The Annals of Probability
2025, Vol. 53, No. 2, 753–796
https://doi.org/10.1214/24-AOP1721

THE CONTACT PROCESS ON DYNAMIC REGULAR GRAPHS: SUBCRITICAL PHASE AND MONOTONICITY

BY BRUNO SCHAPIRA[1,a] AND DANIEL VALESIN[2,b]

[1]*Centre de Mathématiques et Informatique, Aix-Marseille Université,* [a]*bruno.schapira@univ-amu.fr*
[2]*Department of Statistics, University of Warwick,* [b]*daniel.valesin@warwick.ac.uk*

We study the contact process on a dynamic random d-regular graph with an edge-switching mechanism as well as an interacting particle system that arises from the local description of this process called the herds process. Both these processes were introduced in (da Silva, Oliveira and Valesin (2021)); there it was shown that the herds process has a phase transition with respect to the infectivity parameter λ, depending on the parameter ν that governs the edge dynamics. Improving on a result of (da Silva, Oliveira and Valesin (2021)), we prove that the critical value of λ is strictly decreasing with ν. We also prove that, in the subcritical regime, the extinction time of the herds process started from a single individual has an exponential tail. Finally, we apply these results to study the subcritical regime of the contact process on the dynamic d-regular graph. We show that, starting from all vertices infected, the infection goes extinct in a time that is logarithmic in the number of vertices of the graph, with high probability.

1. Introduction. This paper is a follow-up to [12], which studied the contact process on a dynamic random d-regular graph with an edge-flip mechanism introduced in [10]. The work [12] mainly focused on proving the existence of a supercritical regime, where the extinction time of the process grows exponentially with the number of vertices of the graph. Here we show that there is a phase transition between two regimes, where the order of magnitude of the extinction time switches abruptly from logarithmic to exponential, as the infection parameter crosses a critical value. The highlight of our analysis is that it allows us to establish that this critical value of the infection parameter is a strictly monotone function of the rate of the edge-flip mechanism.

1.1. *Contact process on static finite graphs.* The *contact process* on a graph G is an interacting particle system in which the vertices of the graph can be either healthy or infected. Healthy vertices get infected at rate λ times the number of infected neighbors, where $\lambda > 0$ is a fixed parameter of the model, while infected vertices become healthy at rate 1, independently of each other. When the graph G is infinite, a quantity of interest is the *critical rate* $\lambda_c(G)$, defined as the supremum of the values of λ for which the process started from any finite infected set dies out (reaches the all-healthy configuration) almost surely.

In the case when G is finite, the all-healthy configuration is always reached almost surely (regardless of λ). The *extinction time* τ_G is the hitting time of the all-healthy configuration, for the process started from all infected. It has been observed in several cases that, when $(G_n)_{n\geq 1}$ is a sequence of finite graphs which converges locally to some (rooted) infinite graph G_∞, typically the extinction time τ_{G_n} grows logarithmically with n when λ is smaller than $\lambda_c(G_\infty)$ and grows exponentially with n when λ is larger than $\lambda_c(G_\infty)$. For instance, this has been shown when G_n is a d-dimensional cube $\{0,\dots,n\}^d$ [8, 14, 15, 29, 33], a d-regular tree up to height n [11, 36], or in the case which interests us more here, when G_n is a

Received September 2023; revised August 2024.
MSC2020 subject classifications. Primary 60J85, 60K35; secondary 82C22.
Key words and phrases. Contact process, dynamic graph.

random d-regular graph with n vertices [23, 30], in which cases G_∞ is, respectively, $\mathbb{Z}^d$, the canopy tree, and the d-regular tree $\mathbb{T}^d$.

1.2. *Contact process on a dynamical random d-regular graph.* We present now the dynamical version of the random d-regular graph first introduced and studied in [10]. Throughout the paper we fix the degree $d \geq 3$, and whenever we talk about a d-regular graph with n vertices, we assume that nd is even. We allow our graphs to contain loops (edges involving the same vertex twice) and parallel edges (multiple edges between the same two vertices), but we will keep writing "graph," instead of some other terminology such as "multigraph."

Let G be a d-regular graph with n vertices and e, e' be two of its edges; let u, v be the vertices of e and u', v' the vertices of e'. We can define two possible *switches* of these edges by replacing them with either: (1) edges with vertices $\{u, u'\}$ and $\{v, v'\}$ or (2) edges with vertices $\{u, v'\}$ and $\{v, u'\}$. We then define a continuous-time Markov chain $(G_t)_{t\geq 0}$ on the space of d-regular graphs on a fixed set of n vertices, as follows. The initial graph G_0 is distributed according to the uniform distribution on the set of d-regular graphs. Then, given the state G_t at time t, we prescribe that any of the $2 \cdot \binom{|E|}{2}$ possible edge switches occurs on this graph with rate $\frac{\mathsf{v}}{nd}$, where $\mathsf{v} > 0$ is a positive parameter. It is readily seen that the uniform distribution on random d-regular graphs is stationary with respect to this dynamics and, moreover, that any fixed edge is involved in a switch at a rate which converges to v, as $n \to \infty$. It is also interesting to note that a d-regular graph on n vertices sampled uniformly at random is connected with high probability as $n \to \infty$ (see [4], Section 7.6).

We next consider the process $(G_t, \xi_t)_{t\geq 0}$, where $(G_t)_{t\geq 0}$ is as above and $(\xi_t)_{t\geq 0}$ is a contact process evolving on the dynamic graph. As previously mentioned, the process starts from the configuration where all vertices are infected, and our main interest is in the time $\tau_{(G_t)}$ when the process reaches the all-healthy configuration. The following result was proved in [12]. In both this theorem and in Theorem 1.2 below, the probability measure $\mathbb{P}$ includes the randomness of both the random dynamic graph and the contact process.

THEOREM 1.1 ([12]). *For each $\mathsf{v} > 0$, there exists $\bar{\lambda}(\mathsf{v}) \in (0, \lambda_c(\mathbb{T}^d))$ such that the following holds. For any $\lambda > \bar{\lambda}(\mathsf{v})$, there exists $c > 0$ such that*

$$\mathbb{P}\big(\tau_{(G_t)} > \exp\{cn\}\big) \xrightarrow{n\to\infty} 1.$$

Note, in particular, the interesting feature that $\bar{\lambda}(\mathsf{v})$ is strictly smaller than $\lambda_c(\mathbb{T}^d)$, which means that the dynamics of the graph helps the contact process to survive for a longer time than in the static model.

1.3. *Main results.* In this paper we complete the picture by proving the following result.

THEOREM 1.2. *For each $\mathsf{v} > 0$, there exists $\bar{\lambda}(\mathsf{v}) \in (0, \lambda_c(\mathbb{T}^d))$ such that the following holds*:

(i) *For any $\lambda > \bar{\lambda}(\mathsf{v})$, there exists $c > 0$ such that*

$$\mathbb{P}\big(\tau_{(G_t)} > \exp\{cn\}\big) \xrightarrow{n\to\infty} 1.$$

(ii) *For any $\lambda < \bar{\lambda}(\mathsf{v})$, there exists $C > 0$ such that*

$$\mathbb{P}(\tau_{(G_t)} > C \log n) \xrightarrow{n\to\infty} 0.$$

As in the static case, the value $\bar{\lambda}(\mathsf{v})$ corresponds to the critical value for the contact process on a limiting model, which in our case is called the *herds process*. This was introduced and

analyzed in [12], where, in particular, a phase transition delimited by a positive and finite parameter $\bar{\lambda}(\mathsf{v})$ was established. Here we improve upon this result by showing that the herds process exhibits a form of sharp threshold phenomenon, namely, that the tail distribution of the extinction time decays exponentially fast in the whole subcritical regime (see Lemma 3.2 below). An informal description of the herds process is given in the next subsection, and a precise definition is given in Section 2.

Our second result answers a question of [12] concerning the monotonicity of $\bar{\lambda}(\mathsf{v})$.

THEOREM 1.3. *The mapping* $\mathsf{v} \mapsto \bar{\lambda}(\mathsf{v})$ *is strictly decreasing.*

By combining Theorems 1.2 and 1.3, we see that, for any $\mathsf{v}_1 < \mathsf{v}_2$, there exist values of λ such that the contact process on the dynamic graph (G_t) dies out in logarithmic time at parameter values (λ, v_1) and survives exponentially long at parameter values (λ, v_2). A natural direction of investigation would be to show that, when $\mathsf{v}_1 < \mathsf{v}_2$ and $\lambda > \bar{\lambda}(\mathsf{v}_1)$ (so that λ is supercritical for both speeds), the typical ("metastable") density of infection in the graph is larger at (λ, v_2) than it is at (λ, v_1). Using elementary duality properties of the contact process, this problem is related to proving strict monotonicity of the extinction probability of the herds process with respect to v. Since our methods do not involve direct comparison through coupling between processes with different choices of v, we are unable to settle this issue; see also the discussion in Remark 2.4 about this difficulty. In fact, when $0 = \mathsf{v}_1 < \mathsf{v}_2$, a coupling between the two processes entailing stochastic domination can be defined, and this was used in [12] to show that $\bar{\lambda}(\mathsf{v}) < \bar{\lambda}(0) = \lambda_c(\mathbb{T}^d)$, for any $\mathsf{v} > 0$, but this is a very special case.

1.4. *Methods of proof and organization of the paper.* As in [12], the proof of Theorem 1.2 relies on a detailed analysis of the herds process. Informally, this process evolves as a contact process on a family of d-regular trees, where the number of trees also evolves with time. On each tree the process obeys the same rules as the usual contact process, regarding infection and recovery (though we adopt a slight change of terminology: the vertex states "healthy" and "infected" here are called "empty" and "occupied by a particle," respectively). In addition, each edge in any of the existing trees splits the tree into two pieces at a constant rate v. When this happens, the two disjoint pieces of the tree are extended to form two new copies of a d-regular tree.

The value $\bar{\lambda}(\mathsf{v})$ is defined as the threshold for the infection parameter λ above which the process has a positive probability of surviving forever, when starting from a single tree with a single particle. The heart of the proof is to show that, when λ is smaller than this threshold, the probability to survive for a time larger than t decays exponentially fast with t. This is obtained using some coupling argument with a two-type herds process, which allows to show that the expected number of infected particles at time t, denoted (in this section only) $F(\lambda, \mathsf{v}, t)$, is a submultiplicative sequence (as a function of the time parameter); see Section 2.2. Using this, we can define the rate of exponential decay of this function, $\varphi(\lambda, \mathsf{v})$, and then the proof boils down to showing that it is strictly increasing with respect to both parameters. This is obtained via a kind of Russo's formula; see Proposition 3.7, which in our setting is quite involved, compared to the original formula from percolation theory. Moreover, the strict monotonicity of $\varphi(\lambda, \mathsf{v})$ also proves Theorem 1.3.

Another important ingredient is to control the higher moments of the number of infected particles, which requires some serious additional technical work, due to the nonlinearity of these functionals; see Section 3.3. Bounding higher moments is needed to control the total number of particle births (or in the usual contact process terminology, infections) up to the extinction time, which in turn allows one to couple the contact process on a large dynamic d-regular graph with the herds process, up to this extinction time. This coupling argument is explained in Section 4, where we complete the proof of Theorem 1.2.

1.5. *Related works.* The study of the phase transition for long vs. short time extinction for the contact process on finite graphs and the closely related question of metastability in the supercritical regime has been studied intensively over the past years, in particular, as we already mentioned on finite boxes of $\mathbb{Z}^d$ [8, 14–16, 29, 33], finite d-regular trees [11, 36], and random d-regular graphs [23, 30], but also in a number of other examples, such as the configuration model with power law degree distribution [7, 9, 28], or with general degree-distribution [3, 19], preferential attachment graphs [2, 5], Erdős–Rényi random graphs [3], inhomogeneous random graphs [6], or hyperbolic random graphs [25]. There are also results concerning some general classes of finite graph sequences [27, 32].

On the other hand, the study of the contact process on dynamical graphs started more recently; see, for example, [12, 13, 18, 20–22, 26, 31, 34, 35].

The phenomenon identified in our Theorem 1.3—roughly put, "more rapid graph dynamics helps the infection survive"—should not be expected to be universal for the contact process on dynamic graphs. In graphs where degrees are highly heterogeneous, high-degree vertices ("stars") play a crucial role in maintaining the infection when the infection rate is small, and this mechanism can be compromised by rapid breaking of edges. This is the case, for instance, in the power law graphs studied in [21]: in Theorem 1 there, there are situations when increasing the parameter η, which controls the speed of graph updates, causes a fast extinction regime to appear, thus in the opposite direction as in our setting. It would be interesting to investigate situations where degrees are heterogeneous but not as drastically as in a power law random graph, for instance, when degrees are bounded. Presumably, in such a setting, an analysis of a limiting graph dynamics could still be carried out, and our methods involving differential inequalities could be helpful.

2. Preliminaries on the herds process.

In this section we give a formal definition of the herds process and introduce notation. We also present some tools that will be employed in the analysis of this process in later sections, namely, stochastic domination by a pure-birth process and a submultiplicativity inequality for the expectation of the number of particles.

2.1. *Definition and construction.* Throughout this paper we fix $d \geq 3$ and let $\mathbb{T}^d$ denote the infinite d-regular rooted tree. The root is denoted by o, and we write $u \sim v$ when two vertices u and v are neighbors.

DEFINITION 2.1. Let

$$P_{\mathrm{f}}(\mathbb{T}^d) := \{A \subset \mathbb{T}^d : A \text{ is finite and nonempty}\}.$$

We call each $A \in P_{\mathrm{f}}(\mathbb{T}^d)$ a *herd shape* and each $x \in A$ a *particle* of A. Given a herd shape A and an edge $e = \{u, v\}$ of $\mathbb{T}^d$ with u closer to the root than v, in the graph distance of $\mathbb{T}^d$, define

$$A_{e,1} := \{w \in A : w \text{ is closer to } u \text{ than to } v\}, \qquad A_{e,2} := A \setminus A_{e,1}.$$

We say that e is an *active edge of* A if $A_{e,1} \neq \varnothing$ and $A_{e,2} \neq \varnothing$.

DEFINITION 2.2. Define the *set of herd configurations*

$$\mathcal{S} := \left\{\xi : P_{\mathrm{f}}(\mathbb{T}^d) \to \mathbb{N}_0 \text{ with } \sum_A \xi(A) < \infty\right\}.$$

In a herd configuration $\xi \in \mathcal{S}$, $\xi(A)$ is interpreted as the number of herds with shape A. Given $A \in P_{\mathrm{f}}(\mathbb{T}^d)$, we let δ_A denote the herd configuration such that $\delta_A(B) = 1$ if $B = A$, and $\delta_A(B) = 0$ otherwise. An *enumeration* of $\xi \in \mathcal{S}$ is a sequence $A_1, \ldots, A_m \in P_{\mathrm{f}}(\mathbb{T}^d)$ such that $\xi = \sum_{i=1}^m \delta_{A_i}$.

We will generally denote deterministic elements of $\mathcal{S}$ by the letter ξ and random elements of $\mathcal{S}$ by Ξ.

DEFINITION 2.3. The *herds process* $(\Xi_t)_{t\geq 0}$ with birth rate $\lambda > 0$ and splitting rate $\mathsf{v} > 0$ is a continuous-time Markov chain on $\mathcal{S}$ whose dynamics is given by the following description of possible jumps and corresponding rates:

(a) *Death in herds with more than one particle*: For each A with $|A| > 1$ such that $\xi(A) > 0$, and for each $x \in A$, with rate $\xi(A)$, the process jumps from ξ to $\xi - \delta_A + \delta_{A\setminus\{x\}}$.

(b) *Death in herds with one particle*: For each A with $|A| = 1$ such that $\xi(A) > 0$, with rate $\xi(A)$, the process jumps from ξ to $\xi - \delta_A$.

(c) *Birth*: for each A such that $\xi(A) > 0$, and for each $y \in \mathbb{T}^d$ with $y \notin A$, with rate $\lambda \cdot |\{x \in A : x \sim y\}| \cdot \xi(A)$, the process jumps from ξ to $\xi - \delta_A + \delta_{A\cup\{y\}}$.

(d) *Split*: for each A such that $\xi(A) > 0$, and for each active edge e of A, with rate $\mathsf{v}\cdot\xi(A)$, the process jumps from ξ to $\xi - \delta_A + \delta_{A_{e,1}} + \delta_{A_{e,2}}$.

REMARK 2.1. A priori, it could be the case that the above description gave rise to an explosive chain, that is, a chain that jumps infinitely many times in a bounded time interval. So, strictly speaking, the process is only defined up to the explosion time (the infimum of times $t > 0$ such that infinitely many jumps happen in $(0, t)$). However, we will show shortly (see Corollary 2.2 below) that the chain is, in fact, not explosive.

REMARK 2.2. Our choice for the state space $\mathcal{S}$ of the herds process makes it so that, in case there are multiple herds with the same shape A (meaning that $\xi(A) \geq 2$), then these herds are indistinguishable. An alternative choice was made in [12]: there, a state of the process was an index set $\mathcal{J}$ and a mapping from $\mathcal{J}$ to $P_{\mathrm{f}}(\mathbb{T}^d)$ so that each $i \in \mathcal{J}$ represented a different herd. This alternative choice requires heavier notation but has some advantages; for instance, when a particle is born in a herd, it makes sense to consider the herd before and after the birth (since it keeps the same index). Although here we will adopt the leaner description of Definition 2.3, we will sometimes pretend that a richer description is available. For instance, in one of our arguments (see Lemma 3.3), we fix a particle in a herd at time 0, and consider the evolution of the cardinality of the herd containing that particle for times $t \geq 0$.

We let $\mathbb{P}$ be a probability measure under which the herds process is defined and $\mathbb{E}$ the associated expectation. When we want to be explicit about the parameters, we will write $\mathbb{P}_{\lambda,\mathsf{v}}$ and $\mathbb{E}_{\lambda,\mathsf{v}}$. *When no explicit mention regarding the initial configuration is made, we assume it to consist of a single herd with a single particle placed at the root vertex*. We may write $\mathbb{P}_{\lambda,\mathsf{v}}(\cdot|\Xi_0 = \xi)$ (and similarly, $\mathbb{E}_{\lambda,\mathsf{v}}[\cdot|\Xi_0 = \xi]$) to specify some other initial configuration $\xi \in \mathcal{S}$.

DEFINITION 2.4. Given $\xi \in \mathcal{S}$, we let

$$X(\xi) := \sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} |A| \cdot \xi(A), \tag{1}$$

where $|\cdot|$ denotes cardinality; that is, $X(\xi)$ is the total number of particles among all herds in ξ. We also let

$$\mathscr{E}(\xi) := \sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \big|\{\text{active edges of } A\}\big| \cdot \xi(A), \tag{2}$$

the total number of active edges among all herds of ξ. For the herds process $(\Xi_t)_{t\geq 0}$ (started from an arbitrary configuration), we write

$$X_t := X(\Xi_t), \qquad \mathscr{E}_t := \mathscr{E}(\Xi_t), \tag{3}$$

with X and $\mathscr{E}$ as in (1) and (2), respectively.

