Diffusion limit for the partner model at the critical value

Anirban Basak*   Rick Durrett [†]   Eric Foxall [‡]

Abstract

The partner model is an SIS epidemic in a population with random formation and dissolution of partnerships, and with disease transmission only occuring within partnerships. Foxall, Edwards, and van den Driessche [7] found the critical value and studied the subcritical and supercritical regimes. Recently Foxall [4] has shown that (if there are enough initial infecteds I_0) the extinction time in the critical model is of order $\sqrt{N}$. Here we improve that result by proving the convergence of $i_N(t) = I(\sqrt{N}t)/\sqrt{N}$ to a limiting diffusion. We do this by showing that within a short time, this four dimensional process collapses to two dimensions: the number of SI and II partnerships are constant multiples of the the number of infected singles. The other variable, the total number of singles, fluctuates around its equilibrium like an Ornstein-Uhlenbeck process of magnitude $\sqrt{N}$ on the original time scale and averages out of the limit theorem for $i_N(t)$. As a by-product of our proof we show that if τ_N is the extinction time of $i_N(t)$ (on the $\sqrt{N}$ time scale) then τ_N has a limit.

1 Introduction

In the partner model each of N individuals can be susceptible or infected and in a partnership or not. So the system is described by the five quantities S_t and I_t, the number of single susceptible and infected individuals, and SS_t, SI_t, and II_t, the number of partnered pairs of the three possible combinations, at time t. Infected individuals become healthy (and susceptible to re-infection) at rate 1. A susceptible individual with an infected partner becomes infected at rate λ. Partnerships dissolve at rate r_-. Each pair of single individuals forms a partnership at rate r_+/N.

---

*International Centre for Theoretical Sciences, Tata Institute of Fundamental Research, Bengaluru, India, E-mail: anirban.basak@icts.res.in

[†]Duke University, Durham, NC, United States E-mail: rtd@math.duke.edu

[‡]University of Alberta, Edmonton, AB, Canada, E-mail: efoxall@ualberta.ca
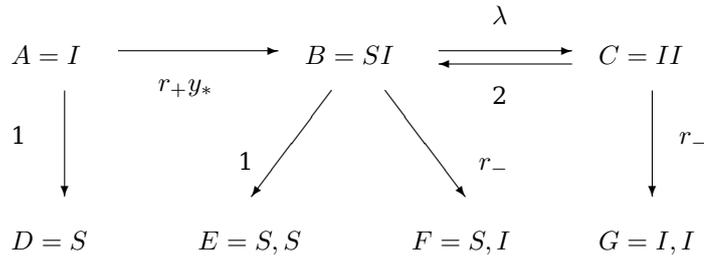
Figure 1: Markov chain for computation of R_0

Foxall, Edwards, and van den Driessche [7] introduced this model and showed that despite the complexity of the model it is possible to find the critical value explicitly. To do this they used the continuous time Markov chain X_t with state space $\{A, B, C, D, E, F, G\}$ and rates as shown in Figure 1. Thinking of a single infected individual in an otherwise susceptible population, we start in state A, and let τ be the first time X_t enters $\{D, E, F, G\}$. The basic reproduction number for the model is

$$R_0 = P_A(X_\tau = F) + 2P_A(X_\tau = G) \tag{1.1}$$

which is the expected number of infected singles at time τ. The critical value of λ is

$$\lambda_c = \sup\{\lambda \geq 0, R_0 \leq 1\}$$

with $\lambda_c < \infty$ if and only if $r_+ > 1 + 1/r_-$. There is an explicit formula for λ_c but it is not very pretty since the formulas for the hitting probabilities are somewhat complicated.

In [7] it was shown that

Theorem 1.1. *If $R_0 < 1$ there are constants T, C so that, from any initial configuration, with high probability the process dies out by time $T + C \log N$. If $R_0 > 1$, then for any $\epsilon > 0$ there are constants T, C, γ, such that from any initial configuration with at least ϵN infected, with probability at least $1 - e^{-\gamma N}$ the process survives for time $e^{\gamma N}$ and the frequencies of the five types $s_t = S_t/N, i_t = I_t/N$, etc. are within ϵ of their equilibrium values $(s_*, i_*, ss_*, si_*, ii_*)$ when $T \leq t \leq e^{\gamma N}$.*

To describe the equilibrium values we need some notation. Let $Y_t = S_t + I_t$ be the number of single individuals and $y_t = Y_t/N$. y_t approaches and remains close to a stationary value y_* which is the unique equilibrium in $(0, 1)$ for the ODE

$$y' = r_-(1 - y) - r_+ y^2$$

We will explain this result in more detail later, see (3.4). The equilibrium frequency of singles, y_*, is the solution of $r_-(1 - y_*) = r_+ y_*^2$.

To find the number of single infecteds in equilibrium we let $i_t = I_t/N$ and note the three events that affect the number of infected singles are

- $I \to S$ at rate $I_t = i_t N$,

- $I + I \to II$ at rate $(r_+/N)\binom{I_t}{2} \approx r_+(i_t^2/2)N$, and

- $S + I \to SI$ at rate $(r_+/N)S_t I_t \approx r_+ i_t s_t N$.

Fixing $i_t = i \in (0, y^*)$ we define probabilities

$$p_S = \frac{1}{z} \qquad p_{II} = \frac{r_+ i}{2z} \qquad p_{SI} = \frac{r_+(y_* - i)}{z}$$

where $z = 1 + r_+(y_* - i/2)$ is the sum of the numerators. Let

$$\Delta_S = -1 \qquad \Delta_{II} = -2 + P_C(X_\tau = F) + 2P_C(X_\tau = G)$$
$$\Delta_{SI} = -1 + P_B(X_\tau = F) + 2P_B(X_\tau = G)$$

be the expected change in the number of single infecteds (at partnership breakup) due to the three events. Finally let

$$\Delta(i) = p_S \Delta_S + p_{II} \Delta_{II} + p_{SI} \Delta_{SI}$$

be the expected change in the number of infecteds per event. In equilibrium $\Delta(i_*) = 0$. Having found i_* and $s_* = y_* - i_*$, it is routine to find ii_*, si_*, and ss_*; see Section 5 of [7] for more details. It's also worth noting that the condition $i_* = 0$ is equivalent to $R_0 = 1$.

The analysis of the critical case was done in a second paper by Foxall [4]. The main result is

Theorem 1.2. *Let V_t be the number of infected vertices at time t. If $R_0 = 1$ then*

- *there are $C, \gamma > 0$ so that from any initial configuration, with probability at least $1 - e^{-\gamma m}$, $V_{mC\sqrt{N}} = 0$, and*

- *if $V_0 \geq \sqrt{N}$ and $y_0 \geq y_* - (\log N)/\sqrt{N}$ there is $c > 0$ so that $V_{c\sqrt{N}} \neq 0$ with probability at least $1 - e^{-c(\log N)^2}$.*

The goal of this paper is to obtain a more complete description of the process in the case $R_0 = 1$. In particular, for the rest of the paper we assume that $R_0 = 1$.

Theorem 1.2 shows that the extinction time τ_N is of order $N^{1/2}$ if $V_0 \geq \sqrt{N}$. By analogy with critical branching processes one might expect the time to be of order N if $V_0 = N$ (same order as the initial values). To explain why $N^{1/2}$ is the right order of magnitude and to indicate what more precise result we would like to prove, we sketch the proof in the following simpler setting.

Example: Contact process on a complete graph with N vertices. Individuals die at rate 1, and give birth at rate β to an offspring that is sent to a randomly chosen vertex, so the number of occupied vertices $X(t)$ is a Markov chain on $\{0, 1, \ldots N\}$ with transition rates

$$q(k, k-1) = k \quad \text{and} \quad q(k, k+1) = \beta k(1 - k/N).$$

The critical value for prolonged survival is $\beta_c = 1$.

Theorem 1.3. *Let $x_N(t) = X(N^{1/2}t)/N^{1/2}$, with $\beta = 1$. Then $x_N(t) \Rightarrow x_t$, the solution of*

$$dx_t = -x_t^2 \, dt + \sqrt{2x_t} dB_t. \tag{1.2}$$

Let $\tau_0(x_N) = \inf\{t : x_N(t) = 0\}$. If $x_N(0) \to \infty$, $\tau_0(x_N) \Rightarrow \tau_0(x)$, the hitting time of 0 for the diffusion process started at ∞.

Proof. With $\beta = 1$ we find

$$\frac{d}{dt} E[X_t] = -(E[X_t^2])/N$$

and using Jensen's inequality, $E[X_t]$ satisfies the differential inequality $y' \leq -y^2/N$. Since $E[X_0] \leq N$ this gives $E[X_t] \leq N/(1+t)$, and using Markov's inequality $P(X_{\epsilon N^{1/2}} \leq \epsilon^{-2} N^{1/2}) \geq 1 - \epsilon$. Letting Let $x_N(t) = X(N^{1/2}t)/N^{1/2}$ this means that $x_N(\epsilon) \leq \epsilon^{-2}$ with probability at least $1 - \epsilon$. The drift of x_t^N is

$$N^{1/2} \frac{1}{N^{1/2}} [-x_N(t)N^{1/2} + x_N(t)N^{1/2}(1 - x_N(t)/N^{1/2})] = -(x_N(t))^2$$

while the diffusivity is

$$N^{1/2} \frac{1}{N} [x_N(t)N^{1/2} + x_N(t)N^{1/2}(1 - x_N(t)/N^{1/2})] \approx 2x_N(t)$$

(these terms are explained in Section 2). The first result then follows from Lemma 2.7 in the next section. Then, as in the proof of Theorem 1.6 in Section 4.7, to show $\tau_0(x_N) \Rightarrow \tau_0(x)$, it is enough to show that for each $\epsilon > 0$ we can find $\delta > 0$, so that if $x_N(t) \leq \delta$ then $x_N(t+\epsilon) = 0$ with probability at least $1 - \epsilon$. This is easy to do once we note that X_t is dominated by the critical branching process $\tilde{X}_t$ in which each particle splits in two, or dies, each at rate one. We know that $P(\tilde{X}_t = 0 \mid \tilde{X}_0 = k) = (1-(1+t)^{-1})^k \geq 1-k/t$, which is at least $1-\epsilon$ if we let $t = k/\epsilon$. Letting $k = \delta N^{1/2}$, the result follows with $\delta = \epsilon^2$. $\square$

In the above example, there are three main steps:

i) Show X_t comes down to $C_\epsilon \sqrt{N}$ within $\epsilon N^{1/2}$ time,
ii) Show that $x_N(t) = X_{N^{1/2}t}/N^{1/2}$ converges to a diffusion,
iii) Show that once $x_N(t)$ is small, it hits zero in a short time.

The corresponding result for the partner model follows the same three steps, but is more complicated because the process is four dimensional and there are two different time scales.

Some notation. To state our results and to avoid confusion between SI and $S \cdot I$ etc we introduce alternative notations that we will use throughout the paper: $J = II$, $K = SI$ and $L = SS$. We refer to the stochastic process (S, I, J, K, L) as the *infection process*. We outline some further notational conventions below.

Asymptotic notation. Let a_N, b_N be sequences of real numbers.

- $a_N = O(b_N)$ if $\limsup_{N\to\infty} |a_N/b_N| < \infty$.
- $a_N = \Omega(b_N)$ if $\liminf_{N\to\infty} |a_N/b_N| > 0$.
- $a_N = o(b_N)$ if $\lim_{N\to\infty} |a_N/b_N| = 0$.
- $a_N = \omega(b_N)$ if $\lim_{N\to\infty} |a_N/b_N| = \infty$.

Moreover, for efficiency of notation we will say that a certain property holds for $o(f(N)) \leq t \leq \omega(f(N))$ if there exist sequences a_N, b_N with $a_N = o(f(N))$ and $b_N = \omega(f(N))$ such that the property holds for all $t \in [a_N, b_N]$.

Rescaling. Throughout the paper, the placement of the time variable, as for example I_t or $I(t)$, is chosen according to notational convenience and does not change the meaning. On the other hand, we will often want to rescale in either time or space, so we introduce the following notation for these purposes. When we need to distinguish the spatial scale, upper case is reserved for the originally defined variables S, I, J, K, L and $Y = S + I$, and lower case denotes the following:

$$s_t^N = S_t/N, \ y_t^N = Y_t/N,$$
$$i_t^N = I_t/\sqrt{N}, \ j_t^N = J_t/\sqrt{N}, \ \text{and} \ k_t^N = K_t/\sqrt{N}.$$

Note that I, J, K are rescaled by $1/\sqrt{N}$, and not by $1/N$ as in Theorem 1.1; as demonstrated by Theorem 1.3 this rescaling is more appropriate when $R_0 = 1$. To distinguish time scales we note that two scales will be relevant: the original time scale that we call fast, and the $\sqrt{N}$ time scale that we call slow, or long. We will use the superscript ^{N} for the fast time scale and the subscript $_N$ for the slow time scale. So, for example,

$$i_t^N = I_t/\sqrt{N} \ \text{and} \ i_N(t) = I(\sqrt{N}t)/\sqrt{N}.$$

This distinction appears in discussion as well as computation: for example, saying that i_N reaches a certain value within $O(1)$ amount of time is the same as saying that i^N reaches that value in $O(N^{1/2})$ time. For the sake of consistency, and to distinguish from limit processes, we will write S^N, I^N, etc. for the originally defined variables, on the original time scale. A few more processes will be introduced later, and they will follow the same notation for time scale; some will have upper and lower case versions, again to denote different spatial scales; the precise scaling is specified in each case.

Times. We will use lower-case s or t for time variables, T to denote a fixed (deterministic) time and τ for stopping times. For an $\mathbb{R}$-valued process X, we let

$$\tau_x^-(X, t) = \inf\{s \geq t \colon X_s \leq x\} \quad \text{and} \quad \tau_y^+(X, t) = \inf\{s \geq t \colon X_s > y\},$$

and if X is $\mathbb{R}_+$-valued we write $\tau_0(X, t)$ for $\tau_0^-(X, t)$. $\tau_x^-(X)$ denotes $\tau_x^-(X, 0)$ and similarly in other cases. In one case we will need the following:

$$\tau_{x,y}(X, t) = \inf\{s \geq t \colon X_t \leq x \text{ or } X_t > y\}.$$

We will also define some labelled and unlabelled times such as $\tau, \tau^*, \tau_1, \tau_2, \ldots$ when proving specific results, if they do not fit the above template, or to save on notation if they are written frequently.

We first describe the limit processes, followed by statements of the main results, and then we provide the workflow. Throughout the paper, if we say a statement holds with high probability (whp), then it has probability tending to 1 as $N \to \infty$.

Deterministic limits. We show in Section 3 that as $N \to \infty$, sample paths of y_t^N and (i_t^N, j_t^N, k_t^N) converge in distribution to deterministic limits y_t and (i_t, j_t, k_t). We find that $\lim_{t \to \infty} y_t = y_*$, the solution of $r_+ y_*^2 = (1 - y_*) r_-$, while (i_t, j_t, k_t) converges as $t \to \infty$ to a point on the ray $(\alpha, \beta, 1)\mathbb{R}_+$ of fixed points for the linear system described in (3.7), where α, β are given by (3.11). This is reminiscent of (multiplicative) state space collapse in queueing networks where a vector of queue lengths are all proportional to one of them. There are many results of this type. For examples, see [8, 1, 21, 20, 19].

Diffusion limits. We will show that the fluctuations $z_t^N = \sqrt{N}(y_t^N - y_*)$ are approximately an Ornstein-Uhlenbeck process

$$dz = -\mu_z z \, dt + \sigma_z \, dB$$

that evolves on the fast time scale, where μ_z, σ_z are some positive constants. On the other hand, the fluctuations of (i^N, j^N, k^N) (once they are close to the ray) occur on the slow time scale. Since it stays close to a ray in phase space, (i^N, j^N, k^N) is effectively one-dimensional, and we will show that $i_N(t) = I^N(\sqrt{N}t)/\sqrt{N}$ converges to the limit in (1.2) but with different constants for the mean and variance. As in the previous result, the hitting time of 0 converges. More precisely, we prove the following results, which are the main goal of this article. As pointed out earlier, $R_0 = 1$ is assumed throughout; for clarity, we recall this assumption in each of our main results.

Theorem 1.4. *Suppose that* $R_0 = 1$, $|(i_N(0), j_N(0), k_N(0)) - (\alpha x, \beta x, x)| = O(N^{-\epsilon})$ *for some* $x, \epsilon > 0$ *and* $|z_N(0)| = O(1)$ *as* $N \to \infty$, *where* α *and* β *are defined in* (3.11). *Then there are constants* $\mu_X, \sigma_X^2 > 0$ *such that for any fixed* $T > 0$, i_N *converges in distribution in* $C[0, T]$ *to the diffusion*

$$dX_t = -\mu_X X_t^2 dt + \sigma_X \sqrt{X_t} dB_t \tag{1.3}$$

started from $X_0 = \alpha x$.

It is possible to compute the constants μ_X, σ_X from our proof; since the expressions are not particularly nice-looking and do not add much insight, we have omitted them. It

is possible the assumptions on the rate of convergence of initial data may be relaxed; to do so one would require a more careful account of transient behavior. Since this paper is already lengthy, we have not pursued this extension.

Our next result builds on Theorem 1.4 and allows the process to start from ∞. In the proof of Theorem 1.5, which is in Section 4, we also point out why it makes sense to start the limiting diffusion (1.3) from ∞.

Theorem 1.5. *Suppose that $R_0 = 1$, $i_N(0) + j_N(0) + k_N(0) \to \infty$ and $y_N(0) \to y_*$ as $N \to \infty$. Then i_N converges in distribution in $C[\epsilon, T]$ for any fixed $0 < \epsilon < T < \infty$ to the diffusion* (1.3), *started from $X_0 = \infty$.*

It is possible that in Theorem 1.5 the assumption $y_N(0) \to y_*$ can be dropped. We do not pursue this direction in this paper. Lastly we show convergence of the hitting time.

Theorem 1.6. *Under the assumptions of Theorem 1.4 or 1.5, for*

$$\tau_0(i_N + j_N + k_N) = \inf\{t : (i_N(t), j_N(t), k_N(t)) = (0, 0, 0)\}$$

we have $\tau_0(i_N + j_N + k_N) \Rightarrow \tau_0(X)$, the time to hit zero for the limiting diffusion (1.3).

The separation of time scales between y^N and the infection variables i^N, j^N, k^N may remind the reader of the work of Kang and Kurtz [12] and Kang, Kurtz, and Popovic [13] on chemical reaction networks. We found that writing our model in their framework does not simplify the difficult aspects of the proof, so we have opted instead to use a general result of [3] for obtaining the diffusion limit, which is tailored to our context in Lemma 2.7.

Workflow. There are seven main steps, described in greater detail in Section 4. In order to make certain estimates it is helpful to define the following additional observables:

i) The positive linear combination $H_t^N = I_t^N + \gamma J_t^N + \eta K_t^N$ and $h_t^N = H_t^N / \sqrt{N}$, where $(1, \gamma, \eta)$, defined by (4.4), is a left eigenvector for the matrix A given by (3.6), that determines the linear system (3.7) for (i_t, j_t, k_t). The variable H^N is helpful to the analysis because the linear terms drop out of the equation for its drift.

ii) The rescaled infection vector $(U^N, V^N, W^N) = (I^N, \gamma J^N, \eta K^N)/H^N$, and its meta-stable equilibrium value (u_*, v_*, w_*) which is given by (7.3).

iii) The quantity $Q^N = \theta_2(U^N - u_*)^2 + \theta_1(V^N - v_*)^2$, where θ_1, θ_2, given by (7.6), are well-chosen positive constants. Q^N measures the deviation of (U^N, V^N, W^N) from equilibrium, so when Q^N is small, (i_t^N, j_t^N, k_t^N) is close to the invariant ray, a property which is essential to obtaining a limiting 1-dimensional equation for the diffusion.

Of course, $H_N(t) = H^N(\sqrt{N}t)$, etc. The main steps, written in terms of the slow time scale, are sketched below in the context of Theorem 1.5, when $h_N(0) = \omega(1)$; for Theorem 1.4, we just need that if $|z_N(0)|, h_N(0)$ and $Q_N(0)$ are small, then they can be kept small for $\omega(1)$ amount of time.

1. **Lemma 4.1.** Show that $|z_N(t)| = O(\sqrt{\log N})$ for $o(1) \le t \le \omega(1)$.
2. **Lemma 4.2.** Show that $h_N(t) = O(\log N)$ for $o(1) \le t \le \omega(1)$.
3. **Lemma 4.3.** Show that $|Q_N(t)| = O(N^{-1/6})$ for $o(1) \le t \le \omega(1)$.
4. **Lemma 4.6.** Show that the integral of $i_N(t)z_N(t)$ averages to zero on finite time intervals.
5. **Lemma 4.7.** Show for any $\epsilon > 0$ there is $C_\epsilon > 0$ so that $P(\tau_{C_\epsilon}^-(h_N) \le \epsilon) \ge 1 - \epsilon$.
6. **Lemma 4.8, Theorems 1.4 and 1.5.** Show convergence of $i_N(t)$ to the diffusion limit.
7. **Lemma 4.9.** Show that $\tau_0(h_N) \to 0$ in probability as $h_N(0) \to 0$, uniformly for large N.

The organization of the rest of the paper is as follows: In Section 2 we gather some probability estimates and limit theorems that are used throughout the paper. In Section 3 we describe the deterministic limits. In Section 4 we state precise lemmas relating to each of the workflow steps, and prove diffusion limits. The latter sections are devoted to proofs of the lemmas.

2 Sample path estimation

In this section we describe some sample path estimates, a diffusion limit theorem, and results on drift and diffusivity of functions of continuous time Markov chains that are used throughout the paper. Any results that are not cited are proved in the Appendix. The natural setting for these results is semimartingales. For an overview of the semimartingale theory that is used here, we refer the reader to [9].

First we recall some standard definitions from semimartingale theory, noting along the way how the present context fits into this framework.

Let $(\Omega, \mathcal{F}, \mathbf{F}, \mathbf{P})$ be a filtered probability space that satisfy the usual conditions. This means that the filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ is right continuous in the sense that $\mathcal{F}_t = \bigwedge_{s>t} \mathcal{F}_s$ for each t, and each $\mathcal{F}_t$ in the filtration $\mathbf{F} = (\mathcal{F}_t)_{t \in \mathbb{R}_+}$ contains the $\mathbf{P}$ null sets of $\mathcal{F}$. In [9] it is also assumed that $\mathcal{F}$ is $\mathbf{P}$ complete; if this is not the case then it is easy to check that completing $\mathcal{F}$ and then $\mathbf{F}$ with respect to null sets does not violate right continuity. In our case, the filtered probability space is that of the finite state continuous-time Markov chain corresponding to the state variables $(S^N, I^N, J^N, K^N, L^N)$, with the completion of the natural filtration. Since such a process is Feller, as shown in [18, I.5], the corresponding filtered space satisfies the usual conditions.

