
- (*Weak survival*) For $\lambda \in (\lambda_1, \lambda_2)$, the infection survives with positive probability, but the root is infected finitely many times almost surely.
- (*Strong survival*) For $\lambda > \lambda_2$, the infection survives, and the root gets infected infinitely many times with positive probability.

A natural interest is then to study the phase diagram of the contact process on Galton–Watson trees. In this paper we establish the necessary and sufficient criterion for $\lambda_1 > 0$. In particular, we provide the first known result for extinction in Galton–Watson trees with unbounded offspring distribution.

THEOREM 1. *Consider the contact process on the Galton–Watson tree with offspring distribution ξ , and suppose that only the root of the tree is initially infected. If ξ has an exponential tail, that is, $\mathbb{E}e^{c\xi} < \infty$ for some $c > 0$, then there exists $\lambda_0 = \lambda_0(\xi) > 0$ such that for all $\lambda < \lambda_0$, the process dies out almost surely.*

Received April 2019; revised January 2020.

MSC2020 subject classifications. Primary 60K35; secondary 05C80.

Key words and phrases. Contact process, epidemics, random graph, Galton–Watson tree, phase transition.

Recently, Huang and Durrett [12] proved that, on Galton–Watson trees, $\lambda_2 = 0$ if the offspring distribution ξ is subexponential, that is, $\mathbb{E}e^{c\xi} = \infty$ for all $c > 0$. Combining Theorem 1 with their result, we have the complete characterization on the existence of extinction on Galton–Watson trees. Moreover, Theorem 1 establishes a stronger version of the following conjecture by Huang and Durrett.

CONJECTURE 2 ([12]). *Suppose that $\mathbb{P}(\xi \geq k) = (1 - p)^k$ for all k larger than some constant K , and consider the contact process on the Galton–Watson tree with offspring distribution ξ . Then, λ_2 , the weak-strong survival threshold, is strictly positive.*

The challenge in understanding the infection time on trees with unbounded degree distributions is that the infection persists for a long time around high-degree vertices, as there are many neighbors from which it can be reinfected. Indeed, it was shown in [2] that the infection will last time $e^{c_\lambda d}$ in a neighborhood of a vertex of degree d with positive probability for some $c_\lambda > 0$. Thus, exponential tails on the degree distribution are needed for there to be few enough high degree vertices in the tree for extinction to be certain.

The next object of interest is the contact process on random graphs. For the contact process on the Erdős–Rényi random graph $\mathcal{G}_{n,d/n}$, no rigorous results were known regarding its phase diagram—whether it shows short or long survival, or both. In this work we prove that, on $\mathcal{G}_{n,d/n}$, the contact process exhibits two different phases depending on λ as a consequence of an analogous criterion on more general random graphs.

We focus on studying the contact process on the random graph with degree distribution μ , which we denote by $G \sim \mathcal{G}(n, \mu)$ (definitions given in Section 2.2). For the contact process on $G \sim \mathcal{G}(n, \mu)$, our main goal is to study how long the process survives in terms of the size of the graph. The second result of this paper establishes the necessary and sufficient criterion for the contact process on $G \sim \mathcal{G}(n, \mu)$ to display the short survival phase. We assume throughout that μ satisfies

$$(1.1) \quad \mathbb{E}_{D \sim \mu} D(D - 2) > 0 \quad \text{and} \quad \mathbb{E}_{D \sim \mu} D^2 < \infty$$

in order to ensure the existence of the giant component and take advantage of the *configuration model*. For details on (1.1), see Section 2.2.

THEOREM 3. *Suppose that μ satisfies (1.1) and there exists some constant $c > 0$ such that $\mathbb{E}_{D \sim \mu} e^{cD} < \infty$. Consider the contact process on $G \sim \mathcal{G}(n, \mu)$ where all vertices are initially infected. Then, there exist constants $0 < \underline{\lambda}(\mu) \leq \bar{\lambda}(\mu) < \infty$ such that the following hold:*

- (1) *For all $\lambda < \underline{\lambda}$, the survival time of the process is at most $n^{1+o(1)}$ -time whp.*
- (2) *For all $\lambda > \bar{\lambda}$, the survival time of the process is $e^{\Theta(n)}$ -time whp.*

THEOREM 4. *Suppose that μ satisfies (1.1) and $\mathbb{E}_{D \sim \mu} e^{cD} = \infty$ for all $c > 0$. Consider the contact process on $G \sim \mathcal{G}(n, \mu)$ where all vertices are initially infected. Then for any fixed $\lambda > 0$, the survival time of the process is $e^{\Theta(n)}$ -time whp.*

REMARK 1.1. In the statements of Theorems 3 and 4 (and Corollary 5 below as well), the notion whp covers the randomness coming from both the choice of graph G and the contact process. Therefore, they should be understood as,

“There exists an event \mathcal{A} over the choice of G which occurs whp, such that the statement holds whp over the contact process on G given $G \in \mathcal{A}$.”

For the case of Erdős–Rényi random graphs, which are contiguous to $\mathcal{G}(n, \mu)$ with $\mu = \text{Pois}(d)$ (see Section 2.2 for details), we can show that the contact process on $\mathcal{G}_{n,d/n}$ exhibits two different phases as a consequence of Theorem 3.

COROLLARY 5. *For any fixed $d > 1$, consider the contact process on $G \sim \mathcal{G}_{n,d/n}$ where all vertices are initially infected. Then, there exist constants $0 < \underline{\lambda}(d) \leq \bar{\lambda}(d) < \infty$ such that the following hold:*

- (1) *For all $\lambda < \underline{\lambda}$, the survival time of the process is at most $n^{1+o(1)}$ -time whp.*
- (2) *For all $\lambda > \bar{\lambda}$, the survival time of the process is $e^{\Theta(n)}$ -time whp.*

REMARK 1.2. One might be interested in studying the contact process on $G \sim \mathcal{G}(n, \mu)$, with the initial condition such that only a single vertex is infected. When a uniformly random vertex in G is infected, initially, while all the other ones are healthy, we will later see that Theorems 3 and 4 continue to hold, if we change “whp” to “with positive probability” at the end of the statements of Theorems 3(2) and 4. To be precise, by “with positive probability,” we mean whp over the choice of $G \sim \mathcal{G}(n, \mu)$, with positive probability over the choice of the initially infected vertex v and with positive probability over the contact process. Proofs are given in Remark 5.5 for exponential distributions and Remark 6.4 for subexponential distributions.

To sum up, we establish a “universality” criterion for the contact process on Galton–Watson trees (resp., random graphs with given degree distributions) on the existence of the phase of extinction (resp., short survival). Our methods do not give sharp estimates on the critical value, and it is an interesting open problem to determine the location of the phase transition. We also believe that the two critical values in Theorems 1 and 3 coincide. Precisely, we conjecture that $\lambda_1(\text{GW}(\mu')) = \lambda_c(\mathcal{G}(n, \mu))$, where:

- $\lambda_1(\text{GW}(\mu'))$ is the death–survival threshold of the Galton–Watson tree with offspring distribution μ' , the size-biased distribution of μ (see Section 2.2 for details).
- $\lambda_c(\mathcal{G}(n, \mu))$ is the short-long survival threshold of $\mathcal{G}(n, \mu)$.

1.1. Related works. In [11], Harris first introduced the contact process on \mathbb{Z}^d and showed that the death–survival threshold $\lambda_c(\mathbb{Z}^d)$ satisfies $0 < \lambda_c(\mathbb{Z}^d) < \infty$ for any d . Building upon this work, the model on \mathbb{Z}^d has been studied intensively, and we refer to Liggett [17] for a survey of results. Pemantle [26] studied the contact process on the infinite d -ary tree \mathbb{T}_d and showed that it exhibits three different phases—extinction, weak survival and strong survival—for $d \geq 3$. This result was later generalized by Liggett [16] for the case $d = 2$. Stacey [29] gave a shorter proof that applies for any $d \geq 2$.

Less is known for the contact process on general Galton–Watson trees. Recently, Huang and Durrett [12] proved that, on Galton–Watson trees, $\lambda_2 = 0$ if the offspring distribution ξ is subexponential. Along with Theorem 1, we now have the complete characterization of the existence of extinction in the contact process on Galton–Watson trees.

There has been considerable work studying the phase transitions of survival times on large finite graphs. Stacey [28] and Cranston et al. [6] studied the contact process on the d -ary tree \mathbb{T}_d^h of depth h starting from the all-infected state, and their results show that the survival time T_h , as $h \rightarrow \infty$, satisfies (i) $T_h/h \rightarrow \gamma_1$ in probability if $\lambda < \lambda_2(\mathbb{T}_d)$; (ii) $|\mathbb{T}_d^h|^{-1} \log \mathbb{E}T_h \rightarrow \gamma_2$ in probability and $T_h/\mathbb{E}T_h \xrightarrow{d} \text{Exp}(1)$ if $\lambda > \lambda_2(\mathbb{T}_d)$, where γ_1, γ_2 are constants depending on d, λ . In [8, 9, 24], similar results were established for the case of the lattice cube $\{1, \dots, n\}^d$.

Recently, in work of Mourrat and Valesin [25] and Lalley and Su [15], it was shown that, for any $d \geq 3$, the contact process on the random d -regular graph, whose initial configuration is the all-infected state, exhibits the following phase transition:

- (*Short survival*) For $\lambda < \lambda_1(\mathbb{T}_d)$, it survives for $O(\log n)$ -time whp.
- (*Long survival*) For $\lambda > \lambda_1(\mathbb{T}_d)$, it survives for $e^{\Theta(n)}$ -time whp.

Moreover, in [15] a “cutoff phenomenon” of the fraction of infected vertices was established. In [25], the same result as above is proven for $G \sim \mathcal{G}(n, \mu)$ with bounded μ (Theorems 1.3 and 1.4). For an unbounded degree distribution μ , Chatterjee and Durrett [5] proved that if μ obeys a power law, then the contact process always displays long survival for any $\lambda > 0$, though their survival time was slightly weaker than exponential ($e^{n^{1-\delta}}$ for any $\delta > 0$). This result was later generalized in [22] to an exponential survival. Our Theorems 3 and 4 extend the aforementioned results to any general μ . In [27], a long survival on general graphs for $\lambda > \lambda_c(\mathbb{Z})$ was settled, with survival time at least $\exp(|G|/\{\log |G|\}^\kappa)$ for any $\kappa > 1$. [4] studied the contact process under similar settings as Theorem 3(2) and Corollary 5(2) with additional assumptions on the degree distribution and showed that the expected survival time is exponentially large in n .

On random graphs with power-law degree distributions, metastability properties on the size of infected vertices were studied in [3, 23]. For other types of random graphs, recently in [20] it was shown that the contact process on random geometric graphs exhibits both short and long survival.

1.2. *Main techniques.* We sketch the ideas in the paper before giving the full proofs. The analysis of the subcritical contact process (i.e., extinction and short survival) relies on three main ideas which we now describe. Here, we assume that the offspring distribution of the Galton–Watson trees and the degree distribution of the random graphs have exponential tails.

► *Modified process: Preventing recoveries at the root.* One main difficulty in studying the contact process on Galton–Watson trees comes from complicated dependencies inside the given tree. To overcome this obstacle, we consider the following modification of the process:

- A vertex is added above the root that is always infected. As such the chain no longer has an absorbing state.
- Recoveries at the root only occur when none of its descendants are infected at the time of recovery. All the other infections and recoveries are the same as the original process.

In the modified process, when the root is infected, the processes inside each subtree from a child of the root behave independently. By relating the stationary probability of the root being uninfected to the extinction time we develop a recursive relationship over the tree height. As a result, we show that the expected survival time of the contact process with small enough λ is bounded by a constant, for any finite-depth Galton–Watson trees.

► *Exponential decay of infection depth: The delayed process.* To relate the finite Galton–Watson trees to the infinite tree, we prove that the probability that the infection goes deeper than depth h decays exponentially in h . To this end, we introduce the *delayed process*, which spends exponentially longer time at states containing deeper infections. Based on a similar argument introduced above, we show that the expected survival time of the delayed process on the Galton–Watson tree is bounded by a constant if λ is small enough. This will imply that, in the original process, the infection can go deeper than h at most with an exponentially small probability in h . Thus, the contact process on the infinite Galton–Watson tree can be regarded as that on a large-depth finite tree, and hence we establish Theorem 1.

► *Coupling the local neighborhoods of $\mathcal{G}(n, \mu)$.* To study the contact process on $G \sim \mathcal{G}(n, \mu)$ exhibiting short survival, we attempt to dominate the local neighborhoods of the graph by Galton–Watson trees, in terms of isomorphic inclusions of graphs. However, some of the local neighborhoods $N(v, r)$ will contain a cycle, and hence we introduce modified

Galton–Watson type processes that contain a cycle and behave similarly as the Galton–Watson trees. After dominating the local neighborhoods of G by the new branching processes, we study the contact process on the latter graphs and bound its survival time based on the aforementioned ideas.

On the other hand, when studying the long survival for $G \sim \mathcal{G}(n, \mu)$, we rely on the existence of what we call *embedded expanders* inside the graph. Roughly speaking, we call a subset W of vertices in G an *embedded expander*, if: (i) all vertices in W have high degree, say, at least M , and (ii) distance R -neighborhood of every subset $W' \subset W$ of at most a certain size intersects with W at more than $2|W'|$ vertices (see a precise definition in Lemma 5.2). As noted above, we expect an infection at a degree M vertex to last for at least time exponential in M .

Intuitively, if a subset W' of an embedded expander W is infected, then the infections inside W' would happen repeatedly for a reasonably long time due to its large degrees, and hence it will not die out whp before infecting its neighbors within distance R . Thus, if W is an embedded expander, then the infection is likely to spread over $2|W'|$ vertices after some time. In Sections 5 and 6 we make this intuition rigorous and prove the existence of an embedded expander inside G . For the latter argument we partially rely on the *cut-off line algorithm* (Definition 7.3), which was introduced in [14], to find the cores of random graphs.

For Theorem 4 we show that if μ is subexponential, then we can find an embedded expander in G such that R is arbitrarily smaller than M . Therefore, even if λ is very small, it will be possible for infections in the embedded expander to travel the distance of R before dying out.

1.3. Organization. The rest of the paper is organized as follows. After we set up notations and review some preliminary facts in Section 2, we prove Theorem 1, Theorem 3(1), Theorem 3(2) in Sections 3, 4 and 5, respectively. In Section 6 we prove Theorems 4. In Section 7 we prove a structural lemma on the embedded expanders mentioned above which plays a crucial role in establishing Theorems 3(2) and 4.

1.4. Notations. For two positive sequences (a_n) and (b_n) , we say that $a_n = O(b_n)$ or $b_n = \Omega(a_n)$ if there exists a constant C independent of n such that $a_n \leq Cb_n$ for all n . If $a_n = O(b_n)$ and $b_n = O(a_n)$, we write $a_n = \Theta(b_n)$.

2. Preliminaries. In this section we set up notation and briefly describe some basic properties of the contact process and random graphs which will be used throughout the paper.

For a graph $G = (V, E)$ (finite or infinite), the contact process on G with infection rate λ is the continuous-time Markov chain on the state space $\{0, 1\}^V$, where 0 (resp., 1) corresponds to the healthy (resp., infected) state. If the initial state is $\mathbf{1}_A$, that is, the vertices in $A \subset V$ are infected, we denote the process by

$$(X_t) \sim \text{CP}^\lambda(G; \mathbf{1}_A).$$

We will frequently use the notation $\mathbf{0}$ for the all-healthy state $\mathbf{0} = \mathbf{1}_\emptyset$ and write $\mathbf{1}_v = \mathbf{1}_{\{v\}}$ if the state has a single infected vertex v . The transition rule of the process can be described as follows:

- X_t becomes $X_t - \mathbf{1}_v$ with rate 1 for each v such that $X_t(v) = 1$.
- X_t becomes $X_t + \mathbf{1}_u$ with rate $\lambda N_t(u)$ for each u with $X_t(u) = 0$, where $N_t(u)$ is the number of neighbors v of u with $X_t(v) = 1$.

We sometimes write $\text{CP}^\lambda(G)$ when the initial condition is unnecessary. For convenience, we usually denote the state space by $\{0, 1\}^G$.

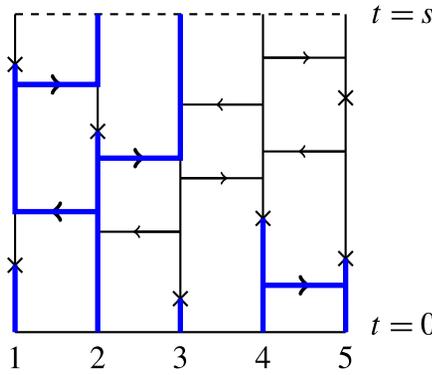


FIG. 1. A realization of the contact process on the interval $V = \{1, \dots, 5\}$ with initial condition $X_0 = \mathbf{1}_V$. The blue lines describe the spread of infection. We see that $X_s = \mathbf{1}_{\{2,3\}}$.

2.1. *Graphical representation of contact processes.* We briefly discuss a coupling method of the contact processes using a *graphical representation* based on Chapter 3, Section 6 of [18]. The idea is to record the infections and recoveries in $\text{CP}^\lambda(G; \mathbf{1}_A)$ on the space-time domain $G \times \mathbb{R}_+$. Define i.i.d. Poisson processes $\{N_v(t)\}_{v \in V}$ with rate 1 and i.i.d. Poisson processes $\{N_{\vec{uv}}(t)\}_{\vec{uv} \in \vec{E}}$ with rate λ , where $\vec{E} = \{\vec{uv}, \vec{vu} : (uv) \in E\}$ is the set of directed edges. Further, we let $\{N_v(t)\}_{v \in V}$ and $\{N_{\vec{uv}}(t)\}_{\vec{uv} \in \vec{E}}$ to be mutually independent. Then, the graphical representation is defined as follows:

1. Initially, we have the *empty* domain $V \times \mathbb{R}_+$.
2. For each $v \in V$, mark \times at the point (v, t) at each event time t of $N_v(\cdot)$.
3. For each $\vec{uv} \in \vec{E}$, add an arrow from (u, t) to (v, t) at each event time t of $N_{\vec{uv}}(\cdot)$.

This gives a geometric picture of $\text{CP}^\lambda(G; \mathbf{1}_A)$, and further provides a coupling of the processes over all possible initial states. Figure 1 tells us how to interpret the infections at time t based on this graphical representation. We point out two lemmas which are easy consequences of the above construction. For proofs, see, for example, [18].

LEMMA 2.1. *Suppose that we have the aforementioned coupling among the contact processes on a graph G . Let T_v and T_G be the first time when $\text{CP}^\lambda(G; \mathbf{1}_v)$ and $\text{CP}^\lambda(G; \mathbf{1}_G)$ reach the all-healthy state $\mathbf{0}$, respectively. Then, we have $T_G = \max\{T_v : v \in G\}$.*

LEMMA 2.2. *For a given graph $G = (V, E)$ and any $A \subset V$, let $(X_t) \sim \text{CP}^\lambda(G; \mathbf{1}_A)$. Consider any (random) subset \mathcal{I} of \mathbb{R}_+ , and define (X'_t) to be the coupled process of (X_t) that has the same initial state, infections and recoveries, except that the recoveries at a fixed vertex v are ignored at times $t \in \mathcal{I}$. Then, for any $t \geq 0$, we have $X_t \leq X'_t$, that is, $X_t(v) \leq X'_t(v)$ for all v .*

2.2. *Random graphs.* Let μ be a probability distribution on \mathbb{N} and n be any integer. The random graph $\mathcal{G}(n, \mu)$ with degree distribution μ is defined by the following procedure:

- Let d_1, \dots, d_n be n i.i.d. samples from μ conditioned on $\{\sum_{i=1}^n d_i \text{ is even}\}$.
- Sample G by taking a simple graph on n vertices with degrees $\{d_i\}_{i=1}^n$ uniformly at random among all possible choices.

Further, we consider a variant of $\mathcal{G}(n, \mu)$ which is constructed as follows:

- Sample d_1, \dots, d_n as above. Here, d_i denotes the number of *half-edges* attached to vertex i .
- Pair all the half-edges uniformly at random.

The resulting graph G is called the configuration model which is denoted by $\mathcal{G}_{\text{cf}}(n, \mu)$. The difference here is that G is not necessarily a simple graph. However, if the second moment of μ is finite, we have the following contiguity between the two models. For details, see, for example, [30], Chapter 7.

LEMMA 2.3 ([13, 30]). *Suppose that $\mathbb{E}_{D \sim \mu} D^2 < \infty$. Then, uniformly in n , we have*

$$\mathbb{P}_{G \sim \mathcal{G}_{\text{cf}}(n, \mu)}(G \text{ is simple}) \in (0, 1).$$

In particular, for any subset A_n of graphs with n vertices,

$$\mathbb{P}_{G \sim \mathcal{G}_{\text{cf}}(n, \mu)}(G \in A_n) \rightarrow 0 \text{ implies } \mathbb{P}_{G \sim \mathcal{G}(n, \mu)}(G \in A_n) \rightarrow 0.$$

Throughout the paper we study the configuration model $\mathcal{G}_{\text{cf}}(n, \mu)$ instead of $\mathcal{G}(n, \mu)$, under the assumption $\mathbb{E}_{D \sim \mu} D^2 < \infty$. Further, it is well known that $G \sim \mathcal{G}_{\text{cf}}(n, \mu)$ (and hence, $G \sim \mathcal{G}(n, \mu)$) whp contains the unique connected component of size linear in n , if and only if $\mathbb{E}_{D \sim \mu} D(D - 2) > 0$ (for details, see [21]). Hence, we always assume $\mathbb{E}_{D \sim \mu} D(D - 2) > 0$ which is the most interesting case for us. Otherwise, the graph decomposes into many small components, and the contact process would not exhibit $e^{\Theta(n)}$ -survival time for any $\lambda > 0$.

2.3. *Local weak convergence.* Given a sequence of random graphs G_n , let $N(v, r)$ be an induced subgraph of G_n consisting of vertices of distance at most r from v . Let \mathbb{P}_n be the distribution of the neighborhood $N(v, r)$ where v is a uniformly chosen vertex of G_n . We say that a random rooted tree \mathcal{T} is the local weak limit of G_n if for any finite r and any rooted tree T of depth at most r ,

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(N(v, r) = T) = \mathbb{P}(\mathcal{T}_r = T),$$

where \mathcal{T}_r is the subtree of the first r generations of \mathcal{T} .