We now state a useful stochastic domination result. The proof involves a quick inspection of transition rates, and we omit it.

LEMMA 2.1 (Domination by pure-birth chain). *For the herds process $(\Xi_t)_{t\geq 0}$ with parameters λ, ν and some initial configuration $\xi \in \mathcal{S}$, let N_t denote the number of birth events until time t. Let $(Z_t)_{t\geq 0}$ be the continuous-time Markov chain on $\mathbb{N}$ with $Z_0 = X(\xi)$ and jump rates*

$$q(i, i+1) = d\lambda i, \qquad q(i, j) = 0 \quad \text{for } j \neq i+1. \tag{4}$$

Then $(N_t)_{t\geq 0}$ is stochastically dominated by $(Z_t - Z_0)_{t\geq 0}$. In particular, $\mathbb{P}(N_t < \infty) \geq \mathbb{P}(Z_t < \infty) = 1$ for any t.

COROLLARY 2.2. *The herds process is nonexplosive.*

PROOF. As above, let N_t denote the number of birth events of the herds process until time t. The number of death events in $[0, t]$ is necessarily smaller than $X(\Xi_0) + N_t$. Moreover, the number of splitting events in $[0, t]$ is at most $\mathscr{E}(\Xi_0) + N_t$ (indeed, births cause the number of active edges to increase by at most one, while deaths and splittings cannot increase the number of active edges). This shows that the total number of jumps of the process in $[0, t]$ is at most $X(\Xi_0) + \mathscr{E}(\Xi_0) + 3N_t$. Since N_t is almost surely finite, we conclude that the number of jumps also is. $\square$

We also have the following important consequence concerning the processes (X_t) and $(\mathscr{E}_t)$ defined in (3).

COROLLARY 2.3. *For any $\lambda > 0$, $\nu > 0$, $T > 0$, and $k \geq 1$, there exists $c > 0$ such that the herds process $(\Xi_t)_{t\geq 0}$ with parameters λ, ν and arbitrary (deterministic) initial configuration ξ satisfies*

$$\mathbb{E}\Big[\max_{0\leq t\leq T} (X_t)^k \,\Big|\, \Xi_0 = \xi\Big] \leq cX(\xi)^k$$

and

$$\mathbb{E}\Big[\max_{0\leq t\leq T} (\mathscr{E}_t)^k \,\Big|\, \Xi_0 = \xi\Big] \leq c\big(X(\xi) + \mathscr{E}(\xi)\big)^k.$$

PROOF. Again, let N_t denote the number of births in the herds process until time t. Note that

$$\max_{0\leq t\leq T} X_t \leq X_0 + N_T \quad \text{and} \quad \max_{0\leq t\leq T} \mathscr{E}_t \leq \mathscr{E}_0 + N_T; \tag{5}$$

to justify the latter, we observe that deaths and splits can only decrease $\mathscr{E}_t$, while a birth can increase $\mathscr{E}_t$ by at most one.

Let $(Z_t)_{t\geq 0}$ be the pure-birth chain of Lemma 2.1, started with $Z_0 = X_0$. By that lemma, $(Z_t - Z_0)_{t\geq 0}$ stochastically dominates $(N_t)_{t\geq 0}$. Then

$$\mathbb{E}\Big[\max_{0\leq t\leq T} (X_t)^k \,\Big|\, \Xi_0 = \xi\Big] \leq \mathbb{E}\big[(X_0 + N_T)^k \,\big|\, \Xi_0 = \xi\big] \leq \mathbb{E}\big[Z_T^k\big],$$

where the last expectation is with respect to the probability measure under which (Z_t) is defined (note that $Z_0 = X_0$ is fixed and deterministic). Next, the law of Z_T is equal to the

law of $\sum_{i=1}^{Z_0} \zeta_T^{(i)}$, where $(\zeta_t^{(1)}), \ldots, (\zeta_t^{(Z_0)})$ are independent pure-birth processes with rates as in (4), each started with a population of one. Using Minkowski's inequality,

$$\mathbb{E}[Z_T^k]^{1/k} \le \sum_{i=1}^{Z_0} \mathbb{E}\big[\big(\zeta_T^{(i)}\big)^k\big]^{1/k} = cX_0,$$

where $c := \mathbb{E}[(\zeta_T^{(1)})^k]^{1/k}$, which by [1], Corollary 1, page 111, is finite and only depends on λ, k and T. This completes the proof of the first bound. For the second one, we start using the second bound in (5),

$$\mathbb{E}\Big[\max_{0\le t\le T} (\mathscr{E}_t)^k \,|\, \Xi_0 = \xi\Big] \le \mathbb{E}\big[(\mathscr{E}_0 + N_T)^k \,|\, \Xi_0 = \xi\big] \le \mathbb{E}\big[(\mathscr{E}_0 + Z_T - Z_0)^k \,|\, Z_0 = X_0\big],$$

and then complete the proof as in the previous case. □

We now present three properties of the herds process (in Lemmas 2.4, 2.5, and 2.6 below). In all three cases, the proof is elementary and omitted.

LEMMA 2.4 (Invariance under tree automorphisms). *Let $\psi : \mathbb{T}^d \to \mathbb{T}^d$ be a graph automorphism. Fix $A \in P_{\mathrm{f}}(\mathbb{T}^d)$, and let $(\Xi_t)_{t\ge0}$ and $(\Xi'_t)_{t\ge0}$ be herds processes started from δ_A and $\delta_{\psi(A)}$, respectively. Then the process*

$$\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \Xi_t(A)\cdot\delta_{\psi(A)}, \quad t\ge 0$$

has the same distribution as $(\Xi'_t)_{t\ge0}$. In particular, the processes $(X(\Xi_t))_{t\ge0}$ and $(X(\Xi'_t))_{t\ge0}$ have the same distribution.

LEMMA 2.5 (Decomposition into independent processes). *Let $\xi \in \mathcal{S}$ with enumeration $\xi = \sum_{i=1}^n \delta_{A_i}$, where $A_1, \ldots, A_n \in P_{\mathrm{f}}(\mathbb{T}^d)$. Then the herds process, started from ξ, has the same distribution as $(\Xi_t^{(1)} + \cdots + \Xi_t^{(n)})_{t\ge0}$, where $(\Xi_t^{(1)})_{t\ge0}, \ldots, (\Xi_t^{(n)})_{t\ge0}$ are independent herds processes, started from $\delta_{A_1}, \ldots, \delta_{A_n}$, respectively.*

Before stating the third property, we define a partial order on $\mathcal{S}$.

DEFINITION 2.5. Given two herd configurations ξ and ξ', we write $\xi \preceq \xi'$ if there exist enumerations

$$\xi = \sum_{i=1}^m \delta_{A_i}, \qquad \xi' = \sum_{j=1}^n \delta_{A'_j}$$

such that $m \le n$ and $A_i \subset A'_i$, for $i = 1, \ldots, m$.

LEMMA 2.6 (Attractiveness). *If $\xi \preceq \xi'$, then $(\Xi_t)_{t\ge0}$, started from ξ, is stochastically dominated (with respect to $\preceq$) by $(\Xi'_t)_{t\ge0}$ started from ξ'.*

We now give a definition pertaining to extinction vs. survival of the herds process and introduce the value $\bar\lambda(\nu)$ that appears in the statements of our main theorems.

DEFINITION 2.6. We say that the herds process *survives* if the event $\{\sum_A \Xi_t(A) > 0\ \forall t\}$ occurs, that is, the process always has particles; otherwise, we say that the process *dies out*. For any $\nu > 0$, we define $\bar\lambda(\nu)$ as the supremum of the values of λ for which the process with parameters λ and ν dies out with probability 1.

It is easy to see that $\bar\lambda(\mathsf{v}) \geq 1/d$. Indeed, when $\lambda < 1/d$, the rate at which existing particles die always exceeds the rate at which new particles are born, so the process eventually reaches the empty configuration. A moment's thought, using, for instance, a comparison with a branching process, shows that $\bar\lambda(\mathsf{v}) < \infty$.

REMARK 2.3. The critical value $\bar\lambda(0)$ is just the critical rate separating between the extinction and the (global) survival regimes on the (static) infinite d-regular tree. It is known that $d \cdot \bar\lambda(0) \xrightarrow{d\to\infty} 1$ (see [24], p. 135, (5.4)). We can then also conclude that, for any $\mathsf{v} > 0$, we have $d \cdot \bar\lambda(\mathsf{v}) \xrightarrow{d\to\infty} 1$, by combining the bound $\bar\lambda(\mathsf{v}) \geq 1/d$ mentioned above with the bound $\bar\lambda(\mathsf{v}) < \bar\lambda(0)$, which was already given in [12], Theorem 3.

2.2. *Submultiplicativity of number of particles.* The goal of this section is to prove the following inequality. Recall that X_t denotes the number of particles in the herds process at time t. Also, recall that, whenever the initial condition of the herds process is omitted (say, as in the expectation in the right-hand side of (6) below), it is equal to $\delta_{\{o\}}$.

PROPOSITION 2.7. *For any $t \geq 0$, $p \geq 1$ and $\xi \in \mathcal{S}$, we have*

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \xi\big] \leq X(\xi)^p \cdot \mathbb{E}\big[X_t^p\big]. \tag{6}$$

Consequently, for any $t, s \geq 0$, $p \geq 1$ and $\xi \in \mathcal{S}$, we have

$$\mathbb{E}\big[X_{t+s}^p \,|\, \Xi_0 = \xi\big] \leq \mathbb{E}\big[X_s^p \,|\, \Xi_0 = \xi\big] \cdot \mathbb{E}\big[X_t^p\big]. \tag{7}$$

This proposition will be a consequence of the following lemma.

LEMMA 2.8. *Let $A, B \in P_{\mathrm{f}}(\mathbb{T}^d)$ be disjoint and $p \geq 1$. Then*

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_{A\cup B}\big]^{1/p} \leq \mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_A\big]^{1/p} + \mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_B\big]^{1/p}.$$

We postpone the proof of this lemma and, for now, show how it implies the proposition.

PROOF OF PROPOSITION 2.7. We claim that, for any $p \geq 1$, $A \in P_{\mathrm{f}}(\mathbb{T}^d)$, and $t \geq 0$,

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_A\big]^{1/p} \leq |A| \cdot \mathbb{E}\big[X_t^p\big]^{1/p}. \tag{8}$$

We prove this by induction on $|A|$. For $|A| = 1$, the above holds with an equality by Lemma 2.4. For the induction step, we assume that $|A| \geq 2$, take $u \in A$ and bound, using Lemma 2.8, Lemma 2.4, and the induction hypothesis,

$$\begin{aligned}\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_A\big]^{1/p} &\leq \mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_{A\setminus\{u\}}\big]^{1/p} + \mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_{\{u\}}\big]^{1/p} \\ &\leq \big(|A\setminus\{u\}|\big) \cdot \mathbb{E}\big[X_t^p\big]^{1/p} + \mathbb{E}\big[X_t^p\big]^{1/p} \\ &= |A| \cdot \mathbb{E}\big[X_t^p\big]^{1/p}.\end{aligned}$$

Now, take $\xi \in \mathcal{S}$ with enumeration $\xi = \sum_{i=1}^m \delta_{A_i}$. Let $(\Xi_t^{(1)}), \ldots, (\Xi_t^{(m)})$ be independent herds processes, started from $\delta_{A_1}, \ldots, \delta_{A_m}$, respectively. By Lemma 2.5 we have

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \xi\big]^{1/p} = \mathbb{E}\left[\left(\sum_{i=1}^m X\big(\Xi_t^{(i)}\big)\right)^p\right]^{1/p}.$$

Minkowski's inequality gives

$$\mathbb{E}\left[\left(\sum_{i=1}^{m} X\big(\Xi_t^{(i)}\big)\right)^p\right]^{1/p} \le \sum_{i=1}^{m} \mathbb{E}\big[X\big(\Xi_t^{(i)}\big)^p\big]^{1/p}. \tag{9}$$

By (8) the right-hand side is smaller than

$$\sum_{i=1}^{m} |A_i| \cdot \mathbb{E}\big[X_t^p\big]^{1/p} = X(\xi) \cdot \mathbb{E}\big[X_t^p\big]^{1/p}.$$

We have thus proved (6). To prove (7), we use the Markov property. Let $s, t \ge 0$ and $\xi \in \mathcal{S}$; we have

$$\begin{aligned}\mathbb{E}\big[X_{t+s}^p \,|\, \Xi_0 = \xi\big] &= \sum_{\xi' \in \mathcal{S}} \mathbb{E}\big[X_{t+s}^p \,|\, \Xi_s = \xi'\big] \cdot \mathbb{P}\big(\Xi_s = \xi' \,|\, \Xi_0 = \xi\big) \\ &\le \sum_{\xi' \in \mathcal{S}} X(\xi')^p \cdot \mathbb{E}\big[X_t^p\big] \cdot \mathbb{P}\big(\Xi_s = \xi' \,|\, \Xi_0 = \xi\big) \\ &= \mathbb{E}\big[X_s^p \,|\, \Xi_0 = \xi\big] \cdot \mathbb{E}\big[X_t^p\big]. \end{aligned}$$

□

To prove Lemma 2.8, we will define an auxiliary process, which informally describes two herds processes that evolve almost independently, except that they share the same splitting events. We start by defining the state space of this process.

DEFINITION 2.7. Let

$$P_{\mathrm{f},2}\big(\mathbb{T}^d\big) := \big\{(A, B) : A, B \subset \mathbb{T}^d, A \cup B \text{ finite and nonempty}\big\}$$

and

$$\mathcal{S}_2 := \left\{\widetilde{\xi} : P_{\mathrm{f},2}\big(\mathbb{T}^d\big) \to \mathbb{N}_0 \text{ with } \sum_{(A,B)} \widetilde{\xi}(A, B) < \infty\right\}.$$

We interpret an element $(A, B) \in P_{\mathrm{f},2}(\mathbb{T}^d)$ as a *two-type herd*, that is, there are two species of particles, one of which occupies A and the other B. We emphasize that A and B need not be disjoint, and one of them, but not both, can be empty. We will also need some projection functions from $\mathcal{S}_2$ to $\mathcal{S}$.

DEFINITION 2.8. We define $\pi, \pi_1, \pi_2 : \mathcal{S}_2 \to \mathcal{S}$ by setting, for $\xi' = \sum_{i=1}^m \delta_{(A_i, B_i)}$:

$$\pi(\widetilde{\xi}) = \sum_{i=1}^{m} \delta_{A_i \cup B_i}, \qquad \pi_1(\widetilde{\xi}) = \sum_{i : A_i \neq \varnothing} \delta_{A_i}, \qquad \pi_2(\widetilde{\xi}) = \sum_{i : B_i \neq \varnothing} \delta_{B_i}.$$

This definition is illustrated in Figure 1.

DEFINITION 2.9. The *two-type herds process* $(\widetilde{\Xi}_t)_{t \ge 0}$ (with rates λ and v) is a continuous-time Markov chain on $\mathcal{S}_2$ with transitions described as follows. At any time $t \ge 0$, and for any (A, B) with $\widetilde{\Xi}_t(A, B) \ge 1$, both A and B are subject, independently of each other, to death and birth mechanisms, as in the original herds process (in particular, if either of them is empty, it stays empty). However, they are subject together to the same splitting events: an edge e is said to be active if $(A \cup B)_{e,1} \neq \varnothing$ and $(A \cup B)_{e,2} \neq \varnothing$. Then a split occurs at any active edge e with rate v, and when this happens, (A, B) is split into the two pairs $(A_{e,1}, B_{e,1})$ and $(A_{e,2}, B_{e,2})$. (In case $A = \varnothing$, we let $A_{e,1} = A_{e,2} = \varnothing$ and similarly for B).

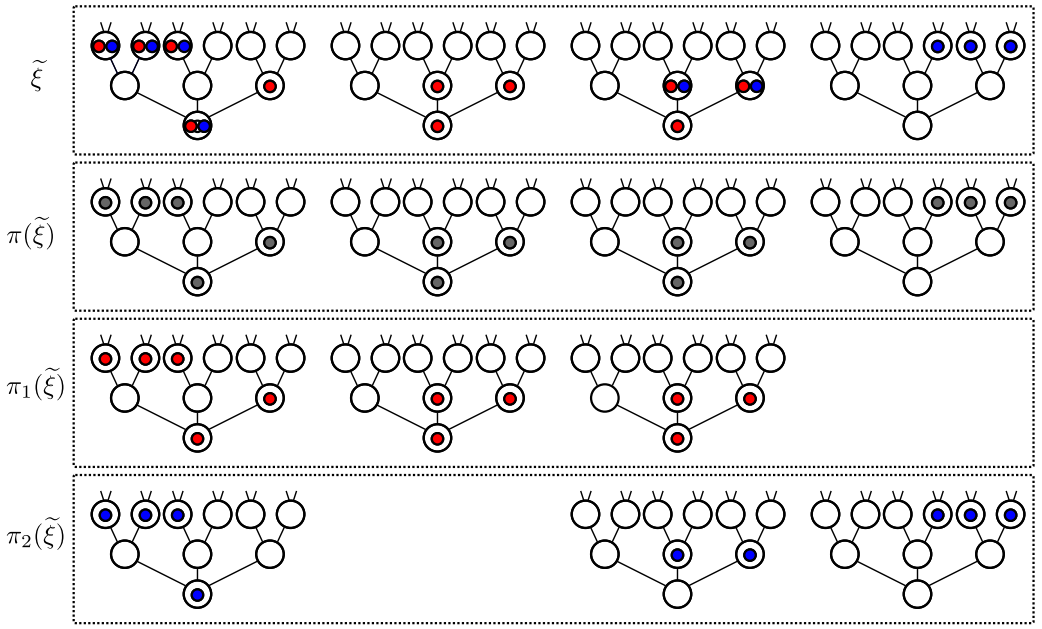

FIG. 1. *A two-type herds configuration $\widetilde{\xi}$ is depicted on top, with the two types represented in blue and red (it is assumed that in the four herds of this configuration, no particle is present other than the ones depicted). The projections $\pi(\widetilde{\xi})$, $\pi_1(\widetilde{\xi})$ and $\pi_2(\widetilde{\xi})$ are shown in the second, third, and fourth rows, respectively.*

We record two observations about the two-type herds process in the following lemma. The proof is done by comparing jump rates, and we omit it.