A procss X is called *optional* if it is measurable with respect to the σ-field (on $\Omega \times \mathbb{R}_+$) generated by all càdlàg adapted processes. All the processes considered here are optional. We assume the reader is familiar with the notions of stopping time, predictable time and process, localization and martingale.

Given a stochastic process X we denote by X_- the left-continuous process obtained from X. We further let $\Delta X = X - X_-$ denote the process of jumps. We say that X has *bounded jumps* if $|\Delta X| \leq c$ a.s. for some constant $c > 0$, and let $\Delta_\star(X)$ denote the infimum of such values of c. X is *quasi-left continuous* (qlc) if $\Delta X_\tau = 0$ a.s. on $\{\tau < \infty\}$ for any predictable time τ.

Given a process A, define the process $\mathrm{Var}\,(A)$ by setting $\mathrm{Var}\,(A)_t(\omega)$ equal to the total variation of the function $s \mapsto A_s(\omega)$ on the interval $[0,t]$. A process A has *finite variation* if $\mathrm{Var}\,(A)_t(\omega) < \infty$ for each t, ω, and is *locally integrable* if it has a localizing sequence (τ_n) such that $E[\mathrm{Var}\,(A)_{\tau_n}] < \infty$ for each n. The *compensator* of a locally integrable process A, denoted A^p, is the unique predictable and locally integrable process such that $A - A^p$ is a local martingale (see [9, I.3.18]).

A *semimartingale* (s-m) X is a process that can be written as $X = X_0 + M + A$, where X_0 is an $\mathcal{F}_0$-measurable random variable, M is a local martingale and A has finite variation. We call a semimartingale *special* if it can be written as

$$X = X_0 + X^m + X^p \tag{2.1}$$

where X^p is the compensator of X and X^m is a uniquely defined local martingale satisfying $X_0^m = 0$. By [9, I.4.24], if X has bounded jumps then it is special and $|\Delta X^m| \leq 2\Delta_\star(X)$, and if it also qlc then using [9, I.2.35] in the proof of [9, I.4.24], we have the more convenient estimate $\Delta_\star(X^m) \leq \Delta_\star(X)$.

Recall that if a martingale M is locally square-integrable then M^2 has a compensator, denoted $\langle M \rangle$ and called the predictable quadratic variation (pqv). Any local martingale M with $M_0 = 0$ and bounded jumps is locally square integrable (see [9, I.4.1]). If X is

a special s-m and X^m is locally square-integrable we will use $\langle X \rangle$ to denote $\langle X^m \rangle$. The following basic estimate is used repeatedly throughout the paper.

Lemma 2.1 ([6, Lemma 3]). *Let X be a semimartingale with bounded jumps and let $a, \phi > 0$.*

$$\text{If} \quad 0 < \phi \Delta_\star(X) \leq 1/2 \quad \text{then} \quad P\left(\sup_{t \geq 0} \{|X_t^m| - \phi \langle X \rangle_t\} \geq a \right) \leq 2e^{-\phi a}. \qquad (2.2)$$

In order to obtain tidy expressions for X^p and $\langle X \rangle$, we define quasi-absolute continuity, together with drift and diffusivity. To motivate the name, note that if X is a special semimartingale with X^m locally square-integrable, then as shown in [6, Lemma 2], X is qlc iff both X^p and $\langle X \rangle$ are continuous.

Definition 2.2. *A special semimartingale X with locally square-integrable martingale part X^m is called* quasi-absolutely continuous *(qac) if both X^p and $\langle X \rangle$ are absolutely continuous. In this case define the* drift $\mu(X) = (\mu_t(X))_t$ *and the* diffusivity $\sigma^2(X) = (\sigma_t(X))_t$ *for Lebesgue-a.e. t by*

$$\mu_t(X) = \frac{d}{dt} X_t^p, \quad \sigma_t^2(X) = \frac{d}{dt} \langle X \rangle_t. \qquad (2.3)$$

Letting $X = (S^N, I^N, J^N, K^N, L^N)$ denote the infection process, we note that the estimates of Lemma 2.1 and the upcoming Lemma 2.5 will be applied to observables of the form $f(X_t^\tau)$, where $f : R \to \mathbb{R}$ is Lipschitz, $R \subset \mathbb{R}^5$ is compact, and τ is a stopping time so that the stopped process $X_t^\tau = X_{t \wedge \tau}$ has $X_t^\tau \in R$ for $t \in \mathbb{R}_+$. Any such process is right continuous and has finite variation so is a s-m, and has bounded jumps since this is the case for X and since f is Lipschitz. Let $q_i(x)$, $i = 1, \ldots, n$ denote the rates of the various transitions ($S + I \to K, K \to J, J \to K$ etc.) and Δ_i their effect (for example, S^N, I^N decreases by 1, K^N increases by 1), and let $Y_t = f(X_t^\tau)$. By writing Y_t as a sum of jumps and using the standard linear and quadratic martingales for Poisson processes, it is easy to show that Y is qac and has

$$\mu_t(Y) = \mathbf{1}(t < \tau) \sum_{i=1}^{n} q_i(X_t)(f(X_t + \Delta_i) - f(X_t)),$$

$$\sigma_t^2(Y) = \mathbf{1}(t < \tau) \sum_{i=1}^{n} q_i(X_t)(f(X_t + \Delta_i) - f(X_t))^2,$$

a fact that we use ubiquitously when invoking Lemma 2.1 below. In a few cases, we are interested in computing the drift for products of such processes, as well as processes of the form $f(X_t^\tau) g(t)$, where $g(t)$ is an absolutely continuous deterministic function (which is easily shown to be a qac s-m). Below we state a result from [6] that allows to find the drift for products of qac s-m.

Lemma 2.3 (Product rule). *Suppose X_t, Y_t are qac semimartingales on a common filtered probability space. Then $\langle X^m, Y^m \rangle$ and $(XY)^p$ exist and are absolutely continuous. Denote $\mu_t(XY) = \frac{d}{dt}(XY)_t^p$. Then*

$$\mu(XY) = \sigma(X, Y) + X_- \mu(Y) + Y_- \mu(X),$$

where $\sigma_t(X, Y) = \frac{d}{dt} \langle X^m, Y^m \rangle_t$.

The next result will allow us to obtain simple estimates for the drift of certain processes. As in [9, I.3.4], if H is optional and A has locally finite variation, we let

$$(H \cdot A)_t(\omega) = \int_0^t H_s(\omega) dA_s(\omega),$$

for $\omega \in \Omega$ such that the above can be evaluated as a Stieltjes integral; a sufficient condition is that H also has finite variation. Of course, if A is absolutely continuous then $H \cdot A$ is differentiable and $\frac{d}{dt}(H \cdot A)_t = H_t \frac{d}{dt} A_t$ for Lebesgue-a.e. t. As noted in [9, I.3.18], if H is predictable and both $A, H \cdot A$ are locally integrable then $H \cdot A^p = (H \cdot A)^p$, which we use below.

Lemma 2.4 (Taylor approximation)**.** *Let X be a qac s-m with bounded jumps and let $f \in C^2(\mathbb{R})$. Then, $f(X)$ is a qac s-m and satisfies the following inequality for Lebesgue-a.e. t:*

$$|\mu_t(f(X)) - f'(X_t)\mu_t(X)| \leq \frac{1}{2}\sigma_t^2(X) \sup_{|x - X_t| \leq \Delta_\star(X)} |f''(x)|.$$

Building on Lemma 2.1 we derive the following simple result, which is the basis of several estimates in this article.

Lemma 2.5 (Drift barrier)**.** *Fix $x > 0$ and let X be a qac s-m on $\mathbb{R}$ with bounded jumps, such that $\Delta_\star(X) \leq x/2$. Suppose there are positive reals $\mu_\star, \sigma_\star^2, C_{\mu_\star}, C_\Delta$ with $\max\{\Delta_\star(X)\mu_\star/\sigma_\star^2, 1/2\} \leq C_\Delta$ so that if $0 < X_t < x$ then*

$$\mu_t(X) \leq -\mu_\star, \quad |\mu_t(X)| \leq C_{\mu_\star} \quad and \quad \sigma_t^2(X) \leq \sigma_\star^2. \tag{2.4}$$

Let $\Gamma = \exp(\mu_\star x/(32 C_\Delta \sigma_\star^2))$. Then we have

$$P\left(\sup_{t \leq \lfloor \Gamma \rfloor x/16 C_{\mu_\star}} X_t \geq x \;\mid\; X_0 \leq x/2\right) \leq 4/\Gamma. \tag{2.5}$$

We will occasionally need to apply Lemma 2.5 to a stopped process, for which the following easy corollary will be helpful.

Corollary 2.6 (Drift barrier with stopped process)**.** *In the setting of Lemma 2.5, let τ be a stopping time and suppose that (2.4) holds assuming $t < \tau$ in addition to $0 < X_t < x$. Then*

$$P\left(\sup_{t \leq \tau \wedge \lfloor \Gamma \rfloor x/16 C_{\mu_\star}} X_t \geq x \;\mid\; X_0 \leq x/2\right) \leq 4/\Gamma.$$

The final result of this section is about identifying the diffusion limit (or an ODE limit) for a stochastic process. Before stating the result let us introduce some more notations. A stochastic process $X_t = (X_{t,1}, \ldots, X_{t,d})$ is an $\mathbb{R}^d$-valued semimartingale if each component is a s-m. The compensator (when it exists) of X_t is defined component-wise and the predictable quadratic variation (when exists) is a $d \times d$ matrix-valued process with $\langle X \rangle_{t,ij} = \langle X_i^m, X_j^m \rangle_t$. Drift and diffusivity are similarly defined. Here we use Theorem 4.1 in [3] to obtain easily checkable conditions for convergence to an ODE or diffusion limit.

Lemma 2.7 (Diffusion limit)**.** *Let X_t^N be a sequence of semimartingales, a be a Lipschitz $d \times d$ matrix-valued function on $\mathbb{R}^d$ and $b : \mathbb{R}^d \to \mathbb{R}^d$ be Lipschitz. Suppose that $\Delta_\star(X^N) \to 0$ as $N \to \infty$. Also assume that for each $T, R > 0$,*

$$\sup_{t \leq T \cap \tau_R^+(|X^N|)} |(X_i^N)^p - \int_0^t b(X_{i,s}^N)ds| \to 0, \tag{2.6}$$

$$\sup_{t \leq T \cap \tau_R^+(|X^N|)} |(\langle X_i^N, X_j^N \rangle_t - \int_0^t a_{ij}(XX^\top)ds| \to 0, \tag{2.7}$$

as $N \to \infty$, where the convergence holds in probability, and $\tau_R^+(|X^N|) = \inf\{t : |X_t^N| \geq R\}$. Suppose $X_0^N \to x \in \mathbb{R}^d$. Then X_t^N converges in distribution to the diffusion process x with

$$x_0 = x \quad and \quad dx = b(x)dt + a(x)dB.$$

In particular, if $a = 0$ then X^N converges to the solution of the ODE system with

$$x_0 = x \quad \text{and} \quad x' = b(x).$$

The above statements also hold if X^N are qac with drift and diffusivity given by functions μ_N, σ_N^2 satisfying

$$\sup_{|x| \leq R} |\mu_N(x) - b(x)|, |\sigma_N^2(xx^\top) - a(xx^\top)| \to 0.$$

3 Deterministic limits

The goal of this section is show that on the fast time scale, y^N and (i^N, j^N, k^N) have deterministic limits, and to compute a relevant eigenvector for later on. In particular, none on the theorems of this section are used elsewhere in the paper; they are simply provided for context. It is worth noting that the limit for (i^N, j^N, k^N) is a linear system, and only comes into force once y^N is close to its equilibrium value y_*.

Recall that, in terms of original notation, $J^N = II$, $K^N = SI$ and $L^N = SS$. From the transition rates of the partner model we get the following equations for the drift (see [7, Section 5] for a detailed derivation):

$$\mu(S^N) = 2r_- L^N + r_- K^N - 2\frac{r_+}{N} \cdot \frac{S^N(S^N - 1)}{2} - \frac{r_+}{N} S^N I^N + I^N$$

$$\mu(I^N) = 2r_- J^N + r_- K^N - 2\frac{r_+}{N} \cdot \frac{I^N(I^N - 1)}{2} - \frac{r_+}{N} S^N I^N - I^N$$

$$\mu(J^N) = -r_- J^N + \frac{r_+}{N} \cdot \frac{I^N(I^N - 1)}{2} - 2J^N + \lambda K^N \tag{3.1}$$

$$\mu(K^N) = -r_- K^N + \frac{r_+}{N} \cdot S^N I^N + 2J^N - (\lambda + 1)K^N$$

$$\mu(L^N) = -r_- L^N + \frac{r_+}{N} \cdot \frac{S^N(S^N - 1)}{2} + K^N$$

Recalling that $Y^N = S^N + I^N$ is the number of unpartnered individuals and considering only the rate of partnership formation and dissolution, we obtain

$$\mu(Y^N) = r_-(N - Y^N) - \frac{r_+}{N} Y^N(Y^N - 1). \tag{3.2}$$

Since $y_t^N = Y_t^N / N$,

$$\mu(y^N) = r_-(1 - y^N) - r_+(y^N)^2 + O(1/N).$$

Since transition rates are $O(N)$ and y^N jumps by $\pm 2/N$ we have $\sigma^2(y^N) \leq CN/N^2 = C/N = o(1)$, for some absolute constant C. Using Lemma 2.7 we obtain the following result.

Theorem 3.1. *If $y_0^N \to y_0$ as $N \to \infty$ then $y_t^N \Rightarrow y_t$, the solution to the initial value problem with y_0 and*

$$\frac{dy}{dt} = r_-(1 - y) - r_+ y^2 \tag{3.3}$$

From the limiting differential equation, we see that in equilibrium

$$0 = r_-(1 - y_*) - r_+ y_*^2. \tag{3.4}$$

Since $S^N + I^N + 2(J^N + K^N + L^N) = N$ and S^N is accounted for by Y^N, the three remaining equations are those for I^N, J^N and K^N. Since Y^N tends towards Ny_* it is helpful to introduce the variable $Z^N = Y^N - Ny_*$ to describe its distance from equilibrium. Most of the terms in these equations are linear, with the exception of terms

involving $S_t I_t$ and $I_t(I_t - 1)$. Writing $S^N = Y^N - I^N = Ny_* + Z^N - I^N$, from (3.1) we find that

$$\begin{pmatrix} \mu(I^N) \\ \mu(J^N) \\ \mu(K^N) \end{pmatrix} = A \begin{pmatrix} I^N \\ J^N \\ K^N \end{pmatrix} + r_+ \frac{Z^N I^N}{N} \begin{pmatrix} -1 \\ 0 \\ 1 \end{pmatrix} + \frac{r_+(I^N)^2}{N} \begin{pmatrix} 0 \\ 1/2 \\ -1 \end{pmatrix} + \frac{r_+ I^N}{N} \begin{pmatrix} 1 \\ -1/2 \\ 0 \end{pmatrix} \quad (3.5)$$

where the matrix A is given by

$$A = \begin{pmatrix} -(r_+ y_* + 1) & 2r_- & r_- \\ 0 & -(r_- + 2) & \lambda \\ r_+ y_* & 2 & -(r_- + \lambda + 1) \end{pmatrix}. \quad (3.6)$$

Since I^N, J^N, K^N jump by $O(1)$ at rate $O(I^N + J^N + K^N)$, the diffusivity matrix has entries of size $O(I^N + J^N + K^N)$.

Results in [4] suggest, and our results will show that for any $\epsilon > 0$, after $\epsilon N^{1/2}$ time I^N, J^N, and K^N are $O(\sqrt{N})$. Looking at (3.5), if $I^N + J^N + K^N$ and Z^N are $o(1)$ then the entries of the diffusivity matrix, as well as the non-linear terms in the drift, are $o(I^N + J^N + K^N)$. Theorem 3.1 implies that if $Z_0^N = o(N)$ then $Z_s^N = o(N)$ uniformly on (fixed) finite time intervals $[0, t]$. Thus, if we rescale to $i_t^N = I_t^N / \sqrt{N}$, $j_t^N = J_t^N / \sqrt{N}$ and $k_t^N = K_t^N / \sqrt{N}$ (in fact any rescaling by $1/f(N)$ where $f(N) \to \infty$ will work but $f(N) = \sqrt{N}$ is the most relevant to the main results) and use Lemma 2.7 we obtain the following result.

Theorem 3.2. *If $y_0^N \to y_*$ and $(i_0^N, j_0^N, k_0^N) \to (i_0, j_0, k_0)$ as $N \to \infty$ then $(i_t^N, j_t^N, k_t^N) \Rightarrow (i_t, j_t, k_t)$, the solution to the initial value problem with (i_0, j_0, k_0) and*

$$\begin{aligned} \frac{di}{dt} &= -(r_+ y_* + 1)i + 2r_- j + r_- k \\ \frac{dj}{dt} &= -r_- j - 2j + \lambda k \\ \frac{dk}{dt} &= r_+ y_* \cdot i + 2j - (r_- + \lambda + 1)k \end{aligned} \quad (3.7)$$

Theorem 3.2 is not used elsewhere in the paper; it only motivates the following calculations. To analyze the limit behavior of (3.7), which is a linear system, we write the condition for an equilibrium as $A(i, j, k)^\mathsf{T} = 0$. To have a nontrivial solution we need $\det(A) = 0$. Expanding around the first row

$$\begin{aligned} \det(A) &= -(r_+ y_* + 1)[(r_- + 2)(r_- + \lambda + 1) - 2\lambda] \\ &\quad - 2r_-(-\lambda)r_+ y_* + r_-(r_- + 2)r_+ y_* \end{aligned} \quad (3.8)$$

For $\det(A) = 0$ we need

$$(r_+ y_* + 1)[2 + (3 + \lambda)r_- + r_-^2] = r_+ y_* r_-(r_- + 2 + 2\lambda) \quad (3.9)$$

In Section 11 we show that (3.9) is equivalent to $R_0 = 1$. This indeed shows that there is a non-trivial equilibrium. We now proceed to find the solution. To have $A(i, j, k)^\mathsf{T} = 0$ we must have

$$\begin{aligned} -(r_+ y_* + 1)i + 2r_- j + r_- k &= 0 \\ -(r_- + 2)j + \lambda k &= 0 \\ r_+ y_* i + 2j - (r_- + \lambda + 1)k &= 0 \end{aligned} \quad (3.10)$$

The second equation implies $j = \lambda k / (r_- + 2)$. Using this in the first equation we want

$$-(r_+ y_* + 1)i + \frac{2r_- \lambda k}{r_- + 2} + r_- k = 0.$$

Solving we see that if

$$\alpha = \frac{r_-}{r_+ y_* + 1}\left(\frac{2\lambda}{r_- + 2} + 1\right) \qquad \text{and} \qquad \beta = \frac{\lambda}{r_- + 2} \qquad (3.11)$$

then the ray $(\alpha z, \beta z, z)$, $z \geq 0$ is invariant for the dynamical system (3.7).

To prove the dynamical system converges to this ray we note that

$$\theta I - A = \begin{pmatrix} \theta + (r_+ y_* + 1) & -2r_- & -r_- \\ 0 & \theta + (r_- + 2) & -\lambda \\ -r_+ y_* & -2 & \theta + (r_- + \lambda + 1) \end{pmatrix}.$$

The eigenvalues of A are the roots of $0 = \det(\theta I - A) = \theta^3 + b_1 \theta^2 + b_2 \theta + b_3$, where

$$b_1 = \text{trace}(-A) = (r_+ y_* + 1) + (r_- + 2) + r_- + (\lambda + 1)$$

and b_2 is the sum of the 2×2 principal minors of $-A$, which is given by

$$(r_+ y_* + 1)(r_- + 2) + (r_- + 2)(r_- + \lambda + 1) - 2\lambda + (r_+ y_* + 1)(r_- + \lambda + 1) - r_- r_+ y_*.$$

Note that the above is positive since each of the two negative terms are cancelled by a part of the positive term that precedes it. Since $b_3 = \det(-A) = 0$ and $b_1 b_2 - b_3 > 0$ one can use the Routh Hurwicz conditions to conclude that the other two eigenvalues have negative real part. Alternatively one can observe that the non-zero roots of the equation $\det(\theta I - A)$ are

$$\frac{-b_1 \pm \sqrt{b_1^2 - 4b_2^2}}{2}.$$

Therefore the dynamical system (3.7) indeed converges to the invariant ray $(\alpha z, \beta z, z)$, $z \geq 0$. A quantitative statement that applies to the infection process is given in Lemma 4.3.

4 Workflow and diffusion limits

In this section we give a precise statement of the lemmas corresponding to the workflow steps outlined at the end of the Introduction, and use these lemmas to prove the main results, Theorems 1.4, 1.5, and 1.6. The lemmas are listed in the same numerical order as in the Introduction, which is also the order in which they will be proved in subsequent sections.

4.1 Step 1: the number of singles Y_t

Recall that (see (3.1)-(3.2)) on the original time scale

$$Y^N \to Y^N + 2 \quad \text{at rate} \quad r_-(N - Y^N)/2$$
$$Y^N \to Y^N - 2 \quad \text{at rate} \quad r_+ Y^N(Y^N - 1)/2N = r_+(Y^N)^2/2N + O(1).$$

If we let $z_t^N = N^{-1/2}(Y_t^N - Ny_*)$, that is, $Y_t^N = Ny_* + N^{1/2}z_t^N$ then

$$z^N \to z^N + 2/N^{1/2} \quad \text{at rate} \quad r_-[(1 - y_*)N - z^N N^{1/2}]/2 \qquad (4.1)$$
$$z^N \to z^N - 2/N^{1/2} \quad \text{at rate} \quad Nr_+(y_* + z^N N^{-1/2})^2/2 + O(1)$$
$$= Nr_+ y_*^2/2 + r_+ y_* z^N N^{1/2} + O(1 \vee (z^N)^2).$$

Since $r_-(1 - y_*) = r_+ y_*^2$,

$$\mu(z^N) = -\mu_z z^N + O((1 \vee (z^N)^2)/N^{1/2}) \quad \text{where} \quad \mu_z = r_- + 2r_+ y_* \qquad (4.2)$$

and

$$\sigma^2(z^N) = \sigma_z^2 + O(N^{-1} \vee z^N/N^{1/2}) \quad \text{where} \quad \sigma_z^2 = 2[r_-(1 - y_*) + r_+ y_*^2]. \tag{4.3}$$

So, if $z^N = o(N^{1/4})$, then using Lemma 2.7, $z_t^N \Rightarrow z_t$ which satisfies

$$dz = -\mu_z z \, dt + \sigma_z \, dB$$

i.e., the limit z_t is an Ornstein-Uhlenbeck process. To control the behavior of z_t^N we will prove the following two facts.

