The Annals of Applied Probability
2015, Vol. 25, No. 6, 3139–3161
DOI: 10.1214/14-AAP1069

EXTINCTION WINDOW OF MEAN FIELD BRANCHING ANNIHILATING RANDOM WALK[1]

BY IDAN PERL[2], ARNAB SEN[3] AND ARIEL YADIN[2]

Ben-Gurion University of the Negev, University of Minnesota and Ben-Gurion University of the Negev

We study a model of growing population that competes for resources. At each time step, all existing particles reproduce and the offspring randomly move to neighboring sites. Then at any site with more than one offspring, the particles are annihilated. This is a nonmonotone model, which makes the analysis more difficult.

We consider the *extinction window* of this model in the finite mean-field case, where there are n sites but movement is allowed to any site (the complete graph). We show that although the system survives for exponential time, the extinction window is logarithmic.

1. Introduction.

1.1. *The model.* Perhaps the most classical population model is the *Galton–Watson branching process*. Originally devised to model the survival of aristocratic patrilineal surnames, the Galton–Watson process may be described as follows: start with one existing particle. At every time step, all existing particles reproduce an independent number of offspring and die out. The main question is then, what is the probability that the system survives forever? By use of generating functions it is fairly simple to analyze this model, and in fact it is well known that in a Galton–Watson process with offspring distribution L, the probability of extinction is given by the unique minimal solution of the equation $s = \mathbb{E}[s^L]$ in the interval $(0, 1]$. Moreover, the solution q satisfies $q = 1$ if and only if $\mathbb{E}[L] \leq 1$; see, for example, [3, 12] for a thorough treatment.

To make matters more interesting, one might add some geometry, by having the particles not only branch (reproduce) but also move in some underlying graph. This is the *branching random walk* model, which is described as follows: start with one particle at some origin vertex o in graph G. At each time step, all existing

Received October 2013; revised August 2014.

[1]Support for this work came from EPSRC Grant EP/GO55068/1.

[2]Supported in part by Grant 2010357 from the United States–Israel Binational Science Foundation (BSF).

[3]Supported in part by NSF Grant DMS-14-06247.

MSC2010 subject classifications. 60J80, 92D25.

Key words and phrases. Branching annihilating random walk, branching process, population models.

particles reproduce an independent number of offspring and die out. All offspring now independently choose a random neighbor of their parent's vertex, and move to that new position. Thus a specific lineage of particles performs a random walk on G. A different way to view this model is as a tree-indexed random walk (see [4, 5] for more on tree-indexed random walks) where the domain tree is the tree of lineage formed by a Galton–Watson process. See the pioneering work of Biggins [6] and the survey by Shi [14].

Both models mentioned above exhibit some sort of monotonicity, enabling coupling arguments. For example, put in an imprecise way, if one has more particles, the branching random walk is more likely to be recurrent. The additional particles only help it return to the origin.

Let us now introduce the model we work with, which we dub *branching-annihilating random walk*, or BARW for short. Start with a single particle at some origin vertex o of a graph G. At each time step, all particles independently reproduce (or branch) into a random number of offspring. These offspring then each choose independently a random neighbor of their parent's vertex and move to that neighbor. (So far, everything is identical to the branching random walk.) Finally, at every vertex at which there is more than one particle, these particles are eliminated (this is the annihilation phase).

BARW is a model for population reproduction in some geometry, with a competition for resources. The annihilation phase can be viewed as there being only enough resources for one particle at every vertex of the underlying graph.

Let us stress that the difficulty in analyzing BARW stems mainly from the lack of monotonicity. Adding particles may on the one hand assist in the ultimate survival of the system, but may also hinder the survival, as these additional particles may compete for resources and annihilate others, resulting in too few particles to survive.

It is most convenient to work with Poisson distributed offspring, so for simplicity we will restrict to this distribution.

DEFINITION 1. Let $\lambda > 1$ be a real number. Let G be a graph, and let $o \in G$ be some vertex. We define *branching-annihilating random walk* on G, starting at o, with parameter λ, or $\mathsf{BARW}_{G,o}(\lambda)$, as the following Markov process on subsets of G.

Let $(L_{t,j})_{t,j=1}^{\infty}$ be i.i.d. Poisson-λ random variables. Start with $B_0 = \{o\}$. For every $t \geq 0$, given $B_t \neq \varnothing$, define B_{t+1} as follows.

Suppose that $B_t = \{x_1, \dots, x_m\}$. For every $1 \leq j \leq m$, let $y_{j,1}, \dots, y_{j,L_{t,j}}$ be independent vertices chosen uniformly from the set $\{y : y \sim x_j\}$ (the neighbors of x_j in G). Define $Z_{t+1} : G \to \mathbb{R}$ by $Z_{t+1}(x) = \sum_{j=1}^{m} \sum_{i=1}^{L_{t,j}} \mathbb{1}_{\{y_{j,i}=x\}}$. This is the number of offspring that have moved to x.

Finally, let $B_{t+1} = \{x : Z_{t+1}(x) = 1\}$. In the case that $B_t = \varnothing$, then $B_{t+1} = \varnothing$ as well.

1.2. *Main questions and results.* As stated above, BARW lacks monotonicity, and thus it is not easy to analyze. However, it seems reasonable to ask the following immediate questions regarding the long-term behavior. Some of these questions are being studied by the authors in a separate work, for the case of G being the infinite d-regular tree.

Suppose G is an infinite transitive graph. If λ is either too big or too small, one may dominate BARW by a sub-critical Galton–Watson process. Thus we are guaranteed extinction in either case. (This is not surprising, as too little offspring do not give a good enough chance of survival, and too many offspring create too much annihilation, thus again ruining the chance of survival.)

The immediate questions that arise regard a super-critical interval of survival:

- Do there exist $\lambda_c^- \le \lambda_c^+$ such that for $\lambda \in (\lambda_c^-, \lambda_c^+)$ there is positive probability of survival forever, and for $\lambda \notin [\lambda_c^-, \lambda_c^+]$ there is extinction a.s.?
- If such an interval exists, what happens at the critical values $\lambda = \lambda_c^-$ and $\lambda = \lambda_c^+$?
- Can λ_c^-, λ_c^+ be identified?

In this paper we consider BARW in the finite graph setting, and specifically on the complete graph. Of course, there is always a positive probability of extinction in one step on a finite graph, so on a finite graph BARW will a.s. die out at some finite time. However, we may consider BARW on a sequence of finite graphs with size tending to infinity, and try to understand asymptotic properties of the process for large graphs.

In this work we consider the mean-field case, where the sequence under consideration is the complete graph on n vertices as $n \to \infty$.

Our first result states that BARW on the complete graph has an exponentially large expected lifetime.

THEOREM 2. *For every $\lambda > 1$ there exists $c = c(\lambda) > 0$ such that the following holds for all $n \in \mathbb{N}$. Consider* BARW *on the complete graph on n vertices, and let $X_t = |B_t|$ be the number of particles at time t. Let*

$$T_0 = \inf\{t \ge 0 : X_t = 0\}.$$

Then, for each $0 < x < n$,

$$\mathbb{E}[T_0 | X_0 = x] \ge ce^{cn}.$$

Our main result regards the "window" of extinction. It is not difficult to see that for BARW on the complete graph on n vertices, the number of particles will oscillate for a long time around the value $\mathrm{eq} := \frac{\log\lambda}{\lambda} n$. We call it the *quasi-stable* state, which is obtained by solving for the state x such that $\mathbb{E}[X_1 | X_0 = x] = x$. Below it the chain has an upward drift whereas there is a downward drift if the chain goes above the quasi-stable state. Our next result considers how long it takes the process to go extinct, once it has been conditioned to do so; that is, how many steps did it take the process to reach 0 particles, at the last excursion it made below the equilibrium point $\frac{\log\lambda}{\lambda}n$?

THEOREM 3. *For every* $\lambda > 1$ *and* $0 < \varepsilon < \frac{\log\lambda}{\lambda}$, *there exists* $C = C(\lambda, \varepsilon) > 0$ *such that the following holds for all* $n \in \mathbb{N}$.

Consider BARW *on the complete graph on n vertices, and let* $X_t = |B_t|$ *be the number of particles at time* t. *Let*

$$T_0 = \inf\{t \ge 0 : X_t = 0\} \quad \textit{and} \quad T^+_{\mathrm{eq}-\varepsilon n} = \inf\left\{t \ge 0 : X_t \ge \frac{\log\lambda}{\lambda}n - \varepsilon n\right\}.$$

Then for each $0 \le x < \frac{\log\lambda}{\lambda}n - \varepsilon n$,

$$C^{-1}\log(1+x) \le \mathbb{E}\big[T_0 | X_0 = x, T_0 < T^+_{\mathrm{eq}-\varepsilon n}\big] \le C\log(1+x).$$

REMARK 4. Though the above theorem holds for any $\lambda > 1$, the conditioned chain $(X_t)_{t\ge 0} | T_0 < T^+_{\mathrm{eq}-\varepsilon n}$ exhibits remarkably different behaviors in two distinct regimes of the parameter λ: (i) λ is close to 1, and (ii) λ is large; see Figures 1 and 2. Our proof is general enough to tackle both regimes simultaneously.

It would be interesting to find out whether, for a fixed n, the expected extinction time of the conditioned chain $\mathbb{E}[T_0 | X_0 = x, T_0 < T^+_{\mathrm{eq}-\varepsilon n}]$ is decreasing with respect to λ.

1.3. *Similar models and further questions.* BARW, or rather a continuous time versions, have been studied before; see, for example, [7, 8, 15]. However, most focus on survival of the process, or stationary measures.

On the other hand, recently there has been considerable interest among the physicists to study the behavior of a finite population evolving under some stochastic dynamics near its extinction time and particularly to find “most probable or optimal path to extinction” [10, 13].

To best of our knowledge this is the first work to study the “extinction window” for BARW; that is, the length of the last path to extinction. As our results show, at least in the mean-field case, this window is *much* smaller than the lifetime of the system, indicating that extinction is a “catastrophic” phenomenon, meaning that it occurs abruptly in a very short time frame.