LEMMA 2.9. *Let $(A, B) \in P_{\mathrm{f}}(\mathbb{T}^d) \times P_{\mathrm{f}}(\mathbb{T}^d)$, and let $(\widetilde{\Xi})_{t\geq 0}$ be the two-type herds process started from $\delta_{(A,B)}$:*

(a) *The processes $(\pi_1(\widetilde{\Xi}_t))_{t\geq 0}$ and $(\pi_2(\widetilde{\Xi}_t))_{t\geq 0}$ are herds processes started from δ_A and δ_B, respectively.*

(b) *The herds process started from $\delta_{A\cup B}$ is stochastically dominated (in the sense of the partial order $\preceq$) by $(\pi(\widetilde{\Xi}_t))_{t\geq 0}$.*

PROOF OF LEMMA 2.8. Fix $p \geq 1$ and disjoint sets $A, B \in P_{\mathrm{f}}(\mathbb{T}^d)$. Let $(\widetilde{\Xi}_t)_{t\geq 0}$ denote a two-type herds process started from $\delta_{(A,B)}$, defined under a probability measure $\widetilde{\mathbb{P}}$ (with expectation operator $\widetilde{\mathbb{E}}$). By Lemma 2.9(b) and the fact that $X(\cdot)$ is monotone with respect to the partial order $\preceq$, we have

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_{A\cup B}\big] \leq \widetilde{\mathbb{E}}\big[X\big(\pi(\widetilde{\Xi}_t)\big)^p\big].$$

Next, noting that $X(\pi(\widetilde{\Xi}_t)) \leq X(\pi_1(\widetilde{\Xi}_t)) + X(\pi_2(\widetilde{\Xi}_t))$, we have

$$\widetilde{\mathbb{E}}\big[X\big(\pi(\widetilde{\Xi}_t)\big)^p\big] \leq \widetilde{\mathbb{E}}\big[\big(X\big(\pi_1(\widetilde{\Xi}_t)\big) + X\big(\pi_2(\widetilde{\Xi}_t)\big)^p\big].$$

Putting these inequalities together (raised to the power $1/p$) and using Minkowski's inequality, we obtain

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_{A\cup B}\big]^{1/p} \leq \widetilde{\mathbb{E}}\big[X\big(\pi_1(\widetilde{\Xi}_t)\big)^p\big]^{1/p} + \widetilde{\mathbb{E}}\big[X\big(\pi_2(\widetilde{\Xi}_t)\big)^p\big]^{1/p}.$$

By Lemma 2.9(a) the right-hand side equals

$$\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_A\big]^{1/p} + \mathbb{E}\big[X_t^p \,|\, \Xi_0 = \delta_B\big]^{1/p}.$$

This completes the proof. □

REMARK 2.4. It is worth mentioning a curious fact here, which is the main reason for introducing the two-type herds process. Let $A, B \in P_{\mathrm{f}}(\mathbb{T}^d)$ be disjoint. Then a natural guess would be that the number of particles in the herds process starting from $\delta_A + \delta_B$ should stochastically dominate the number of particles in the process starting from $\delta_{A\cup B}$, since in the former case there is a priori more space for the particles to spread. However, there seems to be no simple proof of this fact, and it is not even clear whether it should be true or not. In a natural choice of coupling, one would try to map particles of the process started from $\delta_{A\cup B}$ injectively into particles of the process started from $\delta_A + \delta_B$, and make it so that, when a particle of the former process dies or gives birth, its image under the injective mapping does the same. However, this does not work. For instance, when starting from $\delta_{A\cup B}$, it could be that at a later time, a single split would separate more particles than in the process starting from $\delta_A + \delta_B$, which in turn could, after another birth event, give rise to a larger number of particles in the process starting from $\delta_{A\cup B}$.

REMARK 2.5. The notion of two-type herds process can, of course, be generalized to a multitype herds process. Given any integer $k \geq 1$, the k-type herds process is the continuous-time Markov chain on

$$\mathcal{S}_k := \left\{ \xi : P_{\mathrm{f},k}(\mathbb{T}^d) \to \mathbb{N}_0, \text{ with } \sum_{(A_1,\dots,A_k)} \xi(A_1,\dots,A_k) < \infty \right\},$$

where

$$P_{\mathrm{f},k}(\mathbb{T}^d) := \left\{ (A_1,\dots,A_k) : A_1,\dots,A_k \subset \mathbb{T}^d, \bigcup_{j=1}^{k} A_j \neq \varnothing \right\}$$

and where, like in the two-type herds process, every type obeys birth and death mechanisms, as in the original herds process, ignoring the other types, but they all share the same splitting events. Naturally, an analogue of Lemma 2.9 holds as well in this general setting. This remark will be used in the proof of Lemma 3.17 below.

3. Analysis of the herds process through a growth index.

3.1. *Definition and properties of growth index.* We start by stating and proving a continuity result that will be useful throughout this section.

LEMMA 3.1. *For any $t \geq 0$, $(\lambda, \mathsf{v}) \mapsto \mathbb{E}_{\lambda,\mathsf{v}}[X_t]$ is continuous.*

PROOF. Fix $t \geq 0$. We start with a truncation. Assume that the herds process $(\Xi_s)_{s\geq 0}$ starts from $\delta_{\{o\}}$, and let K_t denote the number of jumps performed by the process until time t. Note that K_t equals the number of births (denoted N_t) plus the number of deaths plus the number of splits until time t. The number of deaths is at most $N_t + 1$, and the number of splits is at most N_t. This gives the bound $K_t \leq 3N_t + 1$. We then bound

$$\mathbb{E}_{\lambda,\mathsf{v}}\big[X_t \cdot \mathbb{1}\{K_t > n\}\big] \leq \mathbb{E}_{\lambda,\mathsf{v}}\big[(1+N_t)\cdot \mathbb{1}\{K_t > n\}\big] \leq \mathbb{E}_{\lambda,\mathsf{v}}\big[(1+N_t)\cdot \mathbb{1}\{N_t > (n-1)/3\}\big].$$

Using the stochastic domination by a pure-birth process given by Lemma 2.1, the right-hand side tends to zero as $n \to \infty$, and this convergence is uniform over (v, λ) in sets of the form $[0,\infty) \times [0, \lambda_0]$, for any $\lambda_0 > 0$. Hence, for any $\lambda_0 > 0$,

$$\mathbb{E}_{\lambda,\mathsf{v}}\big[X_t \cdot \mathbb{1}\{K_t \leq n\}\big] \xrightarrow{n\to\infty} \mathbb{E}_{\lambda,\mathsf{v}}[X_t] \quad \text{uniformly over } (\mathsf{v},\lambda) \in [0,\infty)\times[0,\lambda_0]. \tag{10}$$

Next, write

$$\mathbb{E}_{\lambda,\mathsf{v}}\big[X_t \cdot \mathbb{1}\{K_t = n\}\big] = \sum_{m \le n+1} m \cdot \mathbb{P}_{\lambda,\mathsf{v}}(X_t = m, K_t = n).$$

Note that $(\lambda, \mathsf{v}) \mapsto \mathbb{P}_{\lambda,\mathsf{v}}(X_t = m, K_t = n)$ is continuous for any m, n, due to the fact that the jump rates of the herds process depend continuously on λ and v, and there are only finitely many possible trajectories involving n jumps. This implies that

$$(\lambda, \mathsf{v}) \mapsto \mathbb{E}_{\lambda,\mathsf{v}}\big[X_t \cdot \mathbb{1}\{K_t \le n\}\big] \quad \text{is continuous.} \tag{11}$$

The desired continuity statement now follows from (10) and (11). □

For any $p \ge 1$, define the *growth index of order* p as

$$\varphi_p = \varphi_p(\lambda, \mathsf{v}) := \inf_{t \ge 0} \mathbb{E}_{\lambda,\mathsf{v}}\big[X_t^p\big]^{1/t}. \tag{12}$$

In case $p = 1$, we omit the subscript so that

$$\varphi := \varphi_1.$$

Note that we have

$$\varphi_p^t \le \mathbb{E}\big[X_t^p\big], \quad t \ge 0, p \ge 1. \tag{13}$$

Moreover, (7) implies that

$$\mathbb{E}\big[X_{t+s}^p\big] \le \mathbb{E}\big[X_t^p\big] \cdot \mathbb{E}\big[X_s^p\big], \quad t, s \ge 0, p \ge 1, \tag{14}$$

and Fekete's lemma then ensures that

$$\lim_{t \to \infty} \mathbb{E}\big[X_t^p\big]^{1/t} = \varphi_p, \quad p \ge 1. \tag{15}$$

It will also be useful to observe that

$$[1, \infty) \ni p \mapsto \varphi_p^{1/p} \quad \text{is nondecreasing.} \tag{16}$$

To see this, let $p \ge q \ge 1$. We bound, for any $t \ge 0$,

$$\mathbb{E}\big[X_t^p\big] = \mathbb{E}\big[(X_t^q)^{p/q}\big] \ge \mathbb{E}\big[X_t^q\big]^{p/q}$$

by Jensen's inequality. Hence,

$$\big(\mathbb{E}\big[X_t^p\big]^{1/t}\big)^{1/p} \ge \big(\mathbb{E}\big[X_t^q\big]^{1/t}\big)^{1/q}.$$

By taking $t \to \infty$ and using (15), we obtain $\varphi_p^{1/p} \ge \varphi_q^{1/q}$, as desired.

For the rest of this section, we focus on the growth index with $p = 1$. We will analyse higher values of p in Section 3.4.

We state a lemma with an upper bound that, apart from a constant prefactor, matches (13) in the case $p = 1$.

LEMMA 3.2. *There is a constant $C = C(\lambda, \mathsf{v}) > 0$, depending continuously on λ and v, such that, for any $t \ge 0$,*

$$\mathbb{E}[X_t] \le C \cdot \varphi^t. \tag{17}$$

Before proving this, we state and prove an auxiliary result, concerning the expected number of herds containing a single particle, after one time unit of the dynamics has elapsed.

LEMMA 3.3. *There is a constant $\rho = \rho(\lambda, \mathsf{v}) > 0$, depending continuously on λ and v, such that, for any $\xi \in \mathcal{S}$, we have*

$$\mathbb{E}\Big[\sum_{A:|A|=1} \Xi_1(A) | \Xi_0 = \xi\Big] \geq \rho \cdot X(\xi). \tag{18}$$

PROOF. By Lemma 2.5 and the linearity of expectation, it suffices to show that there exists $\rho > 0$ (satisfying the stated continuity property) such that, for any $A \in P_{\mathrm{f}}(\mathbb{T}^d)$,

$$\mathbb{E}\Big[\sum_{A':|A'|=1} \Xi_1(A') | \Xi_0 = \delta_A\Big] \geq \rho \cdot |A|.$$

To prove this, fix $A \in P_{\mathrm{f}}(\mathbb{T}^d)$, and note that the left-hand side above is larger than

$$\sum_{u \in A} \mathbb{E}\big[\Xi_1(\{u\}) | \Xi_0 = \delta_A\big] \geq \sum_{u \in A} \mathbb{P}\big(\Xi_1(\{u\}) > 0 | \Xi_0 = \delta_A\big).$$

It is easy to see that there is a constant $\rho > 0$, depending continuously only on λ and v, such that the probability in the sum in the right-hand side is larger than ρ (for any A and u). This is achieved by prescribing that the particle present at u at time 0 does not die or give birth until time 1, and moreover, this particle becomes separated through successive splits in the edges that are incident to u, of any other particle in its herd. This shows that the right-hand side above is larger than $\rho|A|$, completing the proof. □

PROOF OF LEMMA 3.2. For any $s, t \geq 0$, we have

$$\mathbb{E}[X_{t+s} | \mathcal{F}_s] = \sum_A \Xi_s(A) \cdot \mathbb{E}[X_t | \Xi_0 = \delta_A] \geq \sum_{A:|A|=1} \Xi_s(A) \cdot \mathbb{E}[X_t],$$

where the equality follows from Lemma 2.5 and the Markov property and the inequality from Lemma 2.4. Then by taking expectations, we obtain that

$$\mathbb{E}[X_{t+s}] \geq \mathbb{E}\Big[\sum_{A:|A|=1} \Xi_s(A)\Big] \cdot \mathbb{E}[X_t].$$

In case $s \geq 1$, by Lemma 3.3 and the Markov property, the right-hand side is larger than

$$\rho \cdot \mathbb{E}[X_{s-1}] \cdot \mathbb{E}[X_t].$$

Further, by Proposition 2.7, letting $\kappa := 1/\mathbb{E}[X_1]$, the above is larger than

$$\kappa\rho \cdot \mathbb{E}[X_s] \cdot \mathbb{E}[X_t]$$

(note that κ is strictly positive and depends continuously on λ and v by Lemma 3.1).

Using this recursively, we obtain, for any $t > 0$ and $n \in \mathbb{N}$,

$$\mathbb{E}[X_{nt}] \geq (\kappa\rho)^{n-1} \cdot \mathbb{E}[X_t]^n$$

so that

$$\mathbb{E}[X_t] \leq (\kappa\rho)^{-\frac{n-1}{n}} \cdot \mathbb{E}[X_{nt}]^{\frac{1}{n}} = (\kappa\rho)^{-\frac{n-1}{n}} \cdot \big(\mathbb{E}[X_{nt}]^{\frac{1}{nt}}\big)^t.$$

Using (15), we obtain

$$\mathbb{E}[X_t] \leq \liminf_{n\to\infty} \big((\kappa\rho)^{-\frac{n-1}{n}} \cdot \big(\mathbb{E}[X_{nt}]^{\frac{1}{nt}}\big)^t\big) = C \cdot \varphi^t,$$

with $C = \frac{1}{\kappa\rho}$, proving the lemma. □

COROLLARY 3.4. *The function* $(\lambda, \mathsf{v}) \mapsto \varphi(\lambda, \mathsf{v})$ *is continuous.*

PROOF. Putting together (13) and (17), we have, for any $t > 0$,

$$C(\lambda, \mathsf{v})^{-1/t} \cdot \mathbb{E}_{\lambda,\mathsf{v}}[X_t]^{1/t} \le \varphi(\lambda, \mathsf{v}) \le \mathbb{E}_{\lambda,\mathsf{v}}[X_t]^{1/t}.$$

Then for any $t > 0$, using the continuity of $(\lambda, \mathsf{v}) \mapsto C(\lambda, \mathsf{v})$ and of $(\lambda, \mathsf{v}) \mapsto \mathbb{E}_{\lambda,\mathsf{v}}[X_t]$,

$$\begin{aligned} C(\lambda, \mathsf{v})^{-1/t} \cdot \mathbb{E}_{\lambda,\mathsf{v}}[X_t]^{1/t} &= \lim_{(\lambda',\mathsf{v}')\to(\lambda,\mathsf{v})} \left[C(\lambda', \mathsf{v}')^{-1/t} \cdot \mathbb{E}_{\lambda',\mathsf{v}'}[X_t]^{1/t}\right] \\ &\le \liminf_{(\lambda',\mathsf{v}')\to(\lambda,\mathsf{v})} \varphi(\lambda', \mathsf{v}') \le \limsup_{(\lambda',\mathsf{v}')\to(\lambda,\mathsf{v})} \varphi(\lambda', \mathsf{v}') \\ &\le \lim_{(\lambda',\mathsf{v}')\to(\lambda,\mathsf{v})} \mathbb{E}_{\lambda',\mathsf{v}'}[X_t]^{1/t} = \mathbb{E}_{\lambda,\mathsf{v}}[X_t]^{1/t}. \end{aligned}$$

Noting now that

$$\lim_{t\to\infty} \left(C(\lambda, \mathsf{v})^{-1/t} \cdot \mathbb{E}_{\lambda,\mathsf{v}}[X_t]^{1/t}\right) = \lim_{t\to\infty} \mathbb{E}_{\lambda,\mathsf{v}}[X_t]^{1/t} = \varphi(\lambda, \mathsf{v}),$$

the proof of continuity is complete. □

We now obtain the following simple characterization of the supercritical regime. Recall the definition of $\bar{\lambda}(\mathsf{v})$ from Definition 2.6.

LEMMA 3.5. *For any* $\lambda > 0$ *and* $\mathsf{v} > 0$, *the following are equivalent*:

(a) *The herds process survives with positive probability*;
(b) $\varphi > 1$;
(c) $\mathbb{E}[X_t] \xrightarrow{t\to\infty} \infty$.

Consequently, for any $\mathsf{v} > 0$,

$$\varphi(\bar{\lambda}(\mathsf{v}), \mathsf{v}) = 1. \tag{19}$$

PROOF. The fact that (b) and (c) are equivalent follows from the fact that, for any $t \ge 0$, $\varphi^t \le \mathbb{E}[X_t] \le C\varphi^t$.

Now, if (a) holds, then we must also have $X_t \xrightarrow{t\to\infty} \infty$ with positive probability, as otherwise, using the conditional Borel–Cantelli Lemma and a standard argument, we would get a contradiction. Then (c) follows from Fatou's Lemma.

Conversely, assume that (c) holds. Let $Z_t := \sum_{A:|A|=1} \Xi_t(A)$ denote the number of herds in Ξ_t with a single particle in them. By Lemma 3.3 we have $\mathbb{E}[Z_{t+1}] \ge \rho \cdot \mathbb{E}[X_t]$. It follows that there exists some $T > 0$ such that $\mathbb{E}[Z_T] \ge 2$. Since different herds evolve independently of each other, we deduce that $(Z_{nT})_{n\ge0}$ dominates a supercritical branching process and, therefore, survives forever with positive probability. Since it is dominated by $(X_{nT})_{n\ge0}$, we get that (a) is satisfied.

Having established the equivalence between (a) and (b), the equality (19) follows from the continuity of φ. □

3.2. *Strict monotonicity of growth index.* Our goal in this section is to prove the following result.

PROPOSITION 3.6. *The map* $(\lambda, \mathsf{v}) \mapsto \varphi(\lambda, \mathsf{v})$ *is strictly increasing in both arguments.*

Before discussing the proof of this, let us see how it allows us to prove Theorem 1.3.

PROOF OF THEOREM 1.3. Let $\mathsf{v}' > \mathsf{v} > 0$. We have

$$\varphi\big(\bar{\lambda}(\mathsf{v}'), \mathsf{v}'\big) = 1 = \varphi\big(\bar{\lambda}(\mathsf{v}), \mathsf{v}\big) < \varphi\big(\bar{\lambda}(\mathsf{v}), \mathsf{v}'\big),$$

where the two equalities are given by (19) and the inequality by the strict monotonicity of φ. Again, using the strict monotonicity of φ, we conclude from $\varphi(\bar{\lambda}(\mathsf{v}'), \mathsf{v}') < \varphi(\bar{\lambda}(\mathsf{v}), \mathsf{v}')$ that $\bar{\lambda}(\mathsf{v}') < \bar{\lambda}(\mathsf{v})$. □

The proof of Proposition 3.6 will require several steps. We start with a definition.