Lemma 4.1 (Step 1). *There is a constant $C_{4.1}$ so that with high probability,*

- $|Z_t^N| \leq C_{4.1}$ *for some* $t \leq C_{4.1} \log N$ *and*

- *if* $|z_0^N| \leq (C_{4.1}/2)\sqrt{\log N}$ *then* $|z_t^N| \leq C_{4.1}\sqrt{\log N}$ *for all* $t \leq N$.

To understand the $\sqrt{\log N}$ scaling note that the stationary distribution of the Ornstein-Uhlenbeck process is a normal, whose tail scales roughly like $\exp(-x^2/2\sigma^2)/x$; using the heuristic that the time to reach a rare set scales like the reciprocal of its stationary probability and letting x be a constant times $\sqrt{\log N}$, we find the time to reach level x scales roughly like N^p for constant p, which (with $p = 1$) is the time scale on which we control z_t^N.

4.2 Step 2: a special linear combination of (I^N, J^N, K^N)

In Section 3 we showed that (i^N, j^N, k^N) converges quickly (in $O(1)$ time) to the invariant ray (3.11) for the ODE (3.7). Thus the knowledge of one component determines the other two, provided we have good control on the distance of the triple from the invariant ray. Recall that in the example from the Introduction (contact process on a complete graph at criticality), the negative drift of X_t brings it down to the natural spatial scale for the diffusion, $C_\epsilon N^{1/2}$, within $\epsilon N^{1/2}$ time. This suggests that in our model, we should look for a linear combination of (I^N, J^N, K^N) that has negative drift when it takes values that are $\omega(N^{1/2})$.

Motivated by these observations we introduce the variable $H^N = I^N + \gamma J^N + \eta K^N$ where $(1, \gamma, \eta)A = 0$ and A is given by (3.6). Existence is guaranteed since $\det(A) = 0$, and the desired constants satisfy

$$-(r_+ y_* + 1) + r_+ y_* \eta = 0$$
$$2r_- - (r_- + 2)\gamma + 2\eta = 0 \tag{4.4}$$
$$r_- + \lambda\gamma - (r_- + \lambda + 1)\eta = 0$$

Solving for η in the first equation, and γ in the second, we find that

$$\eta = (r_+ y_* + 1)/r_+ y_*$$
$$\gamma = \left(\frac{2r_-}{r_- + 2} + \eta \cdot \frac{2}{r_- + 2} \right) \tag{4.5}$$

Clearly, $\eta > 1$, which implies $\gamma > 1$. By assumption, $R_0 = 1$, so in the notation of [7] we have $\lambda = \lambda_c$. Since λ is finite it follows from [7, Theorem 2.1] that $r_+ y_* > 1$ and so $\eta < 2$, which easily implies $\gamma < 2$. We record these in a display equation for later use:

$$1 < \eta < 2 \quad \text{and} \quad 1 < \gamma < 2. \tag{4.6}$$

We now compute the drift of H^N, using (3.5). From our choice of linear combination, the linear part drop out, and only the fluctuation term with Z^N, the quadratic part, and the lower order term remain:

$$\mu(H^N) = (\eta - 1)r_+ \frac{Z^N I^N}{N} - (\eta - \gamma/2)r_+ \frac{(I^N)^2}{N} + (1 - \gamma/2)r_+ \frac{I^N}{N}. \tag{4.7}$$

This gets a bit nicer if we rescale in space and time. With $h_N(t) = H^N(\sqrt{N}t)/\sqrt{N}$ we have

$$\mu(h_N) = (\eta - 1)r_+ z_N i_N - (\eta - \gamma/2)r_+ i_N^2 + (1 - \gamma/2)i_N/\sqrt{N} \tag{4.8}$$

From (4.6) we have $\gamma/2 < 1 < \eta$ so the coefficient in the second term is negative. If $i_N = \omega(1)$ the second term dominates the third. Using the bound $|z^N| = O(\sqrt{\log N})$, the second term dominates the first term when $i_N = \omega(\sqrt{\log N})$. Thus, to obtain a closed-form differential inequality for $E[h_N]$ which is useful when $h_N = \omega(\sqrt{\log N})$ it would be enough to show that I^N/H^N is bounded away from zero after a short time, which is done in Section 6. Writing $\tau^-_{N^{1/5}}(H_N) = \inf\{t : H_N(t) \le N^{1/5}\}$ we will prove the following result. Note we are using the slow time scale here.

Lemma 4.2 (Step 2). *Let $\tau = \tau^-_{N^{1/5}}(H_N) \wedge \tau^+_{C_{4.1}\sqrt{\log N}}(|z_N|)$. With high probability,*

- *$h_N(t) \le \frac{1}{2}\log N$ for some $t \le 1/\sqrt{\log N}$ and*

- *if $h_N(0) \le \frac{1}{2}\log N$ then $h_N(t) \le \log N$ for all $t \le N^{1/2} \wedge \tau$.*

The choice of $N^{1/5}$ as a floor on H_N is so that once $H_N \le N^{1/5}$ a branching process approximation can be used to take H_N to 0. See Section 10.

Our reliance on Lemma 4.1 to bound the first part of the drift prevents us from showing h_N comes all the way down to $O(1)$. To obtain the stronger bound we will have to show that the first term in the drift in (4.8) averages out to 0.

4.3 Step 3: (I^N, J^N, K^N) stays close to the invariant ray

To reduce (I^N, J^N, K^N) to a one-dimensional system we let $U^N = I^N/H^N$, $V^N = \gamma J^N/H^N$, and $W^N = \eta K^N/H^N$. The coefficients in V^N and W^N are there to make $U^N + V^N + W^N = 1$. Recalling the definitions of α, β, γ, and η (see (3.11) and (4.4)) we let $u_* = \alpha/d$, $v_* = \beta\gamma/d$, and $w_* = \eta/d$ where $d = \alpha + \beta\gamma + \eta$ which, as the reader will see, is the fixed point for the dynamical system corresponding to (U^N, V^N, W^N). Let $Q^N = \theta_2(U^N - u_*)^2 + \theta_1(V^N - v_*)^2$. This result is stated on the fast time scale, since this is the time scale on which Q^N naturally converges.

Lemma 4.3 (Step 3). *Let $\tau = \tau^-_{N^{1/5}}(H^N) \wedge \tau^+_{C_{4.1}\sqrt{\log N}}(|z^N|) \wedge \tau^+_{\log N}(h^N)$. There is a constant $C_{4.3}$ so that, for any sequence of constants c^Q_N with $N^{-1/6} \le c^Q_N = o(1)$, with high probability,*

- *$\tau^-_{N^{-1/6}/2}(Q^N) \wedge \tau \le C_{4.3}\log N$, and*

- *if $Q^N_0 \le c^Q_N/2$ then $Q^N_t \le c^Q_N$ for all $t \le N \wedge \tau$.*

Steps 1,2 and 3 could be called the "a priori" bounds, since they provide the control needed to implement the averaging result. With this in mind, we make the following definition. Additional constants for both h and Q are specified since we'll need them later.

Definition 4.4. *Let c^Q_N, c^h_N be sequences of constants. Say that there is c^h_N, c^Q_N control at time t, on the slow time scale, if*

$$|z_N(t)| \le C_{4.1}\sqrt{\log N}, \quad h_N(t) \le c^h_N,$$
$$Q_N(t) \le c^Q_N \text{ and } H_N(t) > N^{1/5}.$$

Define the c^h_N, c^Q_N-control time on the slow time scale as

$$\tau_N(c^h_N, c^Q_N, t) = \inf\{s \ge t : \text{ there is not } c^h_N, c^Q_N \text{ control at time } s\}.$$

*Define the control time $\tau_N(\mathtt{ctrl}, t)$ as $\tau_N(\mathtt{ctrl}, t) = \tau_N(\log N, N^{-1/6}, t)$. Define $\tau^N(c_N^h,$
$c_N^Q, t)$ on the fast time scale by $\sqrt{N}\tau_N(c_N^h, c_N^Q, t)$, similarly for $\tau^N(\mathtt{ctrl}, t)$.*

Applying Steps 1-3 (i.e., Lemmas 4.1, 4.2 and 4.3) sequentially in time, we obtain the following result, which is stated on the slow time scale.

Lemma 4.5. *For each $\epsilon > 0$, with high probability*

$$\tau_{N^{1/5}}^-(H_N) \wedge N^{1/2} \le \tau_{N^{1/5}}^-(H_N) \wedge \tau_N(\mathtt{ctrl}, \epsilon).$$

In words, this result says that for any fixed $\epsilon > 0$, so long as H_N has not hit the interval $[0, N^{1/5}]$, then w.h.p. the variables $|z_N|, Q_N$ and h_N have the desired upper bounds (i.e., those specified by Definition 4.4) on the slow time scale, on the time interval $[\epsilon, N^{1/2}]$.

4.4 Step 4: averaging the drift to 0

Recall from Section 4.1 that z^N is approximately an O.U. process that oscillates on the fast time scale, and once (i^N, j^N, k^N) converges on the invariant ray, we expect it to diffuse along that ray on the slow time scale and thus move slowly when viewed on the fast time scale. Thus, it should not be surprising that we have the following result.

Lemma 4.6 (Step 4). *Fix $T < \infty$. Let $L : \mathbb{R}^3 \to \mathbb{R}$ be Lipschitz in the ℓ^1 norm with constant c_L and such that $L(0,0,0) = 0$. Let $\omega(N^{-1/2}) = c_N^h \le \log N$ and $0 < c_N^Q = o(1)$ be sequences of constants and let $\tau_N = \tau_N(c_N^h, c_N^Q, 0)$. With high probability,*

$$\sup_{t \le T}\left|\int_0^{t \wedge \tau_N} z_N(s) L(i_N(s), j_N(s), k_N(s))\, ds\right| = O(c_L(c_N^h)^{1/2}(N^{-1/4} \vee (c_N^h c_N^Q)^{1/2})\log^2 N).$$

4.5 Step 5: Show h_N comes down to $O(1)$

In order to prove Theorem 1.5 we need to show that w.h.p., $h_N(t) = O(1)$ for fixed $t > 0$.

Lemma 4.7 (Step 5). *Let $\tau_C^-(h_N) = \inf\{t\colon h_N(t) \le C\}$. For small enough $\epsilon > 0$,*

$$\lim_{C \to \infty} \limsup_N P(\epsilon/C^3 < \tau_C^-(h_N) \le 1/\epsilon C) = 1.$$

4.6 Step 6: convergence to diffusion

Here we prove Theorem 1.4 and Theorem 1.5. We will need the following result, proved in Section 10, that shows that once H_N hits $[0, N^{1/5}]$, before long the infection process hits 0.

Lemma 4.8. *Suppose that $H_0^N \le N^{1/5}$, and that w.h.p. $\sup_{t \le N^{1/4}} |z_t^N| \le C_{4.1}\sqrt{\log N}$. Then, w.h.p., $\tau_0(H^N) \le N^{1/4}$ and $H_t^N \le N^{.24}$ for all $t \le \tau_0(H^N)$.*

Proof of Theorem 1.4. We use Lemma 2.7 to prove this theorem. Since the jump size of i_N is $O(N^{-1/2}) = o(1)$ it suffices to find a, b and show convergence of the compensator and predictable quadratic variation. By assumption,

    a) $|z_N(0)| = O(1)$ so by Lemma 4.1, for fixed $T > 0$, w.h.p. $|z_N(t)| \le C_{4.1}\sqrt{\log N}$ for all $t \le T$, and

    b) $Q_N(0) = O(N^{-\epsilon})$, so letting $c_N^Q = 2Q_N(0) \vee N^{-1/6}$, Lemma 4.3 shows that for fixed $T > 0$, w.h.p. $Q_N(t) \le c_N^Q$ for all $t \le T \wedge \tau_{N^{1/5}}^-(H_N)$.

Point b) implies that $h_N(t) = (1/u_* + O(Q_N(t)^{1/2}))i_N(t) = (1/u_* + o(1))i_N(t)$ for all $t \le T \wedge \tau_{N^{1/5}}^-(H_N)$. Point a) and Lemma 4.8 imply that w.h.p. for all $\tau_{N^{1/5}}^-(H_N) \le t \le T$,

$i_N(t) \le h_N(t) = O(N^{0.24-1/2}) = o(1)$. Thus if Theorem 1.4 can be proved for h_N for some constants μ_*, σ_*^2 then it holds for i_N with different constants. We recall (4.8):

$$\mu(h_N) = (\eta-1)r_+ z_N i_N - (\eta-\gamma/2)r_+ i_N^2 + (1-\gamma/2)i_N/\sqrt{N} \qquad (4.9)$$

Also $i_N(0), j_N(0), k_N(0) = O(1)$ so the same holds for $h_N(0)$. Letting $c_N^h = \log N$, by Lemma 4.2 and a) above, w.h.p. $h_N(t) \le c_N^h$ for all $t \le T \wedge \tau_{N^{1/5}}^-(H_N)$. Combining a) and b) with the bound on h_N, w.h.p.

$$\tau_N(c_N^h, c_N^Q, 0) \wedge T = \tau_{N^{1/5}}^-(H_N) \wedge T.$$

Using this and Lemma 4.6, for any fixed $T > 0$, w.h.p.

$$\sup_{t \le T} \left| \int_0^{t \wedge \tau_{N^{1/5}}^-(H_N)} (\eta-1)r_+ z_N(s) i_N(s) ds \right| = O(N^{-(\frac{\epsilon}{2} \wedge \frac{1}{4})}(\log^3 N)) = o(1),$$

uniformly for $t \le T$. Since $i_N \le h_N$, w.h.p. for $t \le T \wedge \tau_{N^{1/5}}^-(H_N)$ the third term in the RHS of (4.9) is $O(N^{-1/2}\log N) = o(1)$. Using the bound on Q_N, if $h_N(t) \le R$ for fixed $R > 0$ and $t \le T \wedge \tau_{N^{1/5}}^-(H_N)$ then w.h.p.

$$i_N(t) = (u_* + o(1))h_N(t) = u_* h_N(t) + o(1).$$

Looking back to (4.9), we let

$$b(x) = -\mu_* x^2 \qquad \text{with} \quad \mu_* = (\eta-\gamma/2)r_+ u_*^2,$$

and let $\tau_R^+(h_N) = \inf\{t : h_N(t) \ge R\}$. Recall that h_N^p denotes the compensator of h_N (see Section 2). Since $h_N^p(t) = \int_0^t \mu_s(h_N)ds$, we find that w.h.p.

$$\sup_{t \le T \wedge \tau_R^+(h_N) \wedge \tau_{N^{1/5}}^-(H_N)} \left| h_N^p(t) - \int_0^t b(h_N(s))ds \right| = o(1).$$

Next let us consider the easier case $\tau_{N^{1/5}}^-(H_N) \le t \le T \wedge \tau_R^+(H_N)$. For this range of values of t, from (4.9), and using Lemma 4.1 it easily follows that $\mu(h_N) = o(1)$ and $h_N = o(1)$. So w.h.p.

$$\sup_{\tau_{N^{1/5}}^-(H^N) \le t \le T \wedge \tau_R^+(h_N)} \left| h_N^p(t) - \int_0^t b(h_N(s))ds \right| = o(1).$$

This proves the assertion about the compensator of h_N, as required by Lemma 2.7.

Now to calculate the diffusivity of h we let m index the possible transitions and write

$$\sigma^2(h_N) = \sum_m q_m (\Delta_m h_N)^2,$$

where q_m is the rate of transition m (it is a function of the state) and $\Delta_m h_N$ is the change in h_N at that transition. Note that there are constants c_m so that $(\Delta_m h_N)^2 = c_m/N$. Recall that on the fast (original) time scale the transitions of I^N, J^N, K^N have the following rates:

transition	rate
$I \to S$	I^N
$J \to K$	$2J^N$
$K \to J$	λK^N
$I + I \to J$	$r_+(I^N)^2/N$
$I + S \to K$	$r_+ S^N I^N/N$
$J \to I + I$	$r_- J^N$
$K \to S + I$	$r_- K^N$

Most rates are linear in I^N, J^N, or K^N. Those which are not are the $I + I \to J$ transition and the $I + S \to K$ transition. As we have seen above, w.h.p., for all $t \leq T \wedge \tau_{N^{1/5}}^-(H^N)$, $I_N(t) \leq N^{1/2} \log N$. Therefore the $I+I \to J$ transition has rate $O((\log N)^2)$. By a similar reasoning the $I + S \to K$ transition has rate $r_+(y_* + O(N^{-1/2} \log N))I^N = r_+ y_* I^N + O((\log N)^2)$. Speeding up time by $N^{1/2}$ and writing in lower case, the rates are equal to Ni_N, $2Nj_N$, etc and the error terms have rate $O(\sqrt{N}(\log N)^2)$. Since each of $i_N(t), j_N(t), k_N(t)$ is equal to (constant + $o(1)$)$h_N(t)$ for $t \leq T \wedge \tau_{N^{1/5}}^-(H_N)$, if in addition $h_N(t) \leq R$ then there are constants d_m so that for each m,

$$q_m = Nd_m h_N(1 + o(1)) + O(\sqrt{N}(\log N)^2)) = Nd_m h_N + o(N).$$

Thus there is a constant σ_*^2 so that if $h_N \leq R$ and $t \leq T \wedge \tau_{N^{1/5}}^-(H_N) \wedge \tau_R^+(h_N)$ then w.h.p.

$$\sigma^2(h_N) = \sum_m Nd_m h_N \frac{c_m}{N} + o(1) = \sigma_*^2 h_N + o(1).$$

If $\tau_{N^{1/5}}^-(H_N) \leq t \leq T$ then since w.h.p. $h_N(t) = O(N^{.24-1/2}) = o(1)$, an easy computation gives $\sigma^2(h_N) = o(1)$. This implies the desired convergence of predictable quadratic variation with $a(x) = \sigma_*^2 x$. An application of Lemma 2.7 now shows h_N (and hence i_N) converges to a diffusion of the desired form. □

We now prove Theorem 1.5.

Proof of Theorem 1.5. We first explain why it makes sense to start the limiting diffusion X from ∞. For $C > 0$, let $\tau_C^-(X) = \inf\{t: X_t \leq C\}$. Using Jensen's inequality,

$$\frac{d}{dt} E[X_t] = -\mu E[X_t^2] \leq -\mu(E[X_t])^2,$$

so we find that

$$E[X_t \mid X_0 = x] \leq \frac{1}{\mu t + 1/x}$$

and using Markov's inequality,

$$P(\tau_C^-(X) \geq t \mid X_0 = x) \leq C^{-1} E[X_t \mid X_0 = x] \leq \frac{1}{C\mu t + C/x}.$$

In particular, if $x, C \to \infty$ with $C \leq x$ then $\tau_C^-(X)$ converges in probability to zero. It is then not hard to show that the law of X, conditioned on $X_0 = x$, converges in distribution as $x \to \infty$.

Since the limiting diffusion is continuous, it crosses any level $C > 0$, if it starts from ∞. Thus if we let $\tau_C(X) = \inf\{t : X_t = C\}$, $\tau_C(X) \downarrow 0$ in probability as $C \uparrow \infty$. Since convergence in distribution allows for small time change, given the proof of Theorem 1.4, it is enough to show there are sequences $\epsilon_m \to 0, C_m \to \infty$ so that for each m, w.h.p. there is a $t_m \leq \epsilon_m$ such that

- $h_N(t_m) = C_m + O(N^{-1/2})$,
- $|z_N(t_m)| \leq C_{4.1}\sqrt{\log N}$ and
- $|Q_N(t_m)| \leq N^{-1/6}$.

Letting $C_m = m$ and $\epsilon_m = 1/\epsilon m$ for small $\epsilon > 0$, Lemma 4.7 gives the bound on h_N. Since Lemma 4.7 also gives $t_m \geq \epsilon/m^3$, we may apply Lemma 4.5 to obtain the desired bounds on $|z_N|$ and Q_N. This completes the proof. □

4.7 Step 7: extinction time

To prove a result for the time for the infection to die out, note that $\tau_0(h_N) = \inf\{t\colon h_N(t) = 0\}$ is the first time (on the $N^{1/2}$ time scale) there are no infected individuals.

The continuous mapping theorem makes half of the proof easy. To complete it we need to show that $\tau_0(h_N)$ converges in probability to 0 as $h_N(0) \to 0$, uniformly for large N. This is accomplished by combining Lemma 4.8 with the following result, that by the definition of $\tau(\mathtt{ctrl}, 0)$ implies that if h_N is initially small and if the values of $|z_N|, Q_N$ can be kept under control then within a short time, H_N hits $[0, N^{1/5}]$.

Lemma 4.9. *Fix $\epsilon > 0$. There is $\delta > 0$ so that*

$$P(\tau_N(\mathtt{ctrl}, 0) < \tau^+_{\log N}(h_N) \wedge \epsilon \mid h_N(0) \leq \delta) \geq 1 - \epsilon.$$

Lemmas 4.8 and 4.9 are postponed to Section 10. For now we use them to prove Theorem 1.6.

Proof of Theorem 1.6. Recall that $\mathcal{Q}$ is the law of the limiting diffusion for h_N and $\tau_x(X)$ is the hitting time of $x > 0$ for the limiting diffusion. By the strong Markov property and Blumenthal's 0-1 law, after hitting x, the process will with probability one immediately hit $(0, x)$ and (x, ∞). From this it follows easily that $\tau_x(X) : C \to \mathbb{R}$ is continuous $\mathcal{Q}$-almost surely. Suppose the infection process satisfies the conditions of Theorem 1.4 with $h_N(0) \to x$. Let P_N denote its law and let $\tau^-_x(h_N) = \inf\{t : h_N(t) \leq x\}$. Using the continuous mapping theorem,

$$P_N(\tau^-_x(h_N) > t) \to \mathcal{Q}(\tau_x(X) > t)$$

Let $\epsilon > 0$. If x is small enough then $\mathcal{Q}_x(\tau_0(X) > \epsilon) < \epsilon$ and hence $\mathcal{Q}(\tau_0(X) > t + \epsilon) \leq \mathcal{Q}(\tau_x(X) > t) + \epsilon$. Combining this with the last result and noting that $x \mapsto \tau^-_x(h_N)$ is increasing,

$$\liminf_{N \to \infty} P_N(\tau_0(h_N) > t) \geq \mathcal{Q}(\tau_0(X) > t + \epsilon) - \epsilon.$$

Letting $\epsilon \to 0$ we have half of the desired convergence in distribution.