Our analysis makes heavy use of the fact that on the complete graph, the geometry plays no role, so that BARW can actually be seen as a Markov chain on $\{0, 1, \dots, n\}$, making the model simpler. It would be very interesting to understand the expected lifetime and extinction window in other finite graph settings. More specifically:

QUESTION 5. Let $(G_n, o_n)_n$ be a sequence of finite rooted graphs converging in the local weak topology [2] to a limiting rooted graph (G, o). Consider BARW on G_n with Poisson-λ offspring:

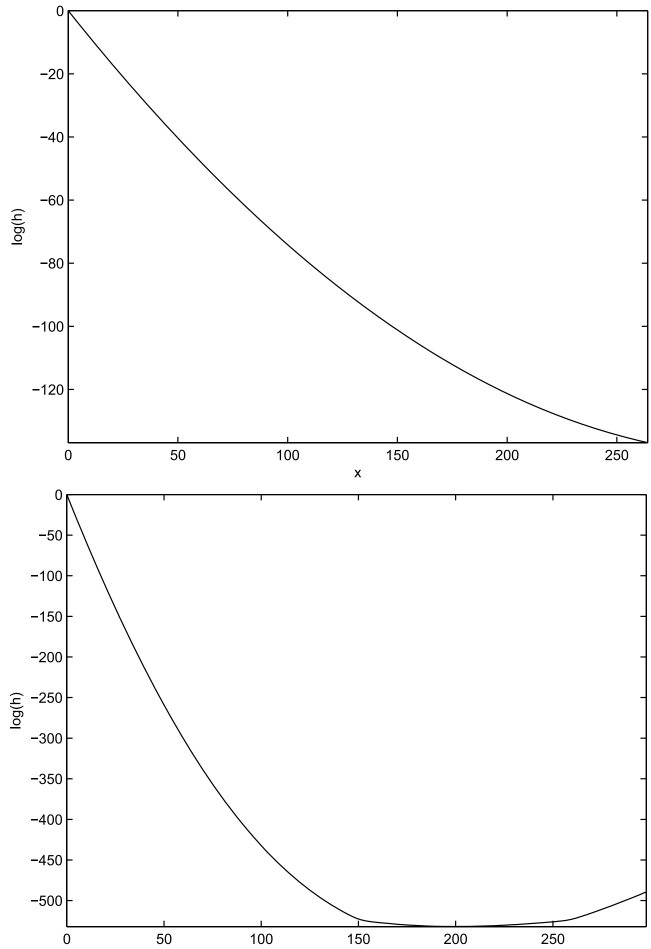

FIG. 1. *Plot of* $\log h(x)$ *vs* x *where* $h(x) = \mathbb{P}_x[T_0 < T^+_{\mathrm{eq}-\varepsilon n}]$ *for* $n = 1200$, $\varepsilon = 0.05$ *and* $\lambda = 1.5$ (*left*) *and* $\lambda = 6$ (*right*). *Note that for* $\lambda = 1.5$, h *is monotonically decreasing*, *but* $\log h$ *is not linear*, *so* h *can not be expressed as* $C\exp(-cx)$. *On the other hand*, *for* $\lambda = 6$, *the function* h *is not even monotone—it first decreases*, *and then it increases near* $\mathrm{eq} - \varepsilon n$.

- Is it true that there exist critical $\lambda_c^- \le \lambda_c^+$ such that if $\lambda \in (\lambda_c^-, \lambda_c^+)$, then the expected lifetime is exponentially large in $|G_n|$, and if $\lambda \notin [\lambda_c^-, \lambda_c^+]$, the expected lifetime is much smaller (perhaps logarithmic)?
- For which λ does BARW on G_n have a logarithmically small extinction window? That is, for which λ does there exist small enough $\eta > 0$ so that conditioned on extinction before reaching above $\eta|G_n|$ particles, the conditioned process has logarithmically small expected lifetime?

The above question is open even for a sequence of finite d-regular graphs with increasing girths (whose local limit is the infinite d-regular tree).

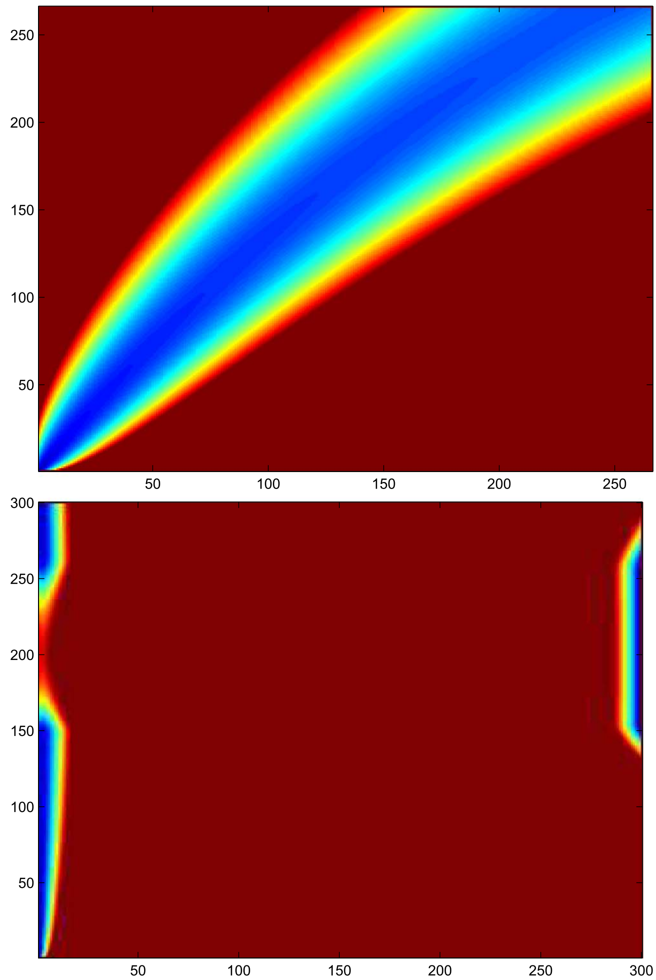

FIG. 2. *Transition probabilities of the* tilted *chain* $P(\cdot,\cdot|T_0 < T^+_{\mathrm{eq}-\varepsilon n})$ *for* $n = 1200$, $\varepsilon = 0.05$ *and* $\lambda = 1.5$ (*left*) *and* $\lambda = 6$ (*right*). *The probabilities are represented by colors—the blue represents high values, and the red represents small values. For* $\lambda = 1.5$, *from any* x *in the tilted chain, the walker goes down by a multiplicative factor with high probability. But for* $\lambda = 6$, *the transition matrix is highly concentrated. For some* x, *the tilted chain goes up with high probability. For some* x, *it goes down with high probability. For a few* x*'s, the transition distribution is bimodal!*

1.4. *Comparison with SIS model and variants.* It has been suggested that the BARW is similar in spirit to the SIS infection model. In the SIS model, all vertices in a graph are either infected or not. The infected vertices infect their neighbors at a certain rate, and every vertex recovers from infection at a different rate, these rates being parameters of the model. The discrete time version of this model may have two interpretations: we may allow only one particle to act at every time step, which is the discrete time backbone of the continuous time chain, or allow all particles to act at the same time. A similar variant may have been used in the BARW model.

It turns out that the SIS and BARW models are sensitive to these kind of local modifications, and we do not see a way to relate them. In the one-particle-at-a-time versions on complete graphs, the chains are birth and death chains, meaning that they are Markov chains on $\{0, 1, \dots, n\}$ with transition probabilities restricting movement only between states at distance 1. This makes the analysis simpler using the available tools for such chains; see, for example, [1], [9], Chapter XVII.5, [11], Chapter 2.4. Let us give a brief account of this analysis.

1.4.1. BARW *one particle at a time.* In the continuous time BARW model, particles die at rate 1 and give off a particle to a uniform vertex at rate $\lambda > 1$. When two particles are at a vertex, they instantly annihilate one another. Consider the number of living individuals and the discrete backbone of this continuous chain as a discrete time Markov process. Note that at each time step either one particle dies or a new one is added, or nothing is changed. If there are x living individuals, with probability $\frac{\lambda}{1+\lambda} \cdot (1 - \frac{x}{n})$, a particle is added to an empty vertex, and the number of individuals increases by 1; with probability $\frac{1}{1+\lambda}$, a living individual dies and the number of total individuals decreases by 1; with remaining probability $\frac{\lambda}{1+\lambda} \cdot \frac{x}{n}$, a particle is added to an occupied vertex, resulting in annihilation, so the number of total living individuals decreases by 1.

To sum up, the transition probabilities of this chain are given by

$$P_B(x, y) = \begin{cases} \dfrac{1}{1+\lambda} + \dfrac{\lambda}{1+\lambda} \cdot \dfrac{x}{n}, & x > 0, y = x - 1, \\ \dfrac{\lambda}{1+\lambda} \cdot \left(1 - \dfrac{x}{n}\right), & x > 0, y = x + 1, \\ 1, & y = x = 0. \end{cases}$$

1.4.2. *SIS one particle at a time.* In the SIS model the difference is that annihilation is replaced by coalescence. Analogously to the above, infected individuals recover with rate 1 and infect a neighbor at rate $\lambda > 1$. So considering the discrete backbone of the total number of infected vertices, with probability $\frac{1}{1+\lambda}$, a vertex recovers and the total number decreases by 1; with probability $\frac{\lambda}{1+\lambda} \cdot \frac{x}{n}$, an infected vertex is infected, resulting in no change to the total number of infected vertices; with probability $\frac{\lambda}{1+\lambda} \cdot (1 - \frac{x}{n})$, a healthy vertex is infected, and the total number increases by 1. The following is a summary of the transition probabilities for the SIS model:

$$P_S(x, y) = \begin{cases} \dfrac{1}{1+\lambda}, & x > 0, y = x - 1, \\ \dfrac{\lambda}{1+\lambda} \cdot \dfrac{x}{n}, & x > 0, y = x, \\ \dfrac{\lambda}{1+\lambda} \cdot \left(1 - \dfrac{x}{n}\right), & x > 0, y = x + 1, \\ 1, & y = x = 0. \end{cases}$$

One now sees that there is an additional drift downward for the BARW model that is not present in the SIS model.