DEFINITION 3.1. Fix the parameters v and λ of the herds process. Given $\xi \in \mathcal{S}$ and $t > 0$, we let

$$g_{\mathsf{v}}(\xi, t) := \sum_{A \in P_{\mathrm{f}}(\mathbb{T}^d)} \xi(A) \sum_{\substack{e \text{ active} \\ \text{edge of } A}} \big(\mathbb{E}_{\lambda,\mathsf{v}}[X_t \,|\, \Xi_0 = \delta_{A_{e,1}} + \delta_{A_{e,2}}] - \mathbb{E}_{\lambda,\mathsf{v}}[X_t \,|\, \Xi_0 = \delta_A]\big), \tag{20}$$

and

$$h_{\lambda}(\xi, t) := \sum_{A \in P_{\mathrm{f}}(\mathbb{T}^d)} \xi(A) \sum_{\substack{x \in A, y \notin A, \\ x \sim y}} \big(\mathbb{E}_{\lambda,\mathsf{v}}[X_t \,|\, \Xi_0 = \delta_{A \cup \{y\}}] - \mathbb{E}_{\lambda,\mathsf{v}}[X_t \,|\, \Xi_0 = \delta_A]\big) \tag{21}$$

(we omit λ from the notation for g_{v}, and we omit v from the notation for h_{λ}).

The functions g_{v} and h_{λ} give a measure of the total impact on the number of particles at time t of splitting the herds of ξ at time zero and having a birth at time zero, respectively.

PROPOSITION 3.7. *For any $T > 0$, we have*

$$\frac{\partial}{\partial \mathsf{v}} \mathbb{E}_{\lambda,\mathsf{v}}[X_T] = \int_0^T \mathbb{E}_{\lambda,\mathsf{v}}\big[g_{\mathsf{v}}(\Xi_t, T - t)\big]\,\mathrm{d}t, \tag{22}$$

and

$$\frac{\partial}{\partial \lambda} \mathbb{E}_{\lambda,\mathsf{v}}[X_T] = \int_0^T \mathbb{E}_{\lambda,\mathsf{v}}\big[h_{\lambda}(\Xi_t, T - t)\big]\,\mathrm{d}t. \tag{23}$$

The proof is postponed to Section 3.3, and we now prove another intermediate result.

LEMMA 3.8. *Let ξ' consist of a single herd with exactly two particles, which are neighbors, and ξ'' consist of two herds, each containing a single particle. Then there exists $\gamma > 0$ (depending continuously on λ and v) such that, for any $t \geq 1$,*

$$\mathbb{E}\big[X_t \,|\, \Xi_0 = \xi''\big] \geq \mathbb{E}\big[X_t \,|\, \Xi_0 = \xi'\big] + \gamma \cdot \mathbb{E}[X_t \,|\, X_0 = \delta_{\{o\}}].$$

PROOF. Fix $v \sim o$. Without loss of generality, we assume that ξ' consists of the herd $\{o, v\}$ and ξ'' consists of the herds $\{o\}$ and $\{v\}$. We define a coupling of the two herds processes (Ξ'_s) and (Ξ''_s) starting, respectively, from ξ' and ξ'': first, we take independent random variables, as follows:

- τ_o and τ_v, both $\sim \mathrm{Exp}(1)$;
- for each $u \sim o$, $\tau_{o,u} \sim \mathrm{Exp}(\lambda)$;
- for each $u \sim v$, $\tau_{v,u} \sim \mathrm{Exp}(\lambda)$;
- $\tau_{\mathrm{split}} \sim \mathrm{Exp}(\mathsf{v})$.

Additionally, let τ' denote the minimum of all these random variables, and $\tau := \min(1, \tau')$. Now, the coupling is defined as follows. In all cases we let $(\Xi'_s, \Xi''_s) = (\xi', \xi'')$ for $s \in [0, \tau)$; the definition of (Ξ'_τ, Ξ''_τ) will be split into several cases, but after time τ, we let Ξ' and Ξ'' continue evolving independently as two herds processes with split rate v. The definition of (Ξ'_τ, Ξ''_τ) is as follows: if $\tau = 1 < \tau'$, then $(\Xi'_\tau, \Xi''_\tau) = (\xi', \xi'')$. Otherwise:

- If $\tau = \tau_o$, then $\Xi'_\tau = \Xi''_\tau = \delta_{\{v\}}$; similarly if $\tau = \tau_v$, then $\Xi'_\tau = \Xi''_\tau = \delta_{\{o\}}$.
- If for some $u \sim o$, $\tau = \tau_{o,u}$, then $\Xi'_\tau = \delta_{\{o,u,v\}}$ and $\Xi''_\tau = \delta_{\{o,u\}} + \delta_{\{v\}}$.
- If for some $u \sim v$, $\tau = \tau_{v,u}$, then $\Xi'_\tau = \delta_{\{o,u,v\}}$ and $\Xi''_\tau = \delta_{\{o\}} + \delta_{\{u,v\}}$.
- If $\tau = \tau_{\text{split}}$, then $\Xi'_\tau = \Xi''_\tau = \delta_{\{o\}} + \delta_{\{v\}}$.

We let $\widehat{\mathbb{P}}$ denote a probability measure under which this coupling is defined and $\widehat{\mathbb{E}}$ be the associated expectation operator. Also, let $(\mathcal{F}_t)_{t\geq 0}$ denote the natural filtration of $(\Xi'_t, \Xi''_t)_{t\geq 0}$.

Fix $t \geq 1$. Using Lemma 2.5 and 2.8 and inspecting all cases concerning τ, it is easy to check that

$$\widehat{\mathbb{E}}\big[X(\Xi''_t)|\mathcal{F}_\tau\big] \geq \widehat{\mathbb{E}}\big[X(\Xi'_t)|\mathcal{F}_\tau\big]. \tag{24}$$

Define the good event $E := \{\tau = \tau_{o,v}\} \cup \{\tau = \tau_{v,o}\}$, and note that, on this event, Ξ''_τ contains Ξ'_τ plus an additional herd with a single particle in it. Hence,

$$\widehat{\mathbb{E}}\big[X(\Xi''_t)|\mathcal{F}_\tau\big] \cdot \mathbb{1}_E = \big(\widehat{\mathbb{E}}\big[X(\Xi'_t)|\mathcal{F}_\tau\big] + F(t-\tau)\big) \cdot \mathbb{1}_E, \tag{25}$$

where $F(s) := \mathbb{E}[X_s | \Xi_0 = \delta_{\{o\}}]$.

We now write

$$\mathbb{E}\big[X_t | \Xi_0 = \xi''\big] = \widehat{\mathbb{E}}\big[X(\Xi''_t)\big] = \widehat{\mathbb{E}}\big[\widehat{\mathbb{E}}\big[X(\Xi''_t)|\mathcal{F}_\tau\big] \cdot \mathbb{1}_E + \widehat{\mathbb{E}}\big[X(\Xi''_t)|\mathcal{F}_\tau\big] \cdot \mathbb{1}_{E^c}\big],$$

and using (24) and (25), we bound the right-hand side from below by

$$\begin{aligned}
&\widehat{\mathbb{E}}\big[\big(\widehat{\mathbb{E}}\big[X(\Xi'_t)|\mathcal{F}_\tau\big] + F(t-\tau)\big) \cdot \mathbb{1}_E + \widehat{\mathbb{E}}\big[X(\Xi'_t)|\mathcal{F}_\tau\big] \cdot \mathbb{1}_{E^c}\big] \\
&\quad \geq \widehat{\mathbb{E}}\big[X(\Xi'_t)\big] + \widehat{\mathbb{E}}\big[F(t-\tau) \cdot \mathbb{1}_E\big] \\
&\quad \geq \mathbb{E}\big[X_t | \Xi_0 = \xi'\big] + \widehat{\mathbb{P}}(E) \cdot \min_{t-1\leq s\leq t} F(s).
\end{aligned}$$

Using (7), we have that $F(s) \cdot F(t-s) \geq F(t)$ for any $s \in [t-1, t]$, which gives

$$\min_{t-1\leq s\leq t} F(s) \geq F(t) \cdot \Big(\min_{0\leq r\leq 1} F(r)\Big)^{-1}.$$

The lemma is thus proved with $\gamma := \widehat{\mathbb{P}}(E) \cdot (\min_{0\leq t\leq 1} F(r))^{-1}$. □

PROOF OF PROPOSITION 3.6. We only prove the strict monotonicity in v; the argument for the strict monotonicity in λ is entirely similar. We start with some basic observations. Let X'_t denote the number of herds in the herds process at time t, which contain exactly two particles, these particles being neighbors. We claim that

$$\mathbb{E}\big[X'_t\big] \geq c \cdot \mathbb{E}[X_{t-1}] \quad \text{for any } t \geq 1, \tag{26}$$

with c some positive constant depending continuously on v. To see this, recall that if we denote by Z_t the number of herds of Ξ_t that contain a single particle, then (as in the proof of Lemma 3.3) one has $\mathbb{E}[Z_{t-1/2}] \geq c_1 \cdot \mathbb{E}[X_{t-1}]$, for some constant $c_1 > 0$, depending continuously on v. Since in any half unit of time a herd with a single particle can be transformed in a herd with two neighboring particles, at a constant price (depending continuously on v), this gives the claim (26).

Recalling the definition of g_{v} in Definition 3.1, by Lemma 2.8 applied with $p = 1$ and Lemma 3.8, we have

$$g_{\mathsf{v}}(\Xi_s, t-s) \geq \gamma \cdot X'_s \cdot \mathbb{E}[X_{t-s}].$$

Together with Proposition 3.7 and (26), this gives, for any $\mathsf{v} > 0$ and any $t \geq 1$, with $F(\mathsf{v}, t) = \mathbb{E}_{\mathsf{v}}[X_t]$,

$$\frac{\partial F}{\partial \mathsf{v}}(\mathsf{v}, t) \geq c\gamma \cdot \int_1^t \mathbb{E}_{\mathsf{v}}[X_{s-1}] \cdot \mathbb{E}_{\mathsf{v}}[X_{t-s}]\, \mathrm{d}s.$$

Applying then Proposition 2.7 gives, for $t \geq 2$,

$$\frac{\partial F}{\partial \mathsf{v}}(\mathsf{v}, t) \geq c' \cdot t \cdot F(\mathsf{v}, t),$$

where $c' > 0$ depends continuously on v (noting that $F(\mathsf{v}, t-1) \geq F(\mathsf{v}, t)/F(\mathsf{v}, 1)$ by Proposition 2.7 again). This implies that, for any fixed $0 < \mathsf{v}_1 < \mathsf{v}_2$,

$$\log \frac{F(\mathsf{v}_2, t)}{F(\mathsf{v}_1, t)} \geq \rho \cdot t \cdot (\mathsf{v}_2 - \mathsf{v}_1),$$

for some $\rho > 0$. Thus, by taking the limit as $t \to \infty$ and using (15), we get

$$\varphi(\mathsf{v}_1, \lambda) \leq e^{-\rho(\mathsf{v}_2 - \mathsf{v}_1)} \cdot \varphi(\mathsf{v}_2, \lambda). \qquad \square$$

3.3. *Proof of derivative formulas.* We now turn to establishing (22) and (23). The proofs are entirely analogous, so we only do the former. A few of the more technical points of the proof are done in the Appendix.

Recall the definition of the function $g_{\mathsf{v}}(\xi, t)$ in (20). We now define the closely related function,

$$\begin{aligned} g_{\mathsf{v},\varepsilon}(\xi, t) := \sum_{A \in P_{\mathrm{f}}(\mathbb{T}^d)} \xi(A) \sum_{\substack{e \text{ active} \\ \text{edge of } A}} \big(&\mathbb{E}_{\lambda, \mathsf{v}+\varepsilon}[X_t \,|\, \Xi_0 = \delta_{A_{e,1}} + \delta_{A_{e,2}}] \\ &- \mathbb{E}_{\lambda,\mathsf{v}}[X_t \,|\, \Xi_0 = \delta_A]\big), \end{aligned} \tag{27}$$

where $\mathsf{v} > 0$ and $\varepsilon > 0$. Note that the expression for $g_{\mathsf{v},\varepsilon}$ only differs from that for g_{v} in that, in the former, the first expectation that appears in the right-hand side is under parameters $\lambda, \mathsf{v}+\varepsilon$ rather than λ, v.

The following integrability result will be useful. The proof is done in the Appendix.

LEMMA 3.9. *For any $T > 0$ and $k \geq 1$, we have*

$$\mathbb{E}_{\lambda,\mathsf{v}}\Big[\max_{0 \leq t \leq T} \big|g_{\mathsf{v}}(\Xi_t, T-t)\big|^k\Big] < \infty \quad \textit{and} \quad \mathbb{E}_{\lambda,\mathsf{v}}\Big[\max_{0 \leq t \leq T} \big|g_{\mathsf{v},\varepsilon}(\Xi_t, T-t)\big|^k\Big] < \infty.$$

The following definition will give an alternative expression for g_{v} and $g_{\mathsf{v},\varepsilon}$ in (28) and (29) below.

DEFINITION 3.2. Given $\xi \in \mathcal{S}$, let $\mathcal{G}(\xi)$ denote the set of all herd configurations ξ' that can be obtained by performing a single split on ξ. Given $\xi' \in \mathcal{G}(\xi)$, let A be the (unique) herd shape that is split into two to obtain ξ' from ξ; let $\mathsf{m}(\xi, \xi') := \xi(A)$.

To further clarify this definition, fix an enumeration $\xi = \sum_{i=1}^m \delta_{A_i}$ of ξ. Then $\mathcal{G}(\xi)$ is the set of $\xi' \in \mathcal{S}$ with

$$\xi' = \sum_{i \neq j} \delta_{A_i} + \delta_{(A_j)_{e,1}} + \delta_{(A_j)_{e,2}},$$

where $j \in \{1, \dots, m\}$ and e is an active edge of A_j. For ξ', as in the above display, we have $\mathsf{m}(\xi, \xi') = \xi(A_j)$.

We now observe that, using Lemma 2.5, we can rewrite

$$g_{\mathsf{v}}(\xi, t) = \sum_{\xi' \in \mathcal{G}(\xi)} \mathsf{m}(\xi, \xi') \cdot \big(\mathbb{E}_{\lambda,\mathsf{v}}[X_t | \Xi_0 = \xi'] - \mathbb{E}_{\lambda,\mathsf{v}}[X_t | \Xi_0 = \xi]\big) \tag{28}$$

and

$$g_{\mathsf{v},\varepsilon}(\xi, t) = \sum_{\xi' \in \mathcal{G}(\xi)} \mathsf{m}(\xi, \xi') \cdot \big(\mathbb{E}_{\lambda,\mathsf{v}+\varepsilon}[X_t | \Xi_0 = \xi'] - \mathbb{E}_{\lambda,\mathsf{v}}[X_t | \Xi_0 = \xi]\big). \tag{29}$$

Fix $\mathsf{v} > 0$, $\lambda > 0$ and $\varepsilon > 0$. We will now construct a coupling $(\mathcal{V}_t, \mathcal{W}_t)_{t \geq 0}$ under a probability measure $\widehat{\mathbb{P}}$ (with dependence on v, λ, ε omitted) so that $(\mathcal{V}_t)$ is a herds process with parameters λ, v, and $(\mathcal{W}_t)$ is a herds process with parameters λ, $\mathsf{v} + \varepsilon$; both of these processes are started from a single herd with a single particle (at the root of $\mathbb{T}^d$).

DEFINITION 3.3 (Coupling $(\mathcal{V}_t, \mathcal{W}_t)$). Take a probability space with probability measure $\widehat{\mathbb{P}}$ under which a herds process $(\mathcal{V}_t)_{t \geq 0}$ with parameters λ, v is defined, started from $\delta_{\{o\}}$. Assume that the split jumps of $(\mathcal{V}_t)$ are given as follows: splitting instructions arise with rate $\mathsf{v} + \varepsilon$ (rather than v), but they are rejected with probability $\frac{\varepsilon}{\mathsf{v}+\varepsilon}$. Let

$$\tau_{\text{sep}} := \inf\{t \geq 0 : \text{a splitting instruction is rejected at time } t\}.$$

We define $(\mathcal{W}_t)_{t \geq 0}$ as follows. For $0 \leq t < \tau_{\text{sep}}$, we set $\mathcal{W}_t = \mathcal{V}_t$. At time τ_{sep}, this process obeys the splitting instruction that was rejected by $(\mathcal{V}_t)$. We then let $(\mathcal{W}_t)_{t \geq \tau_{\text{sep}}}$ continue evolving from $\mathcal{W}_{\tau_{\text{sep}}}$ as a herds process with parameters λ, $\mathsf{v} + \varepsilon$, independent of $(\mathcal{V}_t)_{t \geq \tau_{\text{sep}}}$. Finally, we let

$$\mathcal{X}_t = X(\mathcal{V}_t), \qquad \mathcal{Y}_t = X(\mathcal{W}_t), \quad t \geq 0.$$

For the rest of this section, we fix $T > 0$.

DEFINITION 3.4. We define the process $(\mathcal{A}_t)_{0 \leq t \leq T}$ as

$$\mathcal{A}_t := \mathbb{1}\{\tau_{\text{sep}} \leq t\} \cdot \widehat{\mathbb{E}}[\mathcal{Y}_T - \mathcal{X}_T | \mathcal{F}_{\tau_{\text{sep}}}], \quad 0 \leq t \leq T.$$

That is, in the event $\{\tau_{\text{sep}} > T\}$, we have $\mathcal{A}_t \equiv 0$, whereas in $\{\tau_{\text{sep}} \leq T\}$, this process takes just two values: $\mathcal{A}_t = 0$ for $t \in [0, \tau_{\text{sep}})$ and $\mathcal{A}_t = \widehat{\mathbb{E}}[\mathcal{Y}_T - \mathcal{X}_T | \mathcal{F}_{\tau_{\text{sep}}}]$ for $t \in [\tau_{\text{sep}}, T]$. Our interest in this process stems from the fact that

$$\begin{aligned} \mathbb{E}_{\lambda,\mathsf{v}+\varepsilon}[X_T] - \mathbb{E}_{\lambda,\mathsf{v}}[X_T] &= \widehat{\mathbb{E}}[\mathcal{Y}_T - \mathcal{X}_T] \\ &= \widehat{\mathbb{E}}\big[(\mathcal{Y}_T - \mathcal{X}_T) \cdot \mathbb{1}\{\tau_{\text{sep}} \leq T\}\big] \\ &= \widehat{\mathbb{E}}\big[\widehat{\mathbb{E}}\big[(\mathcal{Y}_T - \mathcal{X}_T) | \mathcal{F}_{\tau_{\text{sep}}}\big] \cdot \mathbb{1}\{\tau_{\text{sep}} \leq T\}\big] = \widehat{\mathbb{E}}[\mathcal{A}_T]. \end{aligned} \tag{30}$$

We now compute the right derivative with respect to time of the conditional expectation of $\mathcal{A}_t$. This lemma is where the function $g_{\mathsf{v},\varepsilon}$ enters the picture.