To get the other half we again fix any arbitrary $\epsilon > 0$. Note that for $\delta > 0$

$$P_N(\tau_0(h_N) > t + \epsilon) \leq P_N(\tau^-_\delta(h_N) > t) + \sup_{h \leq \delta} P_N(\tau_0(h_N) > \epsilon \mid h_N(0) = h).$$

Letting $N \to \infty$, $P_N(\tau^-_\delta(h_N) > t) \to \mathcal{Q}(\tau_\delta(X) > t) \leq \mathcal{Q}(\tau_0(X) > t)$. So it suffices to show that for each $\epsilon > 0$ there is a $\delta > 0$ so that the second term is at most $O(\epsilon)$, uniformly for large N. Since by convergence in distribution, on the $N^{1/2}$ time scale it takes at least $s > 0$ amount of time, w.h.p. as $s \to 0^+$ and $N \to \infty$, for h to reach δ if $h_N(0) \to x > \delta$, using Lemma 4.5 we may assume when taking the above sup that not only $h \leq \delta$ but also that $|z_N(t)|$ and $Q_N(t)$ are small (in the sense of the definition of $\tau(\mathtt{ctrl}, t)$ given in Definition 4.4) for all $t \leq \omega(1) \wedge \tau^-_{N^{1/5}}(H_N)$. The desired bound then follows directly from Lemmas 4.8 and 4.9. This completes the proof. $\qquad \square$

5 Step 1: upper bound on $|z^N|$

Let $C_{4.1}$ be a sufficiently large constant. In this short section we prove Lemma 4.1, in two parts:

- *approach:* show that $|Z^N_t| \leq C_{4.1}$ for some $t \leq C_{4.1} \log N$ w.h.p., then
- *control:* show that if $|z^N_t| \leq (C_{4.1}/2)\sqrt{\log N}$ then w.h.p., $|z^N_t| \leq C_{4.1}\sqrt{\log N}$ for all $t \leq N$.

Approach. Recall from (3.2) that

$$\mu(Y^N) = F_N(Y^N) \quad \text{where} \quad F_N(Y^N) = r_-(N - Y^N) - \frac{r_+}{N}Y^N(Y^N - 1),$$

where, for each N, $F_N : \mathbb{R} \to \mathbb{R}$ is just some function. Let $Y_*^N \in (0, N)$ be the unique value with $F_N(Y_*^N) = 0$. Note that $Y_*^N \neq Ny_*$ since we used $Y^N(Y^N - 1)$ and not $(Y^N)^2$ to compute it. However note that $F_N(Ny_*) = r_+y_*$ and by concavity of F_N, $|F_N'|$ is bounded below by $|F_N'(0)| = r_- - r_+/N$, so

$$|Y_*^N - Ny_*| \leq r_+y^*/(r_- - r_+/N) = O(1),$$

and letting $\tilde{Z}^N = Y^N - Y_*^N$ we have $\tilde{Z}^N - Z^N = Y_*^N - Ny_* = O(1)$, so it is enough to prove the result with $\tilde{Z}^N$ in place of Z^N. Since F_N is concave, if we let $r_N = F_N(0)/Y_*^N$ then for $Y^N \in [0, N]$ with $Y^N \neq Y_*^N$ we have

$$F_N(Y^N)/(Y^N - Y_*^N) \leq -r_N.$$

Letting $r = \frac{1}{2}\liminf_{N\to\infty} r_N$, it follows that $F_N(Y_*^N + \tilde{Z}^N)/\tilde{Z}^N \leq -r$ for large N and $\tilde{Z}^N \neq 0$. If $\tilde{Z}_0^N \geq 1$ then letting $\tau_2^-(\tilde{Z}^N) = \inf\{t : \tilde{Z}_t^N \leq 2\}$ and using the product rule (Lemma 2.3), for $t < \tau_2^-(\tilde{Z}^N)$

$$\mu(e^{rt}\tilde{Z}_t^N) = e^{rt}(r\tilde{Z}_t^N + F(Y_* + \tilde{Z}_t^N)) \leq 0,$$

so $\xi_t = \exp(r(t \wedge \tau_2^-(\tilde{Z}^N)))\tilde{Z}_{t\wedge\tau_2(\tilde{Z}^N)}^N$ is a supermartingale. If $\tau_2^-(\tilde{Z}^N) > t$ then $\xi_t \geq 2e^{rt}$. Moreover $\xi_0 = \tilde{Z}_0^N \leq N$. So

$$P(\tau_2^-(\tilde{Z}^N) > t) \leq P(\xi_t \geq 2e^{rt}) \leq \frac{e^{-rt}}{2}E[\xi_t] \leq e^{-rt}N/2.$$

If $\tilde{Z}_0^N \leq -1$ then letting $\tau_2^-(-\tilde{Z}^N) = \inf\{t : \tilde{Z}_t^N \geq -2\}$, we obtain the same estimate for $P(\tau_2^-(-\tilde{Z}^N) > t)$, so for $\tau_2^-(|\tilde{Z}^N|) = \inf\{t : |Z_t^N| \leq 2\}$, taking a union bound and $t = C \log N$ with $C = 2/r$ we find

$$P(\tau_2^-(\tilde{Z}^N) > (2/r) \log N) \leq 1/N = o(1),$$

which proves the result.

Control. We use Lemma 2.5 to control $z^N = N^{-1/2}Z^N$.

Let $x = (C_{4.1}/2)\sqrt{\log N}$ with $C_{4.1}$ to be determined and let $X = -x + |z^N - x|$, then $\Delta_\star(X) \leq 2/N^{1/2} \leq x/2$ for large N. Since z^N jumps by at most x, if $|z^N| \geq x$ then $\mu(X) = \text{sgn}(z^N)\mu(z^N)$. From the proof of approach, $\mu(z^N)/z^N \leq -r$ if $z^N \neq 0$. If $X \geq x/2$ then $|z^N| \geq 3x/2$ so letting $\mu_\star = 3rx/2$, $\mu(X) \leq -\mu_\star$. If $X \leq x$ then $|z^N| \leq 2x$ so using (4.2)-(4.3), $|\mu(z^N)| \leq Cx$ and $\sigma^2(z^N) \leq C$ for some $C > 0$, so let $C_{\mu_\star}$ be a large enough multiple of x and $\sigma_\star^2$ a large enough constant. Then, $\Delta_\star(X)\mu_\star/\sigma_\star^2 = o(1)$ which allows us to take $C_\Delta = 1$. Then, $\Gamma \geq \exp(\delta L^2)$ and $\lfloor\Gamma\rfloor x/16C_{\mu_\star} \geq \delta\Gamma$ for some $\delta > 0$. Taking $C_{4.1} > 1/\sqrt{\delta}$ makes $\Gamma, \delta\Gamma \geq N$ for large N and the result follows from Lemma 2.5.

6 Step 2: upper bound on h_N

Recall $H^N = I^N + \gamma J^N + \eta K^N$ and $\tau_{N^{1/5}}^-(H^N)$ is the first time that $H^N \leq N^{1/5}$. Recalling (4.7) we see that the negative term in $\mu(H^N)$ involves I^N. Therefore, the first step is to get a lower bound on I^N/H^N. Note the event below is taken to be vacuous if $\tau_{N^{1/5}}^-(H^N) < C_{6.1}$.

Lemma 6.1. *There is a constant $C_{6.1} > 0$ so that with high probability,*

$$H_t^N \leq C_{6.1}I_t^N \quad \text{for all} \quad C_{6.1} \leq t \leq N \wedge \tau_{N^{1/5}}^-(H^N).$$

Proof. Let $\epsilon > 0$ be a small enough constant, then there are two steps:

- *approach*: show that $\tau_\epsilon^+(I^N/H^N) \wedge \tau_{N^{1/5}}^-(H^N) \leq 1/\epsilon$ w.h.p., then
- *control*: show that if $I_0^N/H_0^N \geq \epsilon$ then w.h.p. $I_t^N/H_t^N \geq \epsilon/2$ for all $t \leq N \wedge \tau_{N^{1/5}}^-(H^N)$.

Approach. From equations (3.5) and (4.7), we have

$$\mu(I^N) = -r_+ \frac{Z^N}{N} I^N - (r_+ y_* + 1) I^N + 2r_- J^N + r_- K^N + r_+ \frac{I^N}{N} \tag{6.1}$$

$$\mu(H^N) = (\eta - 1) r_+ \frac{Z^N}{N} I^N - (\eta - \gamma/2) r_+ \frac{(I^N)^2}{N} + (1 - \gamma/2) r_+ \frac{I^N}{N}.$$

Using the product rule of Lemma 2.3 and the Taylor approximation of Lemma 2.4, we find that if $H^N > N^{1/5}$, then since $H^N = \omega(1)$,

$$\mu\left(\frac{I^N}{H^N}\right) = \frac{\mu(I^N)}{H^N} + I^N \mu\left(\frac{1}{H^N}\right) + \sigma\left(I^N, \frac{1}{H^N}\right) \tag{6.2}$$

$$= \frac{\mu(I^N)}{H^N} - \frac{I^N}{(H^N)^2} \mu(H^N) + \sigma^2(H^N) O\left(\frac{I^N}{(H^N)^3}\right) + \sigma\left(I^N, \frac{1}{H^N}\right).$$

The assumption $H^N = \omega(1)$ is used together with the fact $\Delta_\star(H^N) = O(1)$ to obtain $O(I^N/(H^N)^3)$ from the Taylor approximation. Next we show the last two terms are $O(1/H^N)$. To compute $\sigma(I^N, 1/H^N)$ note that I^N or H^N jumps at rate $O(H^N)$, I^N jumps by $O(1)$, and $1/H^N$ jumps by $O(1/(H^N)^2)$ when $H^N = \omega(1)$. Multiplying, we obtain $\sigma(I, 1/H^N) = O(1/H^N)$. By a similar argument $\sigma^2(H^N) = O(H^N)$, and since $I^N \leq H^N$ it follows that $\sigma^2(H^N) I^N/(H^N)^3 = O(1/H^N)$.

If $I^N/H^N \leq \epsilon$ then

$$\max\{\gamma J^N/H^N, \eta K^N/H^N\} \geq (1 - \epsilon)/2.$$

Since $|Z^N|, I^N \leq N$ it follows that $\mu(H^N) = O(I^N)$. From (6.1) and (6.2) we then deduce that if $I^N/H^N \leq \epsilon$ and $H^N = \omega(1)$ then

$$\mu(I^N/H^N) \quad \geq -\epsilon(r_+ + r_+ y_* + 1) + r_- \min(2/\gamma, 1/\eta)(1 - \epsilon)/2 \tag{6.3}$$
$$-O(\epsilon^2) - o(1).$$

If $\epsilon > 0$ is taken small enough, the right-hand side is at least a constant $\mu_0 > 0$. To estimate $\sigma^2(I^N/H^N)$ we note that if $H^N = \omega(1)$ then I^N/H^N jumps at rate $O(H^N)$, by an amount $O(1/H^N)$ when I^N jumps, and an amount $O(I^N/(H^N)^2) = O(1/H^N)$ (since $I^N \leq H^N$) when H^N jumps. Thus $\sigma^2(I^N/H^N) = O(1/H^N)$ when $H^N = \omega(1)$.

To summarize our progress so far, if we let

$$\tau_1 = \inf\{t \colon I_t^N/H_t^N > \epsilon \text{ or } H_t^N \leq N^{1/5}\}$$

and note that $t < \tau_1$ implies $H^N > N^{1/5} = \omega(1)$ and $1/H^N < N^{-1/5}$, then there are constants $\mu_0, C > 0$ such that

$$\mu_t(I^N/H^N) \geq \mu_0 \quad \text{and} \quad \sigma^2(I^N/H^N) \leq CN^{-1/5} \text{ for all } t < \tau_1. \tag{6.4}$$

Define $\xi_t = I_{t \wedge \tau_1}^N/H_{t \wedge \tau_1}^N$. In the notation of Section 2,

$$\xi_t^m = \xi_t - \xi_0 - \xi_t^p \leq \xi_t - \mu_0(t \wedge \tau_1) \quad \text{and} \quad \langle \xi \rangle_t \leq CN^{-1/5} t \wedge \tau_1.$$

So for $a, \phi > 0$

$$\xi_t^m + (a + \phi \langle \xi \rangle_t) \leq \xi_t + a - (\mu_0 - \phi C N^{-1/5}) t \wedge \tau_1. \tag{6.5}$$

Using Lemma 2.1, if $\phi\Delta_\star(\xi) \le 1/2$, then the left-hand side of (6.5) is ≥ 0 for all $t \ge 0$ with probability $\ge 1 - 2e^{-\phi a}$, or in other words,

$$P(\xi_t \ge -a + (\mu_0 - \phi CN^{-1/5})t \wedge \tau_1 \text{ for all } t \ge 0) \ge 1 - 2e^{-\phi a}.$$

Letting $a = \epsilon$ and $\phi = (\mu_0/2C)N^{1/5}$ gives $e^{-\phi a} = o(1)$. Since ξ is stopped if ever $H^N \le N^{1/5}$, it follows that $\Delta_\star(\xi)$, which is $O(1/H^N)$, is a fortiori $O(N^{-1/5})$, which means that $\phi\Delta_\star(\xi) \le 1/2$ if $C > 0$ is taken large enough. Summarizing, we find that

$$P(\xi_t \ge -\epsilon + \mu_0 t \wedge \tau_1/2 \text{ for all } t \ge 0) = 1 - o(1).$$

Since $t > \tau_1$ if ever $\xi_t \ge \epsilon$, it follows that

$$P(\tau_1 > 4\epsilon/\mu_0) = 1 - o(1).$$

Since $4\epsilon/\mu_0$ is a constant and $\tau_1 = \tau^-_{N^{1/5}}(H^N) \wedge \tau^+_\epsilon(I^N/H^N)$, the first part is proved.

Control. We now show that if $I_0^N/H_0^N \ge \epsilon$ then w.h.p. $I_t^N/H_t^N \ge \epsilon/2$ for all $t \le N \wedge \tau^-_{N^{1/5}}(H^N)$. To do so we use Corollary 2.6 to Lemma 2.5. Let $\tau = \tau^-_{N^{1/5}}(H^N)$, let $X_t = \epsilon - I^N(t \wedge \tau)/H^N(t \wedge \tau)$, and let $x = \epsilon/2$. Similarly as for ξ, we find that $\Delta_\star(X) = O(1/H^N) = O(N^{-1/5})$, which is $o(x)$. From (6.1)-(6.2) it is easy to check that $|\mu(I^N/H^N)| = O(1)$ when $H^N = \omega(1)$ so let $C_{\mu_\star}$ be a large constant. From (6.4) we have $\mu_t(X) \le -\mu_0$ and $\sigma_t^2(X) \le CN^{-1/5}$ when $t < \tau$ and $X_t \ge 0$, so let $\mu_\star = \mu_0$ and $\sigma_\star^2 = CN^{-1/5}$ for some $C > 0$. In this way $\Delta_\star(X)\mu_\star/\sigma_\star^2 = O(1)$ which allows us to let C_Δ be a large constant. Then, $\Gamma \ge \exp(\delta N^{1/5})$ and $\lfloor\Gamma\rfloor x/16C_{\mu_\star} \ge \delta\Gamma \ge N$ for some $\delta > 0$ and large N, so Corollary 2.6 gives

$$\lim_{N\to\infty} P\left(\sup_{t<N\wedge\tau^-_{N^{1/5}}(H^N)} I_t^N/H_t^N \le \epsilon/2 \mid I_0^N/H_0^N \ge 3\epsilon/4\right) = 0,$$

which completes the proof. $\qquad\square$

Proof of Lemma 4.2. The result has two parts.

Approach. First we show that

$$\text{w.h.p., } h_N(t) \le \tfrac{1}{2}\log N \text{ for some } t \le 1/\sqrt{\log N}.$$

Note the slow time scale is used. We will need the estimates we've proved so far. Let

$$\tau_1 = \inf\{t \ge C_{4.1}N^{-1/2}\log N : |z_N(t)| > C_{4.1}\sqrt{\log N}\} \quad \text{and}$$
$$\tau_2 = \inf\{t \ge C_{6.1}N^{-1/2} : h_N(t) > C_{6.1}i_N(t) \text{ or } H_N(t) \le N^{1/5}\}.$$

By Lemma 4.1, $\tau_1 > N^{1/2}$ w.h.p., and by Lemma 6.1, $\tau_2 > N^{1/2} \wedge \tau^-_{N^{1/5}}(H_N)$ w.h.p. It is more convenient if we shift the time variable over by $-C_{4.1}N^{-1/2}\log N$ so that both estimates begin to hold at $t = 0$; since $N^{-1/2}\log N = o(1/\sqrt{\log N})$ this will not affect the conclusion.

We recall (4.8):

$$\mu(h_N) = (\eta - 1)r_+ z_N i_N - (\eta - \gamma/2)r_+ i_N^2 + (1 - \gamma/2)i_N/\sqrt{N}$$

From (4.6), $\eta > 1 > \gamma/2$, moreover $i_N \le h_N$, so for $t < \tau_1 \wedge \tau_2$,

$$\mu(h_N) \le \left((\eta - 1)r_+ C_{4.1}\sqrt{\log N} + (1 - \gamma/2)r_+\frac{1}{\sqrt{N}}\right)\frac{h_N}{C_{6.1}} - (\eta - \gamma/2)r_+\frac{(h_N)^2}{C_{6.1}^2}.$$

If $h_N > \frac{1}{2} \log N$ the first term is $o((h_N)^2)$. Since $\eta > \gamma/2$, we see there is $\delta > 0$ so that, if we let $\tau_3 = \tau_1 \wedge \tau_2 \wedge \tau^-_{\frac{1}{2} \log N}(h_N)$, then for $t < \tau_3$,

$$\mu(h_N) \leq -\delta \, (h_N)^2.$$

Next we'd like to set up a differential inequality for h_N; the only trouble is, the drift estimate only holds up to a stopping time. To fix this, let $\psi(t; h)$ denote the solution flow for the differential equation $h'(t) = -\delta h(t)^2$ (i.e., the function with maximal domain containing $\{0\} \times \mathbb{R}$ such that $\partial_t \psi(t; h) = -\delta \psi(t; h)^2$ and $\psi(0; h) = h$) and define the continued process

$$\tilde{h}_N(t) = \psi(t - t \wedge \tau_3; h_N(t \wedge \tau_3)).$$

In words, $\tilde{h}_N$ is equal to h_N up to time τ_3, at which point it evolves according to the flow ψ. It is then clear that for all $t \geq 0$,

$$\mu(\tilde{h}_N) \leq -\delta \, (\tilde{h}_N)^2.$$

Taking expectations and using Jensen's inequality we then find

$$\frac{d}{dt} E[\tilde{h}_N(t)] \leq -\delta \left(E[\tilde{h}_N(t)] \right)^2,$$

which implies $E[\tilde{h}_N(t)] \leq \psi(t; E[h_N(0)])$ (note $\tilde{h}_N(0) = h_N(0)$). Solving the DE we have

$$\psi(t; h) = 1/(\delta t + 1/h) \leq 1/(\delta t).$$

If $\tau_3 > t$ then $\tilde{h}_N(t) = h_N(t) > \frac{1}{2} \log N$, so it follows that

$$P(\tau_3 > t) \leq P(\tilde{h}_N(t) = h_N(t) > \frac{1}{2} \log N) \leq \frac{2}{\log N} E[\tilde{h}_N(t)] \leq 2(\delta t \log N)^{-1}.$$

Taking $t = 1/\sqrt{\log N}$ the above is $o(1)$. As noted above, $\tau_1 \wedge \tau_2 > N^{1/2} \wedge \tau^-_{N^{1/5}}(H_N)$ w.h.p. Since $h_N > \frac{1}{2} \log N$ implies $H_N > N^{1/5}$, if $h_N(s) > \frac{1}{2} \log N$ for all $s \leq t$ then w.h.p. $\tau_3 > t$. It follows that

$$P(h_N(s) > \frac{1}{2} \log N \text{ for all } s \leq 1/\sqrt{\log N}) = P(\tau_3 > 1/\sqrt{\log N}) + o(1) = o(1)$$

and the statement is proved.

Control. We now show that

if $h_N(0) \leq \frac{1}{2} \log N$ then w.h.p. $h_N(t) \leq \log N$ for all $t \leq N^{1/2} \wedge \tau^-_{N^{1/5}}(H_N) \wedge \tau^+_{C_{4.1}}(|z^N|)$.