We shall use the following known convergence of $G(n, \mu)$ to its corresponding Galton–Watson tree. Define the size-biased distribution μ' to be

$$\mu'(k - 1) = \frac{k\mu(k)}{\sum_{i=1}^{\infty} i\mu(i)}, \quad k = 1, 2, \dots$$

Note that if $\mu = \text{Pois}(d)$, then $\mu' = \mu$. Let $\mathcal{T}(\mu) \sim \text{GW}(\mu, \mu')$ be the *size-biased* Galton–Watson tree in which the number of children of the root has distribution μ and the number of children of an i th generation vertex ($i \geq 1$) has distribution μ' . We also stress that the Galton–Watson tree is supercritical if and only if $\mathbb{E}_{D' \sim \mu'} D' > 1$, equivalent to $\mathbb{E}_{D \sim \mu} D(D - 2) > 0$, which we saw above.

LEMMA 2.4 ([7], Section 2.1). *Assume that μ has finite mean. Then, the size-biased Galton–Watson tree $\mathcal{T}(\mu)$ is the local weak limit of $\mathcal{G}(n, \mu)$. The Galton–Watson tree with degree distribution $\text{Pois}(d)$ is the local weak limit of the Erdős–Rényi random graph $\mathcal{G}_{n, d/n}$.*

3. Extinction in Galton–Watson trees. Let ξ be a random variable on \mathbb{N} having an exponential tail, namely, $\mathbb{E} \exp(c\xi) = M < \infty$ for some constants $c, M > 0$. Throughout this section we assume $\mathbb{E} \xi > 1$, which makes $\mathcal{T} \sim \text{GW}(\xi)$, the Galton–Watson tree with offspring distribution ξ , survive forever with positive probability. We also denote the depth- L Galton–Watson tree by $\mathcal{T}_L \sim \text{GW}(\xi)_L$, and its root denoted by ρ .

The goal of this section is to establish Theorem 1. To this end, we prove the following in the next two sections:

- We first show that, for small enough λ , the expected survival time of $\text{CP}^\lambda(\mathcal{T}_L; \mathbf{1}_\rho)$ is bounded by a constant uniform in L .

- Then, we prove that, for small enough λ , the probability that the infection in $\text{CP}^\lambda(\mathcal{T}_L; \mathbf{1}_\rho)$ goes deeper than h decays exponentially in h .

At the end we will combine the two to see that the death–survival threshold λ_1 of the infinite Galton–Watson tree is strictly positive. Moreover, both properties will be essential in Section 4.

3.1. *Expected survival time in finite trees.* In this section we prove the following theorem.

THEOREM 3.1. *Let L be an arbitrary integer and ξ, \mathcal{T}_L be defined as above. Let R_L be the first time when $\text{CP}^\lambda(\mathcal{T}_L; \mathbf{1}_\rho)$ reaches state $\mathbf{0}$. Then, there exist constants $C, \lambda_0 > 0$, depending only on ξ , such that for any $\lambda \leq \lambda_0$ and L , we have $\mathbb{E}R_L \leq C$.*

Let $D \sim \xi$ denote the degree of the root ρ and v_1, \dots, v_D be the children of ρ . Further, let \mathcal{T}_{v_i} denote the subtree of \mathcal{T} rooted at v_i . To establish Theorem 3.1, our attempt is to study the effect of joining the subtrees \mathcal{T}_{v_i} together at ρ , and hence expressing R_L in terms of R_{L-1} . The main difficulty of this approach comes from the fact that the contact process on \mathcal{T}_L does not behave independently on each subtree \mathcal{T}_{v_i} . To overcome this obstacle, we study the contact process in a slightly different setting, that is, by adding a parent ρ^+ above the root ρ which is infected permanently.

DEFINITION 3.2 (Root-added contact process). Let T be a finite tree rooted at ρ . Let T^+ be the tree that has a parent vertex ρ^+ of ρ which is connected only with ρ . The *root-added contact process* on T is the continuous-time Markov chain on the state space $\{0, 1\}^{T^+}$, defined as the contact process on T^+ with ρ^+ set to be infected permanently (hence, we exclude ρ^+ from the state space). That is, ρ^+ is infected initially, and it does not have a recovery clock attached to itself. Let $\text{CP}^\lambda_{\rho^+}(T^+; x_0)$ denote the root-added contact process on T with initial condition $x_0 \in \{0, 1\}^{T^+}$.

By adding a permanently infected parent, we can take advantage of independence between different subtrees as well as the stationary distribution of the process, as briefly discussed in Section 1.2. In the following lemma we formally introduce the “modified process” explained in Section 1.2 and construct a quantitative recursion argument in terms of the tree depth.

LEMMA 3.3. *Let L be an arbitrary integer and ξ, \mathcal{T}_L be defined as above. Define S_L to be the first time when $\text{CP}^\lambda_{\rho^+}(\mathcal{T}_L^+; \mathbf{1}_\rho)$ reaches state $\mathbf{0}$. Then, there exists a constant $\lambda_0 > 0$, depending only on ξ , such that for any $\lambda \leq \lambda_0$ and L , $\mathbb{E}S_L \leq e$.*

PROOF. We build an inductive argument in terms of L , by considering the modified contact process $(\tilde{X}_t) \sim \tilde{\text{CP}}^\lambda_{\rho^+; \rho}(\mathcal{T}_L^+; \mathbf{1}_\rho)$ defined as follows:

- (\tilde{X}_t) is coupled with $(X_t) \sim \text{CP}^\lambda_{\rho^+}(\mathcal{T}_L^+; \mathbf{1}_\rho)$ in the sense that they share the same locations of recovery and infection clocks. In particular, ρ^+ is permanently infected in (\tilde{X}_t) .

- In (\tilde{X}_t) , the recovery at ρ at time s is valid if and only if $\tilde{X}_s = \mathbf{1}_\rho$. Otherwise, we ignore the recovery at ρ . In other words, when there exists an infected vertex other than ρ and ρ^+ , the recovery at ρ is neglected.

Let \tilde{S}_L be the first time when \tilde{X}_t reaches the all-healthy state $\mathbf{0}$. Then, Lemma 2.2 tells us that $S_L \leq \tilde{S}_L$. Assume that we started running (\tilde{X}_t) from $t = 0$. Then, there are two possibilities for the transition to the second state from the initial state $\mathbf{1}_\rho$:

- A. ρ is healed.
- B. ρ infects one of its children, say, v_i .

When **A** happens, then \tilde{S}_L is just the time elapsed until encountering **A**. If we let D denote the number of children of ρ , then probability of the event **A** is $\frac{1}{1+\lambda D}$ and also

$$\mathbb{E}[\tilde{S}_L | \mathbf{A}] = \frac{1}{1 + \lambda D}.$$

On the other hand, when **B** happens, then the recoveries at ρ are neglected until all of its descendants are healthy. Therefore, the infection and recovery occurring inside the subtrees $\mathcal{T}_{v_i} \cup \{\rho\}$ become independent of each other until all of them become completely healthy at the same time. Hence, after the occurrence of **B**, where we have $\mathbf{1}_{\{\rho, v_i\}}$ as its new initial state, (\tilde{X}_t) can be viewed as the product chain (X_t^\otimes) of root-added contact processes, defined as follows:

$$(X_t^\otimes) \sim \text{CP}_\rho^\otimes(\mathcal{T}_L; \mathbf{1}_{v_i}) := \left(\bigotimes_{\substack{j=1 \\ j \neq i}}^D \text{CP}_\rho^\lambda(\mathcal{T}_{v_j}^+; \mathbf{0}) \right) \otimes \text{CP}_\rho^\lambda(\mathcal{T}_{v_i}^+; \mathbf{1}_{v_i}).$$

(Here, for each \mathcal{T}_{v_j} , we view ρ as its permanently infected parent of the root v_i .) Note that this perspective is valid until \tilde{X}_t returns back to $\mathbf{1}_\rho$.

Let \tilde{S}_i^\otimes denote the time that the above product chain on $\bigcup_{j=1}^D \mathcal{T}_{v_j}$ started from the state $\mathbf{1}_{v_i}$ reaches the all-healthy state $\mathbf{0}$. At time $s = \tilde{S}_i^\otimes$, \tilde{X}_s is again in the state $\mathbf{1}_\rho$; hence, it again meets with either **A** or **B** in the next step. Note that in this situation the expected waiting time to encounter either event is $\frac{1}{1+\lambda D}$. Also, define \tilde{S}^\otimes to be the average of \tilde{S}_i^\otimes over all i , recalling that when event **B** occurs, each child v_i is infected with equal probability. Then, if we continue this procedure until \tilde{X}_t reaches $\mathbf{0}$, we get

$$\begin{aligned} \mathbb{E}[\tilde{S}_L | \mathcal{T}_L] &= \sum_{k=0}^{\infty} \left(\frac{\lambda D}{1 + \lambda D} \right)^k \frac{1}{1 + \lambda D} \\ (3.1) \quad &\times \left[(k + 1) \frac{1}{1 + \lambda D} + k \mathbb{E}[\tilde{S}^\otimes | \{\mathcal{T}_{v_i} : i \in [D]\}] \right]. \end{aligned}$$

Simplifying the sum then gives

$$\mathbb{E}[\tilde{S}_L | \mathcal{T}_L] = 1 + \lambda D \mathbb{E}[\tilde{S}^\otimes | \{\mathcal{T}_{v_i} : i \in [D]\}]$$

which implies

$$(3.2) \quad \mathbb{E}[\tilde{S}_L | D] = 1 + \lambda D \mathbb{E}[\tilde{S}^\otimes | D].$$

The next step of the proof is to estimate $\mathbb{E}[\tilde{S}^\otimes | D]$ by relating it to the stationary distributions of the root-added contact processes. Let $\pi^{(D)}$ be the stationary distribution of the product chain $\text{CP}_\rho^\otimes(\mathcal{T}_L)$ (when defining $\pi^{(D)}$, note that the initial state of the process is irrelevant). We also let π_i be the stationary distribution of $\text{CP}_\rho^\lambda(\mathcal{T}_i^+)$. Then, we have

$$\pi^{(D)} = \bigotimes_{i=1}^D \pi_i.$$

For any state x on $\mathcal{T}_L \setminus \{\rho\}$, $\pi^{(D)}(x)$ is proportional to the expected time that the chain $(X_t^\otimes) \sim \text{CP}_\rho^\otimes(\mathcal{T}_L)$ stays at state x . Moreover, the expected time for the chain to stay at $\mathbf{0}$ is

$(\lambda D)^{-1}$, and, after escaping from $\mathbf{0}$, it spends time $\mathbb{E}[\tilde{S}^\otimes | \mathcal{T}_L]$ in expectation before returning back to $\mathbf{0}$. Therefore,

$$(3.3) \quad \pi^{(D)}(\mathbf{0}) = \frac{(\lambda D)^{-1}}{(\lambda D)^{-1} + \mathbb{E}[\tilde{S}^\otimes | \mathcal{T}_L]} = \frac{1}{1 + \lambda D \mathbb{E}[\tilde{S}^\otimes | \mathcal{T}_L]}.$$

Similarly, we have

$$(3.4) \quad \pi_i(\mathbf{0}) = \frac{1}{1 + \lambda \mathbb{E}[S_{L-1} | \mathcal{T}_{v_i}]},$$

where S_{L-1} is the first time when $(X_t^i) \sim \text{CP}_\rho^\lambda(\mathcal{T}_{v_i}^+; \mathbf{1}_{v_i})$ reaches state $\mathbf{0}$. Here, note that S_{L-1} matches with the definition from the statement of this lemma since $\mathcal{T}_{v_i} \sim \text{GW}(\xi)_{L-1}$. Therefore, we obtain that

$$(3.5) \quad 1 + \lambda D \mathbb{E}[\tilde{S}^\otimes | \mathcal{T}_L] = \prod_{i=1}^D (1 + \lambda \mathbb{E}[S_{L-1} | \mathcal{T}_{v_i}]).$$

Since $\{\mathcal{T}_{v_i}\}_{i \geq 1}$ are i.i.d. $\text{GW}(\xi)_{L-1}$ for all i , integrating (3.5) over the randomness of $\{\mathcal{T}_{v_i} : i \in [D]\}$ tells us that

$$1 + \lambda D \mathbb{E}[\tilde{S}^\otimes | D] = (1 + \lambda \mathbb{E}[S_{L-1}])^D \leq \exp\{(\lambda \mathbb{E}[S_{L-1}])D\}.$$

Combining this with (3.2), we get

$$(3.6) \quad \mathbb{E}[\tilde{S}_L | D] \leq \exp\{(\lambda \mathbb{E}[S_{L-1}])D\}.$$

In the last step of the proof, we complete the inductive argument using the fact that $D \sim \xi$ has an exponential tail. Let us set $c, M > 0$ to be the constants satisfying $\mathbb{E} \exp(cD) = M$. When $L = 0$, we trivially have that $\mathbb{E}S_0 = 1$. Define K and λ_0 as

$$K = e \cdot \max\{\log M, 1\}, \quad \lambda_0 = \frac{c}{K}.$$

Suppose that $\mathbb{E}[S_{L-1}] \leq e$. Then, for any $\lambda \leq \lambda_0$, we have

$$\begin{aligned} \mathbb{E}S_L \leq \mathbb{E}\tilde{S}_L &\leq \mathbb{E}_{D \sim \xi} [\exp(\lambda \mathbb{E}[S_{L-1}]D)] = \mathbb{E}_{D \sim \xi} \left[\exp\left(\frac{\lambda \mathbb{E}[S_{L-1}]}{c} \cdot cD\right) \right] \\ &\leq \exp\left\{ \log M \frac{\lambda \mathbb{E}[S_{L-1}]}{c} \right\} \leq e, \end{aligned}$$

where we used Jensen’s inequality to deduce the first inequality in the second line. Finally, an elementary induction argument implies the desired result. \square

PROOF OF THEOREM 3.1. For R_L, S_L defined as in the statement of Theorem 3.1 and Lemma 3.3, respectively, we have $\mathbb{E}R_L \leq \mathbb{E}S_L$ due to Lemma 2.2. Therefore, setting λ_0 as in the proof of Lemma 3.3 and $C = e$, we obtain $\mathbb{E}R_L \leq C$ for all $\lambda \leq \lambda_0$ and L . \square

3.2. Exponential decay of the infection depth. In this section we show that the maximal depth that the infection can reach before dying out decays exponentially.

For any integer L , let $\mathcal{T}_L \sim \text{GW}(\xi)_L$ and \mathcal{T}_L^+ be the graph obtained by adding a new parent root ρ^+ above ρ in \mathcal{T}_L as before. For each state $x \in \{0, 1\}^{\mathcal{T}_L}$, define the depth of x in \mathcal{T}_L^+ to be

$$r(x) = r(x; \mathcal{T}_L^+) = \max\{d(\rho^+, v) : x(v) = 1\}.$$

For $x = \mathbf{0}$, we set $r(\mathbf{0}) = 0$. Consider the root-added process $(X_t) \sim \text{CP}_{\rho^+}^\lambda(\mathcal{T}_L^+; \mathbf{1}_\rho)$ (Definition 3.2), and let S_L be the first time when the process reaches the state $\mathbf{0}$. Let $H = \max\{r(X_t) : t \in [0, S_L]\}$ be the maximal depth that the process reaches during an excursion from $\mathbf{0}$. Our goal in this section is to establish the following theorem and conclude the proof of Theorem 1.

THEOREM 3.4. *Let $L > 0$ be any integer and let \mathcal{T}_L, S_L and H be as above. There exist constants $K, \lambda_0 > 0$, depending only on ξ , such that for all $\lambda \leq \lambda_0, h > 0$ and $m > 0$, we have*

$$\mathbb{P}(H > h \mid \mathcal{T}_L) \leq 2m(K\lambda)^h$$

with probability at least $1 - m^{-1}$ over the choice of \mathcal{T}_L .

In order to control the deepest depth of infection, we introduce the *delayed* contact process.

DEFINITION 3.5 (Delayed contact process). Let \mathcal{S}^+ be a graph rooted at ρ^+ and $\mathcal{S} = \mathcal{S}^+ \setminus \{\rho^+\}$. For any two states $x, y \in \{0, 1\}^{\mathcal{S}}$, let Q_{xy} be the rate of transition from x to y in the contact process $\text{CP}_{\rho^+}^{\lambda}(\mathcal{S}^+)$. For a fixed constant $\theta \in (0, 1)$, the *delayed contact process*, denoted by $\text{DP}_{\rho^+}^{\lambda, \theta}(\mathcal{S}^+; x_0)$, is the continuous-time Markov chain on $\{0, 1\}^{\mathcal{S}}$ with initial state x_0 and transition rate

$$Q_{xy}^{(\theta)} = \theta^{r(x; \mathcal{S}^+)} Q_{xy} = \theta^{r(x)} Q_{xy}.$$

According to the definition, in the delayed contact process we spend exponentially longer time in the states with deeper depths. Let $\pi_{\mathcal{S}}, \nu_{\mathcal{S}}^{\theta}$ denote the stationary distributions of $\text{CP}_{\rho^+}^{\lambda}(\mathcal{S}^+)$ and $\text{DP}_{\rho^+}^{\lambda, \theta}(\mathcal{S}^+)$, respectively. Then,

$$(3.7) \quad \nu_{\mathcal{S}}^{\theta}(x) = \frac{\theta^{-r(x)} \pi_{\mathcal{S}}(x)}{\sum_y \theta^{-r(y)} \pi_{\mathcal{S}}(y)},$$

where the summation is over all possible states $y \in \{0, 1\}^{\mathcal{S}}$.

Suppose we have a lower bound on $\nu_{\mathcal{S}}^{\theta}(\mathbf{0})$. Then, this implies an upper bound on $\pi_{\mathcal{S}}(x)$, by

$$\pi_{\mathcal{S}}(x) = \frac{\theta^{r(x)} \nu_{\mathcal{S}}^{\theta}(x)}{\sum_y \theta^{r(y)} \nu_{\mathcal{S}}^{\theta}(y)} \leq \frac{\theta^{r(x)} \nu_{\mathcal{S}}^{\theta}(x)}{\nu_{\mathcal{S}}^{\theta}(\mathbf{0})},$$

which intuitively infers that it is (exponentially) unlikely to see states of having very deep infections until the process comes back to $\mathbf{0}$. Based on this intuition, we establish the following proposition.

PROPOSITION 3.6. *Let $L > 0$ be any integer and $\mathcal{T}_L \sim \text{GW}(\xi)_L$. Set $\nu_{\mathcal{T}_L}^{\theta}$ to denote the stationary distribution of $\text{DP}_{\rho^+}^{\lambda, \theta}(\mathcal{T}_L^+)$ on the space $\{0, 1\}^{\mathcal{T}_L}$. Then, there exist constants $K, \lambda_0 > 0$, depending only on ξ , such that for all $\lambda \leq \lambda_0$ and L ,*

$$\mathbb{E}[\nu_{\mathcal{T}_L}^{\theta}(\mathbf{0})^{-1}] \leq 2,$$

where θ is given by $\theta = K\lambda$.

PROOF. Let S_L^{θ} be the first time when $(X_t) \sim \text{DP}_{\rho^+}^{\lambda, \theta}(\mathcal{T}_L^+; \mathbf{1}_{\rho})$ reaches state $\mathbf{0}$. We first derive an analog of (3.2) based on the methods from Lemma 3.3. To this end, define $(\tilde{X}_t) \sim \tilde{\text{DP}}_{\rho^+; \rho}^{\lambda, \theta}(\mathcal{T}_L^+; \mathbf{1}_{\rho})$ to be the modification of (X_t) in such a way that:

1. (\tilde{X}_t) shares the same infection and recovery clocks as (X_t) .
2. In (\tilde{X}_t) , healing attempt at ρ is ignored if there exists an infected vertex other than ρ^+ and ρ at that moment.

Let $D \sim \xi$ denote the number of children of ρ , and let $\{\mathcal{T}_{u_i} : i = 1, \dots, D\}$ be the subtrees from the children u_1, \dots, u_D of ρ . Then, when there is an infected vertex other than ρ^+ and ρ in \tilde{X}_t , it can be regarded as the slowed-down version of process $\text{DP}_\rho^{\lambda, \theta}(\mathcal{T}_L)$, where it spends longer time by the factor of θ^{-1} at each state, since the tree \mathcal{T}_L is one depth lower than \mathcal{T}_L^+ . Note that in $\text{DP}_\rho^{\lambda, \theta}(\mathcal{T}_L)$, ρ is the permanently infected parent that has D children.

Let \tilde{S}_L^θ be the first time when \tilde{X}_t becomes $\mathbf{0}$. Also, let \tilde{S}_i^θ be the first time when $\text{DP}_\rho^{\lambda, \theta}(\mathcal{T}_L; \mathbf{1}_{u_i})$ is $\mathbf{0}$, and set \tilde{S}^θ to be the average of \tilde{S}_i^θ over $i = 1, \dots, D$. Then, we can apply the same argument as Lemma 3.3 to this setting and deduce that

$$\begin{aligned} \mathbb{E}[\tilde{S}_L^\theta \mid \mathcal{T}_L] &= \sum_{k=0}^\infty \left(\frac{\lambda D}{1 + \lambda D} \right)^k \frac{1}{1 + \lambda D} \left[\frac{k + 1}{\theta(1 + \lambda D)} + \frac{k}{\theta} \mathbb{E}[\tilde{S}^\theta \mid \mathcal{T}_L] \right] \\ (3.8) \quad &= \frac{1}{\theta} (1 + \lambda D \mathbb{E}[\tilde{S}^\theta \mid \mathcal{T}_L]). \end{aligned}$$

Now we relate these equations with the stationary distributions. Let $\tilde{v}_{\mathcal{T}_L}^\theta, v_{\mathcal{T}_{u_i}}^\theta$ be the stationary distributions of $\text{DP}_\rho^{\lambda, \theta}(\mathcal{T}_L), \text{DP}_\rho^{\lambda, \theta}(\mathcal{T}_{u_i}^+)$, respectively. Further, define $v_{\mathcal{T}_L}^\otimes = \otimes_{i=1}^D v_{\mathcal{T}_{u_i}}^\theta$. In contrast to what we had in Lemma 3.3, we do not necessarily have $\tilde{v}_{\mathcal{T}_L}^\theta = v_{\mathcal{T}_L}^\otimes$.