1.4.3. *Extinction window.* In this subsection, whenever we talk about the BARW and the SIS model, we refer to their one-particle-at-a-time version. For BARW and the SIS model, the quasi-stable states are given by $\mathrm{eq}_B = \frac{\lambda-1}{2\lambda}n$ and $\mathrm{eq}_S = \frac{\lambda-1}{\lambda}n$, respectively. Clearly, for both these chains, the expected extinction time is at least exponential in n, that is, $\mathbb{E}_x[T_0] \ge ce^{cn}$ for any x, since we can find $\delta > 0$ such that between the states 0 and δn each of the chains can be coupled from below with a simple random walk with bias away from zero. Thanks to the standard results on birth and death chains regarding hitting probabilities ([1], [9], Chapter XVII.5, [11], Chapter 2.4), the extinction window is also easy to calculate for these models. Let us first talk about the SIS model. The transition probabilities of the chain conditioned on the event $\{T_0 < T_{\mathrm{eq}_S - \varepsilon n}\}$ can be obtained via Doob's h-transform,

$$\hat{P}_S(x, y) = P_S(x, y)\frac{\mathbb{P}_y[T_0 < T_{\mathrm{eq}_S-\varepsilon n}]}{\mathbb{P}_x[T_0 < T_{\mathrm{eq}_S-\varepsilon n}]}.$$

Let $M = \mathrm{eq}_S - \varepsilon n$. We will use the following standard notation for the jump probabilities of a birth and death chain: $p_x = P_S(x, x + 1)$, $q_x = P_S(x, x - 1)$ and $r_x = P_S(x, x)$. Then $\mathbb{P}_x[T_0 < T_M] = \frac{\varphi(M)-\varphi(x)}{\varphi(M)}$ where $\varphi(x) = \sum_{m=0}^{x-1}\prod_{j=1}^{m}\theta_j$ for $x > 1$ and $\varphi(0) = 0$ and $\theta_x = q_x/p_x$. Note that the tilted chain $\hat{P}_S$ is again a birth and death chain on $0, 1, 2, \dots, M$ with jump probabilities $\hat{p}_x = \hat{P}_S(x, x+1)$, $\hat{q}_x = \hat{P}_S(x, x - 1)$ and $\hat{r}_x = \hat{P}_S(x, x) = r_x$.

We have

$$\begin{aligned}\frac{\hat{p}_x}{\hat{q}_x} &= \frac{p_x}{q_x}\cdot\frac{\sum_{m=x+1}^{M}\prod_{j=1}^{m}\theta_j}{\sum_{m=x-1}^{M}\prod_{j=1}^{m}\theta_j} \\ &= \frac{\theta_{x+1} + \theta_{x+1}\theta_{x+2} + \cdots + \theta_{x+1}\theta_{x+2}\cdots\theta_M}{1 + \theta_x + \theta_x\theta_{x+1} + \cdots + \theta_x\theta_{x+1}\cdots\theta_M}\end{aligned} \tag{1}$$

for $0 < x < M$, $\frac{1}{\lambda} < \theta_x < \frac{1}{1+\lambda\varepsilon}$. Writing $z = \min(x + C_1 \log n, M)$ for sufficiently large C_1, we can approximate the ratio in (1) by

$$\frac{\theta_{x+1} + \theta_{x+1}\theta_{x+2} + \cdots + \theta_{x+1}\theta_{x+2}\cdots\theta_z}{1 + \theta_x + \theta_x\theta_{x+1} + \cdots + \theta_x\theta_{x+1}\cdots\theta_z} + O(n^{-1}). \tag{2}$$

Using the fact that $|\theta_x - \theta_y| \le C_2\frac{|x-y|}{n}$ for all $x, y < M$, we can write (2) as

$$\frac{\theta_x + \theta_x^2 + \cdots + \theta_x^{z-x}}{1 + \theta_x + \theta_x^2 + \cdots + \theta_x^{z-x+1}} + o(1) = \theta_x + o(1),$$

where the error term $o(1)$ is uniform in $0 < x < M$.

Hence, for sufficiently large n, the tilted chain $\hat{P}_S$ can be coupled from above by a lazy simple random walk with holding probability $\frac{1}{1+\lambda}$ and with a bias toward 0. Therefore, we conclude that there exists a constant $C > 0$ such that

$$x \leq \mathbb{E}_x[T_0 | T_0 < T_{\mathrm{eq}_S - \varepsilon n}] \leq Cx \qquad \text{for each } 0 < x < \mathrm{eq}_S - \varepsilon n.$$

We can prove a similar result on the extinction window for the BARW one-particle-at-a-time model following exactly the same arguments as above.

1.4.4. *SIS all particles at once.* As mentioned, in this note we consider BARW with all particles reproducing at once. The analogous SIS version could be defined as follows. At every time step, every infected vertex infects Poi(λ) uniformly chosen neighbors (perhaps some chosen more than once). Vertices not re-infected then recover. This is the same as replacing annihilation in BARW with coalescence. So the SIS model is the same as a branching-coalescing random walk.

When considered on the complete graph, if there are x infected vertices, every vertex receives Poi($\frac{\lambda x}{n}$) infections, so is left infected at the next time step with probability $(1 - e^{-\lambda x/n})$, independently for all vertices. Thus, given that there are x infected vertices at time t, the number of infected vertices at time $t + 1$ has $\mathrm{Bin}(n, (1 - e^{-\lambda x/n}))$ distribution.

Note that the equation $n(1 - e^{-\lambda x/n}) = x$ has exactly two solutions in $[0, n]$, one which is at $x = 0$, and the other being the equilibrium of this model. Since $n(1 - e^{-\lambda x/n}) - x$ is maximized at $x = \mathrm{eq} = n\frac{\log \lambda}{\lambda}$ and since this maximum is positive, we have that the equilibrium of SIS is larger than eq, the equilibrium of BARW.

Analysis of the SIS model's extinction window is another possible future direction of research.

1.5. *Preliminaries and notation.* It will be much simpler to use the following equivalent form of BARW on the complete graph on n vertices. (Here is where the mean-field structure makes the analysis much simpler.) Given that $|B_t| = x$, that is, there are x particles at time t, every particle branches into Poisson-λ particles, and each of these chooses a new vertex, independently, and uniformly among all n vertices. Thus, due to the summability of the Poisson distribution, at the branching phase every vertex receives an independent Poisson-$\frac{\lambda x}{n}$ number of particles. In the annihilation phase only those vertices with exactly one particle survive to the next step, which happens at a given vertex with probability $b(x) := \frac{\lambda x}{n} e^{-\lambda x/n}$.

Thus, we have just shown that if $(X_t)_t$ is the number of existing particles in BARW on the complete graph on n vertices, then $(X_t)_t$ is a Markov chain with transitions given by

$$\mathbb{P}[X_{t+1} = y | X_t = x] = \mathbb{P}\big[\mathrm{Bin}\big(n, b(x)\big) = y\big] = \binom{n}{y} b(x)^y \big(1 - b(x)\big)^{n-y}.$$

This observation will be central in what follows.

We use the notation $\mathbb{P}_x$ and $\mathbb{E}_x$ to denote the probability measure and expectation of BARW on the complete graph on n vertices with (Poissonian) offspring mean λ and with $X_0 = x$.

Let $\lambda > 1$. Consider the Galton–Watson process with offspring distribution $L \sim \mathrm{Poi}(\lambda)$. It is well known that there exists a number $q = q(\lambda) \in (0, 1)$ such that the process dies out with probability q, and that q is the unique fixed point of the equation $s = \mathbb{E}[s^L] = e^{-\lambda(1-s)}$ in $(0, 1)$. Also, since q is the probability of extinction, it is clear that $q(\lambda)$ is a continuous strictly decreasing function of λ; see, for example, [12].

Throughout, we make extensive use of the following inequalities, which are easy to verify:

- For any $t \in (0, 1)$ and $n \in \mathbb{N}$, $e^{-nt} \geq (1 - t)^n$.
- For any $0 \leq t \leq \frac{1}{2}$, we have $\sqrt{1 - 2t} \geq 1 - t - 2t^2$.
- For any $0 \leq t \leq \frac{1}{2}$, we have $1 - t \geq e^{-1+\sqrt{1-2t}}$.
- The last two inequalities can be combined to deduce $1 - t \geq e^{-t(1+2t)}$.

$x \wedge y$ denotes the minimum of x, y, and $x \vee y$ denotes the maximum of x, y.

We also make use of the stopping times

$$T_x^+ = \inf\{t \geq 0 : X_t \geq x\}.$$

Another tool we will use is the following standard large deviations result concerning binomial random variables. For $0 < \xi < 1$,

$$\mathbb{P}\bigl[\mathrm{Bin}(n, b) < \xi nb\bigr] \leq \exp\biggl(-nb \cdot \frac{(1 - \xi)^2}{4}\biggr).$$

2. The extinction time for unconditional chain. In this section we prove Theorem 2.

Let $\tau_{\varepsilon n}^+$ be the *return* time to one of the sites in $[\varepsilon n, n]$,

$$\tau_{\varepsilon n}^+ := \inf\{t \geq 1 : X_t \geq \varepsilon n\}.$$

For the proof of Theorem 2 we do not require the full strength of the following lemma, but it will also be required in the sequel. Recall from Section 1.5 that given $\lambda > 1$, $q = q(\lambda) \in (0, 1)$ is the Poisson dual parameter, that is, the unique number satisfying $\lambda e^{-\lambda} = q e^{-q}$.

LEMMA 6. *Let $0 < \varepsilon < \frac{1}{2\lambda}$ and small enough such that $\lambda e^{-\lambda\varepsilon} > 1$. Let $\lambda_1 = \lambda e^{-\lambda\varepsilon}$, $\lambda_2 = \lambda(1 + 2\lambda\varepsilon)$, and define $q_1 = q(\lambda_1)$, $q_2 = q(\lambda_2)$. Then*

$$\frac{q_2^x - q_2^{\varepsilon n}}{1 - q_2^{\varepsilon n}} \leq g(x) \leq \frac{q_1^x - q_1^n}{1 - q_1^n}, \qquad 0 \leq x < \varepsilon n,$$

where $g(x) := \mathbb{P}_x[T_0 < T_{\varepsilon n}^+]$.

PROOF. Denote $b(x) = \frac{\lambda x}{n} e^{-\lambda x/n}$. On $X_t = x$, we have that $X_{t+1} \sim \mathrm{Bin}(n, b(x))$. So

$$\mathbb{E}[q_1^{X_{t+1}} | X_t = x] = (b(x)q_1 + 1 - b(x))^n = (1 - b(x)(1 - q_1))^n$$
$$\leq e^{-nb(x)(1-q_1)} = [e^{-\lambda(1-q_1)e^{-\lambda x/n}}]^x \leq q_1^x,$$

where the last inequality follows by the definition of q_1. This implies that $(q_1^{X_t})_{t=0}^{T_{\varepsilon n}^+}$ is a supermartingale. We may apply the optional stopping theorem,

$$q_1^x \geq \mathbb{E}[q_1^{X_{T_0 \wedge T_{\varepsilon n}^+}}] = \mathbb{E}[q_1^{X_{T_0}} \mathbb{1}_{\{T_0 < T_{\varepsilon n}^+\}}] + \mathbb{E}[q_1^{X_{T_{\varepsilon n}^+}} \mathbb{1}_{\{T_0 > T_{\varepsilon n}^+\}}] \geq g(x) + (1 - g(x))q_1^n,$$

and therefore $g(x) \leq \frac{q_1^x - q_1^n}{1 - q_1^n}$.