This time, let $\tau_1 = \tau^+_{C_{4.1}}(|z_N|)$, and define τ_2 as in the proof of approach. By Lemma 6.1, $\tau_2 > N^{1/2} \wedge \tau^-_{N_{1/5}}(H_N)$. For the present result we cannot ignore small times, so first we show the conclusion holds for all $t \leq C_{6.1} N^{-1/2}$ without assuming a bound on i_N/h_N. Let $\tau_3 = \tau_1 \wedge \tau^+_{\log N}(h_N)$. From (4.8), if $t < \tau_3$ then since $i_N \leq h_N$, $\eta - \gamma/2 > 0$ and $z_N(t) = O(\sqrt{\log N})$, for large N we have $\mu(h_N(t)) \leq \log(N) h_N(t)$. Since h_N jumps by $O(N^{-1/2})$ at rate $N h_N$ on the slow time scale, $\sigma^2(h_N) \leq C h_N$ for some constant $C > 0$. Let $\xi_t = h_N(t \wedge \tau_3)$, then for $a, \phi > 0$ and all $t \geq 0$,

$$\xi^m_t - a - \phi \langle \xi \rangle_t \geq \xi_t - \xi_0 - a - \left((\log N)^2 + C\phi \log N \right) t.$$

Using Lemma 2.1, if $\phi \Delta_\star(\xi) \leq 1/2$, the LHS is ≤ 0 for all $t \geq 0$ with probability at least $1 - 2e^{-\phi a}$. Letting $\phi = (\log N)/C$ and $a = 1$, $\phi \Delta_\star(\xi) = O(N^{-1/2} \log N) \leq 1/2$ for large enough N and $e^{-\phi a} = o(1)$, so

$$P(\xi_t \leq \xi_0 + 1 + 2t(\log N)^2 \text{ for all } t \geq 0) = 1 - o(1).$$

It follows that $P(\xi_t \leq \log N \text{ for all } t \leq C_{6.1}N^{-1/2}) = 1 - o(1)$, since $\xi_0 \leq \frac{1}{2}\log N$ and $1 + 2t(\log N)^2 \leq \frac{1}{4}\log N$ for all $t \leq C_{6.1}N^{-1/2}$, which implies that

$$P(h_N(t) \leq \frac{3}{4}\log N \text{ for all } t \leq C_{6.1}N^{-1/2} \wedge \tau_1) = 1 - o(1).$$

With the above estimate in hand, to prove the desired result we may assume $h_N(0) \leq \frac{3}{4}\log N$, redefine $\tau_2 = \tau^+_{C_{6.1}}(h_N/i_N) \wedge \tau^-_{N^{1/5}}(H_N)$ and suppose that $\tau_2 > N^{1/2} \wedge \tau^-_{N^{1/5}}(H_N)$ w.h.p. We will use Corollary 2.6. Let $\tau = \tau_1 \wedge \tau_2$, let $X_t = h_N(t \wedge \tau) - \log(N)/2$ and let $x = \log(N)/2$, then $\Delta_\star(X) = O(N^{-1/2}) = o(x)$. Since h_N jumps by $O(N^{-1/2})$ at rate $O(Nh_N)$, if $h_N \leq \log N$ then $|\mu(h_N)| = O(N^{1/2}h_N) = O(N^{1/2}\log N)$ and $\sigma^2(h_N) = O(h_N) = O(\log N)$, so let $C_{\mu_\star} = CN^{1/2}\log N$ and $\sigma_\star^2 = C\log N$ for large C. From the proof of approach, if $t < \tau$ and $h_N(t) \geq \log(N)/2$ then $\mu_t(h_N) \leq -\delta(h_N)^2$, so let $\mu_\star = -\delta(\log N)^2/4$. With these choices $\Delta_\star(X)\mu_\star/\sigma_\star^2 = o(1)$ so let $C_\Delta = 1$. We now find $\mu_\star x/\sigma_\star^2 = \Omega((\log N)^2)$. Therefore, $\Gamma = e^{\Omega((\log N)^2)} = \omega(N)$ and $x/16C_{\mu_\star} = \Omega(N^{-1/2})$. Using Corollary 2.6,

$$P(h_N(t) \geq \log N \text{ for some } t \leq N^{1/2} \wedge \tau \mid h_N(t) \leq \frac{3}{4}\log N) = o(1),$$

and the result follows since $N^{1/2} \wedge \tau = N^{1/2} \wedge \tau^-_{N^{1/5}}(H_N) \wedge \tau_1$ w.h.p. $\qquad\square$

7 Step 3: (I^N, J^N, K^N) stays close to the invariant ray

In this section we prove Lemma 4.3. As in the statement of the lemma, let

$$\tau = \tau^-_{N^{1/5}}(H^N) \wedge \tau^+_{C_{4.1}\sqrt{\log N}}(|z^N|) \wedge \tau^+_{\log N}(h^N).$$

Looking back to (6.2), we showed that if $H^N = \omega(1)$ then

$$\sigma(I^N, 1/H^N) + \sigma^2(H_N)O(I^N/(H^N)^3) = O(1/H^N);$$

it is clear the same estimate holds with J^N, K^N in place of I^N. If, moreover, $H^N, |Z^N| \leq N^{1/2}\log N$, then since $\eta, \gamma \geq 1$ (see (4.6)), $I^N, J^N, K^N \leq H^N$, from (6.2) we obtain the estimate $\mu(H_N) = O((\log N)^2)$. Recalling that $(U^N, V^N, W^N) = (I^N, \gamma J^N, \eta K^N)/H^N$, writing (6.2) but with γJ^N and ηK^N as well, and using the above estimates, we find that if $t < \tau$ then

$$\mu_t\begin{pmatrix} U^N \\ V^N \\ W^N \end{pmatrix} = \frac{1}{H^N_t}\,\mu_t\begin{pmatrix} I^N \\ \gamma J^N \\ \eta K^N \end{pmatrix} + O((\log N)^2N^{-1/5}). \tag{7.1}$$

Let $\mathcal{J}^N$ denote the vector (I^N, J^N, K^N). Again, if $H^N, |Z^N| \leq N^{1/2}\log N$ then from (3.5) we find that $\mu(\mathcal{J}^N) = A\mathcal{J}^N + O((\log N)^2N^{-1/5})$. Letting $D = \text{Diag}(1, \gamma, \eta)$ denote the diagonal matrix with $1, \gamma, \eta$ along the main diagonal, and $\mathcal{V}^N$ denote the vector $(U^N, V^N, W^N)^\top$, so that $\mathcal{V}^N = D\mathcal{J}^N/H^N$. Moreover let $\Lambda = DAD^{-1}$ denote the conjugation of A by D. Combining the estimate on $\mu(\mathcal{J}^N)$ with (7.1) we find that if $t < \tau$ then

$$\mu_t(\mathcal{V}^N) = \frac{1}{H^N_t}D\mu_t(\mathcal{J}^N) + O((\log N)^2N^{-1/5})$$
$$= \Lambda\mathcal{V}^N_t + O((\log N)^2N^{-1/5}) \tag{7.2}$$

To reduce the dimensionality, since from (3.10) we have $A(\alpha, \beta, 1)^\top\mathbb{R} = 0$ it follows that $\Lambda(\alpha, \beta\gamma, \eta)^\top\mathbb{R} = 0$. We select the vector $\mathcal{V}_\star = (u_\star, v_\star, w_\star)$ with $u_\star + v_\star + w_\star = 1$, namely

$$u_\star = \alpha/d \qquad v_\star = \beta\gamma/d \qquad w_\star = \eta/d \qquad \text{where } d = \alpha + \beta\gamma + \eta. \tag{7.3}$$

Then, $W^N - w_* = -(U^N - u_*) - (V^N - v_*)$, so (7.2) can be re-written as

$$\mu_t \begin{pmatrix} U_t^N - u_* \\ V_t^N - v_* \end{pmatrix} = \begin{pmatrix} \Lambda_{11} - \Lambda_{13} & \Lambda_{12} - \Lambda_{13} \\ \Lambda_{21} - \Lambda_{23} & \Lambda_{22} - \Lambda_{23} \end{pmatrix} \begin{pmatrix} U_t^N - u_* \\ V_t^N - v_* \end{pmatrix} + O((\log N)^2 N^{-1/5}) \qquad (7.4)$$

Note that Λ is obtained from A by multiplying entries in rows 2,3 by γ, η and dividing entries in columns 2,3 by γ, η, respectively. Referring to (3.6) we find that

$$\begin{pmatrix} \Lambda_{11} - \Lambda_{13} & \Lambda_{12} - \Lambda_{13} \\ \Lambda_{21} - \Lambda_{23} & \Lambda_{22} - \Lambda_{23} \end{pmatrix} = \begin{pmatrix} -(r_+ y_* + 1 + r_-/\eta) & \left(\frac{2r_-}{\gamma} - \frac{r_-}{\eta} \right) \\ -(\gamma\lambda/\eta) & -(r_- + 2 + \gamma\lambda/\eta) \end{pmatrix} \qquad (7.5)$$

The diagonal entries in the matrix are negative, so the trace is negative. From (4.6), $\gamma/2 < \eta$ which implies $2r_-/\gamma - r_-/\eta > 0$, so the determinant is positive and so both eigenvalues have negative real part (which we already knew from analyzing A). To turn these calculations into control on the distance of $\mathcal{V}$ from $\mathcal{V}_*$ we let

$$\theta_1 = 2r_-/\gamma - r_-/\eta \quad \text{and} \quad \theta_2 = \gamma\lambda/\eta \qquad (7.6)$$

and examine $Q_t^N = \theta_2(U_t^N - u_*)^2 + \theta_1(V_t^N - v_*)^2$.

Approach. First we show that w.h.p., $\tau_{(N^{-1/6}/2)}^-(Q^N) \wedge \tau \le C_{4.3} \log N$. From Lemma 2.3, for a process X we have $\mu(X^2) = 2X\mu(X) + \sigma^2(X)$, and of course, $\sigma^2(X - c) = \sigma^2(X)$ and $\mu(X - c) = \mu(X)$ for any constant c. As noted in the proof of Lemma 6.1, $U^N = I^N/H^N$ jumps by $O(1/H^N)$ at rate $O(H^N)$, so $\sigma^2(U^N) = O(1/H^N)$ and similarly for V^N. Let $a_1 = \min(r_+ y_* + 1 + r_-/\eta, r_- + 2 + \gamma\lambda/\eta)$, so that both diagonal entries in (7.5) are at most $-a_1$. Since the cross-terms cancel (by choice of θ_1, θ_2), from (7.4) we find that for $t < \tau$,

$$\mu_t(Q^N) \le -2a_1\theta_2(U^N - u_*)^2 - 2a_1\theta_1(V^N - v_*)^2 + O((\log N)^2 N^{-1/5}) + O(1/H^N)$$
$$\le -2a_1 Q^N + a_2(\log N)^2 N^{-1/5}$$

for some constant a_2 and large N. It is also not hard to check that $|\mu_t(Q_t^N)| = O(Q_t^N)$ for $t < \tau$, which we will need in a moment. Letting $\tilde{Q} = Q - (a_2/a_1)(\log N)^2 N^{-1/5}$, $\mu_t(\tilde{Q}^N) \le -2a_1\tilde{Q}^N$ for $t < \tau$ so $\xi_t = \exp(2a_1(t \wedge \tau))\tilde{Q}^N(t \wedge \tau)$ is a supermartingale. Since $(a_2/a_1)(\log N)^2 N^{-1/5} \le N^{-1/6}/4$ for large N and since $\xi_0 = \tilde{Q}_0^N \le Q_0^N \le \theta_1 + \theta_2$,

$$P(\tau_{N^{-1/6}/2}^-(Q^N) \wedge \tau > t) \le P(\xi_t \ge e^{2a_1 t} N^{-1/6}/4) \le 4N^{1/6} e^{-2a_1 t}(\theta_1 + \theta_2).$$

Choosing t equal to a large enough multiple of $\log N$, the RHS is $o(1)$.

Control. Now we suppose that c_N^Q are constants with $N^{-1/6} \le c_N^Q = o(1)$ and show that

$$\text{if } Q_0^N \le c_N^Q/2 \text{ then w.h.p. } Q_t^N \le c_N^Q \text{ for all } t \le N \wedge \tau.$$

We will use Corollary 2.6. Let

$$\tau_1 = \tau \wedge \tau_{c_N^Q}^+(Q^N), \quad x = c_N^Q/2 \quad \text{and} \quad X_t = Q_{t \wedge \tau_1}^N - x.$$

As noted above, U^N jumps by $O(1/H^N)$ when $H^N = \omega(1)$. Thus when $H^N = \omega(1)$, $(U^N - u_*)^2$ jumps by $O((U^N - u_*)/H^N + 1/(H^N)^2)$. If $t < \tau$ and $Q^N \le c_N^Q$ then since $H^N \ge N^{1/5}$ and $c_N^Q \ge N^{-1/6} = \omega(1/H^N)$,

$$\frac{U^N - u_*}{H^N} + \frac{1}{(H^N)^2} = O\left(\frac{(c_N^Q)^{1/2}}{H^N} \right) = O(N^{-1/5}(c_N^Q)^{1/2}).$$

An analogous estimate holds for $(V - v_*)^2$, which shows that $\Delta_\star(X) = O(N^{-1/5}(c_N^Q)^{1/2}) = o(x)$. If $Q_t^N \ge c_N^Q/2$ and $t < \tau$, we have $\mu(Q^N) \le -(2a_1 - o(1))c_N^Q/2 \le -\mu_\star$ for large N,

where $\mu_\star = a_1 c_N^Q / 2$. As noted above, if $t < \tau$ then $|\mu(Q_t^N)| = O(Q_t^N)$ so if in addition $Q_t^N \leq c_N^Q$ then $|\mu(Q_t^N)| = O(c_N^Q) \leq C_{\mu_\star} = C c_N^Q$ for large enough constant $C > 0$. Q^N has transition rate $O(H^N)$, and as shown above, if $Q^N \leq c_N^Q$ and $t < \tau$ then Q^N jumps by $O((c_N^Q)^{1/2}/H^N)$ and so $\sigma^2(Q_t^N) = O(c_N^Q/H^N) \leq \sigma_\star^2 = C c_N^Q N^{-1/5}$ for large enough $C > 0$. Since $\Delta_\star(X)\mu_\star/\sigma_\star^2 = O((c_N^Q)^{1/2}) = o(1)$, let $C_\Delta = 1$. Since $\mu_\star x / \sigma_\star^2 = \Omega(c_N^Q N^{1/5}) = \Omega(N^{1/5-1/6})$, we find $\Gamma = \exp(\Omega(N^{1/30})) = \omega(N)$ and $x/16 C_{\mu_\star} = \Omega(1)$. Corollary 2.6 then gives the desired result.

8 Step 4: averaging $z_N i_N$ to 0

In this section we prove Lemma 4.6. Letting $\sigma_z^2 = 4r_-(1-y_*)$, using (4.1) and recalling $r_-(1-y_*) = r_+ y_*^2$, we see

$$z^N \to z^N + 2/N^{1/2} \quad \text{at rate} \quad q_+ = \sigma_z^2 N/8 - r_- z^N N^{1/2}/2$$
$$z^N \to z^N - 2/N^{1/2} \quad \text{at rate} \quad q_- = \sigma_z^2 N/8 + r_+ y_* z^N N^{1/2} + O(1 \vee (z^N)^2).$$

In order to prove an averaging result we'd like to work with a process whose transitions are symmetric on reflection about 0. Thus we define the following process $\tilde{z}^N$, which can be thought of as a spatially discrete Ornstein-Uhlenbeck process. Letting $\mu_z = r_- + 2r_+ y_*$ we let $\tilde{z}^N$ have transitions

$$\tilde{z}^N \to \tilde{z}^N + 2/N^{1/2} \quad \text{at rate} \quad \tilde{q}_+ = \tilde{q}_+(\tilde{z}^N) = \sigma_z^2 N/8 - \mu_z N^{1/2} \tilde{z}^N/4 \qquad (8.1)$$
$$\tilde{z}^N \to \tilde{z}^N - 2/N^{1/2} \quad \text{at rate} \quad \tilde{q}_- = \tilde{q}_-(\tilde{z}^N) = \sigma_z^2 N/8 + \mu_z N^{1/2} \tilde{z}^N/4.$$

Furthermore take $\tilde{z}^N(0) = 2N^{-1/2} \lfloor N^{1/2} z^N(0)/2 \rfloor$ so that $\tilde{z}^N$ takes values in $2N^{-1/2}\mathbb{Z}$. Couple z^N with $\tilde{z}^N$ in the obvious way, i.e. couple jumps of $+2/N^{1/2}$ at the minimum of the two rates and similarly for jumps of $-2/N^{-1/2}$, and let $D^N = z^N - \tilde{z}^N$ with respect to this coupling. The following result controls the size of D^N. The power of $\log N$ in the bound is not optimal but it's good enough and frees us from having to track yet another constant.

Lemma 8.1. *With high probability,*

$$\sup\{ |D_t^N| : t \leq N \wedge \tau_{C_{4.1}\sqrt{\log N}}^+(|z^N|) \} \leq N^{-1/4} \log N.$$

Proof. D^N has the following transitions:

$$D^N \to D^N + 2/N^{1/2} \quad \text{at rate} \quad q_+(D^N) = \max(q_+ - \tilde{q}_+, 0) + \max(\tilde{q}_- - \tilde{q}_-, 0)$$
$$D^N \to D^N - 2/N^{1/2} \quad \text{at rate} \quad q_-(D^N) = \max(\tilde{q}_+ - q_+, 0) + \max(q_- - \tilde{q}_-, 0).$$

Using the fact that $\max(a,0) - \max(-a,0) = a$ for $a \in \mathbb{R}$, we compute

$$\mu(D^N) = 2N^{-1/2}(q_+(D^N) - q_-(D^N))$$
$$= 2N^{-1/2}((q_+ - \tilde{q}_+) - (q_- - \tilde{q}_-))$$

From their definition and the choice of constant $\mu_z = r_- + 2r_+ y_*$,

$$((q_+ - \tilde{q}_+) - (q_- - \tilde{q}_-))$$
$$= q_+ - q_- - (\tilde{q}_+ - \tilde{q}_-)$$
$$= (-r_- - 2r_+ y_*)z^N N^{1/2}/2 - (-\mu_z/4 - \mu_z/4)\tilde{z}^N N^{1/2}/2 + O(1 \vee (z^N)^2)$$
$$= -\mu_z(z^N - \tilde{z}^N)N^{1/2}/2 + O(1 \vee (z^N)^2)$$

and so

$$\mu(D^N) = -\mu_z D^N + O((1 \vee (z^N)^2)/N^{1/2}). \tag{8.2}$$

Computing the diffusivity,

$$\sigma^2(D^N) = (4/N)(|q_+ - \tilde{q}_+| + |q_- - \tilde{q}_-|) \tag{8.3}$$
$$= O(\max(z^N, \tilde{z}^N, (1 \vee (z^N)^2)/N^{1/2})/N^{1/2}.$$

Define $\tau = \tau^+_{C_{4.1}\sqrt{\log N}}(|z^N|) \wedge \tau^+_{N^{-1/4}\log N}(|D^N|)$, and observe that if $t < \tau$ then not only is $|z^N_t| \le C_{4.1}\sqrt{\log N}$ but also

$$|\tilde{z}^N_t| \le |z^N_t| + |D^N_t| \le C_{4.1}\sqrt{\log N} + N^{-1/4}\log N \le 2C_{4.1}\sqrt{\log N}$$

for large N. We use Corollary 2.6 to control $|D^N|$. Let $x = \frac{1}{2}N^{-1/4}\log N$ and let $X = -x + |D^N - x|$, then $\Delta_\star(X) \le 2/N^{1/2} \le x/2$ for large N. Since D^N jumps by at most x, if $|D^N| \ge x$ then $\mu(X) = \mathrm{sgn}(D^N)\mu(D^N)$. Suppose $X \ge x/2$ and $t < \tau$, then $|D^N_t| \ge 3x/2$, and since $(1 \vee (z^N_t)^2)/N^{1/2} = o(D^N_t)$, from (8.2) we have $\mu(D^N_t)/D^N_t \le -\mu_z/2$. Thus, letting $\mu_\star = 3x\mu_z/4$, we find that $\mu_t(X) \le -\mu_\star$. Now suppose that $X \le x$ and $t < \tau$, then $|D^N_t| \le 2x$, and using (8.2)-(8.3) and $|z^N_t|, |\tilde{z}^N_t| = O(\sqrt{\log N})$, $|\mu_t(z^N)| \le 4\mu_z x$ and $\sigma^2_t(D^N) \le CN^{-1/2}\sqrt{\log N}$ for some $C > 0$ and large N, so let $C_{\mu_\star} = 4\mu_z x$ and let $\sigma^2_\star = CN^{-1/2}\sqrt{\log N}$. Then, $\Delta_\star(X)\mu_\star/\sigma^2_\star = o(1)$ which allows us to take $C_\Delta = 1$. Then, $\Gamma = \exp(\Omega((\log N)^{3/2})) = \omega(N)$ and $x/16C_{\mu_\star} = \Omega(1)$, so the result follows from Corollary 2.6. $\qquad\square$

In the context of Lemma 4.6, if $t < \tau_N = \tau_N(c^h_N, c^Q_N, 0)$ then $h_N(t) \le c^h_N \le \log N$. Letting $L_N(t)$ denote $L(i_N(t), j_N(t), k_N(t))$, since $L(0,0,0) = 0$, $i_N, j_N, k_N \le h_N$ and L is Lipschitz with constant c_L, $|L_N(t)| \le c_L c^h_N \le c_L(c^h_N)^{1/2}(\log N)^{1/2}$. Let $\tau_{N,D} = \tau_N \wedge \tau^+_{N^{-1/4}\log N}(|D_N(s)|)$, then

$$\sup_{t \le T}\left|\int_0^{t \wedge \tau_{N,D}} |D^N(s)L_N(s)|ds\right| = O(c_L N^{-1/4}(c^h_N)^{1/2}(\log N)^{3/2}T).$$

Since by definition $\tau_N \le \tau^+_{C_{4.1}\sqrt{\log N}}(|z^N|)$, Lemma 8.1 implies that w.h.p. $\tau_{N,D} = \tau_N$, so it is enough to prove Lemma 4.6 with z^N replaced by $\tilde{z}^N$. Thus, we will prove the following result.

Lemma 8.2. *Let $L_N(s)$ denote $L(i_N(s), j_N(s), k_N(s))$. In the context of Lemma 4.6 except with $\tilde{z}^N$ in place of z^N, with high probability*

$$\sup_{t \le T}\left|\int_0^{t \wedge \tau_N} \tilde{z}_N(s)L_N(s)\,ds\right| = O(c_L(c^h_N)^{1/2}(N^{-1/4} \vee (c^h_N c^Q_N)^{1/2})\log^2 N).$$

We begin by estimating excursions of $\tilde{z}^N$.

Lemma 8.3. *Define $c_* = \sqrt{\sigma^2_z/4\mu_z}$. Let $\tau^*_0 = \inf\{t: \tilde{z}^N_t = 0\}$ and let*

$$\tau^*_1 = \inf\{s > \tau^*_0 : |\tilde{z}^N(s)| > c_*\}$$
$$\tau^*_2 = \inf\{s > \tau^*_1 : \tilde{z}^N(s) = 0\}.$$

- *If $C > 1/\mu_z$ then w.h.p. $\tau^*_0 \le C\log N$.*
- *$E[\tau^*_1 - \tau^*_0] \ge 1/4\mu_z$.*
- *There are constants $\theta, \Theta > 0$ so that $E[\exp(\theta(\tau^*_2 - \tau^*_0))] \le \Theta$ for large N.*
- *With θ, Θ as above, $P\left(\displaystyle\int_{\tau^*_0}^{\tau^*_2} |\tilde{z}^N_s|ds > 3c_*M + (\theta + \sigma^2_*)M^2\right) \le (1 + 3\Theta)e^{-\theta m}.$*

Proof. From (8.1) we compute

$$\mu(\tilde{z}^N) = -\mu_z \tilde{z}^N \quad \text{and} \quad \sigma^2(\tilde{z}^N) = \sigma_z^2.$$

If $\tilde{z}_0^N > 0$ then $\xi_t = e^{\mu_z(t \wedge \tau_0^*)} \tilde{z}_{t \wedge \tau_0^*}^N$ is a supermartingale. Since $\tilde{z}_t^N \geq 2N^{-1/2}$ if $\tau_0^* > t$,

$$P(\tau_0^* > t \mid \tilde{z}_0^N = z) \leq P(\xi_t \geq 2N^{-1/2} e^{\mu_z t} \mid \tilde{\xi}_0 = z) \leq \frac{\sqrt{N} z}{2} e^{-\mu_z t}.$$

By symmetry of $\tilde{z}^N$, the same estimate holds if $-\tilde{z}_0^N = z$. Since $|\tilde{z}^N| \leq N^{1/2} + 2N^{-1/2}$, the first statement follows by taking $t = C \log N$ for $C > 1/\mu_z$.