For each state $x \in \Omega_L := \{0, 1\}^{\cup_{i=1}^D \mathcal{T}_{u_i}}$ of $\text{DP}_\rho^{\lambda, \theta}(\mathcal{T}_L)$, we decompose it into $x = (x_i)_{i=1}^D$, where $x_i \in \Omega_i := \{0, 1\}^{\mathcal{T}_{u_i}}$. Setting $\pi_{\mathcal{T}_{u_i}}$ to be the stationary distribution of $\text{CP}_\rho^\lambda(\mathcal{T}_{u_i}^+)$ and $\pi_{\mathcal{T}_L}^\otimes := \otimes_{i=1}^D \pi_{\mathcal{T}_{u_i}}$, the equation (3.7) implies that

$$\begin{aligned} \tilde{v}_{\mathcal{T}_L}^\theta(x) &= \frac{\theta^{-r(x; \mathcal{T}_L)} \pi_{\mathcal{T}_L}^\otimes(x)}{\sum_{y \in \Omega_L} \theta^{-r(y; \mathcal{T}_L)} \pi_{\mathcal{T}_L}^\otimes(y)} = \frac{\theta^{-r(x; \mathcal{T}_L)} \prod_{i=1}^D \pi_{\mathcal{T}_{u_i}}(x_i)}{\sum_{y \in \Omega_L} \theta^{-r(y; \mathcal{T}_L)} \prod_{i=1}^D \pi_{\mathcal{T}_{u_i}}(y_i)}; \\ (3.9) \quad v_{\mathcal{T}_L}^\otimes(x) &= \prod_{i=1}^D \left[\frac{\theta^{-r(x_i; \mathcal{T}_{u_i})} \pi_{\mathcal{T}_{u_i}}(x_i)}{\sum_{y_i \in \Omega_i} \theta^{-r(y_i; \mathcal{T}_{u_i})} \pi_{\mathcal{T}_{u_i}}(y_i)} \right] \\ &= \frac{\theta^{-\sum_{i=1}^D r(x_i; \mathcal{T}_{u_i})} \prod_{i=1}^D \pi_{\mathcal{T}_{u_i}}(x_i)}{\sum_{y \in \Omega_L} \theta^{-\sum_{i=1}^D r(y_i; \mathcal{T}_{u_i})} \prod_{i=1}^D \pi_{\mathcal{T}_{u_i}}(y_i)}. \end{aligned}$$

Notice that

$$r(x; \mathcal{T}_L) = \max\{r(x_i; \mathcal{T}_{u_i}) : i = 1, \dots, D\} \leq \sum_{i=1}^D r(x_i; \mathcal{T}_{u_i}).$$

Therefore, deeper states tend to have larger weight in $v_{\mathcal{T}_L}^\otimes$ than in $\tilde{v}_{\mathcal{T}_L}^\theta$ which implies that

$$v_{\mathcal{T}_L}^\otimes(\mathbf{0}) \leq \tilde{v}_{\mathcal{T}_L}^\theta(\mathbf{0}).$$

Moreover, we have the following equations as an analog of (3.3), (3.4):

$$(3.10) \quad \tilde{v}_{\mathcal{T}_L}^\theta(\mathbf{0}) = \frac{1}{1 + \lambda D \mathbb{E}[\tilde{S}^\theta \mid \mathcal{T}_L]}, \quad v_{\mathcal{T}_L}^\otimes(\mathbf{0}) = \prod_{i=1}^D \left[\frac{1}{1 + \lambda \mathbb{E}[S_{L-1}^\theta \mid \mathcal{T}_{u_i}]} \right].$$

We combine our discussion with (3.8) to deduce that

$$\mathbb{E}[S_L^\theta \mid \mathcal{T}_L] \leq \mathbb{E}[\tilde{S}_L^\theta \mid \mathcal{T}_L] \leq \frac{1}{\theta} \prod_{i=1}^D (1 + \lambda \mathbb{E}[S_{L-1}^\theta \mid \mathcal{T}_{u_i}]),$$

and hence

$$(3.11) \quad \mathbb{E}[S_L^\theta] \leq \frac{1}{\theta} \mathbb{E}_{D \sim \xi} [\exp(\lambda \mathbb{E}[S_{L-1}^\theta] D)].$$

The final step is to adjust the constants and to deduce the conclusion. Let $c, M > 0$ be constants satisfying $\mathbb{E}_{D \sim \xi} \exp(cD) = M$. We set $K, \lambda_0 > 0$ and θ as

$$K = \max \left\{ \frac{2 \log M}{c \log 2}, 2 \right\}, \quad \lambda_0 = \frac{1}{2K}, \quad \theta = K\lambda,$$

where $\lambda \in (0, \lambda_0]$ is arbitrary. For $L = 0$, we have $\mathbb{E}S_0^\theta = \theta^{-1}$. Suppose that $\mathbb{E}S_{L-1}^\theta \leq 2/\theta$. Then, the right-hand side of (3.11) can be bounded by

$$\begin{aligned} \frac{1}{\theta} \mathbb{E}_{D \sim \xi} [\exp(\lambda \mathbb{E}[S_{L-1}^\theta] D)] &= \frac{1}{\theta} \mathbb{E}_{D \sim \xi} \left[\exp \left(\frac{\lambda \mathbb{E}[S_{L-1}^\theta]}{c} \cdot cD \right) \right] \\ &\leq \frac{1}{\theta} \exp \left\{ \log M \left(\frac{\lambda}{c} \mathbb{E}[S_{L-1}^\theta] \right) \right\} \leq \frac{1}{\theta} M^{\frac{2}{cK}} \leq \frac{2}{\theta}, \end{aligned}$$

where the first inequality is due to Jensen’s inequality. Therefore, for K, λ_0 as above, we have $\mathbb{E}S_L^\theta \leq 2/\theta$ for all $\lambda \leq \lambda_0$ with $\theta = K\lambda$. Finally, note that $v_{\mathcal{T}_L}^\theta(\mathbf{0})$ is given by

$$v_{\mathcal{T}_L}^\theta(\mathbf{0}) = \frac{1}{1 + \lambda \mathbb{E}[S_L^\theta | \mathcal{T}_L]}.$$

Thus, we obtain the desired conclusion by taking expectation over its reciprocal and plugging in the estimate $\mathbb{E}S_L^\theta \leq 2/\theta$. \square

PROOF OF THEOREM 3.4. Let $L > 0$ be an arbitrary integer, and let $\mathcal{T}_L \sim \text{GW}(\xi)_L$. Also, let K, λ_0 be the constants given by Lemma 3.6, and let $\pi_{\mathcal{T}_L}, v_{\mathcal{T}_L}^\theta$ be the stationary distributions of $\text{CP}_{\rho^+}^\lambda(\mathcal{T}_L^+)$ and $\text{DP}_{\rho^+}^{\lambda, \theta}(\mathcal{T}_L^+)$, respectively, with $\theta = K\lambda$.

Set $\Omega = \{0, 1\}^{\mathcal{T}_L}$, and define

$$A := \{x \in \Omega : r(x; \mathcal{T}_L^+) \geq h\}.$$

We first observe that

$$\frac{\pi_{\mathcal{T}_L}(A)}{\pi_{\mathcal{T}_L}(\mathbf{0})} = \frac{\sum_{x \in A} \theta^{r(x; \mathcal{T}_L^+)} \tilde{v}_{\mathcal{T}_L}^\theta(x)}{\tilde{v}_{\mathcal{T}_L}^\theta(\mathbf{0})} \leq \frac{\tilde{v}_{\mathcal{T}_L}^\theta(A)}{\tilde{v}_{\mathcal{T}_L}^\theta(\mathbf{0})} \theta^h.$$

Proposition 3.6 and Markov’s inequality imply that with probability $1 - m^{-1}$ over the choice of \mathcal{T}_L , we have $\tilde{v}_{\mathcal{T}_L}^\theta(\mathbf{0})^{-1} \leq 2m$, and hence for such choices

$$(3.12) \quad \frac{\pi_{\mathcal{T}_L}(A)}{\pi_{\mathcal{T}_L}(\mathbf{0})} \leq 2m(K\lambda)^h.$$

Moreover, if $(X_t) \sim \text{CP}_{\rho^+}^\lambda(\mathcal{T}_L^+)$ hits A , then the expected time needed for X_t to escape from A is at least 1. Indeed, it takes a unit expected time just to heal one infected site of depth at least h . In other words, if we set S_L, H as in the statement and define $\gamma_L(h) := |\{t \in [0, S_L] : X_t \in A\}|$ where $|\cdot|$ denotes the Lebesgue measure, then

$$\mathbb{E}[\gamma_L(h) | H \geq h, \mathcal{T}_L] \geq 1.$$

Combining this with (3.12) tells us that

$$\begin{aligned} \mathbb{P}(H \geq h | \mathcal{T}_L) &\leq \mathbb{E}[\gamma_L(h) | H \geq h, \mathcal{T}_L] \mathbb{P}(H \geq h | \mathcal{T}_L) \leq \mathbb{E}[\gamma_L(h) | \mathcal{T}_L] \\ &\leq \frac{\pi_{\mathcal{T}_L}(A)}{\pi_{\mathcal{T}_L}(\mathbf{0})} \leq 2m(K\lambda)^h, \end{aligned}$$

with probability $1 - m^{-1}$ over the choice of \mathcal{T}_L . \square

We conclude this section by completing the proof of Theorem 1.

PROOF OF THEOREM 1. Let K, λ_0 be given as Theorem 3.4, and set $\lambda \leq \lambda_0$ to be a constant such that $K\lambda < 1$. Let $\delta > 0$ be any given small number, and set h to be the constant satisfying $(K\lambda)^h = \frac{\delta^2}{8}$. Further, let $\mathcal{T} \sim \text{GW}(\xi)$ and ρ be its root.

Define $E(h)$ to be the event that the infection inside $\text{CP}^\lambda(\mathcal{T}; \mathbf{1}_\rho)$ does not go deeper than depth h until dying out. Then, Theorem 3.4 implies that

$$\mathbb{P}(E(h)) \geq 1 - \delta$$

which can be seen by setting $m = \frac{2}{\delta}$.

Let \mathcal{T}_h be the truncated tree of \mathcal{T} at depth h , and couple the processes $\text{CP}^\lambda(\mathcal{T}; \mathbf{1}_\rho)$ and $\text{CP}^\lambda(\mathcal{T}_h; \mathbf{1}_\rho)$ by identifying the recoveries and infections inside \mathcal{T}_h . Then, on $E(h)$, $\text{CP}^\lambda(\mathcal{T}; \mathbf{1}_\rho)$ can be regarded as $\text{CP}^\lambda(\mathcal{T}_h; \mathbf{1}_\rho)$. Let R and R_h be the times when $\text{CP}^\lambda(\mathcal{T}; \mathbf{1}_\rho)$ and $\text{CP}^\lambda(\mathcal{T}_h; \mathbf{1}_\rho)$ reaches $\mathbf{0}$. Then, Theorem 3.1 tells us that

$$\mathbb{E}[R \mid E(h)] = \mathbb{E}[R_h \mid E(h)] \leq \frac{\mathbb{E}R_h}{\mathbb{P}(E(h))} < \infty.$$

Thus, for $(X_t) \sim \text{CP}^\lambda(\mathcal{T}; \mathbf{1}_\rho)$, we have

$$\mathbb{P}(X_t \neq \mathbf{0} \text{ for all } t \geq 0) \leq \delta.$$

Since this holds true for all $\delta > 0$, we conclude that $\lambda_1(\text{GW}(\xi)) \geq \lambda > 0$. \square

4. Short survival in random graphs. We turn our attention to the contact process on random graphs $G \sim \mathcal{G}(n, \mu)$. Throughout the rest of the paper, μ is a probability distribution on \mathbb{N} that satisfies for $D \sim \mu$,

$$\sigma^2 := \mathbb{E}D^2 < \infty \quad \text{and} \quad b := \frac{\mathbb{E}D(D-1)}{\mathbb{E}D} > 1,$$

as discussed in Section 2.2. In this section, in particular, we assume that μ has an exponential tail, that is, $\mathbb{E} \exp(cD) < \infty$ for some $c > 0$. Our goal is to establish Theorem 3(1) by proving the following.

THEOREM 4.1. *Let μ be as above, and $G \sim \mathcal{G}(n, \mu)$. For a vertex $v \in G$, let T_v denote the time when $\text{CP}^\lambda(G; \mathbf{1}_v)$ reaches the state $\mathbf{0}$. Then, there exist events $\mathcal{E}_1, \mathcal{E}_2(G)$ and constants $B, \lambda_0 > 0$, depending on μ , such that the following hold:*

- \mathcal{E}_1 is an event over the random graphs such that $\mathbb{P}(G \in \mathcal{E}_1) = 1 - o(1)$.
- $\mathcal{E}_2(G)$ is an event over the contact process $\text{CP}^\lambda(G)$ such that

$$\mathbb{P}(\mathcal{E}_2 \mid G \in \mathcal{E}_1) = 1 - o(1).$$

- For all $\lambda \in (0, \lambda_0)$, we have

$$\frac{1}{n} \sum_{v \in G} \mathbb{E}[T_v \mid G \in \mathcal{E}] \leq B$$

for all large enough n .

Then, our main theorem follows simply by applying Markov's inequality.

PROOF OF THEOREM 3(1). Let T be the time when $\text{CP}^\lambda(G; \mathbf{1}_G)$ reaches the state $\mathbf{0}$. On the event \mathcal{E}_1 and $\mathcal{E}_2 = \mathcal{E}_2(G)$ given in Theorem 4.1, for any constant $C > 0$ we have

$$\begin{aligned} \mathbb{P}(T > Cn \mid \mathcal{E}_1, \mathcal{E}_2) &= \mathbb{P}\left(\max_{v \in G} T_v > Cn \mid \mathcal{E}_1, \mathcal{E}_2\right) \\ &\leq \sum_{v \in G} \mathbb{P}(T_v > Cn \mid \mathcal{E}_1, \mathcal{E}_2) \leq \frac{B}{C}, \end{aligned}$$

where the first equality is due to Lemma 2.1, and the second is from Markov’s inequality. Since the events \mathcal{E}_1 and \mathcal{E}_2 given $G \in \mathcal{E}_1$ both hold whp, we obtain the conclusion. \square

In the rest of the section, we focus on proving Theorem 4.1. Our proof relies much on the fact that the local neighborhood $N(v, L) := \{u \in G : \text{dist}(u, v) \leq L\}$ of a fixed vertex v roughly looks like a Galton–Watson branching process. Hence, the results from Section 3 will play a huge role in this section as well.

However, since $\mathcal{G}(n, \mu)$ contains cycles with nontrivial probability, we introduce a variant of Galton–Watson trees that can cover the effect of cycles in $\mathcal{G}(n, \mu)$, and develop a delicate coupling argument with the local neighborhood $N(v, L)$. This new branching process will stochastically dominate $N(v, L)$ in terms of isomorphic embeddings of graphs, and hence the contact process will survive for a longer time. The result will then follow by showing Theorem 4.1 for this new graph.

4.1. *Coupling the local neighborhood.* Let $G \sim \mathcal{G}(n, \mu)$, where μ has an exponential tail, and let μ' denote the size-biased distribution of μ . As discussed in Section 2.2, it is well known that the local neighborhood $N(v, L)$ around $v \in G$ behaves roughly as the Galton–Watson process $\text{GW}(\mu, \mu')_L$. However, the standard coupling between the two objects produces an error at least $\Theta(n^{-1})$. Therefore, we consider augmented versions of μ, μ' to stochastically dominate $N(v, L)$ by a larger geometry.

DEFINITION 4.2 (Augmented distribution). Let μ be a probability distribution on \mathbb{N} with an exponential tail. Let $k_0 = \max\{k : \sum_{j \geq k} \sqrt{p_j} \geq 1/2\}$, and $k_{\max} := \max\{k : p_k > 0\}$, with $k_{\max} = +\infty$ if the maximum does not exist. When $k_0 < k_{\max}$, we define the augmented distribution μ^\sharp of μ by

$$\mu^\sharp(j) = \frac{1}{Z} \begin{cases} p_j/2 & \text{if } j \leq k_0; \\ \sqrt{p_j} & \text{if } j > k_0, \end{cases}$$

where $Z = \sum_{j \leq k_0} p_j/2 + \sum_{j > k_0} \sqrt{p_j}$. If $k_0 = k_{\max}$, then we let

$$\mu^\sharp(j) = \frac{1}{Z} \begin{cases} p_j/2 & \text{if } j < k_0; \\ \sqrt{p_j} & \text{if } j = k_0, \end{cases}$$

where $Z = \sum_{j < k_0} p_j/2 + \sqrt{p_{k_0}}$.

We observe some of the basic properties of augmented distributions in the following lemma. The proof is based on elementary applications of estimating large deviation events and is postponed to Appendix (Section A.1) since it is a bit technical and less related with the main theme of the work.

LEMMA 4.3. *Let μ be a probability distribution on \mathbb{N} :*

- (1) *If μ has an exponential tail, then so does μ^\sharp .*

(2) Let D_1, \dots, D_n be n independent samples of μ . For a subset $\Delta \subset [n]$, let $\{p_k^\Delta\}_k$ denote the empirical distribution of $\{D_i\}_{i \in [n] \setminus \Delta}$. With high probability over the choice of D_i 's, $\{p_k^\Delta\}_k$ is stochastically dominated by μ^\sharp , for any $\Delta \in [n]$ with $|\Delta| \leq n/3$.

REMARK 4.4. The i.i.d D_i in the second condition of Lemma 4.3 can be viewed as a degree sequence of $G \sim \mathcal{G}(n, \mu)$. Consider the exploration procedure starting from a single fixed vertex v which, at each step, reveals a vertex adjacent to the current explored neighborhood and the half-edges incident to the new vertex. Then, the second statement says that, when the exploration process revealed $N \leq n/3$ vertices inside the local neighborhood of v , the empirical degree distribution of the $n - N$ unexplored vertices is stochastically dominated by μ^\sharp with high probability.

Using the above properties of augmented distributions, we develop a coupling argument to dominate $N(v, L) \subset G$ by a Galton–Watson type branching process. To this end, we first take account of the effect of emerging cycles in $N(v, L)$.

For a constant $\gamma > 0$, let $\mathcal{A}(\gamma)$ be the event that $N(v, \gamma \log n)$ in G contains at most one cycle for all $v \in G$. The following lemma shows that we typically have $\mathcal{A}(\gamma)$ for some constant γ .

LEMMA 4.5. *There exists $\gamma = \gamma(\mu) > 0$ such that for $G \sim \mathcal{G}(n, \mu)$, $\mathbb{P}(G \in \mathcal{A}(\gamma)) = 1 - o(1)$.*

This is a well-known property that holds true in general for various types of random graphs. Our proof of this lemma will be very similar to that of Lemma 2.1 in [19]. However, it is more technical, due to generality of the model, and hence we postpone the proof to Section A.2.

Fix a constant $\gamma_1 > 0$ satisfying the condition in Lemma 4.5, and let $\mathcal{A} = \mathcal{A}(\gamma_1)$ for convenience. In the following we define two Galton–Watson type branching processes which are used to stochastically dominate $N(v, \gamma_1 \log n)$.

DEFINITION 4.6 (Galton–Watson-on-cycle process). Let s, L be positive integers with $s \geq 2$, and let ξ be a probability distribution on \mathbb{N} . We define the *Galton–Watson-on-cycle process* (in short, *GWC-process*), denoted by $\text{GWC}(\xi; s)_L$, as follows:

1. Let C be a cycle of length s , and distinguish one vertex as the root ρ .
2. On C we add $(s - 1)$ independent $\text{GW}(\xi)_L$ trees, each rooted at a vertex of C except for ρ .

The vertex ρ is called the root of $\text{GWC}(\xi; s)_L$.

DEFINITION 4.7 (Edge-added Galton–Watson process). Let l, s, L be positive integers with $s \geq 2$ and $l \leq L$, and let ξ be a probability distribution on \mathbb{N} . We define $\text{EGW}(\xi; l, s)_L$, the *edge-added Galton–Watson process* (in short, *EGW-process*), as follows:

1. Generate a $\text{GW}(\xi)_L$ tree, conditioned on survival until depth l .
2. At each vertex v at depth l , add an independent $\text{GWC}(\xi; s)_{L-l}$ process rooted at v . Here, we preserve the existing subtrees from v .

Let ξ' be another probability measure on \mathbb{N} . Then, $\text{EGW}(\xi, \xi'; l, s)_L$ denotes the EGW-process where the root has degree distribution ξ , and all the descendants have ξ' . Here, we also add $\text{GWC}(\xi'; s)_{L-l}$ in the second step of the definition.

We now develop an argument showing that the local neighborhood $N(v, L)$ is dominated by a combined law of EGW-processes. In what follows, for two probability measures ν_1 and ν_2 on graphs, we say ν_1 *stochastically dominates* ν_2 and write $\nu_1 \geq_{st} \nu_2$ if there exists a coupling between $S_1 \sim \nu_1$ and $S_2 \sim \nu_2$ such that $S_2 \subset S_1$ in terms of isomorphic embeddings of graphs, that is, there exists an injective graph homomorphism from S_2 into S_1 .

Fix a vertex $v \in G$, and consider its local neighborhood $N(v, L_n)$ where $L_n = \gamma_1 \log n$ with γ_1 as in Lemma 4.5. For each l, s with $s \geq 2$, we define the event $\mathcal{B}_{l,s}(v)$ to be the subevent of \mathcal{A} such that, in addition to \mathcal{A} , $N(v, L_n)$ forms a cycle of length s at distance l from v .

For the given degree distribution μ , let μ' be its size-biased distribution, and $\tilde{\mu} := \mu'_{[1, \infty)}$ denote the distribution μ' conditioned on being in the interval $[1, \infty)$. Let μ^\sharp and $\tilde{\mu}^\sharp$ be the augmented distributions of μ and $\tilde{\mu}$, respectively. Further, let $\eta, \eta_{l,s}$ and η_0 denote the probability measures on rooted graphs describing the laws of $N(v, L_n)$, $\text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_{L_n}$ and $\text{GW}(\mu^\sharp, \tilde{\mu}^\sharp)_{L_n}$, respectively.

LEMMA 4.8. *Under the above setting and for a fixed vertex $v \in G$, we have the following stochastic domination:*

$$\eta \mathbf{1}_{\mathcal{A}} \leq_{st} \sum_{l,s:s \geq 2} b_{s,l} \eta_{s,l} + b_0 \eta_0,$$

where $b_{l,s} = \mathbb{P}(\mathcal{B}_{l,s}(v))$, $b_0 = 1 - \sum_{l,s} b_{s,l}$.

PROOF. We study $N(v, L_n)$, from an exploration procedure point of view, in terms of the breadth-first search algorithm. Initially, before exploring anything we have n vertices with each of them having i.i.d. μ half-edges. The term “explore” means that we match a pair of half-edges and form an edge between their endpoint vertices. For convenience, we initially impose an arbitrary ordering on all half-edges before exploring anything. We consider the following exploration procedure:

- We start from the single vertex v and the half-edges adjacent to it.
- Suppose that we explored up to depth- t neighborhood of v . Let $\partial N(v, t)$ denote the unmatched half-edges on the boundary of $N(v, t)$, and we explore the half-edges in $\partial N(v, t)$ one by one, respecting the aforementioned ordering. During the (t, i) th exploration step for $1 \leq i \leq |\partial N(v, t)|$, the i th half-edge in $\partial N(v, t)$ is paired with a uniformly random unexplored half-edge.