We obtain the lower bound similarly:

$$\mathbb{E}[q_2^{X_{t+1}} | X_t = x] = (b(x)q_2 + 1 - b(x))^n = (1 - b(x)(1 - q_2))^n.$$

Now, $b(x)(1 - q_2) = \frac{\lambda x}{n} e^{-\lambda x/n}(1 - q_2) \leq \lambda \varepsilon e^{-\lambda x/n}(1 - q_2) \leq \lambda\varepsilon < \frac{1}{2}$, so a short calculation gives

$$\mathbb{E}[q_2^{X_{t+1}} | X_t = x] \geq e^{-nb(x)(1-q_2)(1+2b(x)(1-q_2))}$$
$$= [e^{-\lambda e^{-\lambda x/n}(1-q_2)(1+2b(x)(1-q_2))}]^x$$
$$\geq [e^{-\lambda(1-q_2)(1+2\lambda\varepsilon)}]^x = q_2^x.$$

This implies that $(q_2^{X_t})_{t=0}^{T_{\varepsilon n}^+}$ is a submartingale. As before, by the optional stopping theorem,

$$q_2^x \leq \mathbb{E}_x[q_x^{X_{T_0 \wedge T_{\varepsilon n}^+}}] = \mathbb{E}[q_2^{X_{T_0}} \mathbb{1}_{\{T_0 < T_{\varepsilon n}^+\}}] + \mathbb{E}[q_2^{X_{T_{\varepsilon n}^+}} \mathbb{1}_{\{T_0 > T_{\varepsilon n}^+\}}]$$
$$\leq g(x) + (1 - g(x))q_x^{\varepsilon n}$$

and therefore $\frac{q_2^x - q_2^{\varepsilon n}}{1 - q_2^{\varepsilon n}} \leq g(x)$. $\square$

PROOF OF THEOREM 2. Fix $\varepsilon = \varepsilon(\lambda) > 0$ small enough so that:

- It meets the requirements of Lemma 6.
- It satisfies $b(\varepsilon n) \leq b(n)$, or equivalently, $\varepsilon \leq e^{\lambda(1-\varepsilon)}$. It follows that $b(\varepsilon n) \leq b(x)$ for all $x \geq \varepsilon n$.
- It satisfies $\varepsilon n \sqrt{\lambda} \leq nb(\varepsilon n)$, or equivalently, $e^{\lambda\varepsilon} \leq \sqrt{\lambda}$.

Keeping in mind that $\mathbb{P}_y[T_0 < T_{\varepsilon n}^+] = 0$ for any $y \geq \varepsilon n$, by the Markov property we have that $\mathbb{P}_x[T_0 < \tau_{\varepsilon n}^+] \leq \mathbb{P}_x[X_1 < \varepsilon n]$ for all x.

Next, we bound the term $\mathbb{P}_x[X_1 < \varepsilon n]$ using standard large deviations for the binomial distribution. Note that by our choice of ε, for any $x \geq \varepsilon n$ we have that

$\mathbb{E}_x[X_1] = nb(x) \geq nb(\varepsilon n) \geq \varepsilon n\sqrt{\lambda}$. Therefore, for any $x \geq \varepsilon n$,

$$\mathbb{P}_x[T_0 < \tau^+_{\varepsilon n}] \leq \mathbb{P}_x[X_1 < \varepsilon n] \leq \mathbb{P}_x[X_1 < \lambda^{-1/2} \cdot \mathbb{E}_x[X_1]] \leq \exp(-c\varepsilon n), \tag{3}$$

where $c = \sqrt{\lambda} \cdot \frac{(1-\lambda^{-1/2})^2}{4}$.

Note that by Lemma 6 we have that for all $x > 0$, $\mathbb{P}_x[T^+_{\varepsilon n} < T_0] \geq c' := 1 - q_1(1-q_1)^{-1} > 0$ where $q_1 = q(\lambda e^{-\lambda\varepsilon})$. Thus, for $x < \varepsilon n$ we have that

$$\mathbb{E}_x[T_0] \geq c' \cdot \inf_{y \geq \varepsilon n} \mathbb{E}_y[T_0].$$

So it remains to consider $x \geq \varepsilon n$.

By (3) and the strong Markov property, on the event $X_0 \geq \varepsilon n$, the random time T_0 dominates a geometric random variable with success probability $e^{-c\varepsilon n}$. Thus, for all x,

$$\mathbb{E}_x[T_0] \geq c' \cdot e^{c\varepsilon n},$$

which proves the theorem. $\square$

3. Bounds on hitting probabilities.

3.1. *Probability of extinction before going above level εn.* Throughout this subsection we denote $g(x) := \mathbb{P}_x[T_0 < T^+_{\varepsilon n}]$.

Let $\alpha = \alpha(\lambda) \in (0, 1)$ such that the following inequalities hold: $(1-\alpha)\lambda > 1, \lambda e^{-\alpha\lambda} < 1$. This is equivalent to $\frac{\log \lambda}{\lambda} < \alpha < 1 - \lambda^{-1}$ which is possible since $\lambda > 1$.

Next, let $p(x, y)$ be the transition function of our Markov chain. Explicitly, for any $0 \leq x, y \leq n$, $p(x, y) = \mathbb{P}[\mathrm{Bin}(n, b(x)) = y]$. Let $m(x) := \mathbb{E}[\mathrm{Bin}(n, b(x))] = nb(x)$ and $m_0(x) := (1-\alpha)m(x)$.

LEMMA 7. *For any $0 < \varepsilon < \frac{1}{\lambda}$, $0 \leq x < \varepsilon n - 1$ and $0 \leq y \leq m_0(x)$, we have that*

$$p(x+1, y) \leq \gamma \cdot p(x, y),$$

where $\gamma = \gamma_{\varepsilon,\alpha} := e^{-\alpha\lambda e^{-\lambda\varepsilon}(1-\lambda\varepsilon)} < 1$.

PROOF. Recall that $b(x) = \frac{\lambda x}{n} e^{-\lambda x/n}$. The function te^{-t} is increasing for $0 \leq t < 1$, which implies that $b(x)$ is increasing while $\frac{\lambda x}{n} < 1$, and thus increasing as long as $x < \varepsilon n$. It now follows that $\frac{p(x+1,y)}{p(x,y)}$ is increasing in y,

$$\frac{p(x+1, y)}{p(x, y)} = \left(\frac{b(x+1)}{b(x)}\right)^y \left(\frac{1-b(x+1)}{1-b(x)}\right)^{n-y},$$

and since $b(x+1) > b(x)$, this expression is indeed increasing in y. It follows that

$$\max_{0 \leq y \leq m_0(x)} \frac{p(x+1, y)}{p(x, y)} = \frac{p(x+1, m_0(x))}{p(x, m_0(x))}.$$

So we want to bound from above the expression $\frac{p(x+1,m_0(x))}{p(x,m_0(x))}$ with a bound that is independent of x.

It can be simply checked that for

$$f(t) = f_x(t) := m_0(x)\log b(t) + \big(n - m_0(x)\big)\log\big(1 - b(t)\big),$$

we have

$$\log\frac{p(x+1, m_0(x))}{p(x, m_0(x))} = f(x+1) - f(x).$$

So we want to bound $f(t+1) - f(t)$. By the mean value theorem, it will be sufficient to bound $f'(t)$.

Recall that $x + 1 < \varepsilon n$, and $\varepsilon < \lambda^{-1}$, so $b(\cdot)$ is monotone increasing for $t \leq x + 1$. Upon differentiation, we get for all $t \in [x, x+1]$,

$$\begin{aligned} f'(t) &= b'(t)\left(\frac{m_0(x)}{b(t)} - \frac{n - m_0(x)}{1 - b(t)}\right) \leq b'(t)\left(\frac{m_0(x)}{b(x)} - \frac{n - m_0(x)}{1 - b(x)}\right) \\ &= \lambda\left(1 - \frac{\lambda t}{n}\right)e^{-\lambda t/n}\cdot(-\alpha)\cdot\left(1 + \frac{m(x)}{n - m(x)}\right) \leq -\alpha\lambda(1 - \lambda\varepsilon)e^{-\lambda\varepsilon}. \end{aligned}$$

Thus

$$\begin{aligned} \max_{0\leq y\leq m_0(x)}\frac{p(x+1, y)}{p(x, y)} &= \frac{p(x+1, m_0(x))}{p(x, m_0(x))} = e^{f(x+1) - f(x)} \\ &\leq e^{-\alpha\lambda e^{-\lambda\varepsilon}(1-\lambda\varepsilon)} = \gamma_{\varepsilon,\alpha}. \end{aligned}$$

□

LEMMA 8. *There exist constants $\eta = \eta(\lambda) > 0$ and $0 < \beta = \beta(\lambda) < \frac{1}{\lambda}$ such that for any $0 < \varepsilon \leq \eta$ there exists $n_0 = n_0(\varepsilon)$ such that for all $n > n_0$, we have that*

$$g(x+1) \leq \beta g(x) \qquad \forall x \geq 0.$$

PROOF. Recall that $g(x) = \mathbb{P}_x[T_0 < T_{\varepsilon n}^+]$. It follows immediately that $g(x) = 0$ for $x \geq \varepsilon n$. Therefore, we only consider $0 \leq x < \varepsilon n$.