LEMMA 3.10. *For any $t \in [0, T)$, on the event $\{\tau_{\text{sep}} > t\}$, we have*

$$\frac{\mathrm{d}}{\mathrm{d}s} \widehat{\mathbb{E}}[\mathcal{A}_{t+s} | \mathcal{F}_t]\Big|_{s=0+} = \varepsilon \cdot g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T - t). \tag{31}$$

PROOF. We abbreviate (29) by writing

$$g_{\mathsf{v},\varepsilon}(\xi,s)=\sum_{\xi'\in\mathcal{G}(\xi)}\mathsf{m}(\xi,\xi')\cdot\beta(\xi,\xi',s),$$

where

$$\beta(\xi,\xi',s):=\mathbb{E}_{\lambda,\mathsf{v}+\varepsilon}\big[X_s|\Xi_0=\xi'\big]-\mathbb{E}_{\lambda,\mathsf{v}}[X_s|\Xi_0=\xi],\quad \xi\in\mathcal{S},\xi'\in\mathcal{G}(\xi),s>0.$$

Fix $s>0$ so that $t+s\le T$, and fix $\xi\in\mathcal{S}$. Let $\mathsf{N}_{t,t+s}$ denote the number of jumps of $(\mathcal{V}_r,\mathcal{W}_r)_{t\le r\le t+s}$. On the event $\{\tau_{\mathrm{sep}}>t,\ \Xi_t=\xi\}$, we have

$$\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]-\widehat{\mathbb{E}}[\mathcal{A}_t|\mathcal{F}_t]=\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]=\widehat{\mathbb{E}}\big[\mathcal{A}_{t+s}\cdot\mathbb{1}\{\mathsf{N}_{t,t+s}=1\}|\mathcal{F}_t\big]+o(s),$$

where the $o(s)$ term (which, of course, refers to when $s\to0$) comes from events where there are multiple jumps in $(t,t+s]$. The conditional expectation on the right-hand side equals

$$\begin{aligned}&\sum_{\xi'\in\mathcal{G}(\xi)}\widehat{\mathbb{E}}\big[\mathcal{A}_{t+s}\cdot\mathbb{1}\{\mathsf{N}_{t,t+s}=1,\tau_{\mathrm{sep}}\in(t,t+s],\mathcal{W}_{\tau_{\mathrm{sep}}}=\xi'\}|\mathcal{F}_t\big]\\&\quad=\sum_{\xi'\in\mathcal{G}(\xi)}\widehat{\mathbb{E}}\big[\beta\big(\xi,\xi',T-\tau_{\mathrm{sep}}\big)\cdot\mathbb{1}\{\mathsf{N}_{t,t+s}=1,\tau_{\mathrm{sep}}\in(t,t+s],\mathcal{W}_{\tau_{\mathrm{sep}}}=\xi'\}|\mathcal{F}_t\big].\end{aligned}$$

For each $\xi'\in\mathcal{G}(\xi)$, let $E(\xi')$ be the event that:

- $\tau_{\mathrm{sep}}\in(t,t+s]$,
- the first jump of $(\mathcal{V}_r,\mathcal{W}_r)_{t\le r\le t+s}$ is the one that occurs at time τ_{sep}, and
- $\mathcal{W}_{\tau_{\mathrm{sep}}}=\xi'$.

Note that $\{\mathsf{N}_{t,t+s}=1,\tau_{\mathrm{sep}}\in(t,t+s],\mathcal{W}_{\tau_{\mathrm{sep}}}=\xi'\}\subset E(\xi')$, and in fact, the two events only differ because $E(\xi')$ allows for more jumps between τ_{sep} and $t+s$. We have thus proved that, on the event $\{\tau_{\mathrm{sep}}>t,\ \Xi_t=\xi\}$,

$$\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]-\widehat{\mathbb{E}}[\mathcal{A}_t|\mathcal{F}_t]=\sum_{\xi'\in\mathcal{G}(\xi)}\widehat{\mathbb{E}}[\mathbb{1}_{E(\xi')}\cdot\beta\big(\xi,\xi',T-\tau_{\mathrm{sep}}\big)|\mathcal{F}_t]+o(s).$$

The above sum equals

$$\begin{aligned}&\sum_{\xi'\in\mathcal{G}(\xi)}\beta\big(\xi,\xi',T-t\big)\cdot\widehat{\mathbb{P}}\big(E\big(\xi'\big)|\mathcal{F}_t\big)\\&\quad+\sum_{\xi'\in\mathcal{G}(\xi)}\widehat{\mathbb{E}}[\mathbb{1}_{E(\xi')}\cdot\big(\beta\big(\xi,\xi',T-\tau_{\mathrm{sep}}\big)-\beta\big(\xi,\xi',T-t\big)\big)|\mathcal{F}_t].\end{aligned}\tag{32}$$

We will treat the two sums separately. The absolute value of the second sum is bounded by

$$\Big(\sup_{\xi'\in\mathcal{G}(\xi),u\in(t,t+s]}\big|\beta(\xi,\xi',T-u)-\beta\big(\xi,\xi',T-t\big)\big|\Big)\cdot\sum_{\xi'\in\mathcal{G}(\xi)}\widehat{\mathbb{P}}\big(E\big(\xi'\big)|\mathcal{F}_t\big).$$

As $s\to0$, the supremum tends to zero, and the sum is $\mathcal{O}(s)$, since $\mathbb{P}(\tau_{\mathrm{sep}}\in(t,t+s]|\mathcal{F}_t)=\mathcal{O}(s)$, so the whole expression is $o(s)$.

We now turn to the first sum in (32). As $s\to0$, we have

$$\widehat{\mathbb{P}}\big(E\big(\xi'\big)|\mathcal{F}_t\big)=(\mathsf{v}+\varepsilon)\cdot\frac{\varepsilon}{\mathsf{v}+\varepsilon}\cdot\mathsf{m}\big(\xi,\xi'\big)\cdot s+o(s)=\varepsilon s\cdot\mathsf{m}\big(\xi,\xi'\big)+o(s),$$

so the sum equals

$$\varepsilon s\sum_{\xi'\in\mathcal{G}(\xi)}\mathsf{m}\big(\xi,\xi'\big)\cdot\beta\big(\xi,\xi',T-t\big)+o(s)=\varepsilon s\cdot g_{\mathsf{v},\varepsilon}(\xi,T-t)+o(s).$$

We have thus proved that

$$\mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot \big(\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t] - \widehat{\mathbb{E}}[\mathcal{A}_t|\mathcal{F}_t]\big) = \mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot \big(\varepsilon s \cdot g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t) + o(s)\big),$$

so (31) follows. □

Next, we obtain the expression for the derivative with respect to time of the (nonconditional) expectation of $\mathcal{A}_t$.

LEMMA 3.11. *For $t \in [0,T)$, we have*

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\mathbb{E}}[\mathcal{A}_t] = \varepsilon \cdot \widehat{\mathbb{E}}\big[\mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t)\big]. \tag{33}$$

The first step in establishing this lemma is noting that, for $0 \le t < t+s \le T$,

$$\frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}] - \widehat{\mathbb{E}}[\mathcal{A}_t]}{s} = \widehat{\mathbb{E}}\bigg[\mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot \frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]}{s}\bigg],$$

since $\mathcal{A}_{t+s} = \mathcal{A}_t$ on $\{\tau_{\mathrm{sep}} \le t\}$, and $\mathcal{A}_t = 0$ on $\{\tau_{\mathrm{sep}} > t\}$. We would now like to take s to zero (from the right only, at least at first) and use Lemma 3.10, but we need to exchange the limit and the expectation; formally,

$$\lim_{s\to 0+} \widehat{\mathbb{E}}\bigg[\mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot \frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]}{s}\bigg] = \widehat{\mathbb{E}}\bigg[\mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot \lim_{s\to 0+} \frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]}{s}\bigg]. \tag{34}$$

The justification of this exchange is done with a standard dominated convergence argument, but an additional bound is required, so we postpone the full proof of Lemma 3.11 to the Appendix.

PROOF OF (22). Fix $\varepsilon > 0$. Using (30) and Lemma 3.11, we have

$$\frac{\mathbb{E}_{\lambda,\mathsf{v}+\varepsilon}[X_T] - \mathbb{E}_{\lambda,\mathsf{v}}[X_T]}{\varepsilon} = \int_0^T \widehat{\mathbb{E}}\big[\mathbb{1}\{\tau_{\mathrm{sep}} > t\} \cdot g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t)\big]\,\mathrm{d}t.$$

By Fubini's theorem (which we can use since we have the integrability condition given in Lemma 3.9), the right-hand side above equals

$$\widehat{\mathbb{E}}\bigg[\int_0^{\tau_{\mathrm{sep}}\wedge T} g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t)\,\mathrm{d}t\bigg].$$

We write this as

$$\begin{aligned}&\widehat{\mathbb{E}}\bigg[\int_0^T g_{\mathsf{v}}(\mathcal{V}_t, T-t)\,\mathrm{d}t\bigg] \\ &\quad + \widehat{\mathbb{E}}\bigg[\int_0^T \big(g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t) - g_{\mathsf{v}}(\mathcal{V}_t, T-t)\big)\,\mathrm{d}t\bigg] - \widehat{\mathbb{E}}\bigg[\int_{\tau_{\mathrm{sep}}\wedge T}^T g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t)\,\mathrm{d}t\bigg].\end{aligned} \tag{35}$$

Note that, since the law of $(\mathcal{V}_t)$ under $\widehat{\mathbb{P}}$ equals the law of (Ξ_t) under $\mathbb{P}_{\lambda,\mathsf{v}}$, we have

$$\widehat{\mathbb{E}}\bigg[\int_0^T g_{\mathsf{v}}(\mathcal{V}_t, T-t)\,\mathrm{d}t\bigg] = \mathbb{E}_{\lambda,\mathsf{v}}\bigg[\int_0^T g_{\mathsf{v}}(\Xi_t, T-t)\,\mathrm{d}t\bigg].$$

Hence, the proof will be completed once we prove that the second and third expectations in (35) tend to zero as $\varepsilon \to 0$.

It is straightforward to show that, for any $\xi \in \mathcal{S}$ and any $t \in [0,T]$,

$$g_{\mathsf{v},\varepsilon}(\xi, t) \xrightarrow{\varepsilon\to 0} g_{\mathsf{v}}(\xi, t).$$

Combining this with the dominated convergence theorem (using Lemma 3.9), we obtain

$$\widehat{\mathbb{E}}\Big[\int_0^T \big(g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t) - g_{\mathsf{v}}(\mathcal{V}_t, T-t)\big)\,\mathrm{d}t\Big] \xrightarrow{\varepsilon\to 0} 0.$$

Next, using the Cauchy–Schwarz inequality, we bound

$$\begin{aligned}&\Big|\widehat{\mathbb{E}}\Big[\int_{\tau_{\mathrm{sep}}\wedge T}^T g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t)\,\mathrm{d}t\Big]\Big| \\ &\quad\le T\cdot\widehat{\mathbb{E}}\Big[\Big(\max_{0\le t\le T}\big|g_{\mathsf{v},\varepsilon}(\mathcal{V}_t, T-t)\big|^2\Big)\Big]^{1/2}\cdot\widehat{\mathbb{P}}(\tau_{\mathrm{sep}}\le T)^{1/2}.\end{aligned}$$

The expectation on the right-hand side is finite by Lemma 3.9, and it is straightforward to check that $\widehat{\mathbb{P}}(\tau_{\mathrm{sep}}\le T)\xrightarrow{\varepsilon\to 0}0$. This completes the proof. □

3.4. *Analysis of higher moments.* Throughout this section we fix λ and v with $\lambda<\bar\lambda(\mathsf{v})$ (recalling the definition of $\bar\lambda(\mathsf{v})$ in Definition 2.6).

We now analyse the growth index φ_p for p possibly larger than 1. Our main goal is to prove the following.

PROPOSITION 3.12. *If $\lambda<\bar\lambda(\mathsf{v})$, then for every $p\ge 1$, we have*

$$\varphi_p<1 \tag{36}$$

and

$$\sup_{t\ge 0}\mathbb{E}\big[X_t^p\big]<\infty. \tag{37}$$

Note that the case $p=1$ has already been proved: the fact that $\varphi<1$ when $\lambda<\bar\lambda(\mathsf{v})$ is given by (19) and Proposition 3.6, and then (37) follows from $\varphi<1$ and (17).

In order to prove Proposition 3.12, we shall need an intermediate result, which we now state.

LEMMA 3.13. *Let $p\ge 2$. There exists $\mathfrak{C}_p>0$ (depending on λ, v and p) such that, for any $s\ge 48p^2\log(2d)/\mathsf{v}$,*

$$\mathbb{E}\Big[\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)}\Xi_s(A)\cdot|A|^p\Big]\le\mathfrak{C}_p\cdot\big(\mathbb{E}\big[X_s^{3/2}\big]+\varphi^s\big).$$

The proof of this lemma is quite involved, and we postpone it to the next section. We now show how to obtain Proposition 3.12 from it.

PROOF OF PROPOSITION 3.12. Fix $\lambda<\bar\lambda(\mathsf{v})$. Note that (37) follows readily from (36) and (15), so we only need to prove (36).

By (16) it suffices to prove that $\varphi_p<1$ for all $p\in\mathbb{N}$, and the case $p=1$ is already done. We proceed by induction, fixing $p\in\{2,3,\ldots\}$ and assuming that $\varphi_q<1$ for all $q\in\{1,\ldots,p-1\}$.

We first claim that there exist $C,c>0$ (which may depend on p, λ, v), such that, for all $t\ge 0$ and $\xi\in\mathcal{S}$,

$$\mathbb{E}\big[X_t^p\,|\,\Xi_0=\xi\big]\le\Big(\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)}\xi(A)\cdot|A|^p\Big)\cdot\mathbb{E}\big[X_t^p\big]+X(\xi)^p\cdot C\mathrm{e}^{-ct}. \tag{38}$$

This bound is a refinement of Proposition 2.7. While in the proof of that proposition we used Minkowski's inequality, here we expand the pth power of a sum in full and bound the various terms that appear.

To prove (38), let $t \geq 0$ and $\xi \in \mathcal{S}$ with enumeration $\xi = \sum_{i=1}^m \delta_{A_i}$. By Lemma 2.5 $(\Xi_t)_{t\geq 0}$ started from $\Xi_0 = \xi$ has the same distribution as $(\Xi_t^{(1)} + \cdots + \Xi_t^{(m)})_{t\geq 0}$, where $(\Xi_t^{(1)})_{t\geq 0}$, …, $(\Xi_t^{(m)})_{t\geq 0}$ are independent herds processes, with $\Xi_0^{(i)} = \delta_{A_i}$ for each i. In particular,

$$\begin{aligned}\mathbb{E}\big[X_t^p \,|\, \Xi_0 = \xi\big] &= \mathbb{E}\left[\left(\sum_{i=1}^m X\big(\Xi_t^{(i)}\big)\right)^p\right] \\ &= \sum_{\substack{(a_1,\dots,a_m):\\ a_1+\cdots+a_m=p}} \frac{p!}{a_1!\cdots a_m!} \cdot \prod_{\substack{i\in\{1,\dots,m\}:\\ a_i>0}} \mathbb{E}\big[X\big(\Xi_t^{(i)}\big)^{a_i}\big].\end{aligned}$$

By the fact that $\Xi_0^{(i)} = \delta_{A_i}$ and Proposition 2.7, the right-hand side is smaller than

$$\sum_{\substack{(a_1,\dots,a_m):\\ a_1+\cdots+a_m=p}} \frac{p!}{a_1!\cdots a_m!} \cdot \prod_{\substack{i\in\{1,\dots,m\}:\\ a_i>0}} |A_i|^{a_i} \cdot \mathbb{E}\big[X_t^{a_i}\big]. \tag{39}$$

We break this sum as

$$\mathbb{E}\big[X_t^p\big] \cdot \sum_{i=1}^m |A_i|^p + \sum_{\substack{(a_1,\dots,a_m):\\ a_1+\cdots+a_m=p,\\ a_1,\dots,a_m<p}} \frac{p!}{a_1!\cdots a_m!} \cdot \prod_{\substack{i\in\{1,\dots,m\}:\\ a_i>0}} |A_i|^{a_i} \cdot \mathbb{E}\big[X_t^{a_i}\big]. \tag{40}$$

Note that the first sum can be written as

$$\mathbb{E}\big[X_t^p\big] \cdot \sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \xi(A) \cdot |A|^p.$$

Next, the induction hypothesis that $\varphi_a < 1$ for all $a \in \{1, \dots, p-1\}$ together with (15) imply that there exist $C, c > 0$ (not depending on t) such that $\mathbb{E}[X_t^a] \leq Ce^{-ct}$ for any $a \in \{1, \dots, p-1\}$. Hence, the second sum in (40) is smaller than

$$\sum_{\substack{a_1,\dots,a_m\in\mathbb{N}_0:\\ a_1+\cdots+a_m=p,\\ a_1,\dots,a_m<p}} \frac{p!}{a_1!\cdots a_m!} \cdot \prod_{\substack{i\in\{1,\dots,m\}:\\ a_i>0}} |A_i|^{a_i} \cdot C\mathrm{e}^{-ct}.$$

By forgetting the last condition in the summation and increasing the value of C if necessary, this is smaller than

$$C\mathrm{e}^{-ct} \cdot \sum_{\substack{a_1,\dots,a_m\in\mathbb{N}_0:\\ a_1+\cdots+a_m=p}} \frac{p!}{a_1!\cdots a_m!} \cdot \prod_{\substack{i\in\{1,\dots,m\}:\\ a_i>0}} |A_i|^{a_i} = C\mathrm{e}^{-ct} \cdot X(\xi)^p.$$

We have thus proved (38).