Let $N(v, t; i)$ denote the explored neighborhood until (t, i) th exploration step. Also, let $H_t = |\partial N(v, t)|$. During the (t, i) th exploration step, the i th half-edge of $\partial N(v, t)$ seeks for its uniformly random pair from the unexplored half-edges. Therefore, if we have yet explored fewer than $\frac{n}{3}$ vertices, then, after pairing a half-edge, the number of newly added half-edges to $N(v, t; i)$ from $N(v, t; i - 1)$ is stochastically dominated by $\tilde{\mu}^\sharp$, due to Lemma 4.3. This implies that, conditioned on the event that $N(v, t; i)$ does not contain any cycles, $N(v, t; i)$ is stochastically dominated by $\mathcal{T}_{t,i}$, where $\mathcal{T}_{t,i}$ is generated by adding new offsprings according to $\tilde{\mu}^\sharp$ to i vertices of depth t to the Galton–Watson tree $\mathcal{T}_t \sim \text{GW}(\mu^\sharp, \tilde{\mu}^\sharp)_t$.

Define (T, I) to be the index of the exploration step when a cycle is formed. In other words, the I th half-edge in $\partial N(v, T)$ is either paired to a j th half-edge in $\partial N(v, T)$ for some $j > I$ or to one of the newly explored half-edges during the (T, k) th exploration step for some $k < I$. Note that on the event \mathcal{A} , either the unique (T, I) exists or it does not exist up to exploring $N(v, L_n)$.

Suppose that there exists unique valid (T, I) . Let C be the cycle formed at this step and $v(C)$ be the vertex in C that is closest to v . Up to the $(T, I - 1)$ th exploration step, we

can stochastically dominate $N(v, T; I - 1)$ by $\mathcal{T}_{T, I-1}$ as mentioned above. Let $w(C)$ be the vertex in $\mathcal{T}_{T, I-1}$ corresponding to $v(C)$ via an isomorphic embedding of $N(v, T; I - 1)$ into $\mathcal{T}_{T, I-1}$. At (T, I) th exploration step, we add $\mathcal{S} \sim \text{GWC}(\tilde{\mu}^\sharp; |C|)_{L_n}$ at $w(C)$. Note that this GWC-process \mathcal{S} can be coupled with C and its descendants in $N(v, L_n)$ in the sense that each Galton–Watson subtree hanging to the cycle of \mathcal{S} stochastically dominates the corresponding subtree in $N(v, L_n)$ hanging to C .

Let l denote the distance from v to $v(C)$. Completing the rest of the exploration as discussed above, $N(v, L_n)$ is stochastically dominated by $\text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, |C|)_{L_n}$, given that there exists the unique valid (T, I) . This implies that on the event \mathcal{A} , $N(v, L_n)$ is stochastically dominated by a combined law of EGW-processes, and, in this combination, the probability mass of appearance of $\text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_{L_n}$ should be $b_{l,s} = \mathbb{P}(\mathcal{B}_{l,s})$. This concludes the proof of the claimed result. \square

4.2. *Estimating the survival time.* Thanks to Lemma 4.8, we now study the contact process on edge-added Galton–Watson processes. On such graphs, we first show that the expected survival time of the contact process is bounded by a constant, when the infection rate is small enough, as an analog of Theorem 3.1.

PROPOSITION 4.9. *Let l, s, L be any integers such that $s \geq 2$ and $L \geq l$. Let $R_{l,s,L}$ denote the first time when $\text{CP}^\lambda(\mathcal{S}; \mathbf{1}_\rho)$ reaches at state $\mathbf{0}$, where $\mathcal{S} \sim \text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_L$ rooted at ρ . Then, there exist constants $C, \lambda_0 > 0$, depending only on μ , such that for all $\lambda \leq \lambda_0, s, l$ and L , we have $\mathbb{E}R_{l,s,L} \leq C$.*

REMARK 4.10. Since the coupling given in Lemma 4.8 only works until depth $L_n = \gamma_1 \log n$, we need to show that the contact process on edge-added Galton–Watson process does not go deeper than $\gamma_1 \log n$ with probability $1 - o(n^{-1})$. Note that the $o(n^{-1})$ error is needed when applying a union bound over all vertices in order to translate our results to $G \sim \mathcal{G}(n, \mu)$. This will be done in the next section based on Theorem 3.4.

To establish Proposition 4.9, we develop a recursive argument on both s and L to deduce an analog of Lemma 3.3 for GWC- and EGW-processes. The idea will be similar to that of Lemma 3.3 which is to utilize the notion of *root-added contact process* (Definition 3.2). We first extend the result of Lemma 3.3 to the case of GWC-processes: in the following lemma we estimate the time that the *root-added* contact process $\text{CP}_\rho^\lambda(\mathcal{S})$ reaches $\mathbf{0}$, where $\mathcal{S} \sim \text{GWC}(\tilde{\mu}^\sharp; s)_L$ (note that the state space is now $\{0, 1\}^{\mathcal{S} \setminus \{\rho\}}$). Here, we fix the root ρ of \mathcal{S} to be the permanently infected parent. There is a slight difference from the previous root-added contact processes considered in Lemma 3.3, since now the permanently infected parent has two children rather than one. We pick a child v of ρ and study $\text{CP}_\rho^\lambda(\mathcal{S}; \mathbf{1}_v)$.

LEMMA 4.11. *Let s, L be any integers with $s \geq 2$, and let $\mathcal{S} \sim \text{GWC}(\tilde{\mu}^\sharp; s)_L$ be a GWC-process rooted at ρ , with $\tilde{\mu}^\sharp$ as before. Let v be any neighbor of ρ , and let $S_{s,L}$ denote the first time when $\text{CP}_\rho^\lambda(\mathcal{S}; \mathbf{1}_v)$ reaches at state $\mathbf{0}$. Then, there exists a constant $\lambda_0 > 0$, depending only on μ , such that for any $\lambda \leq \lambda_0$ and s, L , $\mathbb{E}S_{s,L} \leq 2e$.*

PROOF. Let us first study the case of $s \geq 3$. We will develop an inductive argument on s , similarly as in Lemma 3.3. Let $(X_t) \sim \text{CP}_\rho^\lambda(\mathcal{S}; \mathbf{1}_v)$, and consider the modified version $(\tilde{X}_t) \sim \tilde{\text{CP}}_{\rho;v}^\lambda(\mathcal{S}; \mathbf{1}_v)$ of (X_t) , defined as:

1. (\tilde{X}_t) is coupled with (X_t) in the sense that they share the same locations of recoveries and infections. In particular, ρ is infected in (\tilde{X}_t) .

2. In (\tilde{X}_t) , the recovery at v at time s is valid if and only if $\tilde{X}_s = \mathbf{1}_v$. Otherwise, we ignore the recovery at v . In other words, when there exists an infected vertex other than ρ and v , the recovery at v is neglected.

The modified process $\tilde{\text{CP}}_{\rho;v}^\lambda(\mathcal{S}; \mathbf{1}_v)$ plays the same role as the $\tilde{\text{CP}}$ -process introduced in the proof of Lemma 3.3, which we now detail. Let $D \sim \tilde{\mu}^\sharp$ to satisfy $\deg(v) = D + 2$, and let u_1, \dots, u_D be the neighbors of v which are not on the cycle of \mathcal{S} . Let \mathcal{T}_{u_i} denote the subtrees branching from u_i , which has the law of i.i.d $\text{GW}(\tilde{\mu}^\sharp)_{L-1}$, and regard $v \in \mathcal{T}_{u_i}^+$ as the permanently infected parent of u_i . Further, call $\mathcal{S}' = \mathcal{S} \setminus \bigcup_{i=1}^D \mathcal{T}_{u_i}$, and define $\text{CP}_{\rho,v}^\lambda(\mathcal{S}')$ to be the contact process on \mathcal{S}' in which ρ and v are set to be infected permanently. As we run the process (\tilde{X}_t) from $t = 0$:

A. The second state of \tilde{X}_t is $\mathbf{0}$ with probability $\frac{1}{1+(D+2)\lambda}$. Here, $D + 2$ comes from $D + 1$ possible new infections from v , and one possible infection from ρ to its child other than v . When this happens, the expected waiting time until the transition to $\mathbf{0}$ is $\frac{1}{1+(D+2)\lambda}$.

B. Otherwise, $\{\rho, v\}$ infects a uniformly random neighbor U before v is healed, and (\tilde{X}_t) then can be regarded as a product chain of $\text{CP}_{\rho,v}^\lambda(\mathcal{S}')$ and $\{\text{CP}_v^\lambda(\mathcal{T}_{u_i}^+) : i \in [D]\}$ with initial state $\mathbf{1}_U$, until \tilde{X}_t returns back to $\mathbf{1}_v$. Denote this product chain by $\text{CP}_{\rho;v}^\otimes(\mathcal{S})$ (whose state space is $\{0, 1\}^{\mathcal{S} \setminus \{\rho, v\}}$). Here, U can be thought of the first infected vertex besides $\{\rho, v\}$ in $\text{CP}_{\rho;v}^\otimes(\mathcal{S}; \mathbf{0})$.

Let $\tilde{S}_{s,L}$ be the first time that $\tilde{\text{CP}}_{\rho;v}^\lambda(\mathcal{S}; \mathbf{1}_v)$ becomes $\mathbf{0}$, and let S^\otimes denote the first time that $\text{CP}_{\rho;v}^\otimes(\mathcal{S}; \mathbf{1}_U)$ reaches all-healthy state except ρ, v . Then, similarly as in Lemma 3.3, the above reasoning implies that

$$\begin{aligned} \mathbb{E}[\tilde{S}_{s,L} | \mathcal{S}] &= \sum_{k=0}^{\infty} \left(\frac{(D+2)\lambda}{1+(D+2)\lambda} \right)^k \frac{1}{1+(D+2)\lambda} \\ (4.1) \quad &\times \left[\frac{k+1}{1+(D+2)\lambda} + k\mathbb{E}[S^\otimes | \mathcal{S}] \right] \\ &= 1 + (D+2)\lambda\mathbb{E}[S^\otimes | \mathcal{S}]. \end{aligned}$$

Therefore, we have

$$(4.2) \quad \mathbb{E}[\tilde{S}_{s,L} | D] = 1 + (D+2)\lambda\mathbb{E}[S^\otimes | D].$$

Now, we take account of the stationary distributions of the above processes to obtain the conclusion. Let π^\otimes, π' and π_i be the stationary distributions of $\text{CP}_{\rho;v}^\otimes(\mathcal{S}), \text{CP}_{\rho,v}^\lambda(\mathcal{S}')$ and $\text{CP}_v^\lambda(\mathcal{T}_{u_i}^+)$, respectively. Then, clearly, $\pi^\otimes = (\otimes_{i=1}^D \pi_i) \otimes \pi'$. We can relate these objects with the running times similarly as (3.3), (3.4), by:

$$\begin{aligned} (4.3) \quad \pi^\otimes(\mathbf{0}) &= \frac{1}{1+(D+2)\lambda\mathbb{E}[S^\otimes | \mathcal{S}]}; \\ \pi_i(\mathbf{0}) &= \frac{1}{1+\lambda\mathbb{E}[S_{L-1} | \mathcal{T}_{u_i}]}; \\ \pi'(\mathbf{0}) &= \frac{1}{1+2\lambda\mathbb{E}[S_{s-1,L} | \mathcal{S}']}, \end{aligned}$$

where S_{L-1} denotes the first time when $\text{CP}_v^\lambda(\mathcal{T}_{u_i}^+; \mathbf{1}_{u_i})$ becomes $\mathbf{0}$ and $S_{s-1,L}$ is the time it takes for $\text{CP}_{\rho,v}^\lambda(\mathcal{S}'; \mathbf{0})$ to return to $\mathbf{0}$ after the first infection, besides $\{\rho, v\}$, occurs. Note that the existence of the infection other than $\{\rho, v\}$ in $\text{CP}_{\rho,v}^\lambda(\mathcal{S}'; \mathbf{0})$ is guaranteed by the condition

$s \geq 3$. Also, notice that \mathcal{S}' can be regarded as $\mathcal{S}'' \sim \text{GWC}(\tilde{\mu}^\sharp; s - 1)_L$, since the processes $\text{CP}_{\rho, v}^\lambda(\mathcal{S}'; \mathbf{1}_w)$ (with w being a neighbor of $\{\rho, v\}$ in \mathcal{S}') and $\text{CP}_{\rho'}^\lambda(\mathcal{S}''; \mathbf{1}_{v'})$ for the root ρ' and one of its neighbor v' of \mathcal{S}'' share the same law. This implies that the notation $S_{s-1, L}$ in (4.3) matches with the definition of it given in the statement of the lemma.

Therefore, combining (4.2) and (4.3) gives that

$$(4.4) \quad \mathbb{E}[S_{s, L} \mid D] \leq \mathbb{E}[\tilde{S}_{s, L} \mid D] \leq (1 + 2\lambda\mathbb{E}[S_{s-1, L}]) (1 + \lambda\mathbb{E}[S_{L-1}])^D.$$

Since we already have a bound for $\mathbb{E}[S_{L-1}]$ due to Lemma 3.3, we deduce the desired result by manipulating (4.4), as in the final step of the proof of Lemma 3.3. Namely, for any $\lambda \leq \min\{\lambda_0, (4e)^{-1}\}$ with λ_0 given as in Lemma 3.3, (4.4) gives us that

$$\mathbb{E}[S_{s, L}] \leq e(1 + 2\lambda\mathbb{E}[S_{s-1, L}]),$$

and hence $\mathbb{E}[S_{s-1, L}] \leq 2e$ implies $\mathbb{E}[S_{s, L}] \leq 2e$.

The case $s = 2$ is simpler, since $\text{GWC}(\tilde{\mu}^\sharp; 2)_L$ is the same as the law of \mathcal{T}_L^+ for $\mathcal{T}_L \sim \text{GW}(\tilde{\mu}^\sharp)_L$, except that the parent ρ^+ of ρ in \mathcal{T}_L^+ is now connected with ρ by two edges. From the contact process point of view, this means that the intensity of infection from ρ^+ to ρ is 2λ and everything else is identical to the case of $\text{GW}(\tilde{\mu}^\sharp)_L$. Hence, the same proof as Lemma 3.3 can be replicated, and we obtain that there exists a constant $\lambda_0 > 0$ such that $\mathbb{E}[S_{2, L}] \leq 2e$ for all $\lambda \leq \lambda_0$. We leave the details of the proof to the reader. \square

PROOF OF PROPOSITION 4.9. It can be proven by the same way as Lemma 4.11. For completeness, we present the proof in the Appendix, Section A.3. \square

4.3. *Proof of Theorem 4.1.* Let l, s, L be arbitrary integers with $s \geq 2$ and $L \geq l$, and consider an edge-added Galton–Watson process $\mathcal{S} \sim \text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_L$.

We can extend the result of Theorem 3.4 and Proposition 3.6 to the case of edge-added Galton–Watson processes. The method will be the same as Proposition 4.9 and Lemma 4.11, appropriately adjusted to the current setting of *delayed* contact process (Definition 3.5). We state the result in the following lemma, whose proof is deferred to the Appendix (Section A.4), since it is similar to the previous proofs but more technical.

LEMMA 4.12. *Let $\mathcal{S} \sim \text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_L$ and v_S^θ be the stationary distribution of $\text{DP}_{\rho^+}^{\lambda, \theta}(\mathcal{S}^+)$ on the space $\{0, 1\}^{\mathcal{S}}$. Then, there exist constants $C, \lambda_0 > 0$, depending only on μ , such that for all $\lambda \leq \lambda_0$, we have $\mathbb{E}[v_S^\theta(\mathbf{0})^{-1}] \leq 2$ for $\theta = C\lambda$.*

Based on Proposition 3.6 and Lemma 4.12, we have an analog of Theorem 3.4 for EGW-processes. Thus, we can complete the proof of Theorem 3(1) by combining the previous results to build up a coupling between the contact processes on local neighborhood $N(v, L)$ and on EGW-processes.

PROOF OF THEOREM 4.1. Let $G \sim \mathcal{G}(n, \mu)$ for any large enough n , and let $\gamma_1 > 0$ be the constant satisfying Lemma 4.5. For each $v \in G$, let \mathcal{A}_v be the event that $N_v := N(v, \gamma_1 \log n)$ in G contains at most one cycle. Then, the proof of Lemma 4.5 tells us $\mathbb{P}(\mathcal{A}_v) \geq 1 - o(n^{-1})$.

Let λ'_0 be the minimum between the λ_0 's given by Proposition 4.9 and Lemma 4.12, C be as in 4.12 and $\theta = C\lambda$ for $\lambda \leq \lambda'_0$. Further, let $(X_t) \sim \text{CP}^\lambda(N_v; \mathbf{1}_v)$ and $H_v := \max\{r(X_t) : t \geq 0\}$, where $r(X_t)$ denotes the maximal depth among the infected sites in X_t . Note that $r(X_t)$ stays 0 after X_t becomes $\mathbf{0}$.

Define the event \mathcal{B}_v as

$$\mathcal{B}_v := \{\mathbb{P}(H_v \geq h \mid N_v) \leq 2n^2(C\lambda)^h, \text{ for all } h\}.$$

Following the same proof as Theorem 3.4 based on Lemma 4.12, we have $\mathbb{P}(\mathcal{B}_v) \geq 1 - n^{-2}$ by dominating N_v by the EGW-processes as Lemma 4.8. Assume that $(C\lambda'_0)^{\gamma_1 \log n} \leq n^{-4}$ by making λ'_0 smaller if needed. Then, the event \mathcal{C}_v given by

$$\mathcal{C}_v = \mathcal{C}_v(G) := \{H_v < \gamma_1 \log n\}$$

satisfies $\mathbb{P}(\mathcal{C}_v | \mathcal{A}_v \cap \mathcal{B}_v) = O(n^{-2})$. Note that $\mathcal{A}_v \cap \mathcal{B}_v$ is an event over the random graph G , while \mathcal{C}_v is an event over the contact process (X_t) given the graph G . By the aforementioned coupling of N_v and the EGW-process, Proposition 4.9 gives that

$$(4.5) \quad \mathbb{E}[T_v | G \in \mathcal{A}_v \cap \mathcal{B}_v, \mathcal{C}_v] \leq B$$

for some constant $B = B(\mu) > 0$.

Define the events

$$\mathcal{E}_1 := \bigcap_{v \in G} (\mathcal{A}_v \cap \mathcal{B}_v) \quad \text{and} \quad \mathcal{E}_2(G) := \bigcap_{v \in G} \mathcal{C}_v.$$

Then, the above discussion shows that $\mathbb{P}(G \in \mathcal{E}_1) = 1 - o(1)$ and $\mathbb{P}((X_t) \in \mathcal{E}_2(G) | G \in \mathcal{E}_1) = 1 - o(1)$, and hence (4.5) holds the same given \mathcal{E}_1 and $\mathcal{E}_2(G)$, namely,

$$\mathbb{E}[T_v | G \in \mathcal{E}_1, \mathcal{E}_2(G)] \leq B,$$

under a possible modification of B if needed. Therefore, by linearity of expectation, summing the above over all $v \in G$ gives the conclusion. \square

PROOF OF COROLLARY 5(1). The statement follows immediately from the contiguity of $\mathcal{G}_{\text{cf}}(n, \text{Pois}(d))$ and $\mathcal{G}_{n,d/n}$ ([14], Theorem 1.1). To be precise, for any subset A_n of graphs with n vertices,

$$\mathbb{P}_{G \sim \mathcal{G}_{\text{cf}}(n, \mu)}(G \in A_n) \rightarrow 0 \quad \text{implies} \quad \mathbb{P}_{G \sim \mathcal{G}_{n,d/n}}(G \in A_n) \rightarrow 0,$$

where $\mu = \text{Pois}(d)$. Since the statement is true whp for $G \sim \mathcal{G}_{\text{cf}}(n, \mu)$, the configuration model, it is also true whp for $G \sim \mathcal{G}_{n,d/n}$. \square

5. Long survival in random graphs: Proof of Theorem 3, part 2.

5.1. A structural lemma. Our main tool to prove long survival time is the following structural lemma whose proof is deferred to Section 7. As mentioned in Section 1.2, we show that the random graph G_n contains a large (α, R) -embedded expander. Once some subset of this expander is infected, it is likely to spread the infection over its R -neighborhood whose size more than doubles the original subset. We define an embedded expander as follows.

DEFINITION 5.1 (Embedded expander). For two positive numbers α and R , we say that a subset of vertices W_0 is an (α, R) -embedded expander of G_n if for every subset $A \subset W_0$ with $|A| \leq \alpha|W_0|$, we have

$$(5.1) \quad |N(A, R) \cap W_0| \geq 2|A|,$$

where $N(A, R)$ is the collection of all vertices in G_n of distance at most R from A .

The following lemma concerns the existence of such an (α, R) -embedded expander in the random graph G_n .

LEMMA 5.2. *Suppose that μ satisfies (1.1) and there exists some constant $c > 0$ such that $\mathbb{E}_{D \sim \mu} e^{cD} < \infty$. Let $G \sim \mathcal{G}(n, \mu)$. There exist positive constants α, β, R, j such that the following holds whp. There exist a subgraph \bar{G}_n of G_n whose maximal degree is at most $2j$ and an (α, R) -embedded expander W_0 of \bar{G}_n with $|W_0| \geq \beta n$.*

5.2. *Proof of Theorem 3, part 2.* We first make a simple observation.

LEMMA 5.3. *Let R be a positive integer constant. Consider the contact process Y_t with infection rate λ on a path of length at most R connecting two vertices v and u . Then, there exist positive constants C and λ_0 , depending only on R , such that for all $\lambda \geq \lambda_0$, we have*

$$\mathbb{P}(u \in Y_{t+C} \mid v \in Y_t) \geq \frac{3}{4}.$$

PROOF. Let C be a sufficiently large constant compared to R . Let \mathcal{A} be the event that the infection on $N(v, 1)$ survives in the entire time interval $[t, t + C]$.

By [5], Lemma 1.1, for sufficiently large λ compared to C ,

$$\mathbb{P}(\mathcal{A} \mid v \in Y_t) \geq \frac{99}{100}.$$

Since there is a path of length at most R from v to u , by [5], Lemma 2.4,

$$\mathbb{P}(u \in Y_{t''} \text{ for some } t'' \in [t', t' + R] \mid N(v, 1) \cap Y_{t'} \neq \emptyset) \geq \frac{1}{e^{6R}},$$

and so, as C is large compared to R , we have

$$\begin{aligned} &\mathbb{P}(u \in Y_{t'} \text{ for some } t' \in [t, t + C + R] \mid v \in Y_t) \\ &\geq \mathbb{P}\left(\mathbf{Bin}\left(\frac{C}{R}, \frac{1}{e^{6R}}\right) \geq 1\right) - \frac{1}{100} \geq \frac{98}{100}. \end{aligned}$$

Assume that $u \in Y_{t'}$ for some $t' \in [t, t + C + R]$. Fix a neighbor u' of u . By [5], Lemma 1.1, again, the contact process on the edge (u, u') survives in the entire interval $[t', t' + 3C]$ with probability at least $\frac{99}{100}$. Since $t + C + R \in [t', t' + 2C]$, there is at least one clock ring in $[t + C + R, t' + 3C]$, and the last clock ring before time $t + 3C$ is an infection clock from u' to u , rather than the recovery clock at u , with probability at least $\frac{99}{100}$. If u' is already infected at that time, u will be infected. Otherwise, u has already been infected and remains infected. In either case, u is infected at time $t + 3C$ with probability at least $\frac{96}{100} \geq \frac{3}{4}$. By replacing C by $C/3$, we complete the proof. \square

Let $\alpha, \beta, j, R, \bar{G}_n$ and W_0 as in Lemma 5.2. It suffices to show that the contact process (X_t) on \bar{G}_n with all vertices infected initially survives for $e^{\Theta(n)}$ -time with probability at least $1 - e^{-\Omega(n)}$ over the contact process. For the rest of this proof, all the vertices, edges, paths and balls are of \bar{G}_n , unless otherwise noted.