We have by the Markov property, for $x + 1 < \varepsilon n$,

$$\begin{aligned} g(x+1) &= \sum_y p(x+1, y)g(y) \\ &\leq \sum_{y\leq m_0(x)} p(x+1, y)g(y) + \sum_{m_0(x)<y<\varepsilon n} p(x+1, y)g(y). \end{aligned} \tag{4}$$

We bound the first term in (4) using Lemma 7:

$$\sum_{y\leq m_0(x)} p(x+1, y)g(y) \leq \gamma_{\varepsilon,\alpha}\cdot\sum_{y\leq m_0(x)} p(x, y)g(y) \leq \gamma_{\varepsilon,\alpha}\cdot g(x).$$

For the second term, we use the upper bound of Lemma 6 to obtain

$$\sum_{m_0(x)<y<\varepsilon n} p(x+1,y)g(y) \le \sum_{m_0(x)<y<\varepsilon n} p(x+1,y)\cdot \frac{q_1^y}{1-q_1^n} \le \frac{q_1^{m_0(x)}}{1-q_1^n}.$$

Also by Lemma 6,

$$\begin{aligned}\frac{q_1^{m_0(x)}}{1-q_1^n}\cdot\frac{1}{g(x)} &\le \frac{1-q_2^{\varepsilon n}}{(1-q_1^n)(1-q_2^{\varepsilon n-x})}\cdot\frac{q_1^{m_0(x)}}{q_2^x}\\ &\le \frac{1}{(1-q_1)(1-q_2)}\cdot\big(q_1^{(1-\alpha)\lambda e^{-\lambda\varepsilon}}/q_2\big)^x.\end{aligned}$$

Note that as $\varepsilon\to 0$ we have that $q_1\to q(\lambda), q_2\to q(\lambda)$ and $e^{-\lambda\varepsilon}\to 1$. Combined with the assumption that $(1-\alpha)\lambda>1$, we can deduce that there exists $\eta'>0$ such that $q_1^{(1-\alpha)\lambda e^{-\lambda\varepsilon}}/q_2$ is bounded away from 1 uniformly in $0<\varepsilon\le\eta'$. Moreover, since $\gamma_{\varepsilon,\alpha}=e^{-\alpha\lambda e^{-\lambda\varepsilon}(1-\lambda\varepsilon)}$, and since we assume that $\lambda e^{-\alpha\lambda}<1$, we may take η' small enough so that for all $0<\varepsilon\le\eta'$ we have $\lambda\gamma_{\varepsilon,\alpha}<1$. Consequently, we can find K large enough (that depends only on η') such that

$$\beta' := \sup_{\varepsilon\le\eta'}\left(\gamma_{\varepsilon,\alpha}+\frac{1}{(1-q_1)(1-q_2)}\cdot\left(\frac{q_1^{(1-\alpha)\lambda e^{-\lambda\varepsilon}}}{q_2}\right)^K\right)<\frac{1}{\lambda}.$$

Plugging all this into (4), we conclude that there exist η' and $K\ge 1$ and $\beta'<\lambda^{-1}$ such that for all $0<\varepsilon\le\eta'$ and for every $K\le x<\varepsilon n-1$, we have $h(x+1)\le\beta' h(x)$. This proves the lemma for $x\ge K$.

As for $0\le x<K$, by Lemma 6 we have

$$\frac{g(x+1)}{g(x)}\le\frac{q_1^{x+1}-q_1^n}{1-q_1^n}\cdot\frac{1-q_2^{\varepsilon n}}{q_2^x-q_2^{\varepsilon n}}\le q_1\cdot\left(\frac{q_1}{q_2}\right)^x\cdot\frac{1}{(1-q_2^{\varepsilon n-K})(1-q_1^n)}.$$

Recall that $q_1>q_2$ so $(q_1/q_2)^x\le(q_1/q_2)^K\to 1$ as $\varepsilon\to 0$. Also, $1/((1-q_2^{\varepsilon n-K})(1-q_1^n))\to 1$ as $n\to\infty$ and $\lambda q_1\to\lambda q(\lambda)<1$ as $\varepsilon\to 0$.

Therefore, we may choose η'' such that for all $0<\varepsilon\le\eta''$, $\lambda q_1\cdot(q_1/q_2)^K<\frac{\lambda q(\lambda)+1}{2}$. Thus there exists $n_0=n_0(\varepsilon)$ such that if $n\ge n_0$, we have

$$\lambda q_1\cdot\left(\frac{q_1}{q_2}\right)^K\cdot\frac{1}{(1-q_2^{\varepsilon n-K})(1-q_1^n)}<1.$$

So we can take

$$\beta''=\sup_{\varepsilon\le\eta''} q_1\cdot\left(\frac{q_1}{q_2}\right)^K\cdot\frac{1}{(1-q_2^{\varepsilon n_0-K})(1-q_1^{n_0})}$$

to obtain that $\lambda\beta''<1$, and for all $0<\varepsilon\le\eta''$, sufficiently large n and $0\le x<K$, we have $g(x+1)\le\beta'' g(x)$. Wrap up by setting $\eta=\min\{\eta',\eta''\}$ and $\beta=\max\{\beta',\beta''\}$. $\square$

3.2. *Probability of extinction before going above level $u \gg \varepsilon n$.*

LEMMA 9 (Uniform lower bound). *Fix $\lambda > 1$. There exists $\kappa = \kappa(\lambda) > 0$ such that for all $0 < u < n$ and $0 < x + 1 < u$,*

$$\mathbb{P}_{x+1}[T_0 < T_u^+] \geq \kappa \mathbb{P}_x[T_0 < T_u^+].$$

PROOF. Let $X = (X_k)_{k\geq 0}$, $Y = (Y_k)_{k\geq 0}$ be two Markov chains starting from $X_0 = x$, $Y_0 = x + 1$, respectively, and with Markov transition kernel $p(\cdot, \cdot)$. Consider the following coupling:

- if $b(x) \leq b(x + 1)$, then let $X_1 \sim \mathrm{Bin}(n, b(x))$, and given X_1,
$$Y_1 = X_1 + \mathrm{Bin}\left(n - X_1, \frac{b(x + 1) - b(x)}{1 - b(x)}\right);$$
- if $b(x) > b(x + 1)$, then let $Y_1 \sim \mathrm{Bin}(n, b(x + 1))$, and given Y_1,
$$X_1 = Y_1 + \mathrm{Bin}\left(n - Y_1, \frac{b(x) - b(x + 1)}{1 - b(x + 1)}\right).$$

Next, given X_k, Y_k for $k \geq 1$, if $X_k = Y_k$, then couple $X_{k+1} = Y_{k+1}$, and otherwise let X_{k+1}, Y_{k+1} evolve independently. Note that $X_1 = Y_1$ implies $X_k = Y_k$ for all $k \in \mathbb{N}$.

By the mean value theorem,

$$\left|b(x + 1) - b(x)\right| \leq \sup_{y\in[x,x+1]} \left|b'(y)\right| = \sup_{y\in[x,x+1]} \left|\frac{\lambda}{n} e^{-\lambda y/n}\left(1 - \frac{\lambda y}{n}\right)\right| \leq \frac{\lambda}{n},$$

and since $b(z) \leq e^{-1}$, we get that $\frac{b(x+1)-b(x)}{1-b(x)}, \frac{b(x)-b(x+1)}{1-b(x+1)} \leq \frac{e\lambda}{(e-1)n}$. We have

$$\begin{aligned}\frac{\mathbb{P}[Y \in \{T_0 < T_u^+\}]}{\mathbb{P}[X \in \{T_0 < T_u^+\}]} &\geq \frac{\mathbb{P}[Y_1 = X_1, X \in \{T_0 < T_u^+\}]}{\mathbb{P}[X \in \{T_0 < T_u^+\}]} \\ &= \mathbb{P}[Y_1 = X_1 | X \in \{T_0 < T_u^+\}].\end{aligned}$$

Now, if $b(x) \leq b(x + 1)$, then as $n \to \infty$,

$$\begin{aligned}&\mathbb{P}[Y_1 = X_1 | X \in \{T_0 < T_u^+\}] \\ &\quad\geq \sum_k \mathbb{P}[X_1 = k | X \in \{T_0 < T_u^+\}] \cdot \mathbb{P}\left[\mathrm{Bin}\left(n - k, \frac{b(x + 1) - b(x)}{1 - b(x)}\right) = 0\right] \\ &\quad\geq \mathbb{P}\left[\mathrm{Bin}\left(n, \frac{e\lambda}{(e - 1)n}\right) = 0\right] \to e^{-e\lambda/(e-1)} > 0.\end{aligned}$$

Similarly when $b(x) > b(x + 1)$,

$$\mathbb{P}[Y_1 = X_1 | X \in \{T_0 < T_u^+\}] \geq \mathbb{P}\left[\mathrm{Bin}\left(n, \frac{e\lambda}{(e - 1)n}\right) = 0\right].$$

We may take $\kappa := \inf_n \mathbb{P}[\mathrm{Bin}(n, \frac{e\lambda}{(e-1)n}) = 0]$ to complete the proof. $\square$

LEMMA 10 (Geometric upper bound). *Fix $\lambda > 1$ and $\varepsilon > 0$ small. Then there exists $\theta = \theta(\lambda, \varepsilon) \in (0, 1)$ such that for all $0 \le x < u \le \mathrm{eq} - \varepsilon n$,*

$$\mathbb{P}_x[T_0 < T_u^+] \le \theta^x.$$

PROOF. Consider the probability generating function of a $\mathrm{Poi}(e^{\lambda\varepsilon})$ random variable. Since $e^{\lambda\varepsilon} > 1$, this function has a unique nontrivial fixed point $0 < \theta < 1$, satisfying $\theta = e^{-e^{\lambda\varepsilon}(1-\theta)}$. [Here $\theta = q(e^{\lambda\varepsilon})$ is the probability of extinction of a Galton–Watson process with Poisson-$e^{\lambda\varepsilon}$ offspring distribution.]