To prove the second statement, let $\tau = \tau_{c_*}^+(|\tilde{z}^N|)$, and note that by the strong Markov property, $\tau_1^* - \tau_0^*$ is equal to distribution to τ conditioned on $\tilde{z}_0^N = 0$. We compute

$$\mu((\tilde{z}^N)^2) = 2\tilde{z}^N \mu(\tilde{z}^N) + \sigma^2(\tilde{z}^N)$$
$$= -2\mu_z(\tilde{z}^N)^2 + \sigma_z^2. \tag{8.4}$$

In particular, $\mu((\tilde{z}^N)^2) \leq \sigma_z^2$, so $(\tilde{z}_{t \wedge \tau}^N)^2 - \sigma_z^2(t \wedge \tau)$ is a supermartingale. Since $(\tilde{z}_0^N)^2 = 0$ and $(\tilde{z}_\tau^N)^2 \geq c_*^2$, using optional stopping

$$E[\tau] \geq E[\tilde{z}_\tau^N]/\sigma_z^2 \geq c_*^2/\sigma_z^2 = 1/4\mu_z.$$

To prove the third statement it suffices to show that for large N, both $P(\tau_1^* - \tau_0^* > t)$ and $P(\tau_2^* - \tau_1^* > t)$ are bounded by some function that decays exponentially in t. We prove the two parts in the order given. Suppose $|\tilde{z}_0^N| \leq c_*$ and let $\tau = \tau_{c_*}^+(|\tilde{z}^N|)$. From (8.4), if $|\tilde{z}^N| \leq c_* = \sqrt{\sigma_z^2/4\mu_z}$ then $\mu((\tilde{z}^N)^2) \geq \sigma_z^2/2$. Moreover if $|\tilde{z}^N| \leq c_*$ then $(\tilde{z}^N)^2$ jumps by $2(2N^{-1/2})\tilde{z}^N - (\tilde{z}^N)^2 \leq 4N^{-1/2}c_*$ for large N, and jumps at rate $\sigma_z^2 N/4$, so $\sigma^2((\tilde{z}^N)^2) \leq 4c_*^2 \sigma_z^2$. Let $X_t = (\tilde{z}_{t \wedge \tau}^N)^2 - (\tilde{z}_0^N)^2$, so that for $\phi > 0$,

$$-X_t^m - \phi\langle X \rangle_t \geq -X_t + (\sigma_z^2/2 - 4c_*^2 \sigma_z^2 \phi)(t \wedge \tau).$$

Let $\phi = 1/16c_*^2$ so that there is $\sigma_z^2/4$ in the above parentheses, noting that $\phi \Delta_\star(X) = O(N^{-1/2}) \leq 1/2$ for large N. Using Lemma 2.1 with $a = 1/\phi = 16c_*^2$, we find that

$$P(\tau > t \text{ and } X_t \leq -16c_*^2 + \sigma_z^2 t/4) \leq 2/e.$$

If $\tau > t$ then $|\tilde{z}_t^N| \leq c_*$, so $X_t \leq 4(c_*)^2$. Denoting the constant $t^* = 80c_*^2/\sigma_z^2$ we find that

$$P(\tau > t^*) \leq 2/e.$$

Using the Markov property to iterate, $P(\tau > kt^*) \leq (2/e)^k$, which proves the first part.

Taking now $\tau = \inf\{t : \tilde{z}^N = 0\}$ and j^* the least multiple of $2N^{-1/2}$ larger than c_*, we bound $P(\tau_2^* - \tau_1^* > t) = P(\tau > t \mid |\tilde{z}_0^N| = 2j^*/\sqrt{N})$. Since this quantity is non-decreasing in j^* and we only need an upper bound we may assume j^* is even, and by symmetry we may assume $\tilde{z}^N > 0$. For any j let $\tau_j = \inf\{t : \tilde{z}_t^N = 2j/\sqrt{N}\}$ and let $\tau_{0,j} = \inf\{t : \tilde{z}_t^N \in \{0, 2j/\sqrt{N}\}\}$. The approach is to condition on the number of commutes from $2j^*/\sqrt{N}$ to $j^*/\sqrt{N}$ and back, before hitting 0. Since $\tilde{z}^N$, stopped at zero, is a supermartingale,

$$P(\tilde{z}_{\tau_{0,j^*}}^N = 0 \mid \tilde{z}_0^N = j^*/\sqrt{N}) \geq 1/2,$$

so the number of commutes is at most geometric$(1/2)$. It is easy to check that if a random variable X has an exponential tail, then a geometric sum (with positive stopping probability) of i.i.d. copies of X itself has an exponential tail. Thus it suffices to show that both

$$P(\tau_{j^*/2} > t \mid \tilde{z}_0^N = 2j^*/\sqrt{N}) \text{ and } P(\tau_{0,j^*} > t \mid \tilde{z}_0^N = j^*/\sqrt{N})$$

are bounded by some function that decays exponentially in t, for large N. Using the supermartingale ξ defined above, it is easy to check that

$$P(\tau_{j^*/2} > t \mid \tilde{z}_0^N = 2j^*/\sqrt{N}) \leq 2e^{-\mu_z t}.$$

Then, since $\tau_{0,j^*} \leq \inf\{t : |\tilde{z}_t^N| \geq 2j^*/\sqrt{N}\}$, using the proof of the bound on $\tau_1^* - \tau_0^*$ we deduce that $P(\tau_{0,j^*} > kt^* \mid \tilde{z}_0^N = j^*/\sqrt{N}) \leq (2/e)^k$. Since j^* is even, in applying the proof we may need to replace c_* with a quantity up to $2N^{-1/2}$ larger, but the only effect is an $o(1)$ change in t^*.

To prove the fourth statement, we first note that

$$\int_{\tau_0^*}^{\tau_2^*} |\tilde{z}_s^N| ds \leq (\tau_1^* - \tau_0^*)c_* + \int_{\tau_1^*}^{\tau_2^*} |\tilde{z}_s^N| ds.$$

Since $\tau_1^* \leq \tau_2^*$, the third statement shows $P((\tau_1^* - \tau_0^*)c_* > c_* M) \leq \Theta e^{-\theta M}$. To bound the second term we use the fact that

$$\int_{\tau_1^*}^{\tau_2^*} |\tilde{z}_s^N| ds \leq (\tau_2^* - \tau_1^*) \sup_{t \in [\tau_1^*, \tau_2^*]} |\tilde{z}_t^N|,$$

then control the latter quantity. For ease of notation we suppose $\tilde{z}_0^N = 2j^*/\sqrt{N}$, let $\tau = \inf\{t : \tilde{z}_t^N = 0\}$ and control $\sup_{t \leq \tau} \tilde{z}_t^N$. Define $X_t = \tilde{z}_{t \wedge \tau}^N$, then X is a supermartingale with $X_0 \leq c_* + 2/\sqrt{N} \leq 2c_*$ for large N, and for $\phi > 0$

$$X_t^m - \phi \langle X \rangle_t \geq X_t - X_0 - \sigma_*^2 \phi(t \wedge \tau).$$

Using Lemma 2.1, if $2\phi/\sqrt{N} \leq 1/2$ then $P(X_t > 2c_* + a + \sigma_*^2 \phi(t \wedge \tau)$ for some $t \geq 0) \leq 2e^{-\phi a}$. Fix $\phi = 1$. Since τ is equal in distribution in $\tau_2^* - \tau_1^*$, which is at most $\tau_2^* - \tau_0^*$, from the third statement $P(\tau > M) \leq \Theta e^{-\theta M}$. Letting $a = \theta M$ we find

$$P(\sup_{s \leq \tau} X_s > 2c_* + (\theta + \sigma_*^2)M) \leq (1 + \Theta)e^{-\theta M}.$$

Combining with the estimate of the first term, and another estimate on $\tau_2^* - \tau_1^*$, we obtain the desired result. $\qquad\square$

Proof of Lemma 8.2. We begin by whittling down the statement of Lemma 8.2 until it is ready to analyze in detail. Define

$$L_N^*(s) = \mathbf{1}(s < \tau_N)L_N(s) + \mathbf{1}(s \geq \tau_N) \lim_{t \uparrow \tau_N} L_N(t).$$

In words, L_N^* is equal to L_N until τ_N, at which point it remains at the last value assumed by L_N before τ_N. Since $L_N^*(s) = L_N(s)$ for $s < \tau_N$, the statement of Lemma 8.2 is unchanged if we replace L_N with L_N^*.

In the same manner as $\tilde{z}_N$ is coupled to $\tilde{z}^N$ as described at the beginning of this section, it can be coupled to the full infection process. Recall the c_N^h, c_N^Q control time $\tau_N = \tau_N(c_N^h, c_N^Q, \epsilon)$, defined in Definition 4.4 and given on the slow time scale. Define $\tau_{N,\tilde{z}} = \inf\{t : |\tilde{z}_N(t)| \geq 2C_{4.1}\sqrt{\log N}\}$. By Lemma 8.1 and the definition of τ_N, w.h.p.

$$\tau_{N,\tilde{z}} \geq \tau_{C_{4.1}\sqrt{\log N}}^+(|z_N|) \geq \tau_N,$$

so it is enough to show the statement of Lemma 8.2 holds with $\tau_{N,\tilde{z}}$ in place of τ_N. As in Lemma 8.3 let $c_* = \sqrt{\sigma_z^2/4\mu_z}$. Define $\tau_0^* = \inf\{t : \tilde{z}_N(t) = 0\}$ and for $j \geq 0$,

$$\tau_{2j+1}^* = \inf\{s > \tau_{2j}^* : |\tilde{z}_N(s)| > c_*\}$$
$$\tau_{2j+2}^* = \inf\{s > \tau_{2j+1}^* : \tilde{z}_N(s) = 0\}.$$

The definition of τ_i^*, $i = 0, 1, 2$ differs from Lemma 8.3 only by the choice of time scale. We show the contribution to the integral up to time τ_0^* can be ignored. By Lemma 8.3, w.h.p. $\tau_0^* \leq (2/\mu_z) N^{-1/2} \log N$. Since $i_N, j_N, k_N \leq h_N$, $L(0,0,0) = 0$ and L has Lipschitz constant c_L in the ℓ^1 norm, $|L_N(s)| \leq c_L h_N(s)$. If $s < \tau_N$ then by definition $h_N(s) \leq c_N^h$. Since L_N^* only sees the values $\{L_N(s) \colon s < \tau_N\}$ it follows that $|L_N^*(s)| \leq c_L c_N^h$ for all $s \geq 0$. Since $|\tilde{z}_t^N| \leq 2C_{4.1}\sqrt{\log N}$ for $t < \tau_{N, \tilde{z}}$, using the above estimate on τ_0^* and $c_N^h \leq (c_N^h \log N)^{1/2}$, with high probability

$$\sup_{t \leq T} \left| \int_0^{t \wedge \tau_0^* \wedge \tau_{N, \tilde{z}}} z_N(s) L_N^*(s)\, ds \right| \leq 2C_{4.1}\sqrt{\log N}\, c_L c_N^h (2/\mu_z) N^{-1/2} \log N$$

$$= O(c_L (c_N^h)^{1/2} N^{-1/2} \log^2 N).$$

Thus, without affecting the conclusion, we may assume that $\tau_0^* = 0$, equivalently, $\tilde{z}_N(0) = 0$. At this point we will also replace $t \wedge \tau_{N, \tilde{z}}$ with t as the upper endpoint; this does not decrease the supremum over $t \leq T$. To summarize thus far, it remains to show that, if $\tilde{z}_N(0) = 0$ and $T > 0$ then w.h.p.,

$$\sup_{t \leq T} \left| \int_0^t \tilde{z}_N(s) L_N^*(s)\, ds \right| = O(c_L (c_N^h)^{1/2} (N^{-1/4} \vee (c_N^h c_N^Q)^{1/2}) \log^2 N). \tag{8.5}$$

For $j, k \geq 0$ define $\xi_j = \int_{\tau_{2j}^*}^{\tau_{2j+2}^*} \tilde{z}_N(s)\, ds$,

$$S_k = \sum_{j=0}^k \xi_j L_N^*(\tau_{2j}^*) \quad \text{and} \quad G_k = \int_{\tau_{2k}^*}^{\tau_{2k+2}^*} |\tilde{z}_N(s)(L_N^*(s) - L_N^*(\tau_{2k}^*))|\, ds.$$

Let $\mathcal{I}$ denote the LHS of (8.5). If K_N is such that w.h.p. $\tau_{2K_N}^* \geq T$, then w.h.p.

$$\mathcal{I} \leq |\sup_{k < K_N} S_k| + \sum_{k < K_N} G_k. \tag{8.6}$$

Since the variables $\{\tau_{2j+2}^* - \tau_{2j}^*\}_{j \geq 0}$ are i.i.d., by Lemma 8.3,

$$E[\tau_{2j+2}^* - \tau_{2j}^*] \geq 1/(4\mu_z \sqrt{N}) = \epsilon/\sqrt{N},$$

and since for random variable $X \geq 0$, $E[\theta^2 X^2/2] \leq E[e^{\theta X}]$, by Lemma 8.3,

$$\mathrm{Var}\,(\tau_{2j+2}^* - \tau_{2j}^*) \leq E[(\tau_{2j+2}^* - \tau_{2j}^*)^2] \leq (2\Theta/\theta^2)/N = C/N.$$

Letting $K_N = 2T/E[\tau_2^* - \tau_0^*] = 2\sqrt{N}T/\epsilon$ we have $E[\tau_{2K_N}^*] = 2T$ and $\mathrm{Var}\,(\tau_{2K_N}^*) \leq K_N C/N = 2CT/\epsilon\sqrt{N}$, so using Chebyshev's inequality,

$$P(\tau_{2K_N}^* < T) \leq \frac{\mathrm{Var}\,(\tau_{2K_N}^*)}{T^2} = O(1/\sqrt{N}),$$

and $\tau_{2K_N}^* \geq T$ w.h.p., as desired. It remains to estimate the terms in (8.6).

For $t \geq 0$ let $\mathcal{F}(t)$ denote the information up to time t. The $\{\xi_j\}_{j \geq 0}$ are i.i.d., and by symmetry of $\tilde{z}^N$, ξ_j and $-\xi_j$ are equal in distribution. Because of this and since $L_N^*(\tau_{2j}^*)$ is $\mathcal{F}(\tau_{2j}^*)$-measurable, S is a discrete time martingale. Using Doob's L^2 maximal inequality for martingales and orthogonality of martingale increments,

$$E[(\sup_{k < K_N} S_k)^2] \leq 4E\left[\left(\sum_{j=0}^{K_N - 1} \xi_j L_N^*(\tau_{2j}^*) \right)^2 \right] \leq c_L^2 (c_N^h)^2 \sum_{j=0}^{K_N - 1} E[\xi_j^2]$$

From Lemma 8.3 and noting the change of time scale, there are $\theta, C > 0$ so that for large M, $P\big(\big(\int_{\tau_0^*}^{\tau_2^*} |\tilde{z}_N(s)| ds\big)^2 > CM^4/N\big) \leq Ce^{-\theta M}$, which implies that

$$E[\xi_j^2] \leq E\left[\left(\int_{\tau_0^*}^{\tau_2^*} |\tilde{z}_N(s)| ds\right)^2\right] \leq C/N \tag{8.7}$$

for some possibly larger $C > 0$. It follows that

$$E[|\sup_{k < K_N} S_k|] \leq (E[(\sup_{k < K_N} S_k)^2])^{1/2}$$
$$= O(c_L c_N^h (K_N/N)^{1/2}) = O(c_L c_N^h N^{-1/4}). \tag{8.8}$$

Noting that $c_N^h \leq (c_n^h \log N)^{1/2}$, then using Markov's inequality we find that

$$\text{w.h.p.} \quad \sup_{k < K_N} S_k = O(c_L (c_N^h)^{1/2} N^{-1/4} \log N).$$

It remains to estimate $\sum_{k < K_N} G_k$. Let $\bar{\xi}_k$ denote $\int_{\tau_{2k}^*}^{\tau_{2k+2}^*} |\tilde{z}_N(s)| ds$. Then

$$G_k \leq \bar{\xi}_k \sup_{s \in [\tau_{2k}^*, \tau_{2k+2}^*]} |L_N^*(s) - L_N^*(\tau_{2k}^*)|$$

so by the Cauchy-Schwarz inequality

$$E[G_k] \leq (E[\bar{\xi}_k^2])^{1/2} \left(E\left[\left(\sup_{s \in [\tau_{2k}^*, \tau_{2k+2}^*]} |L_N^*(s) - L_N^*(\tau_{2k}^*)| \right)^2 \right] \right)^{1/2}. \tag{8.9}$$

The first term is $O(1/\sqrt{N})$ by the latter part of (8.7). Let M_k^L denote the supremum inside the second term. To estimate $E[(M_k^L)^2]$, recall (7.3) and note that if $s < \tau_N$ then

$$(i_N(s), \gamma j_N(s), \eta k_N(s)) = (u_*, v_*, w_*) h_N(s) + O((c_N^Q)^{1/2} c_N^h).$$

It follows that

$$M_k^L = O(c_L \sup_{\tau_{2k}^* \leq s < \tau_{2k+2}^*} |h_N(s \wedge \tau_N) - h_N(\tau_{2k}^*)|) + O(c_L (c_N^Q)^{1/2} c_N^h).$$

Let M_k^h denote the above supremum. Using the simple inequality $(a + b)^2 \leq 2(a^2 + b^2)$,

$$E[(M_k^L)^2] = O(c_L^2 E[(M_k^h)^2] + c_L^2 c_N^Q (c_N^h)^2). \tag{8.10}$$

We estimate M_k^h. If $s < \tau_N$ then since $|z_N(s)| \leq C_{4.1}\sqrt{\log N}$ and $h_N(s) \leq c_N^h$, referring to (4.8) we have

$$\mu_s(h_N) = O(\sqrt{\log N} c_N^h + (c_N^h)^2) = O(c_N^h \log N)$$

and so

$$\int_{\tau_{2k}^*}^{s \wedge \tau_N} |\mu_r(h_N)| \, dr = (s \wedge \tau_N - \tau_{2j}^*) O(c_N^h \log N).$$

Since h_N jumps by $1/\sqrt{N}$ at rate $O(Nh_N)$, if $s < \tau_N$ then $\sigma_s^2(h_N) = O(c_N^h)$.

Applying Doob's L^2-maximal inequality,

$$E\left[\left(\sup_{\tau_{2k}^* \leq s \leq t \wedge \tau_N} h_N(s) - h_N(\tau_{2k}^*) - \int_{\tau_{2k}^*}^{s \wedge \tau_N} \mu_r(h_N) \, dr \right)^2 \mid \mathcal{F}_{\tau_{2k}^*} \right]$$
$$\leq 4E\left[\left(h_N(t \wedge \tau_N) - h_N(\tau_{2k}^*) - \int_{\tau_{2k}^*}^{t \wedge \tau_N} \mu_r(h_N) \, dr \right)^2 \mid \mathcal{F}_{\tau_{2k}^*} \right] \leq (t - \tau_{2k}^*) O(c_N^h).$$

Using again the inequality $(a + b)^2 \leq 2(a^2 + b^2)$, from the above two displays we obtain

$$E\left[\left(\sup_{\tau_{2k}^* \leq s \leq t \wedge \tau_N} |h_N(s) - h_N(\tau_{2k}^*)|\right)^2 | \mathcal{F}_{\tau_{2k}^*}\right] \leq (t - \tau_{2k}^*)O(c_N^h) + (t - \tau_{2k}^*)^2 O((c_N^h)^2 \log^2 N)$$

Letting $t = \tau_{2k+2}^*$ and taking an expectation we find that

$$E[(M_k^h)^2] \leq O(c_N^h/\sqrt{N}) + O(N^{-1}(c_N^h)^2 \log^2 N)$$
$$= O(c_N^h/\sqrt{N}).$$

Using (8.10) we find that

$$(E[(M_k^L)^2])^{1/2} = O(c_L(c_N^h)^{1/2}(N^{-1/4} \vee (c_N^h c_N^Q)^{1/2})).$$

Recalling that $K_N = O(\sqrt{N})$ and $(E[\bar{\xi}_k^2])^{1/2} = O(1/\sqrt{N})$, from (8.9) we have

$$E\left[\sum_{k < K_N} G_k\right] = O(c_L(c_N^h)^{1/2}(N^{-1/4} \vee (c_N^h c_N^Q)^{1/2})).$$

Again, using Markov's inequality we find that

$$\text{w.h.p.} \sum_{k < K_N} G_k = O(c_L(c_N^h)^{1/2}(N^{-1/4} \vee (c_N^h c_N^Q)^{1/2}) \log N)$$

and the proof is complete. $\qquad\square$

9 Step 5: bounding the time for h_N to reach $O(1)$

In this section our goal is to prove Lemma 4.7. Before proving Lemma 4.7, we need one additional estimate, that controls the transient behaviour of z^N.

Lemma 9.1. *Let* $\tau = \inf\{t : |z_t^N| \leq C_{4.1} N^{-1/2}\}$. *Then with high probability*

$$\int_0^\tau |z_t^N| dt = O(|z_0^N|) + O((\log N)^2).$$

Proof. We use Lemma 2.5, and assume $z_0^N > 0$; the case $z_0^N < 0$ is analogous. Let $\mu_\star = x = \log N$. Define $X_t = z_t^N - z_0^N - \int_0^t \mu_s(z^N)ds - \mu_\star t$. We note that $X_0 = 0$, $\mu(X) = -\mu_\star$ and $\sigma^2(X) = \sigma^2(z^N) = O(1)$, by (4.3), so let $\sigma_\star^2$ be a large enough constant. Since $\Delta_\star(X) = 2N^{-1/2}$, $\Delta_\star(X)\mu_\star/\sigma_\star^2 = o(1)$, so we can take $C_\Delta = 1$. Since $|\mu(X)| = \mu_\star$, let $C_{\mu_\star} = \mu_\star$. With these choices $\Gamma = \exp(\Omega(\mu x)) = \omega(N)$ and $x/C_{\mu_\star} = \Omega(1)$. By Lemma 2.5, w.h.p. $X_t \leq x$ for all $t \leq N$.