Let $X_t^0 = X_t \cap W_0$ be the collection of infected vertices of W_0 at time t . We show that, thanks to the expander property of W_0 , with very high probability, after some time C , the number of infected vertices in W_0 increases.

LEMMA 5.4. *Let C and λ_0 be the constants in Lemma 5.3. There exists a positive constant C' , depending only on j and R , such that for all $\lambda \geq \lambda_0$ and for every integer $a \in (0, \alpha\beta n]$,*

$$(5.2) \quad \mathbb{P}\left(|X_{t+C}^0| \leq \frac{5}{4}a \mid |X_t^0| = a\right) \leq 2 \exp\left(-\frac{a}{C'}\right).$$

PROOF. We will use Azuma's inequality. Let $G_{n,t}$ be the induced subgraph of \bar{G}_n on the set $\bigcup_{v \in X_t^0} N(v, R)$. Let $(\hat{X}_{t'})_{t' \in [t, t+C]}$ be the contact process on $G_{n,t}$ with $\hat{X}_t := X_t^0$ (so $(\hat{X}_{t'})$ only uses the infection and recovery clocks of vertices and edges inside $G_{n,t}$). Let

$\hat{X}_{t'}^0 = \hat{X}_{t'} \cap W_0$. Let \mathcal{X} be the infected vertices u of \hat{X}_{t+C}^0 such that there exists $v \in X_t^0$ and a directed path of infection on the graphical representation of $(\hat{X}_{t'})$ from (v, t) to $(u, t + C)$, and the vertices of the path lie entirely in $B(v, R)$. We have

$$\mathcal{X} \subset \hat{X}_{t+C}^0 \subset X_{t+C}^0.$$

Since the maximal degree in \bar{G}_n is at most $2j$, the number of vertices, denoted by a' , in $G_{n,t}$ is at most $a(2j)^{R+1}$. Enumerate the vertices in $G_{n,t}$ by $v_1, \dots, v_{a'}$. For each $i = 0, 1, 2, \dots, a'$, let \mathcal{F}_i be the σ -algebra generated by the randomness of the recovery clocks and infection clocks during time $(t, t + C]$ on the vertices v_1, \dots, v_i and edges connecting them. Let

$$\mathcal{X}_i := \mathbb{E}(|\mathcal{X}| \mid \mathcal{F}_i, X_t^0, |X_t^0| = a).$$

We have $\mathcal{X}_{a'} = \mathcal{X}$. By Azuma’s inequality we have, for every $s > 0$,

$$(5.3) \quad \mathbb{P}(|\mathcal{X}_{a'} - \mathcal{X}_0| \geq s) \leq 2 \exp\left(-\frac{s^2}{2a'K^2}\right),$$

where

$$K := \max_j \|\mathcal{X}_j - \mathcal{X}_{j-1}\|_\infty \leq |B(v_{i+1}, R)| \leq (2j)^{R+1}.$$

From Lemma 5.3 and the expander properties of W_0 as in Lemma 5.2, we obtain

$$(5.4) \quad \mathcal{X}_0 = \mathbb{E}(|\mathcal{X}| \mid |X_t^0| = a) \geq \frac{3}{4}|N(X_t^0, R) \cap W_0| \geq \frac{3a}{2}.$$

Thus, by (5.3) for $s = \frac{a}{4}$ and the fact that $|\mathcal{X}_{a'}| \leq |X_{t+C}^0|$, we obtain (5.2). \square

PROOF OF THEOREM 3(2). For the lower bound on survival time, initially, all vertices in W_0 are infected so $|X_0^0| \geq \beta n$. Let t_1 be the first time that $|X_{t_1}^0| = \alpha\beta n$. At time $t_2 = t_1 + C$, we have $|X_{t_2}^0| \geq \frac{5}{4}\alpha\beta n$ with probability at least $1 - \frac{2}{m^2}$ where $m := \exp(\frac{\alpha\beta n}{2C'})$ by Lemma 5.4. Let t_3 be the first time after t_2 that $|X_{t_3}^0| = \alpha\beta n$ again. Repeating this process m times, we get that the contact process survives until time mC with probability at least $1 - 2/m$ by the union bound, proving the lower bound for Theorem 3(2).

As for the upper bound, observe that, for any time t , the probability that the contact process dies out during the time interval $[t, t + 1]$ is at least the probability that, for each vertex v in G_n , at least one of the infection clocks from a neighbor u of v to v or the recovery clock at v rings in $[t, t + 1]$ and the last clock rings before time $t + 1$ is the recovery clock at v . Thus, the probability that the process dies out during $[t, t + 1]$ is at least $\prod_v \frac{c}{\deg(v)}$. By Cauchy–Schwarz inequality and the fact that whp, the total degrees in G_n is $O(n)$, we have $\prod_v \frac{c}{\deg(v)} \geq e^{-c'n}$ for some small constants c, c' . Therefore, whp, the contact process, dies out before time $e^{2c'n}$. \square

REMARK 5.5. To prove the corresponding result (Remark 1.2) for μ having an exponential tail, when initially, there is only one uniformly chosen vertex v infected in G_n , observe that with positive probability over the choice of v , v belongs to W_0 . Thus, it suffices to condition on this event and show that with positive probability over the contact process, the process survives until time e^{cn} for some constant c . Let $\lambda_{0,R}, C_R$ and C'_R be the constants λ_0, C and C' corresponding to R in Lemmas 5.3 and 5.4, respectively. By (5.1), for any bounded number k we have

$$|N(v, kR) \cap W_0| \geq 2^k$$

for sufficiently large n .

We now show that, for sufficiently large λ , at some time there will be a lot of infected vertices in W_0 . This will then allow to take the union bound of the tail probability occurring in Lemma 5.4. Let k be a sufficiently large constant. Since the number of vertices in $N(v, kR)$ is at most $(2j)^{kR+1} = O_{j,k,R}(1)$, there are $O_{j,k,R}(1)$ edges in $N(v, kR)$. Thus, there exist constants $\lambda_{j,k,R}, t_{j,k,R}$ such that, for all $\lambda \geq \lambda_{j,k,R}$, the probability that each vertex in $N(v, kR)$ is infected before time $t_{j,k,R}$ and that there are no recovery clocks ring before time $t_{j,k,R}$ is at least $3/4$. Hence, with probability at least $3/4$, there exists $t_1 \leq t_{j,k,R}$ at which all vertices in $N(v, kR)$ are infected. This implies $|X_{t_1}^0| \geq 2^k \geq (5/4)^k$.

Conditioning on this event and applying Lemma 5.4, we get that with probability at least

$$1 - \sum_{i=k}^{\infty} 2 \exp\left(-\frac{5^i}{C'R 4^i}\right),$$

there exists a time $t \geq t_1$ at which $|X_t^0| \geq \alpha\beta n$. Since k is a sufficiently large constant, this probability is at least $1/2$. Finally, conditioned on this event, the same argument as in the proof of Theorem 3(2) shows that, starting from this t , the contact process survives until time $e^{-\Omega(n)}$ whp. Altogether, the contact process (X_t) starting from v survives until time $e^{-\Omega(n)}$ with probability at least $1/4 - o(1)$ for all $\lambda \geq \max\{\lambda_{0,R}, \lambda_{j,k,R}\}$ as desired.

REMARK 5.6. Corollary 5(2) follows from Theorem 3(2) in the exact same way as we deduced Corollary 5(1) from Theorem 3(1).

6. Long survival in random graphs: Proof of Theorem 4. In this section, following the same strategy as in the proof of Theorem 3(2), we prove Theorem 4. The following structural lemma is an analog of Lemma 5.2 for subexponential distributions.

LEMMA 6.1. *Suppose that μ satisfies (1.1) and $\mathbb{E}_{D \sim \mu} e^{cD} = \infty$ for all $c > 0$. Let $G \sim \mathcal{G}(n, \mu)$. For any $\delta > 0$, there exist $\alpha, \beta, j, R > 0$ with $R \leq \delta j$ such that the following holds whp. There exists a subgraph \bar{G}_n of G_n whose maximal degree is at most $2j$ and an (α, R) -embedded expander W_0 of \bar{G}_n with $|W_0| \geq \beta n$ and $\deg_{\bar{G}_n} w \geq j/2$ for all $w \in W_0$.*

Fix $\lambda > 0$. Let $\alpha, \beta, j, R, \bar{G}_n, W_0$ be as in Lemma 6.1 where $\delta > 0$ is a sufficiently small constant depending on λ . It suffices to show that the contact process (X_t) on \bar{G}_n , with all vertices infected, initially survives for $e^{\Omega(n)}$ -time with probability at least $1 - e^{-\Omega(n)}$ over the contact process. For the rest of the proof, all the vertices, edges, paths and balls are of \bar{G}_n .

Let X_t^0 be the collection of infected vertices of W_0 at time t . We show the following analog of Lemma 5.3.

LEMMA 6.2. *There exists a constant c , depending only on λ , such that for sufficiently large n and for every $u \in N(v, R) \cap W_0$, we have*

$$(6.1) \quad \mathbb{P}(u \in X_{t+2e^c j}^0 \mid v \in X_t^0) \geq \frac{3}{4}.$$

Assuming Lemma 6.2, we prove the following analog of Lemma 5.4.

LEMMA 6.3. *Let c be the constant in Lemma 6.2. There exists a positive constant C_0' depending only on j and R , such that for every integer $a \in (0, \alpha\beta n]$,*

$$(6.2) \quad \mathbb{P}\left(|X_{t+e^c j}^0| \leq \frac{5}{4}a \mid |X_t^0| = a\right) \leq 2 \exp\left(-\frac{a}{C'}\right).$$

The proof of this lemma is identical to the proof of Lemma 5.4. Using this lemma, the proof of Theorem 4 is identical to that of Theorem 3(2). It remains to prove Lemma 6.2.

PROOF OF LEMMA 6.2. Let Γ be a path of length at most R connecting v and u . Since $v \in W_0$, $N_{v,u} := \{v\} \cup (N(v, 1) \setminus \Gamma)$ contains a star with $j/4$ leaves. Let \mathcal{A} be the event that the infection on $N_{v,u}$ survives in the entire time interval $[t, t + e^{cj}]$ for some constant c , depending only on λ .

By [2], Lemma 5.3, by choosing δ sufficiently small in Lemma 6.1 and using the inequality $j \geq R/\delta \geq 1/\delta$, we have

$$\mathbb{P}(\mathcal{A} \mid v \in X_t^0) \geq \frac{99}{100}.$$

Since there is a path of length at most R from v to u , by [5], Lemma 2.4, there exists a constant $c' > 0$, depending only on λ , such that for any time t' ,

$$\mathbb{P}(u \in X_{t'}^0 \text{ for some } t'' \in [t', t' + R + 1] \mid N_{v,u} \cap X_{t'}^0 \neq \emptyset) \geq c'^{R+1}.$$

Thus,

$$\begin{aligned} &\mathbb{P}(u \in X_{t'}^0 \text{ for some } t' \in [t, t + e^{cj} + R + 1] \mid v \in X_t^0) \\ &\geq \mathbb{P}\left(\text{Bin}\left(\frac{e^{cj}}{R}, c'^{R+1}\right) \geq 1\right) - \frac{1}{100} \\ &\geq \mathbb{P}\left(\text{Bin}\left(\frac{e^{cR/\delta}}{R}, c'^{R+1}\right) \geq 1\right) - \frac{1}{100} \geq \frac{98}{100} \end{aligned}$$

by choosing δ sufficiently small (e.g., $\delta = \frac{c}{100 \log c'^{-1}}$) in Lemma 6.1. Since $R \leq \delta j$, we can replace the interval $[t, t + e^{cj} + R + 1]$ in the above inequality by the bigger interval $[t, t + 2e^{cj}]$.

Assume that $u \in X_{t'}^0$ for some $t' \in [t, t + 2e^{cj}]$. By the third inequality in [5], Lemma 2.3, and Markov's inequality, with probability at least $\frac{99}{100}$, there exists a time $t'' \in [t', t' + e^{cj}]$ such that there are at least $\lambda j/4$ neighbors of u infected at time t'' . By [5], Lemma 2.2, with probability at least $\frac{99}{100}$, there are at least $\lambda j/10$ neighbors of u infected at any time in the time interval $[t'', t'' + 3e^{cj}]$. Since $t + 2e^{cj} \in [t'', t'' + 3e^{cj}]$, the probability that u is infected at time $t + 2e^{cj}$ is at least the probability that the last clock rings before time $t + 2e^{cj}$ is an infection clock rather than the recovery clock at u . Since there are at least $\lambda j/10$ neighbors of u infected at any time in the interval $[t'', t + 2e^{cj}]$, the probability of the above event is

$$\frac{\lambda^2 j/10}{\lambda^2 j/10 + 1} \geq \frac{99}{100}.$$

That completes the proof of Lemma 6.2. \square

REMARK 6.4. To prove the corresponding result (Remark 1.2) for subexponential μ , when initially there is only one uniformly chosen vertex v infected in G_n , observe that with positive probability over the choice of v , v belongs to W_0 . Thus, it suffices to condition on this event and show that with positive probability over the contact process, the process survives until time $e^{\Omega(n)}$.

Since W_0 is an (α, R) -embedded expander, observe that for every $A \subset W_0$ with $|A| \leq \alpha|W_0|/4$, we have

$$|N(A, 2R)| \geq 4|A|.$$

Thus, using the same proof as for Lemmas and 6.2 and 6.3, one can see that there exist constants c' and C'' such that for all integer $a \in (0, \alpha\beta n/4]$,

$$(6.3) \quad \mathbb{P}(|X_{t+e^{c'j}}^0| \leq 2a \mid |X_t^0| = a) \leq 2 \exp\left(-\frac{a}{C''}\right).$$

Let k be the largest number such that $2^k \leq \alpha\beta n/4$. Let \mathcal{A}_0 be the event that $X_0^0 = \{v\} \subset W_0$. For each $i = 1, \dots, k$, let \mathcal{A}_i be the event that

$$|X_{ie^{c'j}}^0| \geq 2^i.$$

Let \mathcal{A}^* be the event that the contact process survives up to time $e^{\Omega(n)}$. We want to show that

$$(6.4) \quad \mathbb{P}(\mathcal{A}^* \mid X_0^0 = \{v\}) = \Omega(1).$$

In fact,

$$\begin{aligned} \mathbb{P}(\mathcal{A}^* \mid X_0^0 = \{v\}) &\geq \mathbb{P}\left(\bigcap_{i=1}^k \mathcal{A}_i \cap \mathcal{A}^* \mid X_0^0 = \{v\}\right) \\ &\geq \prod_{i=1}^k \mathbb{P}(\mathcal{A}_i \mid \mathcal{A}_{i-1}) \mathbb{P}(\mathcal{A}^* \mid \mathcal{A}_k). \end{aligned}$$

By (6.3), for each $i = 1, \dots, k$, $\mathbb{P}(\mathcal{A}_i \mid \mathcal{A}_{i-1}) \geq 1 - 2 \exp(-\frac{2^{i-1}}{C''})$. Finally, by the same argument as in the proof of Theorem 3(2), once there are about $\Theta(n)$ vertices in W_0 infected, the contact process survives for an exponentially long time whp. In other words, $\mathbb{P}(\mathcal{A}^* \mid \mathcal{A}_k) = 1 - o(1)$. Let k' be the smallest number such that $2^{k'-1} \geq 10C''$. Combining all of these inequalities, we obtain

$$\begin{aligned} \mathbb{P}(\mathcal{A}^* \mid X_0^0 = \{v\}) &\geq \frac{98}{100} \prod_{i=2}^{k'} \left(1 - 2 \exp\left(-\frac{2^{i-1}}{C''}\right)\right) \\ &\quad \times \prod_{i=k'+1}^k \left(1 - 2 \exp\left(-\frac{2^{i-1}}{C''}\right)\right) \\ &\geq \frac{98}{100} (1 - 4e^{-10}) \prod_{i=2}^{k'} \left(1 - 2 \exp\left(-\frac{2^{i-1}}{C''}\right)\right) = \Omega(1) \end{aligned}$$

as desired. That completes the proof of Remark 1.2.

7. Proof of the structural lemma. In this section we prove the structural Lemmas 5.2 and 6.1. We start by proving Lemma 6.1 for subexponential μ in Section 7.1. The proof of Lemma 5.2 is very similar and is presented in Section 7.2.

7.1. Proof of Lemma 6.1. Step 1. Preprocessing. In this step we eliminate high-degree vertices in G_n so that the degrees become bounded. This will allow us to control the size of the neighborhoods that we explore in the next steps. We prove in Lemma 7.1 that the elimination does not significantly affect relevant parameters of G_n .

Let $b = \frac{\sum_{l=1}^{\infty} l(l-1)\mu(l)}{\sum_{l=1}^{\infty} l\mu(l)} > 1$ be the branching rate of μ . Let d be the mean $d = \mathbb{E}_{D \sim \mu} D > 1$. For a constant j , consider the graph \bar{G}_n obtained from G_n by deleting all vertices with degree at least $2j + 1$ together with their half-edges and their matches. Let \bar{n} be the number of vertices of \bar{G}_n and $0 \leq d_1 \leq \dots \leq d_{\bar{n}} \leq 2j$ be the degree sequence of vertices in \bar{G}_n .

The branching rate \bar{b} of a (deterministic) degree sequence (d_i) is defined to be the branching rate of the empirical measure generated by (d_i) , namely,

$$\bar{b} = \frac{\sum_l l(l-1)\#\{i : d_i = l\}}{\sum_l l\#\{i : d_i = l\}}.$$

Throughout the proof, ε can be any small constant (e.g., $\varepsilon = 1/2$).

LEMMA 7.1 (Eliminating high-degree vertices). *Let j_0, ε be any positive constants with $\varepsilon < 1$. There exists a positive constant $j \geq j_0$ such that the following hold whp:*

1. *Conditioned on the degree sequence $(d_1, \dots, d_{\bar{n}})$, the edges of \bar{G}_n form a uniformly chosen perfect matching of its half-edges.*

2. *The number of vertices and the total degree in \bar{G}_n (which is twice the number of edges of \bar{G}_n) satisfy*

$$\bar{n} \geq (1 - \varepsilon)n \quad \text{and} \quad d_1 + \dots + d_{\bar{n}} \in (1 - \varepsilon, 1 + \varepsilon)nd.$$

3. *The branching rate \bar{b} of the degree sequence of \bar{G}_n satisfies $\bar{b} \in (1 - \varepsilon, 1 + \varepsilon)b$.*

4. *For all $i, 0 \leq d_i \leq 2j$. The number of vertices with large degree is, as expected,*

$$\#\left\{i \in \{1, \dots, \bar{n}\} : d_i \in \left[\frac{j}{2}, 2j\right]\right\} \geq \varepsilon n \mu[j, 2j].$$

To simplify the notation, for the rest of this Section 7 we define

$$u_j := \mu[j, 2j].$$

PROOF. Since the proof of Items 1–3 is rather standard, we defer it to the [Appendix](#), Section A.5. Here, we only prove Item 4. Choose j large enough such that

$$\mathbb{E}_{D \sim \mu} D \mathbf{1}_{D \geq 2j+1} \leq \varepsilon/4.$$

For each vertex $v \in G_n$, consider the random variable

$$X_v := \deg_{G_n}(v) \mathbf{1}_{\deg_{G_n}(v) \geq 2j+1}.$$

These random variables are independent with mean at most $\varepsilon/4$ and variance bounded by the second moment of μ . By Chebyshev’s inequality, whp

$$\sum_{v \in G_n} X_v \leq \varepsilon n/2.$$

Thus, whp, the total number of removed half-edges from vertices of degree in $[0, 2j]$ in G_n , is at most $\sum_{v \in G_n} X_v \leq \varepsilon n/2$.

By Chernof’sf inequality, whp, the number of half-edges of G_n of vertices of degree in $[0, j)$ and $[j, 2j]$, are $(1 - \varepsilon, 1 + \varepsilon)nd$ and $(1 - \varepsilon, 2 + 2\varepsilon)jnu_j$, respectively. Since j is sufficiently large, ju_j is very small compared to d . The first deleted half-edge has probability roughly $\frac{ju_j}{d}$ to be from vertices of degree in $[j, 2j]$. Ideally, one expects to delete at most $\frac{ju_j}{d}\varepsilon nd$ half-edges from these vertices. We will show that it is the case, namely,

CLAIM 7.2. *The number of half-edges deleted from vertices of degree in $[j, 2j]$ is, at most,*

$$10\varepsilon jnu_j \quad \text{whp.}$$

Assuming the claim, the number of vertices originally with degree in G_n in $[j, 2j]$ and with degree less than $j/2$ in \tilde{G}_n is, at most,

$$\frac{10\varepsilon j n u_j}{j/2} = 20\varepsilon n u_j \quad \text{whp.}$$

By Chernof’sf inequality, in G_n the number of vertices of degree in $[j/2, 2j]$ is in $(1 - \varepsilon, 2 + 2\varepsilon)n u_j$ (where we choose j so that $\mu[j/2, j] \leq (1 + \varepsilon)u_j$). Hence, whp, the number of vertices in \tilde{G}_n with degree in $[j/2, 2j]$ is in $(1 - 21\varepsilon, 2 + 2\varepsilon)n u_j$, completing the proof of Item 4. \square

To prove Claim 7.2, we will use the following cut-off line algorithm to find the random matches of the deleted high-degree vertices in G_n .

DEFINITION 7.3 (Cut-off line algorithm). Given a graph G_n in which each vertex v has degree $d_{G_n}(v)$, a perfect matching of the half-edges of G_n is obtained through the following algorithm:

- Each half-edge of a vertex v is assigned a height uniformly chosen in $[0, 1]$ and is placed on the line of vertex v .
- Set the cut-off line at height 1.
- Pick an unmatched half-edge independent of the heights of all unmatched half-edges and match it to the highest unmatched half-edge. Move the cut-off line to the height of the latter half-edge.