Note that for any $0 \le x < u \le \mathrm{eq} - \varepsilon n$, we have $\lambda e^{-\lambda x/n} \ge e^{\lambda\varepsilon}$, so

$$\begin{aligned}\mathbb{E}[\theta^{X_{k+1}} | X_k = x] &= (1 - b(x)(1-\theta))^n \le e^{-nb(x)(1-\theta)} \\ &= [e^{-\lambda e^{-\lambda x/n}(1-\theta)}]^x \le [e^{-e^{\lambda\varepsilon}(1-\theta)}]^x = \theta^x.\end{aligned}$$

This implies $(\theta^{X_k})_{k=0}^{T_u^+}$ is a supermartingale. Since it is bounded, and $T_0 \wedge T_u^+$ is a.s. finite, we may apply the optional stopping theorem to this supermartingale with $X_0 = x$ for some $0 \le x < u$. We obtain

$$\begin{aligned}\theta^x &\ge \mathbb{E}_x[\theta^{X_{T_0 \wedge T_u^+}}] = \mathbb{E}_x[\theta^{X_{T_0}} \mathbb{1}_{\{T_0 < T_u^+\}}] + \mathbb{E}_x[\theta^{X_{T_u^+}} \mathbb{1}_{\{T_0 > T_u^+\}}] \\ &\ge \mathbb{P}_x[T_0 < T_u^+].\end{aligned}$$

□

3.3. *Coupling with subcritical branching process.* We want to investigate what our process behaves like when conditioned on $T_0 < T_u^+$ for $u = \varepsilon n$ and $u = \mathrm{eq} - \varepsilon n$. Let $\varphi(x) = \varphi_u(x) := \mathbb{P}_x[T_0 < T_u^+]$. We denote the transition matrix of the tilted chain as $p_\varphi(\cdot, \cdot)$ which is obtained by applying Doob's h-transform to the original transition matrix $p(\cdot, \cdot)$ w.r.t. the harmonic function φ. The matrix p_φ is given by

$$p_\varphi(x, y) = \mathbb{P}_x[X_1 = y | T_0 < T_u^+] = \frac{\mathbb{P}_x[X_1 = y, T_0 < T_u^+]}{\mathbb{P}_x[T_0 < T_u^+]} = \frac{\varphi(y) p(x, y)}{\varphi(x)}. \tag{5}$$

LEMMA 11. *Let $0 < u < n$ and $\varphi(x) = \mathbb{P}_x[T_0 < T_u^+]$. Suppose $\varphi(y+1) \le \beta\varphi(y)$ for some $\beta > 0$ and for all $y \ge 0$. Then for any $0 \le x < u$, the probability measure $p_\varphi(x, \cdot)$ is stochastically dominated by the probability measure μ_x, where*

$$\mu_x(y) \propto \beta^y p(x, y), \qquad y \ge 0.$$

PROOF. Let $Y \sim p_\varphi(x, \cdot)$, $Z \sim \mu_x$. We need to show that for any $0 \le x < u$ and $k \ge 0$, $\mathbb{P}[Y \le k] \ge \mathbb{P}[Z \le k]$, or equivalently,

$$\frac{\sum_{y=0}^k p(x, y)\varphi(y)}{\varphi(x)} - \frac{\sum_{y=0}^k p(x, y)\beta^y}{\sum_{z=0}^\infty p(x, z)\beta^z} \ge 0.$$

Since $\varphi(x)$ is harmonic with respect to $p(\cdot,\cdot)$, this is equivalent to showing that

$$\frac{\sum_{y=0}^{k} p(x,y)\varphi(y)}{\sum_{z=0}^{\infty} p(x,z)\varphi(z)} - \frac{\sum_{y=0}^{k} p(x,y)\beta^{y}}{\sum_{z=0}^{\infty} p(x,z)\beta^{z}} \geq 0.$$

So we write

$$\begin{aligned}\sum_{y=0}^{k} p(x,y)\varphi(y)\sum_{z=0}^{\infty} p(x,z)\beta^{z} - \sum_{y=0}^{k} p(x,y)\beta^{y}\sum_{z=0}^{\infty} p(x,z)\varphi(z) \\ = \sum_{y=0}^{k}\sum_{z=k+1}^{\infty} p(x,y)p(x,z)\big(\varphi(y)\beta^{z} - \varphi(z)\beta^{y}\big).\end{aligned}$$

Note that for $0 \leq y < z$, by the assumption,

$$\varphi(z)\beta^{y} \leq \varphi(y)\beta^{z-y}\beta^{y} = \varphi(y)\beta^{z},$$

which implies that each term of the above sum is nonnegative. □

LEMMA 12. *Let $0 < u < n$ and $\varphi(x) = \mathbb{P}_x[T_0 < T_u^{+}]$. Suppose $\varphi(y) \geq \kappa\varphi(y-1)$ for some $\kappa > 0$ and for all $0 < y < u$. Then for any $0 \leq x < u$, the probability measure $p_\varphi(x,\cdot)$ stochastically dominates the probability measure ν_x, where*

$$\nu_x(y) \propto \kappa^{y} p(x,y)\mathbb{1}_{\{y<u\}}.$$

PROOF. The proof is exactly similar to that of Lemma 11 where we replace "∞" in the bounds of the summands by $u-1$. We omit the details. □

LEMMA 13. (a) *Fix $0 < p < 1$. Let $X \sim \mathrm{Bin}(n,p)$ and $Y \sim \mathrm{Poi}(-n\log(1-p))$. Then $X \leq_{\mathrm{st}} Y$.*

(b) *Fix $0 < p_1 < p_2 < 1$. Let $X \sim \mathrm{Bin}(n,p_1)$ and $X \sim \mathrm{Bin}(n,p_2)$. Then for any $m \geq 0$, we have $X|\{X \leq m\} \leq_{\mathrm{st}} Y|\{Y \leq m\}$.*

PROOF. (a) We exhibit a coupling such that $X \leq Y$. Note that $X = X_1 + \cdots + X_n$ where $X_1,\dots,X_n$ are i.i.d. $\mathrm{Ber}(p)$ random variables, and $Y = Y_1 + \cdots + Y_n$ where $Y_1,\dots,Y_n$ are i.i.d. $\mathrm{Poi}(-\log(1-p))$ random variables. So let $Y_1,\dots,Y_n$ be as such, and let $X_j := \mathbb{1}_{\{Y_j>0\}}$. It follows that $X_j \leq Y_j$ and $\mathbb{P}[X_j = 0] = \mathbb{P}[Y_j = 0] = (1-p)$, suggesting that indeed $X_1,\dots,X_n$ are i.i.d. $\mathrm{Ber}(P)$, and $X \leq Y$ a.s.

(b) It suffices to show for any $k \leq m$,

$$\frac{\mathbb{P}[X \leq k]}{\mathbb{P}[X \leq m]} \geq \frac{\mathbb{P}[Y \leq k]}{\mathbb{P}[Y \leq m]},$$

which, in turn, is implied by $\mathbb{P}[X = i]\mathbb{P}[Y = j] \geq \mathbb{P}[X = j]\mathbb{P}[Y = i]$ for all $0 \leq i < j \leq m$. Upon rearrangement of terms, the above is equivalent to

$$\frac{(p_1/(1-p_1))^{i}}{(p_2/(1-p_2))^{i}} \geq \frac{(p_1/(1-p_1))^{j}}{(p_2/(1-p_2))^{j}},$$

which is obviously true. □

LEMMA 14. *Denote the Markov chain conditioned on $T_0 < T_{\varepsilon n}^+$ by $(X'_m)_{m\geq 0}$. There exists $\varepsilon_0 > 0$ and $0 < \bar{\gamma} < 1$ such that for all $\varepsilon < \varepsilon_0$ there exists $n_0 = n_0(\varepsilon)$ such that for all $n > n_0$ the following holds. For any $0 \leq x_0 < \varepsilon n$, we can couple $(X'_m)_{m\geq 0}$ with a sub-critical Galton–Watson process $(W_m)_{m\geq 0}$ having offspring distribution* $\mathrm{Poi}(\bar{\gamma})$, *such that*

$$X'_0 = W_0 = x_0, \qquad X'_m \leq W_m \qquad \forall m \geq 1.$$

PROOF. Recall that the transition matrix of the conditioned chain X' is given by $p_\varphi(\cdot,\cdot)$ with $\varphi(x) = \mathbb{P}_x[T_0 < T_{\varepsilon n}^+]$. It suffices to show that for any nonnegative integers $x \leq w$, $p_\varphi(x,\cdot)$ is stochastically dominated by a $\mathrm{Poi}(w\gamma)$ random variable.

By Lemmas 8 and 11, we know that for small enough $\varepsilon > 0$ and large enough n, $p_\varphi(x,\cdot)$ is stochastically dominated by $\mu_x(\cdot)$. Fix $0 \leq x < \varepsilon n$. Note that

$$\mu_x(y) \propto \beta^y p(x,y) \propto \binom{n}{y}\left(\frac{\beta b(x)}{1-b(x)}\right)^y, \qquad 0 \leq y \leq n,$$

which implies that $\mu_x(\cdot)$ is binomially distributed with n trials and success probability $\theta(x)$ that satisfies

$$\frac{\theta(x)}{1-\theta(x)} = \frac{\beta b(x)}{1-b(x)} \quad \text{or} \quad \theta(x) = \frac{\beta b(x)}{1-b(x)(1-\beta)}.$$

Further, by Lemma 13(a), $\mu_x(\cdot)$ is stochastically dominated by a Poisson random variable with mean $g(x) = -n\log(1-\theta(x))$. Note that $b(x) \leq e^{-1} < \frac{1}{2}$. So $\beta b(x) < 1 - b(x)$, and thus $\theta < \frac{1}{2}$. Also, one easily checks that $-\log(1-t) \leq t + 2t^2$ for $t \in [0, \frac{1}{2}]$. We thus obtain $g(x) \leq n\theta(x)(1 + 2\theta(x))$. Recall that $b(x) = \frac{\lambda x}{n}e^{-\lambda x/n}$ is monotone on $[0, \varepsilon n]$ so for $x < \varepsilon n$, we have $b(x) \leq \lambda\varepsilon$. Hence

$$g(x) \leq x \cdot \frac{\beta\lambda}{1-\lambda\varepsilon} \cdot \left(1 + \frac{2\lambda\varepsilon}{1-\lambda\varepsilon}\right).$$

Denote $\bar{\gamma} := \frac{\beta\lambda}{1-\lambda\varepsilon}(1 + \frac{\lambda\varepsilon}{1-\lambda\varepsilon})$. Note that for $\varepsilon \to 0$ we have that $\bar{\gamma} \to \lambda\beta < 1$ by Lemma 8. So choose ε_0 small enough so that $\bar{\gamma} < 1$. Putting all the ingredients together, we get that for all $0 \leq x < \varepsilon n$,

$$p_\varphi(x,\cdot) \leq_{\mathrm{st}} \mathrm{Poi}(x\bar{\gamma}),$$

and keeping in mind that for any two nonnegative integers $x \leq w$ we have $\mathrm{Poi}(x\bar{\gamma}) \leq_{\mathrm{st}} \mathrm{Poi}(w\bar{\gamma})$, the proof is complete. $\square$

COROLLARY 15. *Fix $\lambda > 1$. There exist $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$ there exists $C > 0$ such that the following holds for all $n \geq 1$:*

$$\mathbb{E}_x[T_0 | T_0 < T_{\varepsilon n}^+] \leq C\log(1+x), \qquad 0 \leq x < \varepsilon n.$$

LEMMA 16. *Denote the Markov chain conditioned on* $T_0 < T^+_{\mathrm{eq}-\varepsilon n}$ *by* $(X''_m)_{m\geq 0}$. *Given* $\lambda > 1$ *and* $0 < \varepsilon < \frac{\log\lambda}{\lambda}$, *there exists* $0 < \underline{\gamma} < 1$ *such that the following holds. For any* $0 \leq x_0 < \mathrm{eq} - \varepsilon n$, *we can couple* $(X''_m)_{m\geq 0}$ *with a subcritical Galton–Watson process* $(V_m)_{m\geq 0}$ *having offspring distribution* $\mathrm{Ber}(\underline{\gamma})$, *such that with probability at least* $1 - e^{-(1-\kappa)^2(\lambda^{-1}\log\lambda-\varepsilon)^2 n}$,

$$X''_0 = V_0 = x_0, \qquad X''_m \geq V_m \qquad \forall 1 \leq m \leq e^{(1-\kappa)^2(\lambda^{-1}\log\lambda-\varepsilon)^2 n},$$

where $\kappa \in (0, 1)$ *is as given in Lemma 9 with* $u = \mathrm{eq} - \varepsilon n$.