Next we show this implies the desired bound. Since $z_t^N > 0$ for $t < \tau$ by assumption, $\mu_t(z^N) \leq -rz_t^N$, with r as in the proof of Lemma 4.1. Thus if $X_t \leq x$ then

$$z_t^N \leq z_0^N - r\int_0^t z_s^N ds + (t + 1)x.$$

Solving by repeated substitution we find $z_t^N \leq z_0^N e^{-rt} + x((1 - e^{-rt})/r + e^{-rt}) = z_0^N e^{-rt} + O(x)$. So $\int_0^t z_s^N ds = O(z_0^N) + O(xt)$. By Lemma 4.1, $\tau \leq C_{4.1} \log N$ w.h.p., and the result follows. $\qquad\square$

Proof of Lemma 4.7. Recall the goal is to estimate $\tau = \tau_C^-(h_N) = \inf\{t : h_N(t) \leq C\}$, as $C \to \infty$, assuming $h_N(0) = \omega(1)$. We may assume $C \geq 1$ so that $h_N(t) \geq 1$ for $t < \tau$. Recall from (4.8) that

$$\mu(h_N) = c_1 z_N i_N - c_2 i_N^2 + c_3 i_N/\sqrt{N},$$

where $c_i, i = 1, 2, 3$ are positive constants, and that $\sigma^2(h_N) = O(h_N)$. Using the Taylor approximation of Lemma 2.4, if $h_N = \omega(N^{-1/2})$ then

$$\left| \mu\left(\frac{1}{h_N}\right) + \frac{1}{h_N^2} \mu(h_N) \right| = \sigma^2(h_N) O\left(\frac{1}{h_N^3}\right) = O\left(\frac{1}{h_N^2}\right),$$

and combining with the previous display,

$$\mu\left(\frac{1}{h_N}\right) = -\frac{c_1 z_N i_N}{h_N^2} + c_2 \frac{i_N^2}{h_N^2} + o(1) + O\left(\frac{1}{h_N^2}\right).$$

Let $x(t) = 1/h_N(t)$ and $\nu = 1/C$, so that $\tau = \inf\{t : x(t) \geq \nu\}$ and $x(t) < \nu$ for $t < \tau$. Since $h_N(0) \to \infty$, $x(0) = o(1)$. From the above display, if $t < \tau$ then

$$x(t)^p = -\int_0^t \frac{c_1 z_N(s) i_N(s)}{h_N^2(s)} ds + \int_0^t c_2 i_N^2(s)/h_N^2(s) ds + o(t) + \int_0^t O\left(\frac{1}{h_N^2}\right) dt. \qquad (9.1)$$

Since $i_N \leq h_N$ and $h_N(t) \geq 1/\nu$ for $t < \tau$, for large N

$$x(t \wedge \tau)^p \leq (c_1/\nu) \int_0^t |z_N(s)| ds + c_4(t \wedge \tau) \qquad (9.2)$$

where c_4 is another positive constant. Let $\tau_1 = \inf\{t: |z_N(t)| \leq C_{4.1} N^{-1/2}\}$. Since z_N has constant sign on $[0, \tau_1]$, using Lemma 9.1 and recalling that $|z_0| = o(N^{1/2})$ by assumption, w.h.p.

$$\int_0^{\tau_1} |z_N(s)| ds = O(N^{-1/2}|z_0|) + O(N^{-1/2}(\log N)^2) = o(1). \qquad (9.3)$$

To estimate the integral on $s \in [\tau_1, t]$, couple z_N to $\tilde{z}_N$ beginning at time τ_1 so that $|D_N(\tau_1)| \leq 2N^{-1/2}$. Noting that $z_N(\tau_1) \leq C_{4.1} N^{-1/2}$, by Lemma 4.1 and 8.1, w.h.p. $|D_N(s)| \leq N^{-1/4} \log N$ for all $s \in [\tau_1, t]$. Let τ_k^* be as in the proof of Lemma 8.2 except beginning with $\tau_0^* = \tau_1$. Since $\tilde{z}_N(\tau_1) \in [0, c_*]$, the value of $\tau_2^* - \tau_0^*$ is not larger than if $\tilde{z}_N(\tau_0^*) = 0$. Following that proof, if $M_1 > 0$ is a large enough constant then w.h.p. as $N \to \infty$ for fixed t, $\tau_{2\lfloor M_1 \sqrt{N} t \rfloor}^* \geq t$. Since $E[|\xi_j|] = O(1/\sqrt{N})$, letting $M_2 = M_1 \limsup_n (\sqrt{N} E[|\xi_j|])$, it follows that for large N,

$$E\left[\sum_{k \leq M_1 \sqrt{N} t} \int_{\tau_{2k}^*}^{\tau_{2k+2}^*} |\tilde{z}_N(s)| ds \right] \leq M_2 t.$$

Using Markov's inequality on the last display, combining with the bound on $|D_N|$, and using the fact that $\tau_{2\lfloor M_1 \sqrt{N} t \rfloor}^* \geq t$ w.h.p.,

$$P\left(\int_{\tau_1}^t |z_N(s)| ds > (N^{-1/4} \log N + M_2/2\epsilon)t \right) \leq \epsilon/2 + o(1).$$

Combining with (9.3), we find that with probability $\geq 1 - \epsilon/2 - o(1)$,

$$x(t \wedge \tau)^p \leq (c_4 + c_1(M_2/2\epsilon + o(1))/\nu)t.$$

Since $1/h_N$ jumps by $O(N^{-1/2}/h_N^2)$ at rate $O(N h_N)$, $\sigma^2(x) = \sigma^2(1/h_N) \leq c_5/h_N^3 = c_5 x^3$ for some $c_5 > 0$. Using Lemma 2.1, if $a > 0$ and $\Delta_\star(x)\phi \leq 1/2$ then with probability at least $1 - 2e^{-\phi a}$, for all $t \geq 0$

$$|x(t \wedge \tau) - x(0) - x(t \wedge \tau)^p| \leq a + \phi c_5 \nu^3(t \wedge \tau). \qquad (9.4)$$

Using the above bound on $x(t)^p$, noting $x(0) = o(1)$ (since $h(0) \to \infty$) and $\nu \le 1$, with probability $\ge 1 - 2e^{-\phi a} - \epsilon$ for large N,

$$x(t \wedge \tau) \le (c_4 + c_1(M_2/2\epsilon + o(1)))/\nu + c_5\phi)t + a + o(1).$$

Let $\epsilon = \nu$, $a = \nu/4$, $\phi = 1/\nu^2$, let $M_3 = 2(c_4 + c_1 M_2 + c_5)$ and let $t = \nu^3/M_3$. If N is large then $M_2/2\epsilon + o(1) \le M_2/\epsilon = M_2/\nu$. Since $\nu \le 1$ by assumption,

$$x(t \wedge \tau) \le \nu/2 + \nu/4 + o(1) < \nu$$

for large N, implying $\tau > t$. Since $\phi a = 1/4\nu$ and $\epsilon = \nu$, we have shown that

$$\lim_{\nu \to 0^+} \limsup_N P(\tau \le \nu^3/M_3) = 0.$$

To obtain the upper bound we need to take one more term in the Taylor series for $1/h_N$. Expanding to third order and noting that $1/h_N$ jumps by $O(N^{-1/2}/h_N^2)$ at rate $O(Nh_N)$,

$$\mu\left(\frac{1}{h_N}\right) = -\frac{c_1 z_N i_N}{h_N^2} + c_2 \frac{i_N^2}{h_N^2} + \frac{1}{h_N^3}\sigma^2(h_N) + O(N^{-1/2}h_N^{-5}). \tag{9.5}$$

Since we only need an upper bound on τ, we wait for an amount of time $\nu \wedge \tau$ before estimating the compensator. To estimate the first term we note that $L(i_N, h_N) = i_N/h_N^2$ is Lipschitz with constant $c_L = 2$, when $h_N \ge 1$. By Lemma 4.5, we may use Lemma 4.6 with $c_h = \log N$ and $c_Q = N^{-1/6}$ to find that w.h.p.,

$$\sup_{\nu \le t \le 1} \left| \int_{\nu \wedge \tau}^{t \wedge \tau} \frac{c_1 z_N(s) i_N(s)}{h_N^2(s)} ds \right| = O(N^{-1/12} \log^3 N) = o(1).$$

By Lemma 6.1, w.h.p. $c_2 i_N(t)^2/h_N(t)^2 \ge c_6$ for all $\nu \wedge \tau \le t \le N^{1/2} \wedge \tau$ (since $H_t \ge N^{1/5}$ for $t \le \tau$), and some $c_6 > 0$. Combining these observations with (9.5) we find that w.h.p.,

$$\int_\nu^{t \wedge \tau} \mu_s(x) \ge (c_6 - o(1))(t \wedge \tau - \nu \wedge \tau).$$

Recalling that $\sigma_t^2(x) \le c_5 \nu^3$ for $t < \tau$, using Lemma 2.1 and the above display, if $a > 0$ and $\Delta_\star(x)\phi \le 1/2$ then with probability at least $1 - 2e^{-\phi a} - o(1)$, for $t \ge \nu$,

$$x(t \wedge \tau) \ge x(\nu \wedge \tau) + \int_{\nu \wedge \tau}^{t \wedge \tau} \mu_s(x)ds - a - \phi c_5 \nu^3 (t \wedge \tau - \nu \wedge \tau)$$
$$\ge ((c_6 - o(1) - \phi c_5 \nu^3)(t \wedge \tau - \nu \wedge \tau) - a.$$

On the above event, if $\tau > t \ge \nu$ then $t \wedge \tau = t$, $\nu \wedge \tau = \nu$ and $x(t \wedge \tau) = x(t) < \nu$. Taking $\phi = \nu^{-2}$ and $a = \nu$, if ν is small enough and N large enough that $c_5 \nu + o(1) \le c_6/2$, then

$$\nu \ge (c_6/2)(t - \nu) - \nu$$

and so $t \le (1 + 2/c_6)\nu$. Since our choice of ϕ, a gives $e^{-\phi a} = e^{-1/\nu} \to 0$ as $\nu \to 0$, it follows that

$$\lim_{\nu \to 0^+} \limsup_N P(\tau > (1 + 2/c_6)\nu) = 0. \qquad \square$$

10 Small values of H^N

In this section our goal is to prove Lemmas 4.8 and 4.9. First we prove Lemma 4.9. Thus, given $\epsilon > 0$, we suppose $h_N(0) \leq \delta$ where δ is to be determined during the course of the proof. Let $a_k = 2^k \delta$ for integer k and recall $\tau_N(\texttt{ctrl}, 0)$, given by Definition 4.4. To prove Lemma 4.9 we will repeat the following step up to time $\tau_N(\texttt{ctrl}, 0)$: start from $a_k + O(N^{-1/2})$ and run the process up to time $\tau_k^* = \tau_k \wedge \tau_N(\texttt{ctrl}, 0)$, where τ_k is the first exit time of h_N from (a_{k-1}, a_{k+1}). Since jump sizes are $O(N^{-1/2})$, if $\tau_k^* = \tau_k$ then $h_N(\tau_k^*) = a_j + O(N^{-1/2})$ for some $j \in \{k-1, k+1\}$. Since $H_N(t) \geq N^{1/5}$ for $t < \tau_N(\texttt{ctrl}, 0)$, $a_k = \Omega(N^{-0.3})$ for the duration of this iterative procedure. We begin by estimating the expectation of the exit times.

Lemma 10.1. *Fix a positive integer $m > 0$. There is a constant $C_{10.1}$ so that if $4^m \delta$ is small enough, $k \leq m$ and $a_k \geq 2N^{-0.3}$, if $h_N(0) \in (a_{k-1}, a_{k+1})$ then $E[\tau_k^*] \leq C_{10.1} a_k$ for large N.*

Proof. We claim that to complete the proof it is enough to find $C_{10.1}, \epsilon > 0$ so that if $h_N(0) \in (a_{k-1}, a_{k+1})$ then

$$P\left(\tau_k^* > \epsilon C_{10.1} a_k\right) \leq 1 - \epsilon. \tag{10.1}$$

Indeed, one can deduce from (10.1) that

$$P\left(\tau_k^* > n\epsilon C_{10.1} a_k\right) \leq (1 - \epsilon)^n,$$

for any positive integer n. Now the proof finishes by noting that

$$E[\tau_k^*] \leq \sum_{n \geq 1} \epsilon C_{10.1} a_k (1 - \epsilon)^n \leq C_{10.1} a_k.$$

Thus it remains to prove (10.1). To do so we use h_N^2. If $t < \tau_N(\texttt{ctrl}, 0)$ then $i_N(t) \leq h_N(t) \leq \log N$, and as shown in the proof of Theorem 1.4, for some constant σ_*^2, $\sigma_t^2(h_N) = (\sigma_*^2 + o(1)) h_N(t)$. Omitting the t, from (4.8) we then have, for constants $c_1, c_2 > 0$,

$$\begin{aligned} \mu(h_N^2) &= 2h_N \mu(h_N) + \sigma^2(h_N) \\ &= 2c_1 z_N i_N h_N - 2c_2 i_N^2 h_N + (\sigma_*^2 + o(1)) h_N + O(N^{-1/2} \log N). \end{aligned} \tag{10.2}$$

Let $c_N^h = 2a_k$, $c_N^Q = N^{-1/6}$ and $L = i_N h_N$, which is Lipschitz with constant $c_L = O(a_k)$ (to see this expand $i'h' - ih = i'(h' - h) + h(i' - i)$). By the assumptions $k \leq m$ and $4^m \delta$ small, $a_k = O(1)$. Using Lemma 4.6 and noting $(c_N^h c_n^Q)^{1/2} = (N^{-1/6} a_k)^{1/2} = \Omega(N^{-(.15 + 1/12)}) = \omega(N^{-1/4})$, for fixed $T > 0$, w.h.p.

$$\sup_{t \leq T} \left| \int_0^{t \wedge \tau_k^*} z_N(s) i_N(s) h_N(s) ds \right| = O(a_k^2 N^{-1/12} \log^2 N) = o(a_k). \tag{10.3}$$

Since $k \leq m$, if $t < \tau_k^*$ then $i_N(t) \leq h_N(t) \leq 2^{m+1} \delta$ so $2c_2 i_N(t)^2 h_N(t) \leq c_2' 4^m \delta^2 h_N(t)$ for some c_2'. Thus if $4^m \delta > 0$ is sufficiently small and $t < \tau_k^* \wedge T$, from (10.2)-(10.3) we obtain

$$(h^2(t))^p \geq (\sigma_*^2 - c_2' 4^m \delta^2 - o(1)) a_{k-1} t \geq (\sigma^2 a_{k-1}/2) t,$$

for all large N. On the other hand, if $h_N \geq N^{-1/2}$, then h_N^2 jumps by $O(N^{-1/2}) h_N + O(N^{-1}) = O(N^{-1/2} h_N)$ at rate $O(N h_N)$. Thus if $t < \tau_k^* \wedge T$ then $\sigma^2(h_N^2(t)) = O(h_N^3(t)) \leq c_3 a_{k+1}^3$ for some $c_3 > 0$. Noting that $a_k = 2a_{k-1} = a_{k+1}/2$ and letting $\phi = \sigma_*^2/(64 c_3 a_k^2)$, if $t < \tau_k^* \wedge T$ then

$$(h_N^2(t))^p - \phi \langle h_N^2 \rangle_t \geq (\sigma_*^2/4 - 8\phi c_3 a_k^2) a_k t \geq (\sigma_*^2 a_k/8) t.$$

Since for $t < \tau_k^*$, $h_N^2(t) \le a_{k+1}^2 \delta^2 = 4a_k^2\delta^2$, we find that w.h.p.

$$(h_N^2(t))^p - (h_N^2(t) - h_N^2(0)) - \phi\langle(h_N^2)^m\rangle_t \ge (\sigma_*^2 t/8 - 4a_k\delta^2)a_k.$$

Letting $t = T = a_k$, $a = \sigma^2 a_k^2/16$ and taking $\delta \le \sigma_*^2/32$, if $\tau_k^* > t$ then

$$h_N^2(t) - h_N^2(0) - (h_N^2(t))^p \le -(a + \phi\langle(h_N^2)^m\rangle_t).$$

Since $\Delta_\star(h_N^2) = O(N^{-1/2}h_N) = O(N^{-1/2}a_k)$,

$$\Delta_\star(h_N^2)\phi = O(N^{-1/2}/a_k) = O(N^{-0.5+0.3}) = o(1),$$

using the fact that $a_k = \Omega(N^{-0.3})$. Using Lemma 2.1 it follows that

$$P(\tau_k^* > a_k) \le 1 - \epsilon$$

with $\epsilon = 1 - 2e^{-\phi a}$ and $\phi a = (\sigma_*^2)^2/(2^{10}c_3)$. Now the proof of the claim (10.1) follows by setting $C_{10.1} = 1/\epsilon$. This finishes the proof of the lemma. $\qquad\square$

Next, we estimate the exit probabilities.

Lemma 10.2. *Fix a positive integer $m > 0$ and $\epsilon > 0$. Let δ, τ_k, τ_k^* be as in Lemma 10.1 and its proof. Assume $h_N(0) = a_k + O(N^{-1/2}) \ge N^{-0.3}$. Then for large N and all $k \le m$,*

$$P(h_N(\tau_k^*) \ge a_{k+1}) \le 1/3 + \epsilon.$$

Proof. Using the facts that τ_k is the exit time from $(a_{k-1}, a_{k+1}) = (a_k/2, 2a_k)$, $\tau_k^* \le \tau_k$ and h_N jumps by $O(N^{-1/2})$, for any $T > 0$

$$E[h_N(\tau_{k^*} \wedge T)] \ge 2a_k P(h_N(\tau_k^* \wedge T) \ge 2a_k) + (a_k/2)(1 - P(h_N(\tau_k^* \wedge T) \ge 2a_k)) - O(N^{-1/2}).$$

Rearranging and noting $N^{-1/2} = o(a_k)$,

$$\frac{3}{2}P(h_N(\tau_k^* \wedge T) \le 2a_k) \le \frac{1}{a_k}E[h_N(\tau_k^* \wedge T)] - \frac{1}{2} + o(1). \tag{10.4}$$

As with the derivation of (10.2), for $t < \tau_k^*$,

$$\mu_t(h_N) = c_1 z_N(t)i_N(t) - c_2 i_N(t) + O(N^{-1/2}\log N),$$

and so

$$E[h_N(\tau_k^* \wedge T)] - E[h_N(0)] = E\left[\int_0^{\tau_k^* \wedge T}(c_1 z_N(s)i_N(s) - c_2 i_N(s)^2 + O(N^{-1/2}\log N))ds\right]. \tag{10.5}$$

Using Lemma 4.6 with $L = i_N$, $c_N^h = 2a_k$ and $c_N^Q = N^{-1/6}$ and noting as in the previous proof that $(c_N^h c_N^Q)^{1/2} = \omega(N^{-1/4})$,

$$\int_0^{\tau_k^* \wedge T} z_N(s)i_N(s)\,ds = O\left(a_k N^{-1/12}\log^2 N\right) = o(a_k).$$

Now, the second term on the RHS of (10.5) is negative and since $a_k = \Omega(N^{-0.3})$ the third term is $o(a_k)$. So $E[h_N(\tau_k^* \wedge T)] \le E[h_N(0)] + o(a_k) = a_k(1 + o(1))$, noting that $h_N(0) = a_k + O(N^{-1/2}) = a_k + o(a_k)$. Combining with (10.4),

$$P(h_N(\tau_k^* \wedge T) \ge 2a_k) \le (2/3)(1 - 1/2 + o(1)) \le 1/3 + o(1).$$

Using Lemma 10.1 we see that we can take $T > 0$ large enough so that $P(\tau_k^* > T)) \le \epsilon/2$, uniformly for $k \le m$. Combining with the above display gives the desired result. $\qquad\square$

Equipped with Lemma 10.1 and Lemma 10.2 we now prove Lemma 4.9.

Proof of Lemma 4.9. We first recall the following two facts about a simple random walk on $\mathbb{Z}$ with probability $p < 1/2$ of increasing by 1 and $(1 - p)$ of decreasing by 1, at each time step.

1. Starting from 0 the probability to ever reach $k > 0$ is $(p/(1 - p))^k$, and
2. starting from k, the expected number of jumps out of k is equal to $1/(1 - 2p)$.

Fix a positive integer m and suppose $h_N(0) \leq \delta = a_0$. Using Lemma 10.2 and by comparison it follows that, uniformly for $k \leq m$

1. For any $k > 0$, the probability that h_N reaches $[a_k, \infty)$ is at most $(1/2 + o(1))^k$.
2. If $h_N(t) = a_k + O(N^{-1/2})$, the expected number of times we perform the step of exiting (a_{k-1}, a_{k+1}) before time $\tau_N(\texttt{ctrl}, 0)$ is at most $3 + o(1)$.

Using point 2 and summing over the expected number of exits from each level (and adding one for the initial exit, since $h_N(0)$ may not be equal to $a_k + O(N^{-1/2})$ for some k) to find that

$$E[\tau_N(\texttt{ctrl}, 0) \wedge \tau^+_{2^m \delta}(h_N)] \leq C_{10.1}(a_0 + 3 \sum_{k=-\infty}^{0} a_k + 3 \sum_{k=1}^{m} (1/2 + o(1))^k a_k$$

$$= C_{10.1}\delta(1 + 3 + C_m) = D_m \delta$$

where C_m, D_m are constants that depend only on m. Using Markov's inequality,

$$P(\tau_N(\texttt{ctrl}, 0) \wedge \tau^+_{2^m \delta}(h_N) \geq MD_m\delta) \leq 1/M.$$

Let $M = 2/\epsilon$ so the above probability is at most $\epsilon/2$. Since $\tau_N(\texttt{ctrl}, 0) \leq \tau^-_{N^{-0.3}}(h_N)$, using point 1, $P(\tau_N(\texttt{ctrl}, 0) \geq \tau^+_{2^m \delta}(h_N)) \to 0$ as $m \to \infty$, so take m large enough that this probability is at most $\epsilon/2$. Combining,

$$P(\tau_N(\texttt{ctrl}, 0) \geq 2D_m\delta/\epsilon \wedge \tau_{2^m \delta}(h_N)) \leq \epsilon.$$

Take $\delta > 0$ small enough that $2D_m\delta/\epsilon \leq \epsilon$. Since $\tau^+_{2^m \delta}(h_N) \leq \tau^+_{\log N}(h_N)$ for large N, the result is proved. $\square$

Next we prove Lemma 4.8.

Proof of Lemma 4.8. Recall that $\tau = \inf\{t : H^N_t = 0 \text{ or } H^N_t \geq N^{0.24}\}$ and define

$$\tau' = \tau \wedge \tau^+_{C_{4.1}\sqrt{\log N}}(|z^N|).$$

Note that $H^N_0 \leq N^{1/5}$ and that $\tau' \wedge N^{1/4} \geq \tau \wedge N^{1/4}$ w.h.p. by assumption. We will show that i) $\tau' \leq N^{1/4}$ w.h.p. and that ii) $H_{\tau'} < N^{.24}$ w.h.p. To deduce the result from these, first combine i) with the second assumption to find that w.h.p.

$$\tau \wedge N^{1/4} \leq \tau' \wedge N^{1/4} = \tau',$$

and that since $\tau' \leq \tau$, w.h.p. $\tau = \tau' \leq N^{1/4}$. Then, note that H_τ is either $\geq N^{.24}$ or is equal to 0, so that if $\tau = \tau'$ and $H_{\tau'} < N^{.24}$, then $H_\tau = 0$.