Figure 2 illustrates the algorithm.

PROOF OF CLAIM 7.2. For each half-edge of a vertex in G_n with degree at least $2j + 1$, we choose their matches according to the above algorithm. Assume that at the end of this process of deleting half-edges of vertices with degree at least $2j + 1$, the cut-off line is at height $h \in (0, 1)$. Note that a half-edge of a vertex whose degree in G_n lies in $[j, 2j]$ is deleted if and only if it is above the cut-off line. We show that whp, $h \geq 1 - 4\varepsilon$. Indeed, Item 2 implies that the number of half-edges of G_n is at least $(1 - \varepsilon)nd/2$, and in choosing the heights of these half-edges the number of half-edges with heights above $1 - 4\varepsilon$ is at least

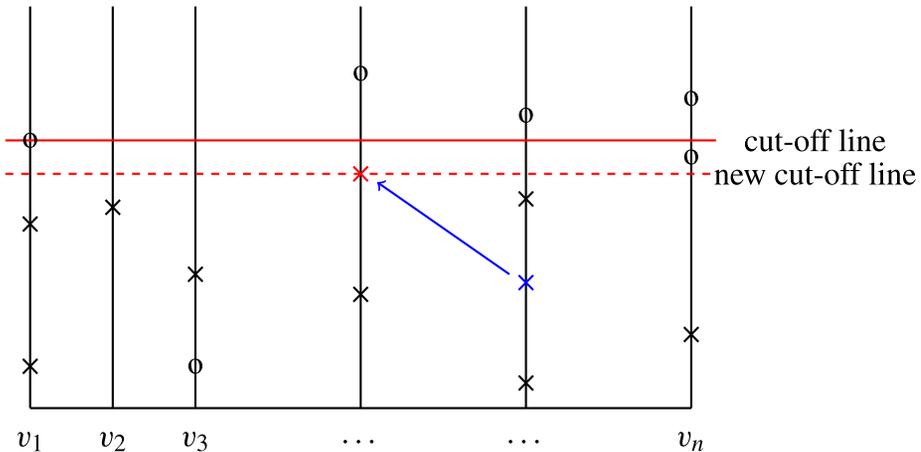


FIG. 2. The circles “o” represent matched half-edges and the crosses “x” represent unmatched half-edges. The blue half-edge is chosen and matched to the red half-edge which is the highest unmatched half-edge. Then, the cut-off line is moved to the new cut-off line (dashed).

$2\varepsilon(1 - 2\varepsilon)nd$ whp, by Chernoff’s inequality, which contradicts the event that we only delete at most εnd half-edges altogether. Hence, the cut-off line is above $1 - 4\varepsilon$ whp. Since the number of half-edges with degree in G_n belonging to $[j, 2j]$ is at most $(1 + \varepsilon)2jnu_j$ whp, the number of half-edges above $1 - 4\varepsilon$ is at most $8\varepsilon(1 + \varepsilon)jnu_j \leq 10\varepsilon jnu_j$ as claimed. That proves Claim 7.2. \square

Step 2. Exploration. Let W be the set of vertices in \bar{G}_n whose degrees belong to $[j/2, 2j]$. We shall find the desired (α, R) -embedded expander W_0 inside W . In this step we explore the R -neighborhoods of these high-degree vertices in W .

After having preprocessed the graph G_n to obtain \bar{G}_n , the remaining randomness is the perfect matching of the half-edges in \bar{G}_n . In this step we condition on the preprocessing step and write the probability in terms of the randomness of the perfect matching in \bar{G}_n . We run the following exploration process to perform some matchings of the half-edges of \bar{G}_n . Let R and τ be some large (bounded) numbers to be chosen (they are chosen in (7.15)):

1. For each vertex $v \in W$, set $r_v = 1$.
2. If $r_v < R$ for all $v \in W$, explore the neighborhood $B_{\bar{G}_n}(v, r_v)$ simultaneously for all $v \in W$; noting that if $u \in B_{\bar{G}_n}(v, r_v) \cap B_{\bar{G}_n}(v', r_{v'})$, we stop exploring the branch starting at u . If $r_v \geq R$ for some $v \in W$, the process terminates. Otherwise, go to (3).
3. For each vertex $v \in W$, if $N(v, r_v)$ intersects at most 100τ other balls $B_{\bar{G}_n}(v', r_{v'})$ ($v' \in W$), set $r_v := r_v + 1$. Otherwise, keep r_v intact. If none of the r_v is increases in this step, the process terminates. Otherwise, go back to (2).

We show that, when the exploration process terminates, the number of vertices $v \in W$ at which the process stops before reaching radius R is insignificant. For that, we choose R and τ so that the expected number of vertices of W that lie in a neighborhood $N(v, 2R)$ is small compared to τ (see (7.1)). And so, it is unlikely that the different neighborhoods $N(v, R)$ intersect frequently.

LEMMA 7.4. *Let R and τ be positive numbers bounded by some constants and satisfying*

$$(7.1) \quad \frac{\bar{b}^{2R-1} j^2 u_j}{d} \leq \frac{\tau}{10}.$$

The number of $v \in W$ with $r_v = R$ is at least $0.98|W|$ whp.

Note that \bar{b} is a random variable and so are R and τ . Nevertheless, we will choose R and τ so that they are bounded.

PROOF. For this proof we write $N(v, r_v)$ for $B_{\bar{G}_n}(v, r_v)$ for simplicity. We first show that for each $v \in W$, it is likely that $r_v = R$; more specifically,

$$(7.2) \quad \mathbb{P}(r_v = R) \geq 0.99.$$

Indeed, if $r_v < R$, then $N(v, R)$ intersects more than 100τ other balls $N(u, R)$ ($u \in W$) which implies that $N(v, 2R)$ contains more than 100τ elements of W . By Markov’s inequality,

$$\mathbb{P}(r_v < R) \leq \frac{\mathbb{E}|N(v, 2R) \cap W|}{100\tau}.$$

By Items 2 and 4 of Lemma 7.1, we have

$$\mathbb{E}|N(v, 2R) \cap W| \leq 2j\bar{b}^{2R-1} \frac{2j(2 + \varepsilon)nu_j}{(1 - \varepsilon)nd} \leq \tau,$$

where we used (7.1). Thus, we get (7.2).

It remains to show that whp, at least $0.98|W|$ vertices v satisfy $r_v = R$. We derive this from (7.2) and Azuma’s inequality. Let X be the number of vertices $v \in W$ with $r_v = R$. By (7.2) and Item 4 of Lemma 7.1,

$$\mathbb{E}X \geq 0.99|W| \geq 0.99(1 - \varepsilon)nu_j.$$

Enumerate the vertices of \bar{G}_n by $u_1, \dots, u_{\bar{n}}$. Let \mathcal{F}_i be the σ -algebra generated by the matchings of the half-edges of vertices u_1, \dots, u_i . We will apply Azuma’s inequality to the martingale $\mathbb{E}(X - \mathbb{E}X | \mathcal{F}_i), i = 0, \dots, \bar{n}$. Since the maximal degree in \bar{G}_n is $2j$, we have that, for every i ,

$$(7.3) \quad |\mathbb{E}(X - \mathbb{E}X | \mathcal{F}_{i+1}) - \mathbb{E}(X - \mathbb{E}X | \mathcal{F}_i)| \leq 4j \max_{u \in \bar{G}_n} |N(u, 2R)| \leq 4j(2j)^{2R},$$

where in the first inequality, we observed that for any fixed matching of the half-edges of vertices u_1, \dots, u_i and any two different matchings of the half-edges of u_{i+1} , there exists a bijection between the extensions of these matchings into perfect matchings of \bar{G}_n such that the number of different matchings are at most $4j$. Using Azuma’s inequality and the fact that j and R are constants, we obtain

$$(7.4) \quad \mathbb{P}(|X - \mathbb{E}X| \geq \varepsilon \mathbb{E}X) \leq \exp\left(-\frac{\varepsilon^2(\mathbb{E}X)^2}{2\bar{n}(4j(2j)^{2R})^2}\right) = \exp(-\Omega(n)).$$

Thus, whp, $X \geq (1 - \varepsilon)\mathbb{E}X \geq 0.98|W|$, as stated. \square

Step 3. Finding W_0 . We now find the desired embedded expander W_0 . Our strategy is roughly as follows. In step 2 we have explored the R -neighborhoods of the high-degree vertices in W . We shall show (in Lemma 7.5) that most of these neighborhoods have a lot of unmatched half-edges. If we think about a new graph in which each of these neighborhoods acts as a single vertex with high degree, then a high-degree core of this new graph corresponds to the desired W_0 .

Consider a new graph G'_n with vertex set $V' \cup V''$ where V'' are the vertices of \bar{G}_n that have not been touched in the exploration step 2 and each element of V' is a ball $N(v, r_v)$ in step 2 whose half-edges are the unmatched half-edges of $N(v, r_v)$. If there is an unmatched half-edge that belongs to at least two balls, we choose one such ball at random and associate this half-edge to that ball. The remaining randomness is the uniform perfect matching of the half-edges in G'_n . We show that many vertices in V' have high degree. Note that $|V'| = |W|$.

LEMMA 7.5 (V' has high degree). *There exist positive constants $\varepsilon', \varepsilon''$ and R_0 , depending only on μ , such that for all bounded positive numbers R_1, R, τ satisfying*

$$(7.5) \quad \begin{aligned} R_0 &\leq \min\{R_1, R - R_1\}, & 800\tau &\leq \varepsilon'^2(\bar{b}(1 - \varepsilon''))^{R_1-1}j, \\ \frac{\bar{b}^{2R_1-1}j^2u_j}{d} &\leq \frac{1}{10^4}, & \frac{\bar{b}^{2R-1}j^2u_j}{d} &\leq \frac{\tau}{10}, \end{aligned}$$

the number of vertices in V' with degree at least M is at least $\frac{\varepsilon'}{2}|V'|$ whp where

$$(7.6) \quad M = \frac{\varepsilon'^3(\bar{b}(1 - \varepsilon''))^{R-1}j}{8}.$$

We note that in (7.5), the last two inequalities are for Lemma 7.4 to hold. The constant R_1 is mainly for technical reasons. The condition that $800\tau \leq \varepsilon'(\bar{b}(1 - \varepsilon''))^{R_1-1}j$ is there so that when we ignore at most 100τ possible common branches, the number of remaining branches is still significant.

PROOF. First, we will show that:

CLAIM 7.6. *For every $v \in W$, with probability at least ε' , the number of half-edges on the boundary of $N(v, R)$ that do not belong to any of the balls $N(v', R)$, $v' \in W \setminus \{v\}$ is at least M .*

Indeed, let u_1, \dots, u_l be the vertices of distance R_1 from v . Since $\frac{\bar{b}^{2R_1-1} j^2 u_j}{d} < \frac{1}{10^4}$, by Lemma 7.4, with probability at least 0.99, $N(v, R_1)$ does not intersect any other balls $N(v', R_1)$.

Since $\frac{\bar{b}^{2R-1} j^2 u_j}{d} < \frac{\tau}{10}$, by Lemma 7.4, with probability at least 0.99, $N(v, R)$ intersects at most 100τ other balls $N(v', R)$, $v' \in W$. Consider the branches $B(u_h, R - R_1)$ consisting of vertices at distance r from v and distance $r - R_1$ from u_h for $R_1 \leq r \leq R$. Conditioned on u_1, \dots, u_l , it suffices to show that with probability at least 0.9, for any choices of sets $A \subset [1, \dots, l]$ with $|A| \leq 100\tau$,

$$(7.7) \quad \left| \bigcup_{h \in [1, \dots, l] \setminus A} \partial B(u_h, R - R_1) \right| \geq M.$$

Letting $Z_{R'}$ be the number of children in the R' th generation of the corresponding size-biased Galton–Watson process, we have for any $\varepsilon'' \in (0, 1)$, $\frac{Z_{R'}}{(\bar{b}(1-\varepsilon''))^{R'}}$ converges almost surely to some random variable Z (with Z not identically 0 and taking values in $[0, \infty]$) as $R' \rightarrow \infty$ (see, e.g., [1], pages 24–29). For some sufficiently small constant ε' (that only depends on μ), we have

$$\mathbb{P}(Z \geq 4\varepsilon') \geq 4\varepsilon'.$$

Thus, for a sufficiently large R_0 and $R_0 \leq R_1, R_0 \leq R - R_1$, we have

$$\mathbb{P}(l \geq \varepsilon'(\bar{b}(1 - \varepsilon''))^{R_1-1} j) \geq \varepsilon'$$

and

$$\mathbb{P}(|\partial B(u_h, R - R_1)| \geq \varepsilon'(\bar{b}(1 - \varepsilon''))^{R-R_1}) \geq \varepsilon'.$$

We can choose ε'' so small that $\bar{b}(1 - \varepsilon'') > 1$. Under the event that $l \geq \varepsilon'(\bar{b}(1 - \varepsilon''))^{R_1-1} j$, let X_h ($h = 1, \dots, l$) be the indicator of the event that $|\partial B(u_h, R - R_1)| \geq \varepsilon'(\bar{b}(1 - \varepsilon''))^{R-R_1}$. When R_0 is sufficiently large, l is also large. Thus, with probability at least 0.9, at least $\varepsilon'l/4$ indices $h \in [1, \dots, l]$ have $X_h = 1$. Under this event, since $100\tau \leq \varepsilon'l/8$ by (7.5), for any choices of sets $A \subset [1, \dots, l]$ with $|A| \leq 100\tau$, there are at least $\varepsilon'l/8$ indices $h \notin A$ with $X_h = 1$ which implies

$$\sum_{h \in [1, \dots, l] \setminus A} |\partial B(u_h, R - R_1)| \geq \varepsilon'^2(\bar{b}(1 - \varepsilon''))^{R-R_1} l/8.$$

Since j, \bar{b} and R are bounded by some constant, for each v the boundaries $\partial B(u_h, R - R_1)$ are disjoint with probability at least 0.9, proving (7.7) and Claim 7.6.

Next, by using Azuma’s inequality the same way that we used it in proving Lemma 7.4, we obtain that whp; there are at least $(1 - \varepsilon)\varepsilon'|W|$ vertices v satisfying the event in Claim 7.6 simultaneously. Combining this with Lemma 7.4 completes the proof of Lemma 7.5. \square

Note that the (random) edges of G'_n form a uniformly chosen perfect matching of its half-edges. On the half-edges of G'_n , consider the following coloring scheme on the half-edges:

- For each vertex v with at least M half-edges in G'_n , choose exactly M half-edges among them uniformly at random and color them blue.

- Perform the uniform random matching among all half-edges in G'_n , and let W'_1 be the induced subgraph on vertices with degree at least M .
- Let \mathfrak{K} be the number of blue edges, formed by two blue half-edges, and for each $v \in W'_1$, let $\text{deg}_b(v)$ be the number of blue edges adjacent to v .

Conditioned on \mathfrak{K} , we see that the distribution of $\{\text{deg}_b(v)\}_{v \in W'_1}$ is given by

$$(7.8) \quad \{B_v\}_{v \in W'_1}, B_v \sim \text{i.i.d. Bin}(M, \theta) \text{ conditioned on } \sum_{v \in W'_1} B_v = 2\mathfrak{K}.$$

Note that, due to the conditioning on the sum of $\{B_v\}$, their distribution is well defined regardless of the specific value of θ . However, for explicitness, we let

$$\theta := \frac{2\mathfrak{K}}{M|W'_1|}.$$

LEMMA 7.7. *Let ε' be as in Lemma 7.5. With high probability,*

$$(7.9) \quad \theta \geq \frac{\varepsilon' u_j M}{60d}.$$

PROOF. By Item 4 of Lemma 7.1, Lemmas 7.4 and 7.5, the number of vertices in W'_1 is at least $\frac{\varepsilon'|W|}{2} \geq \frac{\varepsilon' n u_j}{2}$.

Thus, the total number of blue half-edges in W'_1 is at least $\frac{\varepsilon' n u_j M}{2}$ whp, while the total degree in G'_n is at most $2nd$ by Item 2 of Lemma 7.1. Let

$$\theta_0 = \frac{\#\{\text{number of blue half-edges in } W'_1\}}{\#\{\text{number of half-edges in } G'_n\}} =: \frac{d_{W'_1}}{d_{G'_n}} \geq \frac{\varepsilon' u_j M}{4d}.$$

Since $d_{W'_1} = M|W'_1|$, it suffices to show that whp, $\mathfrak{K} \geq \theta_0 d_{W'_1} / 12$. Indeed, splitting the set of blue half-edges of W'_1 into two parts of equal size A and B (independent of their heights). We perform the cut-off line algorithm 7.3 to match the half-edges of A first. Since at least $d_{W'_1} / 4 = \theta_0 d_{G'_n} / 4$ highest half-edges in G'_n have been matched during this step, the cut-off line is below $1 - \frac{\theta_0}{5}$ whp (otherwise, by Chernoff’s inequality, the number of half-edges above the cut-off line is at most $\theta_0 d_{G'_n} / 4$). By Chernoff’s inequality again, the number of half-edges of B that lie above $1 - \theta_0 / 5$ is at least $\theta_0 d_{W'_1} / 12$ whp. Since all edges between A and B are inside W'_1 , $\mathfrak{K} \geq \theta_0 d_{W'_1} / 12$ whp. \square

Lemma 7.7 allows to find a high-degree core of W'_1 .

LEMMA 7.8. *Let M and θ be as in Lemmas 7.5 and 7.7. Assume that $\theta M \geq 100$. Then, whp, W'_1 contains a subgraph W'_0 with the following properties:*

- The number of vertices in W'_0 is at least $|W'_1| / 2$.
- Each vertex in W'_0 has degree at least $\frac{\theta M}{20}$ inside W'_0 ; in other words, W'_0 is an $\frac{\theta M}{20}$ -core.
- Each vertex in W'_0 has degree at most $(2j)^R$.

PROOF. Let $s = \frac{\theta M}{20}$. The last property follows from the fact that the maximum degree of vertices in V' is $(2j)^R$. For the rest of this proof, we only look at the blue half-edges in W'_1 that are matched to another blue half-edge in W'_1 . To find W'_0 , we use the cut-off line algorithm 7.3 to find a uniform perfect matching of these half-edges of W'_1 as follows. Each of these half-edges is reassigned a height uniformly chosen in $[0, 1]$. If there is a vertex in

W'_1 with less than s unmatched half-edges (equivalently, less than s half-edges below the cut-off line), match its half-edges to the highest unmatched half-edges and move the cut-off line accordingly. Remove this vertex. Repeat this step until there are no such vertices left.

Let W'_0 be the set of remaining vertices. It remains to show that $|W'_0| \geq \frac{|W'_1|}{2}$ whp. Note that, since $\mathfrak{K} = \Theta(n)$, the probability that $\sum_{v \in W'_1} B_v = 2\mathfrak{K}$ in (7.8) happens with probability $\Omega(n^{-C})$. In the rest of this proof, the tail probabilities are exponentially small in n without conditioning on the event $\sum_{v \in W'_1} B_v = 2\mathfrak{K}$. And so, we can forget about the conditioning and assume that the number of internal half-edges of each vertex $v \in W'_1$ has degree distribution $\mathbf{Bin}(M, \theta)$.

We show that, after the removal, the cut-off line is above $2/3$ whp. Assuming this, we have

$$|W'_0| \geq |W'_1| - N',$$

where N' is the number of vertices in W'_1 having less than s half-edges below the line $2/3$. By Lemma 7.7 the number of internal half-edges of each vertex $v \in W'_1$ has degree distribution $\mathbf{Bin}(M, \theta)$. Thus, the distribution of the number of its half-edges that lie below the line $2/3$ is $\mathbf{Bin}(M, \frac{2\theta}{3})$. So, the probability that v has less than s half-edges below the line $2/3$ is at most

$$\mathbb{P}\left(\mathbf{Bin}\left(M, \frac{2\theta}{3}\right) \leq \frac{\theta M}{20}\right) \leq \exp\left(-\frac{\theta M}{12}\right) \leq \frac{1}{200},$$

where we used the Chernoff inequality and the assumption $\theta M \geq 100$. By Chernoff's inequality we have whp, $N' \leq \frac{|W'_1|}{100}$. And so, $|W'_0| \geq \frac{|W'_1|}{2}$ as desired.

Now, we prove that, after the removal process above, the cut-off line is above $2/3$ whp. Let a be the number of removed vertices ($0 \leq a \leq |W'_1|$). The total number of matched half-edges is at most $2as$ because each time we remove a vertex, at most $2s$ half-edges are matched. Thus, the total number of half-edges above the cut-off line is at most $2as \leq 2s|W'_1|$.

On the other hand, given a vertex v with degree distribution $\mathbf{Bin}(M, \theta)$, the distribution of the number of its half-edges that lie above the line $2/3$ is $\mathbf{Bin}(M, \frac{\theta}{3})$. By Chernoff's inequality, the number of half-edges of W'_1 above the line $2/3$ is at least $M\theta|W'_1|/4 = 5s|W'_1| > 2s|W'_1|$. Thus, whp, the cut-off line is above $2/3$, completing the proof of Lemma 7.8. \square

Next, we show that W'_0 is an embedded expander.