PROOF. By Lemma 9 and Lemma 12, for any $0 < x < \mathrm{eq} - \varepsilon n$, the transition distribution $p_\varphi(x, \cdot)$ stochastically dominates $\nu_x(y) \propto \binom{n}{y}(\frac{b(x)\kappa}{1-b(x)})^y \mathbb{1}_{\{y<\mathrm{eq}-\varepsilon n\}}$. In other words,

$$Y|\{Y < \mathrm{eq} - \varepsilon n\} \leq_{\mathrm{st}} p_\varphi(x, \cdot),$$

where Y is distributed as $\mathrm{Bin}(n, \theta(x))$, where $\theta(x) = \frac{\kappa b(x)}{1-b(x)(1-\kappa)}$. By Lemma 13(b) and from the simple inequality $\theta(x) \geq \frac{\kappa x}{n}$, we further have $Z|\{Z < \mathrm{eq} - \varepsilon n\} \leq_{\mathrm{st}} p_\varphi(x, \cdot)$, where Z is distributed as $\mathrm{Bin}(n, \frac{\kappa x}{n})$. Clearly, $\sum_{i=1}^x Z_i \leq_{\mathrm{st}} Z$ where Z_i are i.i.d. $\mathrm{Bin}(\lfloor\frac{n}{x}\rfloor, \frac{\kappa x}{n})$. We can find $\underline{\gamma} < 1$ such that for any $n \geq 1$ and any $0 < x < \mathrm{eq} - \varepsilon n$,

$$\mathbb{P}[Z_i = 0] = \left(1 - \frac{\kappa x}{n}\right)^{\lfloor n/x\rfloor} \leq 1 - \underline{\gamma}.$$

Consequently, Z_i stochastically dominates $\mathrm{Ber}(\underline{\gamma})$ and hence $\mathrm{Bin}(x, \underline{\gamma}) \leq_{\mathrm{st}} Z$.

On the other hand, by Hoeffding's inequality,

$$\mathbb{P}[Z \geq \mathrm{eq} - \varepsilon n] \leq \exp\bigl(-2(1-\kappa)^2\bigl(\lambda^{-1}\log\lambda - \varepsilon\bigr)^2 n\bigr).$$

Thus, on the event $\{Z < \mathrm{eq} - \varepsilon n\}$, which happens with probability at least $1 - \exp(-2(1-\kappa)^2(\lambda^{-1}\log\lambda - \varepsilon)^2 n)$, the distribution $p_\varphi(x, \cdot)$ stochastically dominates $\mathrm{Bin}(x, \underline{\gamma})$. So, a simple union bound allows us to couple the conditioned chain X'' with a subcritical Galton–Watson process having offspring distribution $\mathrm{Ber}(\underline{\gamma})$ so that with probability at least $1 - \exp(-(1-\kappa)^2(\lambda^{-1}\log\lambda - \varepsilon)^2 n)$, the subcritical Galton–Watson process is dominated by X'' for the first $\exp((1-\kappa)^2(\lambda^{-1}\log\lambda - \varepsilon)^2 n)$ steps. This completes the proof. □

COROLLARY 17 (Lower bound on transition window). *Given* $\lambda > 1$ *and* $0 < \varepsilon < \frac{\log\lambda}{\lambda}$, *there exists* $C > 0$ *such that the following holds for all* $n \geq 1$:

$$\mathbb{E}_x\bigl[T_0 | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr] \geq C^{-1}\log(1 + x), \qquad 0 \leq x < \mathrm{eq} - \varepsilon n.$$

PROOF. By Lemma 16, the conditioned chain X'' can be coupled with the subcritical Galton–Watson process V with mean offspring $\underline{\gamma}$ such that $X''_0 = V_0 = x$

and $X''_t \geq V_t$ for all $1 \leq t \leq e^{cn}$ with probability at least $1 - e^{-cn}$. Let S be the time of extinction for the process V. It follows from the standard theory of branching process that $\mathbb{E}_x[S] \geq c' \log(1+x)$ for all $x \geq 0$. On the other hand,

$$\mathbb{P}_x[S > t] \leq x\mathbb{E}_1[V_t] \leq x\underline{\gamma}^t.$$

By choosing $t = D \log n$ with a sufficiently large constant $D = D(\underline{\gamma}) > 0$, we obtain that $E_x[S\mathbb{1}_{\{S \leq D\log n\}}] \geq \frac{c'}{2} \log(1+x)$ for all $0 \leq x \leq n$. By the above coupling,

$$\mathbb{E}_x\big[T_0 | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\big] \geq \mathbb{E}_x[S\mathbb{1}_{\{S \leq D\log n\}}] - D \log n \cdot e^{-cn},$$

which completes the proof. $\square$

4. Upper bound on extinction window. Throughout this section we set $\varepsilon > 0$ and $h(x) = \mathbb{P}_x[T_0 < T^+_{\mathrm{eq}-\varepsilon n}]$.

The next lemma bootstraps the result from Corollary 15.

LEMMA 18. *Given $\lambda > 1$ and $\varepsilon > 0$, there exist $C > 0$ and $\eta > 0$ such that for all $0 \leq x < \eta n$,*

$$\mathbb{E}_x\big[T_0 | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\big] \leq C \log(1+x).$$

PROOF. By Lemmas 9 and 10, there exist $\kappa < 1$ and $\theta < 1$ such that for all $0 \leq x < \mathrm{eq} - \varepsilon n$, $\kappa^x \leq h(x) \leq \theta^x$. We choose $r > 1$ so that $\theta^{r/2} < \frac{\kappa}{2}$.

By Corollary 15, there exist $C > 0$, $\eta' > 0$ small enough such that for all $x < \eta' n$,

$$\mathbb{E}_x\big[T_0 | T_0 < T^+_{\eta' n}\big] \leq C \log(1+x).$$

Then, taking $\eta = \eta'/r$, by the strong Markov property, for any $0 \leq x < \eta n$,

$$\begin{aligned}\mathbb{P}_x\big[T^+_{\eta' n} < T_0 < T^+_{\mathrm{eq}-\varepsilon n}\big] &\leq \sup_{x \geq \eta' n} h(x) \leq \theta^{r\eta n} < \left(\frac{\kappa}{2}\right)^{2\eta n} \\ &\leq \inf_{x \leq \eta n} h(x)^2 \cdot 2^{-2\eta n}.\end{aligned}$$

Note that since for $x \leq \mathrm{eq} - \varepsilon n$,

$$nb(x) = x \cdot \lambda e^{-\lambda x/n} \geq e^{\lambda\varepsilon} \cdot x,$$

we have that if $X_k < u - \varepsilon n$, then $\mathbb{E}[X_{k+1}|X_k] \geq e^{\lambda\varepsilon} \cdot X_k$. Let $A = e^{\lambda\varepsilon/2} > 1$, and let $p = \exp(-\frac{(A-1)^2}{4}) < 1$. By standard large deviations of binomial random variables,

$$\begin{aligned}\mathbb{P}_x[X_1 \leq Ax] \leq \mathbb{P}_x\big[X_1 \leq A^{-1}\mathbb{E}_x[X_1]\big] &\leq \exp\left(-\frac{(1-A^{-1})^2}{4} \cdot \mathbb{E}_x[X_1]\right) \\ &\leq \exp\left(-\frac{(A-1)^2}{4} \cdot x\right) = p^x.\end{aligned}$$

Let $m > 1$ be an integer such that $A^m \geq e$. Then

$$\mathbb{P}_x\big[T_{ex}^+ \leq m\big] \geq \mathbb{P}_x\big[\forall 1 \leq j \leq m \wedge T_{ex}^+, X_j \geq AX_{j-1}\big] \geq \big(1 - p^x\big)^m.$$

This holds for any x. Thus inductively, for $k = \lceil \log n \rceil$,

$$\begin{aligned}\mathbb{P}_x\big[T_{\mathrm{eq}-\varepsilon n}^+ \leq km\big] &\geq \mathbb{P}_x\big[T_{ex}^+ \leq m\big] \cdot \inf_{y \geq ex} \mathbb{P}_y\big[T_{\mathrm{eq}-\varepsilon n}^+ \leq (k-1)m\big] \\ &\geq \cdots \geq (1-p)^{m \log n}.\end{aligned}$$

Again this holds for all $x > 0$. Thus $\frac{1}{km} \cdot (T_0 \wedge T_{\mathrm{eq}-\varepsilon n}^+)$ is dominated by a geometric random variable of mean $(1-p)^{-m \log n}$. So we conclude that

$$\mathbb{E}_x\big[\big(T_0 \wedge T_{\mathrm{eq}-\varepsilon n}^+\big)^2\big] \leq (km)^2 \cdot 2(1-p)^{-2\log n},$$

which is polynomial in n as $n \to \infty$.