Now, (38) and the Markov property imply that, for any $s, t \geq 0$,

$$\mathbb{E}\big[X_{t+s}^p\big] \leq \mathbb{E}\left[\sum_A \Xi_s(A) \cdot |A|^p\right] \cdot \mathbb{E}\big[X_t^p\big] + \mathbb{E}\big[X_s^p\big] \cdot C\mathrm{e}^{-ct}. \tag{41}$$

Using the bound of Lemma 3.13, increasing the constant C if necessary, for any s, t large enough, we then have

$$\mathbb{E}\big[X_{t+s}^p\big] \leq C\big\{\mathbb{E}\big[X_s^{3/2}\big] \cdot \mathbb{E}\big[X_t^p\big] + \varphi^s \cdot \mathbb{E}\big[X_t^p\big] + \mathbb{E}\big[X_s^p\big] \cdot \mathrm{e}^{-ct}\big\}. \tag{42}$$

We also bound

$$\begin{aligned}\mathbb{E}[X_s^{3/2}] &= \mathbb{E}[X_s^{3/2} \cdot \mathbb{1}\{X_s > 0\}] \\ &\le \mathbb{E}[(X_s^{3/2})^{4/3}]^{3/4} \cdot \mathbb{P}(X_s > 0)^{1/4} \\ &\le \mathbb{E}[X_s^2]^{3/4} \cdot \mathbb{P}(X_s \neq 0)^{1/4} \le \mathbb{E}[X_s^2]^{3/4} \cdot \mathbb{E}[X_s]^{1/4} \le C\mathbb{E}[X_s^p]^{3/4} \cdot \varphi^{s/4},\end{aligned}$$

where the first inequality is Hölder's, the second inequality follows from $\mathbb{1}\{X_s \neq 0\} \le X_s$, and the third inequality follows from $X_s^2 \le X_s^p$ and (17). We use this bound in (42), together with the fact that $\varphi < 1$, to obtain

$$\mathbb{E}[X_{t+s}^p] \le C\{\mathbb{E}[X_s^p]^{3/4} \cdot \mathbb{E}[X_t^p] \cdot \mathrm{e}^{-cs} + \mathbb{E}[X_t^p] \cdot \mathrm{e}^{-cs} + \mathbb{E}[X_s^p] \cdot \mathrm{e}^{-ct}\} \tag{43}$$

for suitable choices of c, C.

Now, assume for a contradiction that $\varphi_p \ge 1$. In that case, by (13) we have $\mathbb{E}[X_r^p] \ge 1$ for all $r \ge 0$, and from (43) we obtain, for large enough s, t with $s \le t$,

$$\mathbb{E}[X_{t+s}^p] \le C\mathbb{E}[X_s^p] \cdot \mathbb{E}[X_t^p] \cdot \mathrm{e}^{-cs}.$$

Using this recursively, we have that, for all sufficiently large s and all $n \in \mathbb{N}$,

$$\mathbb{E}[X_{ns}^p] \le C^n \cdot \mathbb{E}[X_s^p]^n \cdot \mathrm{e}^{-cns}.$$

Taking both sides to the power $\frac{1}{ns}$ and letting $n \to \infty$ (using (15)) gives

$$\varphi_p \le C^{1/s} \cdot \mathbb{E}[X_s^p]^{1/s} \cdot \mathrm{e}^{-c}.$$

Now, letting $s \to \infty$ and again using (15) yields $\varphi_p \le \varphi_p \cdot \mathrm{e}^{-c}$, a contradiction. □

Before we turn to the proof of Lemma 3.13, we want to give a consequence of Proposition 2.7, namely, Corollary 3.15 below, which will be useful in dealing with the contact process on dynamic graphs. For $t \ge 0$, we denote by N_t the number of birth events in the herds process up to time t, as in Lemma 2.1. Also, let $N_\infty := \lim_{t\to\infty} N_t$.

LEMMA 3.14. *Let $p \ge 1$. For any $t \in [0, \infty]$ and $\xi \in \mathcal{S}$, we have*

$$\mathbb{E}[N_t^p \mid \Xi_0 = \xi] \le X(\xi)^p \cdot \mathbb{E}[N_t^p].$$

Consequently, for any $s \in [0, \infty)$ and $t \in [0, \infty]$,

$$\mathbb{E}[(N_{s+t} - N_s)^p] \le \mathbb{E}[X_s^p] \cdot \mathbb{E}[N_t^p]. \tag{44}$$

PROOF. First, recall Remark 2.5, which, in particular, shows that one can dominate the number of particles in a herds process starting from ξ by the total number of particles in a multitype herds process with $\ell := X(\xi)$ types, starting from the configuration ξ' where each particle in ξ represents a distinct type. For this auxiliary process, let $N_t^{(i)}$ denote the number of births of particles of type i by time t, for $i \in \{1, \dots, \ell\}$. Then

$$\mathbb{E}[N_t^p \mid \Xi_0 = \xi] \le \mathbb{E}\left[\left(\sum_{i=1}^{\ell} N_t^{(i)}\right)^p\right] \le \left(\sum_{i=1}^{\ell} \mathbb{E}[(N_t^{(i)})^p]^{1/p}\right)^p,$$

where we have used Minkowski's inequality. Now, since each type evolves a usual herds process, we have $\mathbb{E}[(N_t^{(i)})^p] = \mathbb{E}[N_t^p]$ for every i, so the right-hand side above equals $\ell^p \cdot \mathbb{E}[N_t^p]$, as required. □

COROLLARY 3.15. *If $\lambda < \bar{\lambda}(\nu)$, then $\mathbb{E}[N_\infty^p] < \infty$ for all $p \ge 1$.*

PROOF. Fix $p \geq 1$. For any $x > 1$, let $t_x := \frac{p+1}{|\log \varphi|} \cdot \log x$ (note that $\varphi < 1$ since $\lambda < \bar{\lambda}(\mathsf{v})$, by Lemma 3.5 and Proposition 3.6). Letting τ denote the extinction time of the herds process, we first bound

$$\mathbb{P}(N_\infty > x) \leq \mathbb{P}(\tau > t_x) + \mathbb{P}(N_{t_x} > x).$$

The first term on the right-hand side is easy to bound,

$$\mathbb{P}(\tau > t_x) \leq \mathbb{E}[X_{t_x}] \overset{(17)}{\leq} C\varphi^{t_x} = Cx^{-(p+1)}.$$

Next, we bound

$$\begin{aligned}\mathbb{P}(N_{t_x} > x) &\leq \sum_{k=0}^{\lfloor t_x \rfloor - 1} \mathbb{P}\left(N_{k+1} - N_k > \frac{x}{t_x}\right) \\ &\leq \sum_{k=0}^{\lfloor t_x \rfloor - 1} \frac{\mathbb{E}[(N_{k+1} - N_k)^{p+1}]}{(x/t_x)^{p+1}} \leq t_x^{p+2} \cdot x^{-(p+1)} \cdot \sup_{t \geq 0} \mathbb{E}\big[(N_{t+1} - N_t)^{p+1}\big].\end{aligned}$$

Now, the supremum on the right-hand side is finite by (37) and (44). We have thus proved that $\mathbb{P}(N_\infty > x) \leq C(\log x)^{p+2} x^{-(p+1)}$ for some $C > 0$, which implies that N_∞ has finite pth moment. □

3.5. *Bound on moments of herd sizes*: *Proof of Lemma 3.13.* We will need several preliminary results, starting with the following.

LEMMA 3.16. *For any $\alpha \geq 1$ and $p \geq 1$, there exists $C_{\alpha,p} > 0$ (depending on λ, v, α and p) such that the following holds. Fix $u \in \mathbb{T}^d \setminus \{o\}$, and let $(\widetilde{\Xi}_t)_{t \geq 0}$ be a two-type herds process with rates λ and v, started from a single herd containing a type-1 particle at the root and a type-2 particle at u, that is, $\widetilde{\Xi}_0 = \delta_{\{o\},\{u\}}$. Then, letting $t_\alpha := 12\log(\alpha)/\mathsf{v}$, we have*

$$\mathbb{E}\left[\left(\sum_{(A,B):A \neq \varnothing} \widetilde{\Xi}_{t_\alpha}(A,B) \cdot |B|\right)^p\right] < C_{\alpha,p} \cdot \alpha^{-\mathrm{dist}(o,u)}. \tag{45}$$

PROOF. Fix λ, v, and α, let $t_\alpha := 12\log(\alpha)/\mathsf{v}$, and fix u as in the statement. For $t \geq 0$, let Y_t denote the number of herds in $\widetilde{\Xi}_t$ that contain both types, that is,

$$Y_t := \sum_{\substack{(A,B):\\ A \neq \varnothing,\\ B \neq \varnothing}} \widetilde{\Xi}_t(A,B).$$

It will be useful to note that

$$s \leq t \quad \Longrightarrow \quad \{Y_s = 0\} \subseteq \{Y_t = 0\}. \tag{46}$$

Noting that

$$\left(\sum_{(A,B):A \neq \varnothing} \widetilde{\Xi}_{t_\alpha}(A,B) \cdot |B|\right)^p = \left(\sum_{(A,B):A \neq \varnothing} \widetilde{\Xi}_{t_\alpha}(A,B) \cdot |B|\right)^p \cdot \mathbb{1}\{Y_{t_\alpha} > 0\}$$

and using the Cauchy–Schwarz inequality, the left-hand side of (45) is smaller than

$$\mathbb{E}\left[\left(\sum_{(A,B):A \neq \varnothing} \widetilde{\Xi}_{t_\alpha}(A,B) \cdot |B|\right)^{2p}\right]^{1/2} \cdot \mathbb{P}(Y_{t_\alpha} \geq 1)^{1/2}.$$

Using domination by a pure-birth process, the first term in the product above can be bounded by a finite constant that only depends on λ, p, and α. We now show that $\mathbb{P}(Y_{t_\alpha} \geq 1)$ is smaller than $C\alpha^{-2\mathrm{dist}(o,u)}$ for some $C > 0$.

For $i \in \{1, 2\}$ and $t \geq 0$, let $K_t^{(i)}$ denote the set of vertices of $\mathbb{T}^d$ that have been occupied by a type-i particle in some herd for some time $s \leq t$, that is,

$$K_t^{(1)} := \{v \in \mathbb{T}^d : \widetilde{\Xi}_s(A, B) > 0 \text{ for some } s \leq t \text{ and } (A, B) \text{ with } v \in A\},$$

$$K_t^{(2)} := \{v \in \mathbb{T}^d : \widetilde{\Xi}_s(A, B) > 0 \text{ for some } s \leq t \text{ and } (A, B) \text{ with } v \in B\}.$$

We have that $K_0^{(1)} = \{o\}$, $K_0^{(2)} = \{u\}$ and $t \mapsto K_t^{(1)}$ and $t \mapsto K_t^{(2)}$ are both nondecreasing (with respect to set inclusion). Moreover, $K_t^{(1)}$ and $K_t^{(2)}$ are connected subsets of $\mathbb{T}^d$, since they only grow by the inclusion of vertices neighboring vertices that are already present. Also, note that, as long as these sets stay disjoint, there can be at most one herd containing both types. In other words, letting

$$\sigma := \inf\{t : K_t^{(1)} \cap K_t^{(2)} \neq \varnothing\},$$

we have, for any $t \geq 0$,

$$\{\sigma > t\} \subseteq \{Y_s \leq 1 \text{ for all } s \leq t\}. \tag{47}$$

Next, let

$$\ell = \mathrm{dist}(o, u),$$

and let $o = u_0 \sim u_1 \sim \ldots \sim u_\ell = u$ be the vertices of $\mathbb{T}^d$ in the geodesic from o to u. Let

$$u' := u_{\lfloor \ell/3 \rfloor}, \qquad \sigma^{(1)} := \inf\{t \geq 0 : u' \in K_t^{(1)}\}$$

and

$$u'' := u_{\lfloor 2\ell/3 \rfloor}, \qquad \sigma^{(2)} := \inf\{t \geq 0 : u'' \in K_t^{(2)}\}.$$

It is easy to see that

$$\min(\sigma^{(1)}, \sigma^{(2)}) \leq \sigma.$$

Putting this together with (47), we obtain

$$\{Y_s \geq 2 \text{ for some } s \leq t_\alpha\} \subseteq \{\sigma \leq t_\alpha\} \subseteq \{\min(\sigma^{(1)}, \sigma^{(2)}) \leq t_\alpha\}, \tag{48}$$

so we can bound

$$\begin{aligned}
\mathbb{P}(Y_{t_\alpha} \geq 1) &\overset{(46)}{=} \mathbb{P}(Y_s \geq 1 \text{ for all } 0 \leq s \leq t_\alpha) \\
&\overset{(48)}{\leq} \mathbb{P}(\sigma^{(1)} \leq t_\alpha) + \mathbb{P}(\sigma^{(2)} \leq t_\alpha) \qquad (49) \\
&\quad + \mathbb{P}(\min(\sigma^{(1)}, \sigma^{(2)}) > t_\alpha, Y_s = 1 \text{ for } 0 \leq s \leq t_\alpha). \qquad (50)
\end{aligned}$$

We now bound the three terms on the right-hand side separately.

Let us first consider the probability in (50). For any t, if $\min(\sigma^{(1)}, \sigma^{(2)}) > t$ and $Y_t = 1$, then a split in any of the edges

$$\{u_{\lfloor \ell/3 \rfloor}, u_{\lfloor \ell/3 \rfloor + 1}\}, \{u_{\lfloor \ell/3 \rfloor + 1}, u_{\lfloor \ell/3 \rfloor + 2}\}, \ldots, \{u_{\lfloor 2\ell/3 \rfloor - 1}, u_{\lfloor 2\ell/3 \rfloor}\}$$

separates the two types permanently, causing Y to drop to zero. This observation gives

$$\begin{aligned}
\mathbb{P}(\min(\sigma^{(1)}, \sigma^{(2)}) > t_\alpha, Y_s = 1 \text{ for } 0 \leq s \leq t_\alpha) &\leq \exp\{-\mathsf{v} \cdot t_\alpha \cdot (\lfloor 2\ell/3 \rfloor - \lfloor \ell/3 \rfloor)\} \\
&\leq \alpha^{-2\ell},
\end{aligned}$$

where the second inequality follows from the definition of t_α.

We now turn to the two probabilities in (49). We only bound the first one; by a symmetry argument, the same bound will then apply to the second. Let $(W_t)_{t\geq 0}$ be a growth process on $(\mathbb{N}_0)^{\mathbb{T}^d}$ defined as follows. We let $W_0(o) = 1$ and $W_0(u) = 0$ for $u \neq o$. We interpret $W_t(v) = m$ as saying that there are m particles at v at time t. Then we define the dynamics by prescribing that independently, for any $v \sim w$, a particle at v gives birth at a particle at w with rate λ (and particles never die). In particular, $(\sum_v W_t(v))_{t\geq 0}$ is a pure-birth process in which the birth rate is $d\lambda$. It is easy to see that the set-valued process $(\{u \in \mathbb{T}^d : W_t(u) > 0\})_{t\geq 0}$ stochastically dominates $(K_t^{(1)})_{t\geq 0}$, and in particular,

$$\mathbb{P}\big(\sigma^{(1)} \leq t_\alpha\big) \leq \mathbb{P}\big(W_{t_\alpha}(u') > 0\big) \leq \mathbb{E}\big[W_{t_\alpha}(u')\big]. \tag{51}$$

We now claim that, for any $t \geq 0$ and $v \in \mathbb{T}^d$,

$$\mathbb{E}\big[W_t(v)\big] = e^{d\lambda t} \cdot p_t(v), \tag{52}$$

where $p_t(v) := \mathbb{P}(\mathcal{Z}_t = v)$, with $(\mathcal{Z}_t)_{t\geq 0}$ the continuous-time random walk on $\mathbb{T}^d$, which starts at the root at time zero and jumps from any vertex to any neighboring vertex with rate λ. This is a standard many-to-one formula; see [17], and see also [24], Proposition 1.21, page 40, for a proof of an analogous inequality on $\mathbb{Z}^d$. For completeness we include a brief proof. It is simple to verify (52) using the observations that $(t, v) \mapsto \mathbb{E}[W_t(v)]$ is the solution to

$$\begin{cases} \dfrac{\mathrm{d}}{\mathrm{d}t} g(t, v) = \lambda \displaystyle\sum_{w\sim v} g(t, w), \quad t \geq 0, v \in \mathbb{T}^d, \\ g(0, \cdot) = \delta_o(\cdot), \end{cases}$$

and that the right-hand side of (52) solves this equation by direct computation.

Putting together (51) and (52), we have

$$\mathbb{P}\big(\sigma^{(1)} \leq t_\alpha\big) \leq e^{d\lambda t_\alpha} \cdot p_{t_\alpha}(u') \leq e^{d\lambda t_\alpha} \cdot \mathbb{P}\big(\mathrm{Poi}(d\lambda t_\alpha) > \lfloor \ell/3 \rfloor\big),$$

where $\mathrm{Poi}(d\lambda t_\alpha)$ represents a random variable with the Poisson distribution with parameter $d\lambda t_\alpha$, which is the law of the number of jumps of $(\mathcal{Z}_t)$ until time t_α. Since the tail of the Poisson distribution is lighter than exponential, the right-hand side above is smaller than $C\alpha^{-2\ell}$ for some $C > 0$, uniformly in ℓ. This concludes the proof. □

LEMMA 3.17. *For any $p \geq 1$, there exists $C'_p > 0$ (depending on λ, v and p) such that the following holds. Fix $B_0 \subseteq \mathbb{T}^d \setminus \{o\}$, and let $(\widetilde{\Xi}_t)_{t\geq 0}$ be a two-type herds process with rates λ, v started from $\widetilde{\Xi}_0 = \delta_{\{o\},B_0}$. Also, let $T_p := 12p\log(2d)/\mathsf{v}$. Then (uniformly over the choice of B_0)*

$$\mathbb{E}\Big[\Big(\sum_{(A,B):A\neq\varnothing} \widetilde{\Xi}_{T_p}(A, B) \cdot \big(|A| + |B|\big)\Big)^p\Big] \leq C'_p.$$

PROOF. Let

$$R := \sum_{(A,B):A\neq\varnothing} \widetilde{\Xi}_{T_p}(A, B) \cdot \big(|A| + |B|\big);$$

note that R is the total number of particles in $\widetilde{\Xi}_{T_p}$ that belong to herds that contain type-1 particles.