LEMMA 7.9. *Let M and θ be as in Lemmas 7.5 and 7.7. Assume that $\theta M \geq 100$. There exists a constant $\alpha > 0$ such that whp, W'_0 is an $(\alpha, 1)$ -embedded expander.*

PROOF. Let $N = |W'_0|$, $s = \frac{\theta M}{20}$. Note that $N = \Theta(n)$ and $s \geq 5$ is a constant. It suffices to show that, for any subset A of the vertex set of W'_0 of size αN , the size of $N(A, 1)$ is at least twice that of A . In other words, whp, for every $m \leq \alpha N$ and subsets of vertices A, B with $|A| = m, |B| = 2m$, the neighbors of A are not contained fully in B . Fix two sets A and B with $|A| = m, |B| = 2m$. By Lemma 7.8 the number of half-edges in A is at least ms , the number of half-edges in B is $b \leq (2j)^R 2m$ and the total number of half-edges in W'_0 is $c \geq Ns$. The probability that all the neighbors of A belong to B is at most

$$\frac{b}{c-1} \frac{b-1}{c-3} \cdots \frac{b-ms+1}{c-2ms+1} \leq \left(\frac{b}{Ns-2ms+1}\right)^{ms} \leq \left(\frac{4(2j)^R m}{Ns}\right)^{ms}.$$

Taking the union bound over m and choices of A, B , we get that the probability that W'_0 is not an $(\alpha, 1)$ -embedded expander is at most

$$(7.10) \quad \sum_{m=1}^{\alpha N} \binom{N}{2m}^2 \left(\frac{4(2j)^R m}{Ns}\right)^{ms} \leq \sum_{m=1}^{\log n} \frac{C \log^{s-4} n}{N^{s-4}} + \sum_{m=\log n}^{\alpha N} (C\alpha^{s-4})^m$$

for some constant C , depending only on j, R and s . Choosing $\alpha \leq \frac{1}{2C}$ makes the RHS of (7.10) of order $o(1)$. This completes the proof of Lemma 7.9. \square

PROOF OF LEMMA 6.1. Since W'_0 is a subset of V' , each of its vertices is a ball $N(v, r_v)$ in \bar{G}_n , a subgraph of G_n . Let W_0 be the collection of all such centers v . Clearly, $|W_0| = |W'_0| = \Theta(n)$ and W_0 is an $(\alpha, 2R + 1)$ -embedded expander of G_n . To finish the proof of Lemma 6.1 for subexponential degree distributions, it remains to show that there exists a choice of j, R, τ satisfying the assumptions of the previous lemmas (in particular, Lemmas 7.4, 7.5, 7.8 and 7.9). In other words, for any given positive constants R_0, δ and ε' (δ and ε' can be arbitrarily small and R_0 can be arbitrarily large), we show that there exist a constant j and random variables R, R_1, τ such that the following conditions hold:

$$(7.11) \quad R \leq \delta j,$$

$$(7.12) \quad R_0 \leq \min\{R_1, R - R_1\},$$

$$(7.13) \quad \frac{\bar{b}^{2R_1-1} j^2 u_j}{d} \leq \frac{1}{10^4}, \quad 800\tau \leq \varepsilon'^2 (\bar{b}(1 - \varepsilon''))^{R_1-1} j,$$

$$\frac{\bar{b}^{2R-1} j^2 u_j}{d} \leq \frac{\tau}{10},$$

$$(7.14) \quad \frac{\varepsilon'^7 (\bar{b}(1 - \varepsilon''))^{2R-2} j^2 u_j}{64 \cdot 60d} \geq 100.$$

Note that (7.11) comes merely from the statement of Lemma 5.2. For given j and \bar{b} , we define τ, R and R_1 by the following equations so that (7.13) holds automatically:

$$(7.15) \quad \frac{\bar{b}^{2R_1-1} j^2 u_j}{d} = \frac{1}{10^4}, \quad 800\tau = \varepsilon'^2 (\bar{b}(1 - \varepsilon''))^{R_1-1} j, \quad \frac{\bar{b}^{2R-1} j^2 u_j}{d} = \frac{\tau}{10}.$$

Since μ has finite second moment, $\lim_{j \rightarrow \infty} j^2 u_j = 0$ and so (7.12) holds when j is sufficiently large. Since μ is subexponential, $u_j \geq e^{-\varepsilon_0 j}$ for any constant $\varepsilon_0 > 0$ and for large j . By choosing ε_0 to be small compared to δ , (7.11) holds. The inequality (7.14) holds automatically. This completes the proof of Lemma 6.1 for subexponential distributions μ . \square

REMARK 7.10. Observe that in the above subgraph W_0 is also an $(\alpha, 2R + 1)$ -embedded expander of the graph \bar{G}_n with $\deg_{\bar{G}_n} w \geq j/2$ for all $w \in W_0$.

7.2. Proof of Lemma 5.2. In Section 7.1 we used the assumption that μ is subexponential only in the last step of choosing the parameters R, R_1 and τ . In particular, we used this assumption to obtain that j can be an arbitrarily large constant while $u_j := \mu[j, 2j]$ remains positive and $u_j \geq e^{-\varepsilon_0 j}$. The inequality $u_j \geq e^{-\varepsilon_0 j}$ is only used to show that, with the choice of parameters R, R_1, τ as in (7.15), (7.11) holds as stated in Lemma 6.1. Here, when μ has an exponential tail, Lemma 5.2 does not assert that $R \leq \delta j$, and so there is no need for $u_j \geq e^{-\varepsilon_0 j}$.

Thus, if μ has an infinite support, one can still find an arbitrarily large constant j for which $u_j > 0$. This is, therefore, enough for the rest of the proof of Section 7.1 to follow, proving Lemma 5.2 for such μ .

If the support of μ is finite, let j be the largest integer in the support of μ . Let u_j be an arbitrarily small constant with $u_j \leq \mu[j, 2j] = \mu(j)$ (u_j could be much smaller than $\mu(j)$). Since the degrees in G_n are already bounded, there is no need to run the Step 1 of preprocessing the graph as in Section 7.1. So, for this case, $\bar{G}_n = G_n, \bar{n} = n, \bar{b} = b$ and so on. For the exploration, Step 2, we choose W by assigning each vertex in G_n of degree j to W independently with probability $\frac{u_j}{\mu(j)}$. The rest of the proof follows without any changes. In (7.15) we choose u_j to be sufficiently small so that (7.12) holds and so does (7.14). \square

APPENDIX

A.1. Proof of Lemma 4.3. Here, we prove Lemma 4.3 which is based on an elementary analysis of large deviation events.

PROOF OF LEMMA 4.3. The first statement follows directly from the Cauchy–Schwarz inequality: If we have $\mathbb{E}_{D \sim \mu} \exp(3\varepsilon D) < \infty$ for some $\varepsilon > 0$, then

$$\left(\sum_k e^{\varepsilon k} \sqrt{p_k} \right)^2 \leq \left(\sum_k e^{3\varepsilon k} p_k \right) \left(\sum_k e^{-\varepsilon k} \right) < \infty.$$

For the second statement, let n be a given large enough integer, and define

$$k_n = \min \left\{ k : \sum_{j \geq k} p_j \leq \frac{1}{n \log \log n} \right\}.$$

Let D_i for $i = 1, \dots, n$ be i.i.d samples from μ . We start by studying the empirical distribution of the D_i . First, by a simple union bound, the definition of k_n implies that

$$\mathbb{P}(\exists i \in [n] : D_i \geq k_n) \leq \frac{1}{\log \log n} = o(1).$$

Moreover, since μ has an exponential tail, there exists a constant $C > 0$ depending on μ such that $k_n \leq C \log n$. Recall the definition of $k_0 = \max\{k : \sum_{j \geq k} \sqrt{p_j} \geq 1/2\}$. Our next goal is to show that, with high probability, the number of i such that $D_i = k$ is at most $\frac{1}{2}n\sqrt{p_k}$ for all $k_0 \leq k \leq k_n$. We consider two possible cases of k as follows:

1. For k such that $p_k \leq (n \log^2 n)^{-1}$, Markov’s inequality implies that

$$(A.16) \quad \mathbb{P}(|\{i : D_i = k\}| \geq 1) \leq (\log^2 n)^{-1}.$$

2. For k such that $p_k \geq (n \log^2 n)^{-1}$, we use the following large deviation estimate for binomials (Corollary 22.9 of [10]): for $c > 1$,

$$\mathbb{P}(\mathbf{Bin}(n, p) \geq cnp) \leq \exp\{-np(c \log c + 1 - c)\}.$$

This gives that

$$(A.17) \quad \begin{aligned} &\mathbb{P}\left(|\{i : D_i = k\}| \geq \frac{1}{2}n\sqrt{p_k}\right) \\ &\leq \exp\left(\frac{1}{2}n\sqrt{p_k}(1 - 2\sqrt{p_k} + \log(2\sqrt{p_k}))\right) \leq \exp(-n^{1/3}). \end{aligned}$$

Since $k_n \leq C \log n$, applying a union bound on (A.16), (A.17) tells us that

$$(A.18) \quad \mathbb{P}\left(\exists k_0 \leq k \leq k_n : |\{i : D_i = k\}| \geq \frac{1}{2}n\sqrt{p_k}\right) = o(1).$$

When μ satisfies $k_0 < k_{\max}$, (A.18) implies that the empirical distribution of $\{D_i : i \in [n]\}$ is stochastically dominated by μ^\sharp , since $\mu^\sharp(k) \geq \sqrt{p_k}$ by the definition of k_0 . On the other hand, if $k_0 = k_{\max}$, the stochastic domination becomes trivial because we only augment the weight of k_{\max} in μ . Since taking out any $n/3$ entries from $[n]$ can only increase each probability mass of the empirical distribution of $\{D_i : i \in [n]\}$ by a factor of $3/2$, with high probability we have for each $k_0 \leq k$,

$$|\{i : D_i = k\}| \leq \frac{3}{4}n\sqrt{p_k},$$

and hence we conclude the second statement of Lemma 4.3. \square

A.2. Proof of Lemma 4.5. In this section we prove Lemma 4.5. We use Lemma 4.3 to bound the probability of $N(v, c \log n)$ having at least two cycles.

PROOF OF LEMMA 4.5. Let v be an arbitrary vertex in $G \sim \mathcal{G}(n, \mu)$, fixed before we explore the matchings of half-edges. We again study the local neighborhood $N(v, L)$ by exploration process, particularly in terms of the breadth-first search perspective. We start exploring from the single vertex v , and, at time s , we explore all the vertices of distance s from v , based on what we explored until time $s - 1$. Let V_s be the collection of vertices explored at time s , and set $X_s = |V_s|$.

We will bound the probability of discovering at least two cycles during the exploration process until depth $L = c \log n$ (c will be determined later). Let μ^\sharp be the augmented distribution (Definition 4.2), and let $\tilde{\mu}^\sharp := (\mu^\sharp)'_{[1, \infty)}$ denote its size-biased distribution conditioned on being inside the interval $[1, \infty)$. Also, let $\mathcal{T} \sim \text{GW}(\mu^\sharp, \tilde{\mu}^\sharp)_L$, and let Y_s be the number of vertices in \mathcal{T} at depth s . Then, Lemma 4.3 implies that there exists a coupling between $(X_s)_{s \leq L}$ and $(Y_s)_{s \leq L}$ in such a way that $X_s \leq Y_s$ for all $s \leq L$, as long as $\sum_{s=0}^L X_s \leq n/3$. Define B to be the event that $\sum_{s=0}^L X_s \leq n/3$. On B we clearly have

$$\sum_{v \notin \bigcup_{s \leq L} V_s} \text{deg}(v) \geq \frac{2n}{3}.$$

Assume that, when moving from V_s to V_{s+1} , we pair the half-edges adjacent to V_s one by one. Let H_s be the number of unpaired half-edges adjacent to the vertices in V_s . For $s \leq L$ and $1 \leq i \leq H_s$, let $\mathcal{F}_{s,i}$ be the σ -algebra generated by the exploration process until pairing the $(i - 1)$ th half-edge. Set $W_{s,i}$ to be the collection of unpaired half-edges at that moment. Further, let $A_{s,i}$ be the event that the i th half-edge adjacent to V_s is paired with a half-edge in $W_{s,i}$. Then, clearly,

$$\mathbb{P}(A_{s,i} \mid \mathcal{F}_{s,i}, B) \leq \frac{3|W_{s,i}|}{2n}.$$

We bound the size of $W_{s,i}$ based on the following observations:

1. Since the exploration of half-edges in H_{s-1} for $s \geq 1$ is done independently step by step, we can stochastically dominate $|H_s|$ by i.i.d $\zeta_j \sim \tilde{\mu}^\sharp$ as

$$|H_s| \leq_{\text{st}} \sum_{j=1}^{Y_s} \zeta_j \stackrel{d}{=} Y_{s+1}.$$

2. Since $\zeta \sim \tilde{\mu}^\sharp$ satisfies $\zeta \geq 1$, we can bound $|W_{s,i}|$ similarly by

$$|W_{s,i}| \leq_{\text{st}} \sum_{j=1}^{H_s} \zeta_j \leq_{\text{st}} \sum_{j=1}^{Y_{s+1}} \zeta_j \stackrel{d}{=} Y_{s+2}.$$

Therefore, combining above argument gives that

$$\sum_{j=1}^{H_s} \mathbf{1}_{A_{s,j}} \leq_{\text{st}} \text{Bin}\left(Y_{s+1}, \frac{3Y_{s+2}}{2n}\right),$$

where the LHS is conditioned on the event B . Notice that the LHS of the above inequality stochastically dominates the number of cycles formulated during the exploration of depth from s to $s + 1$. Hence, the number of cycles in $N(v, L)$ conditioned on B is stochastically dominated by

$$(A.19) \quad \text{Bin}\left(\sum_{s=1}^{L+1} Y_s, \frac{3Y_{L+2}}{2n}\right).$$

Now, we bound the size of Y_s to conclude our argument. Let $\zeta \sim \mu^\sharp$ and $\zeta' \sim \tilde{\mu}^\sharp$, and let ε, M be the constants that satisfy $\max\{\mathbb{E}e^{\varepsilon\zeta}, \mathbb{E}e^{\varepsilon\zeta'}\} \leq M$. Set K to be a large constant such that $M^{1/K} = e^\varepsilon$. Then, we observe that Y_s/K^s has an exponential tail for all s , since

$$\begin{aligned} \mathbb{E}\left[\exp\left(\varepsilon \frac{Y_s}{K^s}\right)\right] &= \mathbb{E}\left[\mathbb{E}_{\zeta'}\left[\exp\left(\varepsilon \frac{\zeta'}{K^s}\right)\right]^{Y_{s-1}}\right] \\ (A.20) \qquad \qquad \qquad &\leq \mathbb{E}[M^{Y_{s-1}/K^s}] = \mathbb{E}\left[\exp\left(\varepsilon \frac{Y_{s-1}}{K^{s-1}}\right)\right], \end{aligned}$$

where the inequality is due to Jensen’s inequality. Iterating this $(s - 1)$ -times gives that the LHS is bounded by e^ε . Set $c > 0$ to be the constant satisfying $K^{c \log n + 3} = n^{1/6}$. Based on the above observation, we bound the quantity (A.19) as follows:

$$\begin{aligned} \mathbb{P}\left(\mathbf{Bin}\left(\sum_{s=1}^{L+1} Y_s, \frac{3Y_{L+2}}{2n}\right) \geq 2\right) &\leq \mathbb{P}\left(\mathbf{Bin}\left(n^{1/5}, \frac{3}{2}n^{-4/5}\right) \geq 2\right) \\ (A.21) \qquad \qquad \qquad &+ \mathbb{P}\left(\sum_{s=1}^{L+2} Y_s \geq n^{1/5}\right). \end{aligned}$$

It is easy to see that the first term in the RHS is bounded by $o(n^{-1})$. The second term can be bounded using (A.20). Namely,

$$\begin{aligned} \mathbb{P}\left(\sum_{s=1}^{L+2} Y_s \geq n^{1/5}\right) &\leq e^{-\varepsilon n^{1/30}} \mathbb{E}\left[\exp\left(\varepsilon K^{-(L+3)} \sum_{s=1}^{L+2} Y_s\right)\right] \\ &\leq e^{-\varepsilon n^{1/30}} \mathbb{E}\left[\exp\left(\varepsilon K^{-(L+3)} \sum_{s=1}^{L+1} Y_s\right) \cdot \exp(\varepsilon K^{-(L+2)} Y_{L+1})\right]. \end{aligned}$$

Iterating this $(L + 1)$ more times, we obtain

$$\begin{aligned} \mathbb{P}\left(\sum_{s=1}^{L+2} Y_s \geq n^{1/5}\right) &\leq e^{-\varepsilon n^{1/30}} \mathbb{E}\left[\exp\left(\varepsilon \left(\sum_{s=1}^{L+2} K^{-s}\right)\right)\right] \\ &\leq \exp(\varepsilon(1 - n^{1/30})) = o(n^{-1}), \end{aligned}$$

as long as $K \geq 2$. Applying our estimates to (A.21), we conclude the desired result. \square

A.3. Proof of Proposition 4.9. The proof follows the same technique as Lemma 4.11.

PROOF OF PROPOSITION 4.9. Let $s \geq 2$ and L be any integers, and we build up an inductive argument starting from $l = 0$.

Let $\mathcal{S}_0 \sim \text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; 0, s)_L$, and $\rho^+ \in \mathcal{S}_0^+$ be the parent of ρ as before. Define $S_{0,s,L}$ to be the first time when $(X_t) \sim \text{CP}_{\rho^+}^\lambda(\mathcal{S}_0^+; \mathbf{1}_\rho)$ reaches state $\mathbf{0}$. Similarly, as in Lemmas 3.3 and 4.11 we consider $(\tilde{X}_t) \sim \tilde{\text{CP}}_{\rho^+; \rho}^\lambda(\mathcal{S}_0^+; \mathbf{1}_\rho)$, which is coupled with (X_t) in such a way that they share the same infection and recovery clocks, except that in (\tilde{X}_t) , the recovery at ρ is ignored if at that time there exists an infected vertex other than ρ and ρ^+ . Letting $D \sim \mu^\sharp$ be $D + 3 = \text{deg}(\rho; \mathcal{S}_0^+)$, $\mathcal{T}_{u_1}, \dots, \mathcal{T}_{u_D}$ be the i.i.d $\text{GW}(\tilde{\mu}^\sharp)_{L-1}$ subtrees from the children of ρ and \mathcal{S}' be the $\text{GWC}(\tilde{\mu}^\sharp; s)_L$ process that also hangs at ρ , we obtain the following by repeating the same argument in Lemma 4.11:

$$(A.22) \qquad \mathbb{E}[S_{0,s,L} \mid D] \leq (1 + \lambda \mathbb{E}[S_{L-1}])^D (1 + 2\lambda \mathbb{E}[S_{s,L}]),$$

where $S_{L-1}, S_{s,L}$ are as in the statements of Lemmas 3.3 and 4.11, respectively.

For general $l \neq 0$, we first develop the same argument in terms of $S'_l \sim \text{EGW}(\tilde{\mu}^\sharp; l, s)_L$. Let $S'_{l,s,L}$ be the first time when $\text{CP}_{\rho^+}^\lambda((S'_l)^+; \mathbf{1}_\rho)$ reaches at $\mathbf{0}$. For $D' \sim \tilde{\mu}^\sharp$, denoting $D' + 1 = \text{deg}(\rho; (S'_l)^+)$, the subgraphs of descendants from the children of ρ consist of D' i.i.d $\text{EGW}(\tilde{\mu}^\sharp; l - 1, s)_L$ processes. Therefore, repeating the previous reasoning gives that

$$(A.23) \quad \mathbb{E}[S'_{l,s,L} | D'] \leq (1 + \lambda \mathbb{E}[S'_{l-1,s,L}])^{D'}$$

Finally, the subgraphs of descendants (from children of ρ) of $S_l \sim \text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_L$ for $l \geq 1$ with $\text{deg}(\rho) = D$ consist of D i.i.d $\text{EGW}(\tilde{\mu}^\sharp; l - 1, s)_L$ processes. Therefore, we deduce that $S_{l,s,L}$, the first time when $\text{CP}_{\rho^+}^\lambda(S_l; \mathbf{1}_\rho)$ reaches $\mathbf{0}$, satisfies

$$(A.24) \quad \mathbb{E}[S_{l,s,L} | D] \leq (1 + \lambda \mathbb{E}[S'_{l-1,s,L}])^D$$

Here, the law of D follows the conditional distribution of μ^\sharp being inside the interval $[1, \infty)$.

Combining the three equations (A.22), (A.23) and (A.24), we obtain that there exists a constant λ_0 such that for all $\lambda \leq \lambda_0, l, s$ and L ,

$$\mathbb{E}[S_{l,s,L}] \leq 2e,$$

by manipulating the constants in the same way as Lemma 4.11. Then, the standard coupling between contact processes tells us that $R_{l,s,L} \leq_{\text{st}} S_{l,s,L}$ which concludes the proof. \square

A.4. Proof of Lemma 4.12. To establish Lemma 4.12, we first prove the result for GWC-processes. Let $\tilde{\mu}^\sharp$ be as in Section 4. Namely, $\tilde{\mu}^\sharp$ is the augmented distribution of $\mu'_{[1,\infty)}$, where $\mu'_{[1,\infty)}$ is the size-biased distribution of μ conditioned on being in $[1, \infty)$.

LEMMA A.11. *Let $\mathcal{S} \sim \text{GWC}(\tilde{\mu}^\sharp; s)_L$ and v_S^θ be the stationary distribution of $\text{DP}_{\rho^+}^{\lambda,\theta}(\mathcal{S})$ on the space $\{0, 1\}^{\mathcal{S} \setminus \{\rho\}}$ which is the delayed contact process on \mathcal{S} with ρ set to be infected permanently. Then, there exist constants $C, \lambda_0 > 0$, depending only on μ , such that for all $\lambda \leq \lambda_0$ and s, L with $s \geq 2$, we have $\mathbb{E}[v_S^\theta(\mathbf{0})^{-1}] \leq 2$ for $\theta = C\lambda$.*

PROOF. Let v, v' be the two neighbors of ρ in \mathcal{S} . Let S_v^θ denote the first time when $\text{DP}_{\rho^+}^{\lambda,\theta}(\mathcal{S}; \mathbf{1}_v)$ reaches $\mathbf{0}$, and define $S_{v'}^\theta$ analogously. Then, we set $S_L^\theta = \frac{1}{2}(S_v^\theta + S_{v'}^\theta)$.

As before, we build up an inductive argument on s . The case $s = 2$ is essentially the same as Proposition 3.6, since $\mathcal{S} \sim \text{GWC}(\tilde{\mu}^\sharp; 2)_L$ can be thought of as \mathcal{T}_L^+ with $\mathcal{T}_L \sim \text{GW}(\tilde{\mu}^\sharp)_L$, where ρ^+ in \mathcal{T}_L^+ is connected with ρ by a double-edge. Thus, the same proof of Proposition 3.6 can be applied, and we leave the details to the reader.

The general case $s \geq 3$ is also similar to the previous arguments of Propositions 4.9 and 3.6, but there is a subtle difference in comparing the stationary distributions which makes the current case more technical. As before, we start with introducing a modified process as follows.

Let v be a neighbor of ρ in \mathcal{S} , and $(X_t) \sim \text{DP}_{\rho^+}^{\lambda,\theta}(\mathcal{S}; \mathbf{1}_v)$. Define $(\tilde{X}_t) \sim \tilde{\text{DP}}_{\rho;v}^{\lambda,\theta}(\mathcal{S}; \mathbf{1}_v)$ as:

1. (\tilde{X}_t) has the same infection and recovery clocks at (X_t) .
2. In (\tilde{X}_t) , any recovery attempt at v is ignored if there exists an infected vertex other than ρ and v at that moment.

If $\{\rho, v\}$ infects a (random) neighbor U before v is healed, then (\tilde{X}_t) behaves as $\text{DP}_{\rho,v}^{\lambda,\theta}(\mathcal{S}; \mathbf{1}_U)$ (meaning that we fix both ρ, v to be infected forever), until \tilde{X}_t comes back to $\mathbf{1}_v$. Let $\tilde{S}_{s,L}^\theta$ denote the first time when $\tilde{\text{DP}}_{\rho;v}^{\lambda,\theta}(\mathcal{S}; \mathbf{1}_v)$ becomes $\mathbf{0}$, and \tilde{S}^θ be the first time it takes for $\text{DP}_{\rho,v}^{\lambda,\theta}(\mathcal{S}; \mathbf{0})$ to return to $\mathbf{0}$ after infecting a (random) vertex U other than ρ and v .