Showing that $\tau' \leq N^{1/4}$ w.h.p. The idea is to approximate (I^N, J^N, K^N) by a multi-type continuous time branching process. Such processes are characterized by having transition rates that are a homogeneous linear function of the variables. In (I^N, J^N, K^N) there are two non-linear interactions: $I + I$ and $S + I$ partnership formation. If $t < \tau'$ then since $I^N_t \leq H^N_t \leq N^{.24}$, the rate at which a pair of single I form a partnership is

$O((I^N)^2/N) = O(N^{0.48-1}) = o(N^{-1/4})$. Therefore, with high probability no such events occur on the interval $[0, \tau' \wedge N^{1/4}]$. $S + I$ partnerships form at rate $O(S^N I^N/N)$. If $t < \tau'$ then using the bound on $|z^N|$ (and omitting t),

$$S^N I^N/N = (Y^N - I^N)I^N/N = y_* I^N + z^N I^N/\sqrt{N} - (I^N)^2/N$$
$$= y_* I + O(\sqrt{\log(N)/N} I^N) + o(N^{-1/4})$$
$$= y_* I^N + o(N^{-1/4}).$$

Thus, if we pretend the rate is $y_* I^N$ (i.e., generate transitions using two independent sources of randomness, one with rate $y_* I^N$ and the other with rate $o(N^{-1/4})$), then w.h.p. the process so obtained will be identical to the original process on the time interval $[0, \tau' \wedge N^{1/4}]$.

Thus, the continuous time three-type (Markov) branching process $(\mathcal{I}, \mathcal{J}, \mathcal{K})$ obtained by letting $(\mathcal{I}_0, \mathcal{J}_0, \mathcal{K}_0) = (I_0^N, J_0^N, K_0^N)$ and ignoring $I + I$ transitions and the non-linear part of $S + I$ transitions is such that

$$P((\mathcal{I}_t, \mathcal{J}_t, \mathcal{K}_t) \neq (I_t^N, J_t^N, K_t^N) \text{ for some } t \leq \tau' \wedge N^{1/4}) = o(1).$$

If $(I^N, J^N, K^N) = (0, 0, 0)$ then $t \geq \tau'$, so it would be enough to show the extinction time of $(\mathcal{I}, \mathcal{J}, \mathcal{K})$ is $o(N^{1/4})$ w.h.p. We first extract an embedded one-type CMJ (non-Markov) branching process. To do this, we note the following two points:

1. *Initial decay:* each initial particle of type $\mathcal{J}, \mathcal{K}$ decays at rate $\geq r_-$ (regardless of its type) into $0, 1$ or 2 type $\mathcal{I}$ particles before ever producing additional particles of other types, and

2. *Reproduction cycle:* each type $\mathcal{I}$ particle follows the evolution described in Figure 1, yielding $0, 1$ or 2 type $\mathcal{I}$ particles upon reaching the set $\{D, E, F, G\}$.

Since there are initially $O(N^{0.2})$ particles, by point 1., with high probability, within constant times $\log N$ time every initial particle of type $\mathcal{J}, \mathcal{K}$ has turned into $0, 1$ or 2 type $\mathcal{I}$ particles. Thus, since $\log N = o(N^{1/4})$, we may assume all initial particles have type $\mathcal{I}$.

Point 2. says that we can use the Markov chain described in Figure 1 to determine the timing and number of type $\mathcal{I}$ offspring of each type $\mathcal{I}$ particle. The offspring distribution has mean $R_0 = 1$ and is supported on the set $\{0, 1, 2\}$. Referring to Figure 1, the waiting time to produce offspring is the absorption time at $\{D, E, F, G\}$ starting from A, which is at most exponential$(1 + r_+ y_*)$ + exponential(r_-).

In the embedded one-type process of type $\mathcal{I}$ particles, the set of descendants of any particle forms a critical Galton-Watson tree. We recall a couple of facts that hold for such trees, when the offspring distribution has finite variance. The height of the tree (maximum distance to the root) is greater than or equal to n with probability $O(1/n)$, and the total number of vertices is greater than or equal to n with probability $O(n^{-1/2})$. Thus with $O(N^{0.2})$ initial particles, with high probability the tallest tree has height $O(N^{0.22})$ and the sum of tree sizes is $O(N^{0.44})$. In particular there are in total $O(N^{0.44})$ offspring production events. Since the waiting time for offspring is at most the sum of two exponential random variables with fixed constant rates, the longest waiting time for offspring is whp bounded by constant times $\log N$. Bounding the waiting times by their maximum and noting the height bound, whp the process dies out within $O(N^{0.22} \log N) = o(N^{1/4})$ amount of time.

Showing that $H_{\tau'} < N^{.24}$ w.h.p. Using the bound on $|z^N|$ and the fact that $\gamma/2 < \eta$ and $I^N \leq H^N$, from (4.7) we find that for $t < \tau'$,

$$\mu_t(H^N) \leq \frac{a_1}{2}\left(\sqrt{\frac{\log N}{N}} + O(1/N)\right) I_t^N \leq a_1 \sqrt{\frac{\log N}{N}} H_t^N \qquad (10.6)$$

for some constant $a_1 > 0$. For $b > 0$ to be determined, letting $f(x) = e^{-bx}$ and using Taylor's theorem, for fixed x, y we get

$$f(y) - f(x) = (-b(y-x) + \frac{b^2}{2}(y-x)^2)e^{-bx} - \frac{b^3(y-x)^3}{3!}e^{-bz}$$

where z is between x and y. Applying this to e^{-bH^N}, since $\Delta_\star(H^N) = O(1)$ we will have $e^{-bz} = e^{-b(x+O(1))}$ and $(y-x)^3 = O(1)$ across jump times of H^N. Noting that transition rates are bounded by $O(H^N)$ and multiplying by those transition rates, upon summing over y we obtain

$$\mu(e^{-bH^N}) = \left(-b\mu(H^N) + \frac{b^2}{2}\sigma^2(H^N) + O(b^3 H^N e^{O(b)})\right)e^{-bH^N}.$$

Since $\sigma^2(H^N) = \Omega(H^N)$, applying (10.6) we obtain

$$\mu(e^{-bH^N}) \geq \left(-a_1\sqrt{\frac{\log N}{N}} + \Omega(b) + O(b^2 e^{O(b)})\right)bH^N e^{-bH^N}.$$

Taking $b = N^{-0.22}$ and noting that $b^2 e^{O(b)} = o(b)$ we deduce $\mu(e^{-bH^N})$ is non-negative for large N, so the process $\xi_t = e^{-bH^N_{t\wedge\tau'}}$ is a submartingale. From the definition of τ', the fact that the jumps are of size $O(1)$ and the fact that $e^{-bH^N} \leq 1$, we use the optional stopping theorem to obtain

$$P(H_{\tau'} \geq N^{0.24})\exp(-b(N^{0.24} + O(1))) + P(H_{\tau'} < N^{.24})$$
$$\geq E[e^{-bH_{\tau'}}] \geq E[e^{-bH_0}].$$

Since $H_0 \leq N^{0.2}$, $bH_0 = o(1)$ and $bN^{0.24} = \omega(1)$, so

$$P(H_{\tau'} < N^{.24}) \geq e^{-o(1)} - e^{\omega(1)} = 1 - o(1). \qquad \square$$

11 Computing R_0

In this section we show that (3.9) is equivalent to the condition $R_0 = 1$. To do this we recall (1.1), namely the definition of R_0:

$$R_0 = P_A(X_\tau = F) + 2P_A(X_\tau = G)$$

where τ is the hitting time of $\{D, E, F, G\}$ for the Markov chain with rates drawn in Figure 1. To calculate R_0 we let

$$f(x) = P_x(X_\tau = F) + 2P_x(X_\tau = G)$$

and note that $f(D) = 0$, $f(E) = 0$, $f(F) = 1$, and $f(G) = 2$. By considering what happens on the first jump from each state we see that

$$f(A) = \frac{r_+ y_*}{1 + r_+ y_*}f(B), \tag{11.1}$$

$$f(B) = \frac{\lambda}{\lambda + 1 + r_-}f(C) + \frac{r_-}{\lambda + 1 + r_-} \cdot 1, \tag{11.2}$$

$$f(C) = \frac{2}{2 + r_-}f(B) + \frac{r_-}{2 + r_-} \cdot 2. \tag{11.3}$$

The equations (11.2) and (11.3) can be rewritten as

$$\frac{\lambda + 1 + r_-}{\lambda}f(B) = f(C) + \frac{r_-}{\lambda},$$

$$-\frac{2}{2 + r_-}f(B) = -f(C) + \frac{2r_-}{2 + r_-}.$$

Adding these last two equations we have

$$\left(\frac{\lambda + 1 + r_-}{\lambda} - \frac{2}{2 + r_-}\right) f(B) = r_- \left(\frac{1}{\lambda} + \frac{2}{2 + r_-}\right).$$

Adding the fractions in the parentheses we have

$$\frac{(2 + r_-)(\lambda + 1 + r_-) - 2\lambda}{\lambda(2 + r_-)} \cdot f(B) = r_- \frac{r_- + 2 + 2\lambda}{\lambda(2 + r_-)}.$$

Therefore we deduce

$$f(B) = r_- \frac{r_- + 2 + 2\lambda}{(2 + r_-)(\lambda + 1 + r_-) - 2\lambda} = r_- \frac{r_- + 2 + 2\lambda}{2 + (3 + \lambda)r_- + r_-^2}$$

where we have used the simplification of the denominator used in going from (3.8) to (3.9). Using (11.1) now we have

$$f(A) = \frac{r_+ y_*}{1 + r_+ y_*} \cdot r_- \frac{r_- + 2 + 2\lambda}{2 + (3 + \lambda)r_- + r_-^2}$$

The expression on the RHS above will be equal to 1 when

$$(r_+ y_* + 1)(2 + (3 + \lambda)r_- + r_-^2) = r_-(r_+ y_*)(r_- + 2 + 2\lambda)$$

which is the same as (3.9).

When $R_0 = 1$, $f(A) = 1$, so $f(B) = (1 + r_+ y_*)/r_+ y_*$ and

$$f(C) = \frac{2r_-}{2 + r_-} = \frac{2}{2 + r_-} \cdot f(B). \tag{11.4}$$

12 Appendix

Here we prove any uncited results from Section 2.

Proof of Lemma 2.4. By Itô's lemma [9, Theorem I.4.57],

$$f(X_t) = f(X_0) + (f'(X_-) \cdot X)_t + \frac{1}{2}(f''(X_-) \cdot \langle X^c \rangle)_t + \sum_{s \le t}(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s),$$

where X^c is the continuous part of X^m. Since X has bounded jumps, both X and $\langle X^c \rangle$ are locally integrable. Furthermore, since f, f', f'' are continuous, both are locally bounded, so since X is locally bounded, $f(X), f'(X_-), f''(X_-)$ are locally integrable. Taking the compensator of both sides (and noting that $\langle X^c \rangle$ is its own compensator),

$$f(X_t)^p - (f'(X_-) \cdot X^p)_t = \frac{1}{2}(f''(X_-)\langle X^c \rangle)_t + \left(\sum_{s \le t}(f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s)\right)^p. \tag{12.1}$$

To obtain the result we take the derivative, but we need to estimate the last term more carefully. By a Taylor expansion, the term under the sum is at most $\frac{1}{2}(\Delta X_s)^2 \sup_{|x - X_{s-}| \le \Delta_\star(X)} |f''(x)|$. Thus

$$\left|\sum_{t < s \le t+h} f(X_s) - f(X_{s-}) - f'(X_{s-})\Delta X_s\right| \le \frac{R(h)}{2} \sum_{t < s \le t+h} (\Delta X_s)^2,$$

where

$$R(h) = \sup\{|f''(x)| : |x - X_{s-}| \le \Delta_\star(X) \text{ for some } t < s \le t + h\}.$$

Moreover,

$$\left(\sum_{t < s \leq t+h} (\Delta X_s)^2 \right)^p = \langle X^d \rangle_{t+h} - \langle X^d \rangle_t$$

where X^d is the discontinuous part of X^m. Since X is right-continuous,

$$\lim_{h \to 0^+} R(h) = \sup_{|x - X_t| \leq \Delta_\star(X)} |f''(x)|.$$

Let Q_t denote the right-hand side of (12.1). Since $\langle X \rangle = \langle X^c \rangle + \langle X^d \rangle$, it follows that

$$|Q_{t+h} - Q_t| \leq \frac{1}{2} R(h)(\langle X \rangle_{t+h} - \langle X \rangle_t).$$

Then, dividing by h and letting $h \to 0$ and referring again to (12.1) we obtain the desired result, since $\frac{d}{dt} f(X_t)^p = \mu_t(f(X))$, $\frac{d}{dt} X_t^p = \mu(X)$, $\frac{d}{dt} \langle X \rangle_t = \sigma_t^2(X)$, and $X_{t^-} = X_t$ for a.e. t, since $\{t : \Delta X_t \neq 0\}$ is countable. □

Proof of Lemma 2.5. Suppose $|X_0 - x/2| \leq \Delta_\star(X)/2$. Let $\tau = \inf\{t : |X_t - x/2| \geq x/2\}$. If $t \leq \tau$ and $\phi > 0$ then it follows that

$$X_t^m - \phi \langle X \rangle_t \geq X_t - X_0 + (\mu_\star - \phi \sigma_\star^2) t.$$

Let $\phi = \min(\mu_\star/\sigma_\star^2, 1/(2\Delta_\star(X)))$. Hence, $X_t^m - \phi \langle X \rangle_t \geq X_t - X_0$ for any $t \leq \tau$. Next set $a = (x - \Delta_\star(X))/2$. Therefore, $X_\tau \geq x$ implies that $X_\tau - X_0 \geq a$. Since $\phi \Delta_\star(X) \leq 1/2$, $\phi \geq \mu_\star/2C_\Delta \sigma_\star^2$, and $a \geq x/4$, we can apply Lemma 2.1 to conclude that

$$P(X_\tau \geq x \mid |X_0 - x/2| \leq \Delta_\star(X)/2) \leq 2 \exp(-\mu_\star x/8C_\Delta \sigma_\star^2). \tag{12.2}$$

Since $|\mu_t(X)| \leq C_{\mu_\star}$ for $t < \tau$ and $|X_\tau - X_0| \geq (x - \Delta_\star(X))/2$, it follows that

$$|X_\tau^m| - \phi \langle X \rangle_\tau \geq (x - \Delta_\star(X))/2 - (C_{\mu_\star} + \phi \sigma_\star^2)\tau \geq (x - \Delta_\star(X))/2 - 2C_{\mu_\star}\tau,$$

where the last step follows from the fact that $\phi \sigma_\star^2 \leq \mu_\star \leq C_{\mu_\star}$. Using the same value of ϕ, but changing the value of a to $a = (x - \Delta_\star(X))/4 \geq x/8$, let $t = (x - \Delta_\star(X))/8C_{\mu_\star} \geq x/16C_{\mu_\star}$. Then $\tau \leq t$ implies that $|X_\tau^m| - \phi \langle X \rangle_\tau \geq a$. Therefore, using Lemma 2.1 we further deduce

$$P(\tau \leq x/16C_{\mu_\star} \mid |X_0 - x/2| \leq \Delta_\star(X)/2) \leq 2 \exp(-\mu_\star x/16C_\Delta \sigma_\star^2). \tag{12.3}$$

Thus taking a union bound, from (12.2)-(12.3), we deduce

$$P(\tau \leq x/16C_{\mu_\star} \quad \text{or} \quad X_\tau \geq x \mid |X_0 - x/2| \leq \Delta_\star(X)/2) \leq 4 \exp(-\mu_\star x/16C_\Delta \sigma_\star^2). \tag{12.4}$$

Iterating the estimate $\lfloor \Gamma \rfloor$ times, alternately stopping the process when $|X_t - x/2| \leq \Delta_\star(X)/2$ and $|X_t - x/2| \geq x/2$, the result follows from a union bound.

Indeed, setting $\tau_0 = 0$,

$$\tau_{2j-1} = \inf\{t > \tau_{2j-2} : |X_t - x/2| \leq \Delta_\star(X)/2\},$$

and

$$\tau_{2j} = \inf\{t > \tau_{2j-1} : |X_t - x/2| \geq x/2\},$$

for $j = 1, 2, \ldots$, from (12.4) we see

$$P(X_{\tau_{2j}} \geq x \text{ or } \tau_{2j} - \tau_{2j-1} \leq x/16C_{\mu_\star} | \tau_{2j-1} < \infty) \leq 4 \exp(-\mu_\star x/16C_\Delta \sigma_\star^2).$$

Hence, taking a union bound, as mentioned above, (note that $\tau_{2j-1} = \infty$ for some j automatically implies $\sup_t X_t \leq x$) we see

$$P(X_{\tau_{2j}} \geq x \text{ for some } j \leq \lfloor \Gamma \rfloor, \text{ or } \tau_{2\lfloor \Gamma \rfloor} \leq \lfloor \Gamma \rfloor x/16C_{\mu_\star} | X_0 \leq x/2) \leq 4\lfloor \Gamma \rfloor \exp(-\mu_\star x/16C_\Delta \sigma_\star^2).$$

Since $\Delta_\star(X) < x$ it can be easily checked that if $X_{\tau_{2j}} < x$ for all $j \leq \lfloor \Gamma \rfloor$ then $X_t < x$ for all $t \leq \tau_{2\lfloor \Gamma \rfloor}$. Now the desired probability bound is immediate. This completes the proof. □

Proof of Corollary 2.6. We use the following "continuation trick". Define a new process $\tilde{X}$ by

$$\tilde{X}_t = X_{t \wedge \tau} - \mu_\star(t - t \wedge \tau).$$

In words, $\tilde{X}$ is equal to X up to time τ, at which point it decreases at a fixed deterministic speed $\mu_\star$. Since $t \mapsto \tilde{X}_t$ is continuous on $[\tau, \infty)$, $\Delta_\star(\tilde{X}) \leq \Delta_\star(X)$. Moreover, it is easy to check that $\tilde{X}$ satisfies (2.4) assuming only $0 < \tilde{X}_t < x$ (i.e., without assuming $t < \tau$). Thus, Lemma 2.5 applies to $\tilde{X}$. Since $X_t = \tilde{X}_t$ for $t \leq \tau$, the result follows. □

Proof of Lemma 2.7. We show the conditions (4.1)-(4.7) of Theorem 4.1 of [3, Chapter 7] are satisfied. The fact that a and b are Lipschitz ensures the martingale problem for the limit process is well-posed. $(X^N)^p$ and $\langle X^N \rangle$ play the role of B_N and A_N respectively. Since $v^\top \langle X^N \rangle_t v = \langle v^\top X^N v \rangle_t$ for vector v and the latter is $\mathbb{R}$-valued, the process $\langle X^N \rangle$ is non-negative definite. We also note that $X_i - X_i^p$ and $X_i X_j - \langle X \rangle_{ij}$ are local martingales as required by (4.1) and (4.2).

Since the jump size of X^N tends to 0 (4.3) is satisfied. Applying [9, Lemma I.4.24] we see that jump size of X^p and $\langle X \rangle$ also converge to 0, which gives (4.4) and (4.5). The requirements of (4.6) and (4.7) are immediate from (2.6)-(2.7) above. The rest is straightforward, so we omit the details. □

References

[1] Bramson, M. (1998) State space collapse with application to heavy traffic limits for multiclass queueing networks. *Queueing Systems.* 30, 89–140. MR-1663763

[2] Durrett, R. (2010) *Probability: Theory and Examples.* Fourth Edition, Cambridge U. Press. MR-2722836

[3] Ethier, S.N., and Kurtz, T.G. (1986) *Markov Processes: Characterization and Convergence.* John Wiley and Sons, New York. MR-0838085

[4] Foxall, E. (2016) Critical behaviour of the partner model. *Ann. Appl. Probab.*, 26(5), 2824–2859. MR-3563195

[5] Foxall, E. (2016) Stochastic calculus and sample path estimation for jump processes. *arXiv preprint arXiv:1607.07063.*

[6] Foxall, E. (2017) The naming game on the complete graph. *arXiv preprint arXiv:1703.02088.*

[7] Foxall, E., Edwards, R., and van den Driessche, P. (2016) Social contact processes and the partner model. *Ann. Appl. Probab.* 26(3), 1297–1328. MR-3513591

[8] Harrison, J.M. and Van Mieghem, J.A. (1997) Dynamic control of Brownian networks: state space collapse and equivalent workload formulations. *Ann. Appl. Probab.* 7, 747–771. MR-1459269

[9] Jacod, J. and Shiryaev, A.N. (2002). *Limit theorems for stochastic processes.* Springer Science and Business Media. MR-0959133

[10] Jagers, P. and Nerman, O. (1984) The growth and composition of branching populations. *Advances in Applied Prob.* 16(2) 221–259. MR-0742953

[11] Kallenberg, O. (1997) *Foundations of modern probability.* Springer Science and Business Media. MR-1464694

[12] Kang, H.W., and Kurtz, T.G. (2013) Separation of time-scales and model reduction for stochastic reaction networks. *Ann. Appl. Probab.* 23(2), 529–583. MR-3059268

[13] Kang, H.W., Kurtz, T.G., and Popovic (2014) Central limit theorems and diffusion approximations for multiscale Markov chain models. *Ann. Appl. Probab.* 24(2), 721–759. MR-3178496

[14] Karlin, S., and Taylor, H.M. (1975) *A First Course in Stochastic Processes*, Second Edition. Academic Press, New York. MR-0356197

[15] Kurtz, T. G. (1970) Solutions of ordinary differential equations as limits of pure jump Markov processes. *J. Appl. Prob.* 7(1), 49–58. MR-0254917

[16] Kurtz, T.G. (1971) Limit theorems for sequences of jump Markov processes approximating ordinary differential processes. *Journal of Applied Probability.* 8(2), 344–356. MR-0287609

[17] Kurtz, T.G. (1978) Strong approximation theorems for density-dependent Markov chains. *Stochastic Processes and their Applications.* 6(3), 223–240. MR-0464414

[18] Protter, P. (2005) *Stochastic Integration and Differential Equations*, Second Edition. Springer-Verlag. MR-2273672

[19] Shah, D. and Wischik, D. (2012). Switched networks with maximum weight policies: fluid approximation and multiplicative state space collapse. *Ann. Appl. Probab.* 22(1), 70–127. MR-2932543

[20] Stolyar, A. L. (2004). MaxWeight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *Ann. Appl. Probab.* 14(1), 1–53. MR-2023015

[21] Williams, R. J. (1998) Diffusion approximations for open multiclass queueing metworks: Sufficient conditions involving state space collapse. *Queueing Systems: Theory and Applications.* 30(1), 27–88. MR-1663759