We enumerate $\{o\} \cup B_0 = \{u_1, \ldots, u_m\}$, with $u_1 = o$ and $m = |B_0| + 1$. We now define a multitype herds process, as described in Remark 2.5. This new process, denoted $(\widetilde{\widetilde{\Xi}}_t)_{t\geq 0}$, is taken in the same probability space that we have been considering, has rates λ, v, and m types. It starts with a single herd, with a type-1 particle at $u_1 = o$, a type-2 particle at u_2, …, and a

type m particle at u_m. In analogy with R, we let R' denote the total number of particles in $\tilde{\tilde{\Xi}}_{T_p}$ that belong to herds that contain type-1 particles. That is, if we use an m-tuple $(A_1, \dots, A_m)$ to represent a multitype herd shape, then

$$R' := \sum_{(A_1,\dots,A_m):A_1\neq\varnothing} \tilde{\tilde{\Xi}}_{T_p}(A_1, \dots, A_m) \cdot \big(|A_1| + \cdots + |A_m|\big).$$

Similarly as in Lemma 2.9, we see that R is stochastically dominated by R'; in particular,

$$\mathbb{E}\big[R^p\big]^{1/p} \le \mathbb{E}\big[(R')^p\big]^{1/p} \le \sum_{j=1}^{m} \mathbb{E}\big[(R'_j)^p\big]^{1/p},$$

where

$$R'_j := \sum_{(A_1,\dots,A_m):A_1\neq\varnothing} \tilde{\tilde{\Xi}}_{T_p}(A_1, \dots, A_m) \cdot |A_j|.$$

Note that R'_1 is just the total number of type-1 particles in $\tilde{\tilde{\Xi}}_{T_p}$. If we ignore all types except for type 1 in $(\tilde{\tilde{\Xi}}_t)$, we obtain a (one-type) herds process started from $\delta_{\{o\}}$; hence, $\mathbb{E}[(R'_1)^p] = \mathbb{E}[X^p_{T_p}] < \infty$ by Corollary 2.3. Next, for $j \neq 1$, if we ignore all types except for types 1 and j in $(\tilde{\tilde{\Xi}}_t)$, we see a two-type herds process started from $\delta_{\{o\},\{u_j\}}$, and Lemma 3.16 (with $\alpha = (2d)^p$) and the definition of T_p imply that $\mathbb{E}[(R'_j)^p] \le C_p(2d)^{-p\cdot\mathrm{dist}(o,u_j)}$. We then have

$$\sum_{j=1}^{m} \mathbb{E}\big[(R'_j)^p\big]^{1/p} \le \mathbb{E}\big[X^p_{T_p}\big]^{1/p} + C_p^{1/p} \sum_{j=2}^{m} (2d)^{-\mathrm{dist}(o,u_j)}.$$

The second sum on the right-hand side is smaller than

$$\sum_{u\in\mathbb{T}^d} (2d)^{-\mathrm{dist}(o,u)} \le \sum_{i=1}^{\infty} d^i \cdot (2d)^{-i} = 1.$$

We have thus proved that $\mathbb{E}[R^p] \le (\mathbb{E}[X^p_{T_p}]^{1/p} + C_p^{1/p})^p$, so the proof is complete. □

LEMMA 3.18. *Let $p \ge 1$, and let $C''_p := \max(C'_p, \mathbb{E}[X^p_{T_p}])$, where C'_p and T_p are as in Lemma 3.17. Then for any $\xi \in \mathcal{S}$, the herds process $(\Xi_t)_{t\ge0}$ with rates λ, ν satisfies*

$$\mathbb{P}\big(\exists A : |A| \ge x, \Xi_{T_p}(A) > 0 | \Xi_0 = \xi\big) \le C''_p \cdot \frac{X(\xi)}{x^p}, \quad x > 0.$$

PROOF. By using Lemma 2.5 and a union bound, it suffices to prove that, for any $A_0 \in P_{\mathrm{f}}(\mathbb{T}^d)$,

$$\mathbb{P}\big(\exists A : |A| \ge x, \Xi_{T_p}(A) > 0 | \Xi_0 = \delta_{A_0}\big) \le C''_p \cdot \frac{|A_0|}{x^p}, \quad x > 0. \tag{53}$$

We will prove this by induction on $|A_0|$. For the case where $|A_0| = 1$, we bound the left-hand side above by

$$\mathbb{P}(X_{T_p} \ge x | \Xi_0 = \delta_{\{o\}}) \le \frac{\mathbb{E}[X^p_{T_p}]}{x^p}$$

by Markov's inequality.

Now, assume that $|A_0| \ge 2$, and let $u \in A_0$. Let $B_0 := A_0\backslash\{u\}$. We consider a two-type herds process $(\tilde{\Xi}_t)_{t\ge0}$ with rates λ, ν started from $\delta_{\{u\},B_0}$. Using Lemma 2.9(b), the left-hand side of (53) is smaller than

$$\mathbb{P}\big(\exists (A, B) : |A| + |B| \ge x, \tilde{\Xi}_{T_p}(A, B) > 0\big).$$

In turn, this is smaller than

$$\mathbb{P}\big(\exists (A,B) : A \neq \varnothing, |A| + |B| \geq x, \widetilde{\Xi}_{T_p}(A,B) > 0\big) \tag{54}$$

$$+ \mathbb{P}\big(\exists (A,B) : |B| \geq x, \widetilde{\Xi}_{T_p}(A,B) > 0\big). \tag{55}$$

The probability in (54) is smaller than

$$\mathbb{P}\Big(\sum_{(A,B):A\neq\varnothing} \Xi_{T_p}(A,B) \cdot \big(|A| + |B|\big) \geq x \Big) \leq \frac{C'_p}{x^p}$$

by Markov's inequality and Lemma 3.17. By Lemma 2.9(a) the probability in (55) can be expressed using a one-type herds process; it equals

$$\mathbb{P}\big(\exists B : |B| \geq x, \Xi_{T_p}(B) > 0 | \Xi_0 = \delta_{B_0}\big),$$

which is smaller than $C''_p |B_0|/x^p$ by the induction hypothesis (since $|B_0| = |A_0| - 1$). We have then proved that

$$\mathbb{P}\big(\exists A : |A| \geq x, \Xi_{T_p}(A) > 0 | \Xi_0 = A_0\big) \leq \frac{C'_p}{x^p} + \frac{C''_p(|A_0|-1)}{x^p} \leq \frac{C''_p |A_0|}{x^p}. \qquad \square$$

PROOF OF LEMMA 3.13. Let $q = 4p^2$, and fix $s \geq T_q = 12q \log(2d)/\mathrm{v}$. Let E_1 be the event that there is some herd at time s whose number of particles is larger than $X_{s-T_q}^{\frac{1}{2p}}$, that is,

$$E_1 := \big\{\text{there exists } A \in P_{\mathrm{f}}(\mathbb{T}^d) \text{ with } |A| \geq X_{s-T_q}^{\frac{1}{2p}} \text{ and } \Xi_s(A) > 0\big\}.$$

On E_1^c we bound

$$\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \Xi_s(A) \cdot |A|^p \leq X_{s-T_q}^{1/2} \cdot \sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \Xi_s(A) \leq X_{s-T_q}^{1/2} \cdot X_s. \tag{56}$$

On E_1 we bound

$$\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \Xi_s(A) \cdot |A|^p \leq X_s^p.$$

These bounds give

$$\mathbb{E}\Big[\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \Xi_s(A) \cdot |A|^p \Big] \leq \mathbb{E}\big[X_s \cdot X_{s-T_q}^{1/2}\big] + \mathbb{E}\big[X_s^p \cdot \mathbb{1}_{E_1}\big]. \tag{57}$$

We treat the two expectations on the right-hand side separately. For the first one, we start by bounding

$$\mathbb{E}\big[X_s \cdot X_{s-T_q}^{1/2}\big] = \mathbb{E}\big[\mathbb{E}[X_s|\mathcal{F}_{s-T_q}] \cdot X_{s-T_q}^{1/2}\big] \leq \mathbb{E}\big[X_{s-T_q}^{3/2}\big] \cdot \mathbb{E}[X_{T_q}], \tag{58}$$

where the inequality follows from (6) and the Markov property. We now claim that there exists $C > 0$ (not depending on s) such that

$$\mathbb{E}\big[X_{s-T_q}^{3/2}\big] \leq C\mathbb{E}\big[X_s^{3/2}\big]. \tag{59}$$

To see this, first note that each particle that is alive at time $s - T_q$ will stay alive until time s with probability $\rho := \mathrm{e}^{-T_q}$. Then, defining the event

$$E_2 := \Big\{ X_s \geq \frac{\rho}{2} \cdot X_{s-T_q} \Big\},$$

we have, for any $m \in \mathbb{N}$,

$$\mathbb{P}(E_2|\mathcal{F}_{s-T_q}) \cdot \mathbb{1}\{X_{s-T_q} = m\} \geq \mathbb{P}\left(\mathrm{Bin}(m, \rho) \geq \frac{\rho}{2}m\right)$$

(and in case $m = 0$, the left-hand side is trivially equal to 1). Hence,

$$\mathbb{P}(E_2|\mathcal{F}_{s-T_q}) \geq \beta := \inf_{m\geq 1} \mathbb{P}\left(\mathrm{Bin}(m, \rho) \geq \frac{\rho}{2}m\right) > 0.$$

Then

$$\begin{aligned}\mathbb{E}\big[X_s^{3/2}\big] &\geq \mathbb{E}\big[X_s^{3/2} \cdot \mathbb{1}_{E_2}\big] \geq \frac{\rho}{2} \cdot \mathbb{E}\big[X_{s-T_q}^{3/2} \cdot \mathbb{1}_{E_2}\big] \\ &= \frac{\rho}{2} \cdot \mathbb{E}\big[X_{s-T_q}^{3/2} \cdot \mathbb{P}(E_2|\mathcal{F}_{s-T_q})\big] \geq \frac{\rho}{2} \cdot \beta \cdot \mathbb{E}\big[X_{s-T_q}^{3/2}\big],\end{aligned}$$

proving (59) with $C := \frac{2}{\beta\rho}$. With (58) and (59) and putting together all the terms that do not depend on s in a sufficiently large constant C, we have proved that

$$\mathbb{E}\big[X_s \cdot X_{s-T_q}^{1/2}\big] \leq C\mathbb{E}\big[X_s^{3/2}\big].$$

We now turn to the second term on the right-hand side of (57). We first bound the conditional expectation, given $\mathcal{F}_{s-T_q}$, with Cauchy–Schwarz,

$$\mathbb{E}\big[X_s^p \cdot \mathbb{1}_{E_1}|\mathcal{F}_{s-T_q}\big] \leq \mathbb{E}\big[X_s^{2p}|\mathcal{F}_{s-T_q}\big]^{1/2} \cdot \mathbb{P}(E_1|\mathcal{F}_{s-T_q})^{1/2}.$$

Using (6) and the Markov property,

$$\mathbb{E}\big[X_s^{2p}|\mathcal{F}_{s-T_q}\big] \leq X_{s-T_q}^{2p} \cdot \mathbb{E}\big[X_{T_q}^{2p}\big].$$

Using Lemma 3.18 (and again the Markov property) and recalling that $q = 4p^2$, we have

$$\mathbb{P}(E_1|\mathcal{F}_{s-T_q}) \leq C_q'' \cdot \frac{X_{s-T_q}}{(X_{s-T_q}^{1/(2p)})^q} = C_q'' \cdot X_{s-T_q}^{1-2p}.$$

Then, using that X_{s-T_q} is a nonnegative integer and hence $X_{s-T_q}^{1/2} \leq X_{s-T_q}$, we get

$$\mathbb{E}\big[X_s^p \cdot \mathbb{1}_{E_1}|\mathcal{F}_{s-T_q}\big] \leq \big(C_q'' \cdot \mathbb{E}\big[X_{T_q}^{2p}\big]\big)^{1/2} \cdot X_{s-T_q},$$

so

$$\mathbb{E}\big[X_s^p \cdot \mathbb{1}_{E_1}\big] \leq \big(C_q'' \cdot \mathbb{E}\big[X_{T_q}^{2p}\big]\big)^{1/2} \cdot \mathbb{E}[X_{s-T_q}] \leq \big(C_q'' \cdot \mathbb{E}\big[X_{T_q}^{2p}\big]\big)^{1/2} \cdot C\varphi^{s-T_q},$$

where the second inequality follows from (17). Putting together all the constants that do not depend on s, this gives

$$\mathbb{E}\big[X_s^p \cdot \mathbb{1}_{E_1}\big] \leq C\varphi^s.$$

Altogether, this gives

$$\mathbb{E}\Bigg[\sum_{A\in P_{\mathrm{f}}(\mathbb{T}^d)} \Xi_s(A) \cdot |A|^p\Bigg] \leq C \cdot \big(\mathbb{E}\big[X_s^{3/2}\big] + \varphi^s\big),$$

as wanted. □

4. Extinction time on a random d-regular graph with switching. The goal of this section is to prove Theorem 1.2. As mentioned in the Introduction, the first part of the theorem was already proved in [12], so we only prove the second part here.

This section is organized as follows. In Section 4.1 we give a detailed construction of the dynamic random graph and the contact process which co-evolves on this graph. After doing so, we argue that a version of the usual self-duality relation of the contact process is satisfied here. Due to this relation, in proving quick extinction, it suffices to study the process started from a single infection. We start this study in Section 4.2, where we introduce an exploration process, which reveals the contact process but only reveals partial information about the graph. In Section 4.3 we prove a general Markov chain lemma, which allows us to couple this exploration process with (a projection of) the herds process. Finally, in Section 4.4 we take advantage of this coupling to give the proof of Theorem 4.4.

4.1. *Preliminaries*: *Dynamic graph*, *joint evolution*, *duality.* Let us define the class of graphs in which our dynamic graph process takes values. Fix $n \in \mathbb{N}$, and let $V_n := [n] = \{1, \dots, n\}$. The set $\{(u, a) : u \in V_n,\ a \in \{1, \dots, d\}\}$ is called the *set of half-edges*. Given a perfect matching of the set of half-edges (i.e., a bijection σ from this set to itself with no fixed points and equal to its own inverse), we can obtain a (multi)graph by setting V_n as the set of vertices and prescribing that each pair $\{(u, a), (u', a')\}$ with $(u', a') = \sigma(u, a)$ corresponds to an edge between u and u'. Let $\mathcal{G}_n$ denote the set of all multigraphs that can be obtained in this way. Deterministic elements of $\mathcal{G}_n$ will typically be denoted by g, whereas random elements of $\mathcal{G}_n$ will be denoted by G or G_t (in the case of a process).

Fix $\mathsf{g} \in \mathcal{G}_n$. A *switch code* for g is a triple $\mathsf{m} = (e_1, e_2, \eta)$, where e_1, e_2 are distinct edges of g and $\eta \in \{-, +\}$. Fix such a switch code, with $e_1 = \{(u, a), (u', a')\}$ and $e_2 = \{(v, b), (v', b')\}$ so that $(u, a) < (u', a')$ and $(v, b) < (v', b')$ in lexicographic order. In case $\eta = +$, we let $\Gamma^{\mathsf{m}}(\mathsf{g})$ be the graph obtained from g by replacing e_1 and e_2 by the edges $\{(u, a), (v, b)\}$ and $\{(u', a'), (v', b')\}$ (and keeping all other edges intact). In case $\eta = -$, we instead replace e_1, e_2 by $\{(u, a), (v', b')\}$ and $\{(u', a'), (v, b)\}$.

The random graph process $(G_t)_{t \geq 0}$ is the continuous-time Markov chain on $\mathcal{G}_n$, which jumps from g to g' with rate $\upsilon_n := \frac{\mathsf{v}}{nd}$ if $\mathsf{g}' = \Gamma^{\mathsf{m}}(\mathsf{g})$ for some switch code m of g (and rate 0 otherwise). This chain is reversible with respect to the uniform measure on $\mathcal{G}_n$. We will always start the graph dynamics from this distribution.

We now fix $\lambda > 0$ and define the contact process $(\xi_t)_{t \geq 0}$ with infection rate λ on the dynamic graph (G_t). Although we could do so by describing the jump rates of the joint Markov chain $(G_t, \xi_t)_{t \geq 0}$, we will instead use a Poisson graphical construction. We take a probability space with probability measure $\mathbb{P}$ in which the process (G_t) is defined, and also (independently of (G_t)) the following Poisson point processes (all independent) are defined:

- for each vertex u, a Poisson point process R^u on $[0, \infty)$ with intensity 1 (recovery times);
- for each half-edge (u, a), a Poisson point process $R^{(u,a)}$ with intensity λ (transmission times).

Naturally, when $t \in R^u$, vertex u goes to state 0 (if not already there) at time t. Moreover, when $t \in R^{(u,a)}$, there is a transmission from u to the vertex v that owns the half-edge to which (u, a) is matched in G_t (so that if u was in state 1 just before t, then v goes to state 1, if not already there, at t). For each $A \subset V_n$, we let $(\xi_t^A)_{t \geq 0}$ be the contact process on (G_t) with $\xi_0^A = \mathbb{1}_A$ and obtained from this graphical construction (as usual, the graphical construction gives us contact processes started from all possible initial configurations, all coupled in a single probability space and respecting the monotonicity of set inclusion).

The usual duality relation

$$\mathbb{P}(\xi_t^A \cap B \neq \varnothing) = \mathbb{P}(\xi_t^B \cap A \neq \varnothing) \quad \text{for all } t \geq 0,\ A, B \subset V_n \tag{60}$$

holds in this context, but it is important to note that the above probabilities are annealed in the graph environment. Let us briefly prove (60). Fix $t \geq 0$ and $A, B \subset V_n$. Letting $(\mathsf{g}_s)_{0\leq s\leq t}$ be a possible realization of the trajectory of $(G_s)_{0\leq s\leq t}$, we have

$$\begin{aligned}&\mathbb{P}\big(\xi_t^A \cap B \neq \varnothing | (G_s)_{0\leq s\leq t} = (\mathsf{g}_s)_{0\leq s\leq t}\big)\\ &\quad = \mathbb{P}\big(\xi_t^B \cap A \neq \varnothing | (G_s)_{0\leq s\leq t} = (\mathsf{g}_{t-s})_{0\leq s\leq t}\big).\end{aligned}$$

This is verified using a standard argument involving infection paths and time-reversibility of Poisson processes. Integrating this equality over the choice of $(\mathsf{g}_s)_{0\leq s\leq t}$ and using the fact that $(G_s)_{0\leq s\leq t}$ has the same law as $(G_{t-s})_{0\leq s\leq t}$ gives (60).

Letting $\bar{u} \in V_n$ be arbitrary (and deterministic) and writing $\xi_t^{\bar{u}}$ instead of $\xi_t^{\{\bar{u}\}}$, we have

$$\mathbb{P}\big(\xi_t^{V_n} \neq \varnothing\big) \leq \sum_{u\in V_n} \mathbb{P}\big(\xi_t^{V_n}(u) = 1\big) = n \cdot \mathbb{P}\big(\xi_t^{V_n}(\bar{u}) = 1\big) = n \cdot \mathbb{P}\big(\xi_t^{\bar{u}} \neq \varnothing\big), \tag{61}$$

where the equalities follow from symmetry and duality, respectively. Due to (61), the analysis of the extinction time of the contact process started from all vertices infected can be reduced to the analysis of the extinction time of the contact process from a single infection at the arbitrary vertex $\bar{u}$. For the rest of this section, $\bar{u}$ remains fixed.

Next, we describe an *exploration process*, which only reveals partial information about $(G_t)_{t\geq 0}$, namely, only the matching of half-edges at certain points in time on a need-to-know basis imposed by the transmission times of the contact process.