Setting $D \sim \tilde{\mu}^\sharp$ to be $\deg(v) = D + 2$, the same reasoning as (3.1), (4.1) implies that

$$\begin{aligned}
 \mathbb{E}[\tilde{S}_{s,L}^\theta | \mathcal{S}] &= \sum_{k=0}^\infty \left(\frac{(D+2)\lambda}{1+(D+2)\lambda} \right)^k \frac{1}{1+(D+2)\lambda} \\
 &\times \left[\frac{k+1}{\theta(1+(D+2)\lambda)} + k\mathbb{E}[\tilde{S}^\theta | \mathcal{S}] \right] \\
 &= \frac{1}{\theta} (1+(D+2)\lambda) \mathbb{E}[\tilde{S}^\theta | \mathcal{S}].
 \end{aligned}
 \tag{A.25}$$

Now, we take account of the stationary measures to compare the running times. We first set up some notations as follows:

- ν'_S and π'_S are the stationary distribution of $\text{DP}_{\rho,v}^{\lambda,\theta}(\mathcal{S})$ and $\text{CP}_{\rho,v}^\lambda(\mathcal{S})$, respectively.
- $\mathcal{T}_{u_1}, \dots, \mathcal{T}_{u_D}$ denote the subtrees from the children u_1, \dots, u_D of v outside the cycle.

Note that these subtrees are i.i.d $\text{GW}(\tilde{\mu}^\sharp)_{L-1}$.

- Set $\tilde{\mathcal{S}} = \mathcal{S} \setminus \bigcup_{i=1}^D \mathcal{T}_{u_i}$. $\text{DP}_{\rho,v}^{\lambda,\theta}(\tilde{\mathcal{S}})$ denotes the delayed contact process on $\tilde{\mathcal{S}}$ that fixes both ρ, v to be infected permanently which has the depth $r(x; \tilde{\mathcal{S}})$ computed with respect to ρ . In particular, all possible states of $\text{DP}_{\rho,v}^{\lambda,\theta}(\tilde{\mathcal{S}})$ have depth $r(x)$ at least one, since v is always infected.

- $\nu_{\mathcal{T}_{u_i}}$ and $\nu_{\tilde{\mathcal{S}}}$ are the stationary distributions of $\text{DP}_v^{\lambda,\theta}(\mathcal{T}_{u_i}^+)$ and $\text{DP}_{\rho,v}^{\lambda,\theta}(\tilde{\mathcal{S}})$, respectively.

Moreover, $\pi_{\mathcal{T}_{u_i}}$ and $\pi_{\tilde{\mathcal{S}}}$ denote the stationary distributions of $\text{CP}_v^\lambda(\mathcal{T}_{u_i}^+)$ and $\text{CP}_{\rho,v}^\lambda(\tilde{\mathcal{S}})$, respectively. Also, set

$$\nu_S^\otimes = \left(\bigotimes_{i=1}^D \nu_{\mathcal{T}_{u_i}} \right) \otimes \nu_{\tilde{\mathcal{S}}}.$$

Note that $\pi'_S = \left(\bigotimes_{i=1}^D \pi_{\mathcal{T}_{u_i}} \right) \otimes \pi_{\tilde{\mathcal{S}}}$. Keeping in mind that $r(\mathbf{0}; \mathcal{S}) = 1$ in $\text{DP}_{\rho,v}^{\lambda,\theta}(\mathcal{S})$, we obtain by using (3.9) that

$$\nu'_S(\mathbf{0}) \geq \nu_S^\otimes(\mathbf{0}).
 \tag{A.26}$$

Moreover, observe that if we merge ρ and v in $\tilde{\mathcal{S}}$ into a single vertex ρ' , then the resulting graph $\tilde{\mathcal{S}}'$ satisfies $\tilde{\mathcal{S}}' \sim \text{GWC}(\tilde{\mu}^\sharp; s-1)_L$, and we can consider the natural one-to-one correspondence between the two state spaces $\{0, 1\}^{\tilde{\mathcal{S}} \setminus \{\rho, v\}}$ and $\{0, 1\}^{\tilde{\mathcal{S}}' \setminus \{\rho'\}}$. Thus, we can regard them as

$$\Omega = \{0, 1\}^{\tilde{\mathcal{S}} \setminus \{\rho, v\}} = \{0, 1\}^{\tilde{\mathcal{S}}' \setminus \{\rho'\}}.$$

For any $x \in \Omega \setminus \{\mathbf{0}\}$, note that

$$r(x; \tilde{\mathcal{S}}) \in \{r(x; \tilde{\mathcal{S}}'), r(x; \tilde{\mathcal{S}}') + 1\}.$$

In particular, $r(x; \tilde{\mathcal{S}}) - 1 \leq r(x; \tilde{\mathcal{S}}')$. Further, we have $r(\mathbf{0}; \tilde{\mathcal{S}}) = 1$ and $r(\mathbf{0}; \tilde{\mathcal{S}}') = 0$. This implies that if $\nu_{\tilde{\mathcal{S}}'}$ denotes the stationary distribution of $\text{DP}_{\rho'}^{\lambda,\theta}(\tilde{\mathcal{S}}')$, then

$$\nu_{\tilde{\mathcal{S}}'}(\mathbf{0}) \leq \nu_{\tilde{\mathcal{S}}}(\mathbf{0}).$$

Therefore, combining with (A.26), we have

$$\nu'_S(\mathbf{0}) \geq \left(\prod_{i=1}^D \nu_{\mathcal{T}_{u_i}}(\mathbf{0}) \right) \cdot \nu_{\tilde{\mathcal{S}}'}(\mathbf{0}).
 \tag{A.27}$$

We can relate the quantities in (A.27) with the running times of the delayed processes. Let S_{L-1}^θ be the first time when $\text{DP}_{\rho}^{\lambda,\theta}(\mathcal{T}_{u_i}; \mathbf{1}_{u_i})$ returns to $\mathbf{0}$. Similarly, let $v_1, v_2 \notin \{\rho, v\}$ be the

two neighbors of $\{\rho, v\}$ in \mathcal{S}' , let $S_{v_i}^\theta$ be the first time when $\text{DP}_{\rho,v}^{\lambda,\theta}(\mathcal{S}'; \mathbf{1}_{v_i})$ reaches $\mathbf{0}$ and observe that $S_{s-1,L}^\theta \stackrel{d}{=} \frac{1}{2}(S_{v_1}^\theta + S_{v_2}^\theta)$, where the definition of $S_{s-1,L}^\theta$ is given in the beginning of the proof. Continuing similarly as (4.3), (3.10), we get that

$$\begin{aligned} v'_S(\mathbf{0}) &= \frac{1}{1 + (D + 2)\lambda\theta\mathbb{E}[\tilde{S}^\theta \mid \mathcal{S}]}; \\ \text{(A.28)} \quad v_{\mathcal{T}_{u_i}}(\mathbf{0}) &= \frac{1}{1 + \lambda\mathbb{E}[S_{L-1}^\theta \mid \mathcal{T}_{u_i}]}; \\ v_{\tilde{\mathcal{S}}'}(\mathbf{0}) &= \frac{1}{1 + 2\lambda\mathbb{E}[S_{s-1,L}^\theta \mid \tilde{\mathcal{S}}]}, \end{aligned}$$

where the additional factor of θ in the first identity comes from the fact that $r(\mathbf{0}; \mathcal{S}) = 1$ in $\text{DP}_{\rho,v}^{\lambda,\theta}(\mathcal{S})$. Plugging these into (A.27) and using (A.25), we obtain that

$$\mathbb{E}[S_{s,L}^\theta \mid D] \leq \mathbb{E}[\tilde{S}_{s,L}^\theta \mid D] \leq \frac{1}{\theta}(1 + \lambda\mathbb{E}[S_{L-1}^\theta])^D(1 + 2\lambda\mathbb{E}[S_{s-1,L}^\theta]).$$

Arguing similarly as Lemma 4.11 and Proposition 3.6, we deduce that there exist constants $C, \lambda_0 > 0$, depending on μ , such that for all $\lambda \leq \lambda_0$ and s, L with $s \geq 2$, $\mathbb{E}[S_{s,L}^\theta] \leq 3/\theta$ for $\theta = C\lambda$. Setting C to satisfy $C \geq 3$ and applying this to the right-hand side of the above equation (which is written in terms of $(s - 1, L)$) gives the desired conclusion. \square

PROOF OF LEMMA 4.12. To finish the proof of Lemma 4.12, we argue similarly as Proposition 4.9. Namely, we establish the result for $\text{EGW}(\tilde{\mu}^\sharp; l, s)_L$ and then extend it to the general case $\text{EGW}(\mu^\sharp, \tilde{\mu}^\sharp; l, s)_L$. In both steps we appeal to the same technique as Proposition 3.6 which is simpler than what is done here for the GWC-processes. We omit the details due to similarity. \square

A.5. Proof of Lemma 7.1, Items 1–3. Item 1 follows from the definition of G_n that its edges are obtained from a uniformly chosen perfect matching of the half-edges.

For Item 2, choose j large enough such that

$$\delta := \mathbb{E}_{D \sim \mu} D \mathbf{1}_{D \geq 2j+1} \leq \varepsilon/4.$$

For each vertex $v \in G_n$, consider the random variable

$$X_v := \text{deg}_{G_n}(v) \mathbf{1}_{\text{deg}_{G_n}(v) \geq 2j+1}.$$

These random variables are independent with mean δ and variance bounded by the second moment of μ . By Chebyshev’s inequality, whp

$$\sum_{v \in G_n} X_v \leq \varepsilon n/2.$$

Thus, whp, the total number of removed half-edges, is at most $2 \sum_{v \in G_n} X_v \leq \varepsilon n$, and so is the number of removed vertices. Thus, $\bar{n} \geq (1 - \varepsilon)n$. Applying Chernoff’s inequality to the random variables $\bar{X}_v := \text{deg}_{G_n}(v) \mathbf{1}_{\text{deg}_{G_n}(v) \leq 2j}$, we obtain that whp,

$$\sum_{v \in G_n} \bar{X}_v \in (1 - \varepsilon, 1 + \varepsilon)nd.$$

Combining this with the fact that the total number of deleted half-edges is at most εnd whp, we get

$$(1 - 2\varepsilon)nd \leq \sum_{v \in G_n} \bar{X}_v - \varepsilon n \leq d_1 + \dots + d_{\bar{n}} \leq \sum_{v \in G_n} \bar{X}_v \leq (1 + \varepsilon)nd$$

completing the proof of Item 2.

To prove Item 3, let $\varepsilon' = \min\{\varepsilon, \varepsilon(b - 1)\}$. Let k_0 be a large constant such that for all $h \geq k_0$, the branching rate of $\mu_{[0,h]} \in (1 - \varepsilon', 1 + \varepsilon')b$, namely

$$(A.29) \quad \frac{\mathbb{E}_{D \sim \mu} D(D-1) \mathbf{1}_{0 \leq D \leq h}}{\mathbb{E}_{D \sim \mu} D \mathbf{1}_{0 \leq D \leq h}} \in (1 - \varepsilon', 1 + \varepsilon')b,$$

$$(A.30) \quad \begin{aligned} \mathbb{E}_{D \sim \mu} D \mathbf{1}_{0 \leq D \leq k_0} &\in (1 - \varepsilon', 1 + \varepsilon')d, \quad \text{and} \\ \mathbb{E}_{D \sim \mu} D^2 \mathbf{1}_{0 \leq D \leq k_0} &\geq \frac{2}{\varepsilon'} \mathbb{E}_{D \sim \mu} D^2 \mathbf{1}_{k_0 < D}. \end{aligned}$$

Let $k \geq k_0$ be such that

$$(A.31) \quad \mathbb{E}_{D \sim \mu} D^2 \mathbf{1}_{k_0 < D \leq k} \geq \frac{2}{\varepsilon'} \mathbb{E}_{D \sim \mu} D^2 \mathbf{1}_{k < D},$$

which k exists because of the boundedness of $\mathbb{E}_{D \sim \mu} D^2$. Note that (A.31) implies that

$$(A.32) \quad \mathbb{E}_{D \sim \mu} (D \mathbf{1}_{k_0 < D \leq k}) \geq \frac{2}{\varepsilon'} \mathbb{E}_{D \sim \mu} (D \mathbf{1}_{k < D}).$$

We now show that for all constant $j \geq k$, $\bar{b} \in (1 - \varepsilon', 1 + \varepsilon')b$ whp. Let E_l and \bar{E}_l be the number of half-edges attached to vertices of degree l in G_n and \bar{G}_n , respectively. We need to show that whp

$$(A.33) \quad \bar{b} = \frac{\sum_{l=0}^{2j} (l-1) \bar{E}_l}{\sum_{l=0}^{2j} \bar{E}_l} \in (1 - \varepsilon, 1 + \varepsilon)b.$$

Since j is a constant and $E_l = l \sum_{v \in G_n} \mathbf{1}_{\deg_{G_n}(v)=l}$, by Chernoff's inequality, whp we have

$$(A.34) \quad \begin{aligned} \sum_{l=0}^{2j} E_l &\in (1 - \varepsilon', 1 + \varepsilon') \sum_{l=0}^{2j} l \mu(l)n \quad \text{and} \\ \sum_{l=0}^{2j} (l-1) E_l &\in (1 - \varepsilon', 1 + \varepsilon') \sum_{l=0}^{2j} l(l-1) \mu(l)n. \end{aligned}$$

This together with (A.29) and (A.30) give

$$\frac{\sum_{l=0}^{2j} (l-1) E_l}{\sum_{l=0}^{2j} E_l} \in (1 - 3\varepsilon', 1 + 3\varepsilon')b \quad \text{and} \quad \sum_{l=0}^{2j} l \mu(l) \in (1 - \varepsilon', 1 + \varepsilon')d.$$

Since the total number of removed half-edges is at most $\varepsilon'nd$ whp,

$$\sum_{l=0}^{2j} \bar{E}_l \in \sum_{l=0}^{2j} E_l + (-\varepsilon', 0)nd \subset (1 - 2\varepsilon', 1 + 2\varepsilon') \sum_{l=0}^{2j} E_l.$$

From this and (A.34), (A.33) reduces to proving that

$$\sum_{l=0}^{2j} l \bar{E}_l \in (1 - \varepsilon', 1 + \varepsilon') \sum_{l=0}^{2j} l E_l.$$

The upper bound is straightforward. To prove the lower bound, let N_l be the number of vertices of degree l in G_n . Since the number of deleted half-edges is at most $\sum_{l=2j+1}^{\infty} l N_l$,

we have by Markov’s inequality, (A.32) and Chernoff’s inequality, whp

$$\begin{aligned}
 \sum_{l=0}^{2j} (E_l - \bar{E}_l) &\leq \sum_{l=2j+1}^{\infty} lN_l \leq \frac{1}{\varepsilon'} \mathbb{E} \sum_{l=2j+1}^{\infty} lN_l \\
 &\leq \frac{1}{2} \mathbb{E} \sum_{l=k_0+1}^{2j} lN_l \leq \sum_{l=k_0+1}^{2j} lN_l.
 \end{aligned}
 \tag{A.35}$$

Thus, we have

$$\sum_{l=0}^{2j} l(E_l - \bar{E}_l) \leq \sum_{l=k_0+1}^{2j} l^2 N_l \leq \varepsilon' \sum_{l=0}^{2j} l^2 N_l = \varepsilon' \sum_{l=0}^{2j} lE_l,
 \tag{A.36}$$

where the first inequality follows from (A.35) and the fact that the left-hand side of (A.36) is largest when the deleted half-edges counted in $\sum_{l=0}^{2j} (E_l - \bar{E}_l)$ are drawn from vertices of highest degrees possible and the second inequality follows from the Chernoff inequality and (A.30). That completes the proof of (A.33) and hence Item 3. \square

Acknowledgments. We thank the anonymous referee for fruitful comments on improving the expositions of the manuscript.

S. Bhamidi is supported by NSF Grants DMS-1613072, DMS-1606839 and ARO Grant W911NF-17-1-0010.

D. Nam is supported by a Samsung scholarship.

A. Sly is supported by NSF Grant DMS-1352013, Simons Investigator grant and a MacArthur Fellowship.

REFERENCES

- [1] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes. Die Grundlehren der mathematischen Wissenschaften* **196**. Springer, New York. [MR0373040](#)
- [2] BERGER, N., BORGS, C., CHAYES, J. T. and SABERI, A. (2005). On the spread of viruses on the Internet. In *Proceedings of the Sixteenth Annual ACM–SIAM Symposium on Discrete Algorithms* 301–310. ACM, New York. [MR2298278](#)
- [3] CAN, V. H. and SCHAPIRA, B. (2015). Metastability for the contact process on the configuration model with infinite mean degree. *Electron. J. Probab.* **20** Art. ID 26. [MR3325096](#) <https://doi.org/10.1214/EJP.v20-3859>
- [4] CATOR, E. and DON, H. (2018). Explicit bounds for critical infection rates and expected extinction times of the contact process on finite random graphs. Preprint. Available at [arXiv:1804.03753v2](https://arxiv.org/abs/1804.03753v2).
- [5] CHATTERJEE, S. and DURRETT, R. (2009). Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.* **37** 2332–2356. [MR2573560](#) <https://doi.org/10.1214/09-AOP471>
- [6] CRANSTON, M., MOUNTFORD, T., MOURRAT, J.-C. and VALESIN, D. (2014). The contact process on finite homogeneous trees revisited. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** 385–408. [MR3249416](#)
- [7] DEMBO, A. and MONTANARI, A. (2010). Gibbs measures and phase transitions on sparse random graphs. *Braz. J. Probab. Stat.* **24** 137–211. [MR2643563](#) <https://doi.org/10.1214/09-BJPS027>
- [8] DURRETT, R. and LIU, X. F. (1988). The contact process on a finite set. *Ann. Probab.* **16** 1158–1173. [MR0942760](#)
- [9] DURRETT, R. and SCHONMANN, R. H. (1988). The contact process on a finite set. II. *Ann. Probab.* **16** 1570–1583. [MR0958203](#)
- [10] FRIEZE, A. and KAROŃSKI, M. (2016). *Introduction to Random Graphs*. Cambridge Univ. Press, Cambridge. [MR3675279](#) <https://doi.org/10.1017/CBO9781316339831>
- [11] HARRIS, T. E. (1974). Contact interactions on a lattice. *Ann. Probab.* **2** 969–988. [MR0356292](#) <https://doi.org/10.1214/aop/1176996493>
- [12] HUANG, X. and DURRETT, R. (2018). The contact process on random graphs and Galton–Watson trees. Preprint. Available at [arXiv:1810.06040](https://arxiv.org/abs/1810.06040).

- [13] JANSON, S. (2009). The probability that a random multigraph is simple. *Combin. Probab. Comput.* **18** 205–225. MR2497380 <https://doi.org/10.1017/S0963548308009644>
- [14] KIM, J. H. (2006). Poisson cloning model for random graphs. In *International Congress of Mathematicians, Vol. III* 873–897. Eur. Math. Soc., Zürich. MR2275710
- [15] LALLEY, S. and SU, W. (2017). Contact processes on random regular graphs. *Ann. Appl. Probab.* **27** 2061–2097. MR3693520 <https://doi.org/10.1214/16-AAP1249>
- [16] LIGGETT, T. M. (1996). Multiple transition points for the contact process on the binary tree. *Ann. Probab.* **24** 1675–1710. MR1415225 <https://doi.org/10.1214/aop/1041903202>
- [17] LIGGETT, T. M. (1999). *Stochastic Interacting Systems: Contact, Voter and Exclusion Processes. Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]* **324**. Springer, Berlin. MR1717346 <https://doi.org/10.1007/978-3-662-03990-8>
- [18] LIGGETT, T. M. (2005). *Interacting Particle Systems. Classics in Mathematics*. Springer, Berlin. MR2108619 <https://doi.org/10.1007/b138374>
- [19] LUBETZKY, E. and SLY, A. (2010). Cutoff phenomena for random walks on random regular graphs. *Duke Math. J.* **153** 475–510. MR2667423 <https://doi.org/10.1215/00127094-2010-029>
- [20] MÉNARD, L. and SINGH, A. (2016). Percolation by cumulative merging and phase transition for the contact process on random graphs. *Ann. Sci. Éc. Norm. Supér. (4)* **49** 1189–1238. MR3581814 <https://doi.org/10.24033/asens.2307>
- [21] MOLLOY, M. and REED, B. (1995). A critical point for random graphs with a given degree sequence. *Random Structures Algorithms* **6** 161–179. MR1370952 <https://doi.org/10.1002/rsa.3240060204>
- [22] MOUNTFORD, T., MOURRAT, J.-C., VALESIN, D. and YAO, Q. (2016). Exponential extinction time of the contact process on finite graphs. *Stochastic Process. Appl.* **126** 1974–2013. MR3483744 <https://doi.org/10.1016/j.spa.2016.01.001>
- [23] MOUNTFORD, T., VALESIN, D. and YAO, Q. (2013). Metastable densities for the contact process on power law random graphs. *Electron. J. Probab.* **18** Art. ID 103. MR3145050 <https://doi.org/10.1214/EJP.v18-2512>
- [24] MOUNTFORD, T. S. (1993). A metastable result for the finite multidimensional contact process. *Canad. Math. Bull.* **36** 216–226. MR1222537 <https://doi.org/10.4153/CMB-1993-031-3>
- [25] MOURRAT, J.-C. and VALESIN, D. (2016). Phase transition of the contact process on random regular graphs. *Electron. J. Probab.* **21** Art. ID 31. MR3492935 <https://doi.org/10.1214/16-EJP4476>
- [26] PEMANTLE, R. (1992). The contact process on trees. *Ann. Probab.* **20** 2089–2116. MR1188054
- [27] SCHAPIRA, B. and VALESIN, D. (2017). Extinction time for the contact process on general graphs. *Probab. Theory Related Fields* **169** 871–899. MR3719058 <https://doi.org/10.1007/s00440-016-0742-0>
- [28] STACEY, A. (2001). The contact process on finite homogeneous trees. *Probab. Theory Related Fields* **121** 551–576. MR1872428 <https://doi.org/10.1007/s004400100149>
- [29] STACEY, A. M. (1996). The existence of an intermediate phase for the contact process on trees. *Ann. Probab.* **24** 1711–1726. MR1415226 <https://doi.org/10.1214/aop/1041903203>
- [30] VAN DER HOFSTAD, R. (2017). *Random Graphs and Complex Networks, Vol. 1. Cambridge Series in Statistical and Probabilistic Mathematics* **43**. Cambridge Univ. Press, Cambridge. MR3617364 <https://doi.org/10.1017/9781316779422>