Let $\mathcal{E}(u)$ denote the event $\{T_0 < T_u^+\}$. Putting everything together, we obtain that for $0 < x < \eta n$,

$$\begin{aligned}&\mathbb{E}_x\big[T_0 | \mathcal{E}(\mathrm{eq} - \varepsilon n)\big] \\ &\quad = \mathbb{E}_x\big[T_0 \mathbb{1}_{\mathcal{E}(\eta' n)} | \mathcal{E}(\mathrm{eq} - \varepsilon n)\big] \\ &\qquad + \mathbb{E}_x\big[T_0 \wedge T_{\mathrm{eq}-\varepsilon n}^+ \cdot \mathbb{1}_{\{T_{\eta' n}^+ < T_0 < T_{\mathrm{eq}-\varepsilon n}^+\}} | \mathcal{E}(\mathrm{eq} - \varepsilon n)\big] \\ &\quad \leq \mathbb{E}_x\big[T_0 | \mathcal{E}(\eta' n)\big] \cdot \frac{\mathbb{P}_x[\mathcal{E}(\eta' n)]}{\mathbb{P}_x[\mathcal{E}(\mathrm{eq} - \varepsilon n)]} \\ &\qquad + \frac{\sqrt{\mathbb{E}_x[(T_0 \wedge T_{\mathrm{eq}-\varepsilon n}^+)^2] \cdot \mathbb{P}_x[T_{\eta' n} < T_0 < T_{\mathrm{eq}-\varepsilon n}^+]}}{h(x)} \\ &\quad \leq C \log(1+x) + km\sqrt{2}(1-p)^{-\log n} \cdot 2^{-\eta n} \leq C' \log(1+x),\end{aligned}$$

for some constant $C' > 0$. $\square$

LEMMA 19. *Given $\lambda > 1$ and $\varepsilon, \delta > 0$, there exists a constant $C = C(\varepsilon, \lambda, \delta) > 0$ such that for all $0 \leq x < \mathrm{eq} - \varepsilon n$,*

$$\mathbb{E}_x\big[H | T_0 < T_{\mathrm{eq}-\varepsilon n}^+\big] \leq C,$$

where $H := \sum_{k=0}^{\infty} \mathbb{1}_{\{\delta n < X_k < \mathrm{eq} - \varepsilon n\}}$ is the total time spent by X in the interval $(\delta n, \mathrm{eq} - \varepsilon n)$.

PROOF. We start with the observation that for $0 < x < \mathrm{eq} - \varepsilon n$, since $\lambda e^{-\lambda \mathrm{eq}/n} = 1$,

$$\frac{nb(x)}{x} = \lambda e^{-\lambda x/n} \geq \lambda e^{-(\lambda(\mathrm{eq} - \varepsilon n))/n} = e^{\lambda \varepsilon} > 1.$$

Choose $m = m(\varepsilon, \lambda, \delta) \geq 1$ large enough such that $\delta e^{(\lambda\varepsilon/2)m} > \frac{\mathrm{eq}-\varepsilon n}{n}$. Call a step k of the Markov chain X *unusual* if $\delta n < X_k < \mathrm{eq} - \varepsilon n$ and $X_{k+1} < e^{\lambda\varepsilon/2} X_k$. By standard large deviations of binomial random variables, for $0 < \xi < 1$, $\mathbb{P}[\mathrm{Bin}(n, b) < \xi nb] \leq \exp(-nb \cdot \frac{(1-\xi)^2}{4})$. Since $nb(X_k) \geq e^{\lambda\varepsilon} X_k$, the probability of step k being unusual is bounded by

$$\mathbb{P}\bigl[\mathrm{Bin}\bigl(n, b(X_k)\bigr) < e^{-\lambda\varepsilon/2} nb(X_k) | \delta n < X_k < \mathrm{eq} - \varepsilon n\bigr] \leq \exp\biggl(-\bigl(e^{\lambda\varepsilon/2} - 1\bigr)^2 \cdot \frac{\delta n}{4}\biggr).$$

Note that if $\delta n < X_j < u - \varepsilon n$ and for all $k = j, j+1, \dots, j+m-1$ the step k is *not* unusual, then by our choice of m,

$$X_{j+m} \geq e^{(\lambda\varepsilon/2)m} X_j > \frac{\mathrm{eq} - \varepsilon n}{\delta n} \cdot \delta n = \mathrm{eq} - \varepsilon n.$$

That is, if all steps $j, j+1, \dots, j+m-1$ are not unusual, then $T_0 > T^+_{\mathrm{eq}-\varepsilon n}$. Thus $T_0 < T^+_{\mathrm{eq}-\varepsilon n}$ implies that every time j that $\delta n < X_j < \mathrm{eq} - \varepsilon n$, we must have that there exists $j \leq k \leq j + m - 1$ such that k is an unusual step. In conclusion, for any $0 < x < \mathrm{eq} - \varepsilon n$,

$$\begin{aligned}\mathbb{P}_x\bigl[T_0 < T^+_{\mathrm{eq}-\varepsilon n}, H > d\bigr] &\leq \mathbb{P}_x\bigl[X \text{ takes at least } \lfloor d/m \rfloor \text{ unusual steps}\bigr] \\ &\leq \exp\biggl(-\bigl(e^{\lambda\varepsilon/2} - 1\bigr)^2 \cdot \frac{\delta}{4} \cdot \biggl\lfloor \frac{d}{m} \biggr\rfloor \cdot n\biggr).\end{aligned}$$

On the other hand, by Lemma 9, $h(x) \geq \kappa^n$. Combining the above two observations, we obtain that for any $0 < x < \mathrm{eq} - \varepsilon n$,

$$\mathbb{P}_x\bigl[H \geq d | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr] \leq e^{(K_1 - dK_2)n}$$

for constants K_1 and K_2 which are functions of δ, m, ε and λ. Now the assertion of the lemma follows immediately from the representation

$$\mathbb{E}_x\bigl[H | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr] = \sum_{d=0}^{\infty} \mathbb{P}_x\bigl[H \geq d | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr].$$ □

PROOF OF THEOREM 3. The lower bound is established in Corollary 17.

Let us now prove the upper bound. Let $\eta > 0$ be as in Lemma 18. For $x < \eta n$, Theorem 3 follows directly from Lemma 18. For $x \geq \eta n$, by the strong Markov property,

$$\mathbb{E}_x\bigl[T_0 | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr] \leq \mathbb{E}_x\bigl[T^-_{\eta n} | T_0 < T_{\mathrm{eq}-\varepsilon n}\bigr] + \sup_{y < \eta n} \mathbb{E}_y\bigl[T_0 | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr]. \tag{6}$$

We can bound $\mathbb{E}_x[T^-_{\eta n} | T_0 < T_{\mathrm{eq}-\varepsilon n}] \leq 1 + \mathbb{E}_x[H | T_0 < T_{\mathrm{eq}-\varepsilon n}]$ where $H = \sum_{k=0}^{\infty} \mathbb{1}_{\{\eta n < X_k < \mathrm{eq}-\varepsilon n\}}$, and hence by Lemma 19, $\mathbb{E}_x[T^-_{\eta n} | T_0 < T_{\mathrm{eq}-\varepsilon n}] \leq C_1$.

Therefore, from (6) and by Lemma 18,

$$\mathbb{E}_x\bigl[T_0 | T_0 < T^+_{\mathrm{eq}-\varepsilon n}\bigr] \leq C_1 + C \log(\eta n) \leq C' \log(1 + x),$$

which completes the proof of Theorem 3. □

Acknowledgments. This work was initiated when the second and third authors were in the Statistical Laboratory, University of Cambridge. Thanks goes to Nathanaël Berestycki for making this possible. The authors also thank the anonymous referee for helpful comments and suggestions.

REFERENCES

[1] ALDOUS, D. and FILL, J. (2002). Reversible Markov chains and random walks on graphs. Available at http://www.stat.berkeley.edu/~aldous/RWG/book.html.

[2] ALDOUS, D. and STEELE, J. M. (2004). The objective method: Probabilistic combinatorial optimization and local weak convergence. In *Probability on Discrete Structures*. *Encyclopaedia Math*. *Sci*. **110** 1–72. Springer, Berlin. MR2023650

[3] ATHREYA, K. B. and NEY, P. E. (1972). *Branching Processes*. Springer, New York. MR0373040

[4] BENJAMINI, I. and PERES, Y. (1994). Markov chains indexed by trees. *Ann*. *Probab*. **22** 219–243. MR1258875

[5] BENJAMINI, I. and PERES, Y. (1994). Tree-indexed random walks on groups and first passage percolation. *Probab*. *Theory Related Fields* **98** 91–112. MR1254826

[6] BIGGINS, J. D. (1977). Martingale convergence in the branching random walk. *J*. *Appl*. *Probab*. **14** 25–37. MR0433619

[7] BRAMSON, M., DING, W. D. and DURRETT, R. (1991). Annihilating branching processes. *Stochastic Process*. *Appl*. **37** 1–17. MR1091690

[8] BRAMSON, M. and GRAY, L. (1985). The survival of branching annihilating random walk. *Z*. *Wahrsch*. *Verw*. *Gebiete* **68** 447–460. MR0772192

[9] FELLER, W. (1950). *An Introduction to Probability Theory and Its Applications*. *Vol*. *I*. Wiley, New York. MR0038583

[10] FORGOSTON, E., BIANCO, S., SHAW, L. B. and SCHWARTZ, I. B. (2011). Maximal sensitive dependence and the optimal path to epidemic extinction. *Bull*. *Math*. *Biol*. **73** 495–514. MR2780552

[11] LEVIN, D. A., PERES, Y. and WILMER, E. L. (2009). *Markov Chains and Mixing Times*. Amer. Math. Soc., Providence, RI. MR2466937

[12] LYONS, R. and PERES, Y. Probability on trees and networks. Available at http://php.indiana.edu/~rdlyons/prbtree/prbtree.html.

[13] SCHWARTZ, I. B., FORGOSTON, E., BIANCO, S. and SHAW, L. B. (2011). Converging towards the optimal path to extinction. *J*. *R*. *Soc*. *Interface* **8** 1699–1707.

[14] SHI, Z. (2011). Random walks and trees. In *X Symposium on Probability and Stochastic Processes and the First Joint Meeting France–Mexico of Probability*. *ESAIM Proc*. **31** 1–39. EDP Sci., Les Ulis. MR2857100

[15] SUDBURY, A. (1990). The branching annihilating process: An interacting particle system. *Ann*. *Probab*. **18** 581–601. MR1055421

I. PERL
A. YADIN
DEPARTMENT OF MATHEMATICS
BEN-GURION UNIVERSITY OF THE NEGEV
BEER SHEVA 8410501
ISRAEL
E-MAIL: perli@post.bgu.ac.il
yadina@math.bgu.ac.il

A. SEN
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MINNESOTA 55455
USA
E-MAIL: arnab@math.umn.edu