4.2. *Exploration process.* Let

$$\mathcal{P} := \big\{\{(u,a),(u',a')\} : u,u' \in V_n, a,a' \in \{1,\dots,d\}, (u,a) \neq (u',a')\big\}$$

be the set of all potential edges of our random graph. A set $\mathcal{E} \subset \mathcal{P}$ is called *independent* if any two elements of $\mathcal{E}$ have no half-edge in common. Let

$$\mathscr{P} := \{\mathcal{E} \subset \mathcal{P} : \mathcal{E} \text{ is independent}\}. \tag{62}$$

In the same probability space where (G_t) and the graphical construction of the contact process are defined, we now define a process $(\mathscr{E}_t)_{t\geq 0}$ taking values in $\mathscr{P}$. Intuitively, $\mathscr{E}_t$ represents a set of edges that are known to be part of G_t, having been revealed by an exploration induced by the contact process activity. This process will have the following features:

(P1) It starts from $\mathscr{E}_0 = \varnothing$.
(P1) For any t, every element of $\mathscr{E}_t$ is an edge of G_t.
(P3) The pair $(\xi_t^{\bar{u}}, \mathscr{E}_t)_{t\geq 0}$ is a Markov chain.
(P4) For any t, conditionally on $(\xi_t^{\bar{u}}, \mathscr{E}_t)$, the distribution of the edges of G_t apart from those in $\mathscr{E}_t$ is uniform. More precisely, the pairing in G_t of the half-edges in the set $\{(u,a) : (u,a)$ not in any edge of $\mathscr{E}_t\}$ is uniformly distributed among all possibilities.

In order to define the exploration, we will need auxiliary times. Let T_1 be the first time in which there is a transmission mark at a half-edge emanating from $\bar{u}$; let $\mathscr{E}_t = \varnothing$ for all $t \in [0, T_1)$. In case $T_1 < \infty$, say, due to a transmission mark at the half-edge $(\bar{u}, a)$, we reveal the half-edge (v,b) to which $(\bar{u},a)$ is paired in G_{T_1} and include the edge $\{(\bar{u},a),(v,b)\}$ in $\mathscr{E}_{T_1}$.

Now, assume that we have already defined stopping times (with respect to the filtration of (G_t) and the graphical construction) $T_1 \leq T_2 \leq \cdots \leq T_k$ and that we have defined $\mathscr{E}_t$ for $0 \leq t \leq T_k$. In case $T_k = \infty$, set $T_{k+1} = \infty$; from now on, assume that $\{T_k < \infty\}$ occurs. Let T_{k+1} be the first time $t > T_k$ when:

- either a switch occurs involving at least one edge of $\mathscr{E}_{T_k}$ (call this *Case 1*),

- or "the contact process tries to use an unexplored edge," that is, a transmission mark appears at a half-edge emanating from some vertex of ξ_t, and this half-edge is not part of an edge of $\mathscr{E}_{T_k}$ (*Case 2*).

Note that Case 1 is triggered by a switch mark and Case 2 by a transmission mark; none of them is triggered by a recovery mark. We set $\mathscr{E}_t = \mathscr{E}_{T_k}$ for $t \in (T_k, T_{k+1})$, and $\mathscr{E}_{T_{k+1}}$ is defined as follows:

- In Case 1 there are two subcases. First, assume that the switch at time T_{k+1} involves an edge e of $\mathscr{E}_{T_k}$ and another edge outside $\mathscr{E}_{T_k}$. Then we let $\mathscr{E}_{T_{k+1}} = \mathscr{E}_{T_k} \setminus \{e\}$. Now, assume that the switch at time T_{k+1} involves two edges e, e' of $\mathscr{E}_{T_k}$, transforming them into the two new edges e'', e'''. We then set $\mathscr{E}_{T_{k+1}} = (\mathscr{E}_{T_k} \setminus \{e, e'\}) \cup \{e'', e'''\}$.
- In Case 2 we reveal the half-edge that is matched at time t to the half-edge having the transmission mark at that time; letting e be the edge formed by these two half-edges, we let $\mathscr{E}_{T_{k+1}} = \mathscr{E}_{T_k} \cup \{e\}$.

This completes the description of the exploration, and it should be clear that properties (P1), (P2), (P3), and (P4) listed earlier are indeed satisfied.

Our next step is to use the exploration process as a tool to couple the contact process $(\xi_t^{\bar{u}})$ with a herds process.

4.3. *A Markov chain lemma and its application.* We now prove a general result about coupling two continuous-time Markov chains.

LEMMA 4.1. *Let $\mathcal{X}_1$ and $\mathcal{X}_2$ be countable sets, and let $r_1 : \mathcal{X}_1 \times \mathcal{X}_1 \to [0, \infty)$ and $r_2 : \mathcal{X}_2 \times \mathcal{X}_2 \to [0, \infty)$ be functions defining the jump rates for continuous-time (nonexplosive) Markov chains on $\mathcal{X}_1$ and $\mathcal{X}_2$, respectively. Assume that there is a subset $\mathcal{X}_1' \subseteq \mathcal{X}_1$ and functions $\Psi : \mathcal{X}_1' \to \mathcal{X}_2$ and $f : \mathcal{X}_1' \to [0, \infty)$ such that the following two conditions hold*:

$$\sum_{y \in \mathcal{X}_1 \setminus \mathcal{X}_1'} r_1(x, y) \le f(x) \quad \text{for all } x \in \mathcal{X}_1' \tag{63}$$

and

$$\sum_{\substack{z \in \mathcal{X}_2, \\ z \neq \Psi(x)}} \left| r_2\big(\Psi(x), z\big) - \sum_{y \in \Psi^{-1}(z)} r_1(x, y) \right| \le f(x) \quad \text{for all } x \in \mathcal{X}_1'. \tag{64}$$

Fix $\bar{x} \in \mathcal{X}_1'$. Then there exists a coupling $(\mathcal{A}_t, \mathcal{B}_t)_{t \ge 0}$ on $\mathcal{X}_1 \times \mathcal{X}_2$ with the following properties:

(a) *$\mathcal{A}_0 = \bar{x}$, and $(\mathcal{A}_t)_{t \ge 0}$ is a Markov chain on $\mathcal{X}_1$ with jump rates r_1*;
(b) *$\mathcal{B}_0 = \Psi(\bar{x})$ and $(\mathcal{B}_t)_{t \ge 0}$ is a Markov chain on $\mathcal{X}_2$ with jump rates r_2*;
(c) *letting*

$$\sigma := \inf\big\{t : \mathcal{B}_t \neq \Psi(\mathcal{A}_t)\big\}$$

and, for any $a > 0$,

$$T_a := \inf\big\{t : f(\mathcal{A}_t) > a\big\},$$

we have, for any $t > 0$,

$$\mathbb{P}(\sigma \le t \wedge T_a) \le 2at.$$

PROOF. We will define a continuous-time Markov chain $(\mathcal{A}_t, \mathcal{B}_t, \mathcal{W}_t)_{t \ge 0}$ taking values in the set

$$\big\{(x, \Psi(x), 1) : x \in \mathcal{X}_1'\big\} \cup \big\{(x, y, 0) : x \in \mathcal{X}_1, y \in \mathcal{X}_2\big\}, \tag{65}$$

starting from $(\bar{x}, \Psi(\bar{x}), 1)$. The pair $(\mathcal{A}_t, \mathcal{B}_t)$ will satisfy the properties in the statement. The third coordinate process $(\mathcal{W}_t)$ will be a nondecreasing process (it jumps at most once, from 1 to 0) with the property that, for all $t \le \inf\{s : \mathcal{W}_s = 0\}$, we have $\mathcal{B}_t = \Psi(\mathcal{A}_t)$. So we interpret $\mathcal{W}_t$ as the indicator of the event that "the coupling still works at time t."

In order to define this chain, we need to specify the jump rates. When the third coordinate equals zero (meaning that the coupling is already broken), the first and second coordinates move independently, according to the chains defined by r_1 (on $\mathcal{X}_1$) and r_2 (on $\mathcal{X}_2$), respectively. More precisely, from any triple of the form $(x, y, 0)$, the chain jumps as follows:

- For each $x' \in \mathcal{X}_1$, it jumps to $(x', y, 0)$ with rate $r_1(x, x')$.
- For each $y' \in \mathcal{X}_2$, it jumps to $(x, y', 0)$ with rate $r_2(y, y')$.

We now need to specify the jump rates from points in the first set in the union in (65). In order to do so, we first introduce some notation. For each $x \in \mathcal{X}_1'$, we let

$$[x] := \Psi^{-1}\big(\Psi(x)\big) \subset \mathcal{X}_1'.$$

For each $x \in \mathcal{X}_1$ and each $S \subset \mathcal{X}_1$, we write

$$r_1(x, S) := \sum_{y \in S} r_1(x, y).$$

Now, fix $x \in \mathcal{X}_1'$. The following list describes all the possible jumps that the chain can take from $(x, \Psi(x), 1)$ (this starting location is kept fixed throughout the list), and their respective rates:

- *$\mathcal{A}$ and $\mathcal{B}$ jump together, stay coupled*: for each $y \in \mathcal{X}_1'\backslash[x]$, jump to $(y, \Psi(y), 1)$ with rate
$$r_1(x, y) \cdot \frac{r_1(x, [y]) \wedge r_2(\Psi(x), \Psi(y))}{r_1(x, [y])};$$
- *$\mathcal{A}$ jumps alone inside $[x]$, stay coupled*: for each $y \in [x]\backslash\{x\}$, jump to $(y, \Psi(y), 1) = (y, \Psi(x), 1)$ with rate $r_1(x, y)$;
- *$\mathcal{A}$ jumps alone leaving $\mathcal{X}_1'$, break coupling*: for each $y \in \mathcal{X}_1\backslash\mathcal{X}_1'$, jump to $(y, \Psi(x), 0)$ with rate $r_1(x, y)$;
- *$\mathcal{A}$ jumps alone inside $\mathcal{X}_1'$, break coupling*: for each $y \in \mathcal{X}_1'\backslash[x]$, jump to $(y, \Psi(x), 0)$ with rate
$$r_1(x, y) \cdot \left(1 - \frac{r_1(x, [y]) \wedge r_2(\Psi(x), \Psi(y))}{r_1(x, [y])}\right);$$
- *$\mathcal{B}$ jumps alone, breaks coupling*: for each $z \in \mathcal{X}_2$, jump to $(x, z, 0)$ with rate
$$r_2\big(\Psi(x), z\big) - \big(r_1\big(x, \Psi^{-1}(z)\big) \wedge r_2\big(\Psi(x), z\big)\big).$$

It is straightforward to check that the marginal rates for $(\mathcal{A}_t)$ and $(\mathcal{B}_t)$ are correct, so that items (a) and (b) in the statement of the lemma hold.

For each $x \in \mathcal{X}_1'$, let $\mathcal{R}(x)$ denote the rate at which $(\mathcal{A}_t, \mathcal{B}_t, \mathcal{W}_t)$ jumps from $(x, \Psi(x), 1)$ to the set $\mathcal{X}_1 \times \mathcal{X}_2 \times \{0\}$, where the coupling is broken. From the above rates, and then using (63) and (64), it can be seen that

$$\mathcal{R}(x) = r_1\big(x, \mathcal{X}_1\backslash\mathcal{X}_1'\big) + \sum_{\substack{z \in \mathcal{X}_2, \\ z \neq \Psi(x)}} \big|r_1\big(x, \Psi^{-1}(z)\big) - r_2\big(\Psi(x), z\big)\big| \le 2f(x). \tag{66}$$

Next, let $\sigma' := \inf\{t : \mathcal{W}_t = 0\}$, and recall that $T_a := \inf\{t : f(\mathcal{A}_t) > a\}$. The process

$$M_t := \mathbb{1}\big\{\sigma' \le t \wedge T_a\big\} - \int_0^{t \wedge \sigma' \wedge T_a} \mathcal{R}(\mathcal{A}_s)\,\mathrm{d}s, \quad t \ge 0 \tag{67}$$

is a martingale. Let us give an informal justification of this (a formal proof is not too difficult, but we omit it for brevity). We argue that

$$\mathbb{E}\big[M_{t+\epsilon}|(\mathcal{A}_s,\mathcal{B}_s,\mathcal{W}_s)_{0\le s\le t}\big]=M_t+o(\epsilon)\quad\text{as }\epsilon\to 0+, \tag{68}$$

which implies that $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbb{E}[M_{t+\epsilon}|(\mathcal{A}_s,\mathcal{B}_s,\mathcal{W}_s)_{0\le s\le t}]|_{\epsilon=0+}=0$. Assume we are given the information of the trajectory of the chain until time t. First, assume that neither σ' nor T_a have passed by time t. In that case the indicator in (67) is still equal to zero. The rate at which it may jump from zero to one in the immediate future is $\mathcal{R}(\mathcal{A}_t)$: this would happen if the coupling got broken, with $\mathcal{W}$ jumping from 1 to 0. This value $\mathcal{R}(\mathcal{A}_t)$ is the same as the rate with which the integral in (67) is increasing in this scenario, so the integral compensates the indicator function, and we obtain (68). Now, assume instead that either σ' or T_a has already passed by time t. In that case the indicator in (67) is no longer going to change (it is either going to stay equal to 0 or equal to 1 forever after t), and the integral is also not changing after t.

Now, for any $t\ge 0$,

$$\begin{aligned}0=M_0=\mathbb{E}[M_t]&=\mathbb{P}\big(\sigma'\le t\wedge T_a\big)-\mathbb{E}\bigg[\int_0^{t\wedge\sigma'\wedge T_a}\mathcal{R}(\mathcal{A}_s)\,\mathrm{d}s\bigg]\\&\overset{(66)}{\ge}\mathbb{P}\big(\sigma'\le t\wedge T_a\big)-2at.\end{aligned}$$

Now, recalling that $\sigma:=\inf\{t:\mathcal{B}_t\ne\Psi(\mathcal{A}_t)\}$, we have that $\sigma'\le\sigma$, so

$$\mathbb{P}(\sigma\le t\wedge T_a)\le\mathbb{P}\big(\sigma'\le t\wedge T_a\big)\le 2at.$$ □

In the application we have in mind for this lemma, the first Markov chain is the pair $(\xi_t,\mathscr{E}_t)$ consisting of the contact process and the exploration process in the random dynamic graph, as described in the previous subsection (recall that this pair is a Markov chain). The second Markov chain is a certain function of the herds process. We will need to give some definitions for both, as well as for the mapping Ψ between them.

4.3.1. *First Markov chain*: *Contact and exploration process*. The process $(\xi_t,\mathscr{E}_t)_{t\ge0}$ (contact process and exploration process on the dynamic random graph (G_t)) takes values in the state space

$$\mathcal{X}_1:=\big\{(A,\mathcal{E}):A\subset[n],\mathcal{E}\in\mathscr{P}\big\},$$

where we recall the definition of $\mathscr{P}$ in (62). We denote by $r_1(\cdot,\cdot)$ the function giving the jump rates of this chain.

Given $(A,\mathcal{E})\in\mathcal{X}_1$, we define the *graph induced by* $(A,\mathcal{E})$, denoted by $\mathrm{Graph}(A,\mathcal{E})$, as follows. First, enumerate $\mathcal{E}=\{e_1,\dots,e_m\}$, with

$$e_1=\big\{(u_1,a_1),(u_1',a_1')\big\},\qquad\dots,\qquad e_m=\big\{(u_m,a_m),(u_m',a_m')\big\}.$$

Then $\mathrm{Graph}(A,\mathcal{E})$ is the graph with vertex set $A\cup\{u_1,u_1',\dots,u_m,u_m'\}$ and edge set $\mathcal{E}$.

4.3.2. *Second Markov chain*: *Herds process modulo automorphisms*. Recall the definition of the set $P_{\mathrm{f}}(\mathbb{T}^d)$ of herd shapes from Definition 2.1. For $A\in P_{\mathrm{f}}(\mathbb{T}^d)$, define

$$[A]:=\left\{\begin{array}{l}A'\in P_{\mathrm{f}}(\mathbb{T}^d):\text{there is a graph isomorphism }\psi:\mathbb{T}^d\to\mathbb{T}^d\\\quad\text{such that }\psi(A)=A'\end{array}\right\}.$$

This decomposes $P_{\mathrm{f}}(\mathbb{T}^d)$ into equivalence classes.

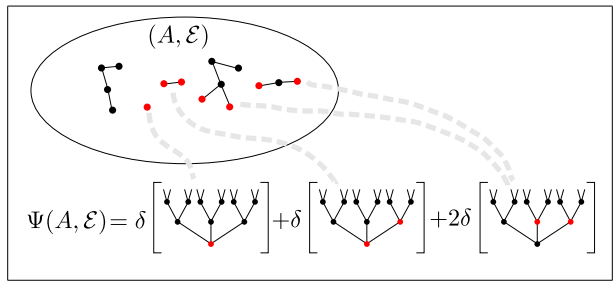

FIG. 2. *Illustration of the mapping* Ψ. *Above, the pair* $(A,\mathcal{E})$ *is depicted (red vertices are those that belong to* A, *and black vertices are those that belong to* $\mathrm{Graph}(A,\mathcal{E})$ *but not to* A). *Below, the herd configuration modulo automorphisms* $\Psi(A,\mathcal{E})$ *is shown. Note that one of the connected components of* $\mathrm{Graph}(A,\mathcal{E})$ *has no counterpart in* $\Psi(A,\mathcal{E})$, *because none of its vertices belongs to* A.

Recall the definition of the set $\mathcal{S}$ of herd configurations $\mathcal{S}$ from Definition 2.2. Given $\xi\in\mathcal{S}$, define $[\xi]:\{[A]:A\in P_{\mathrm{f}}(\mathbb{T}^d)\}\to\mathbb{N}_0$ by setting

$$[\xi]([A])=\sum_{A'\in[A]}\xi(A').$$

We then define

$$\mathcal{X}_2:=\{[\xi]:\xi\in\mathcal{S}\},$$

the *set of herd configurations modulo automorphisms*. Letting $(\Xi_t)_{t\geq0}$ be the herds process, we note that, by Lemma 2.4, the process $([\Xi_t])_{t\geq0}$ is a Markov chain on $\mathcal{X}_2$. We let $r_2(\cdot,\cdot)$ denote the function giving the jump rates of this chain.

4.3.3. *The mapping* Ψ *and the error bound* f. Now that we have defined the pairs $(\mathcal{X}_1, r_1)$ and $(\mathcal{X}_2, r_2)$ that we will use in our application of Lemma 4.1, we will also define the sets $\mathcal{X}_1'$ and the functions $\Psi:\mathcal{X}_1'\to\mathcal{X}_2$ and $f:\mathcal{X}_1'\to[0,\infty)$ that appear in the assumptions of that lemma.

We start with

$$\mathcal{X}_1':=\{(A,\mathcal{E})\in\mathcal{X}_1:\mathrm{Graph}(A,\mathcal{E})\text{ is a forest}\}. \tag{69}$$

The mapping Ψ is easy to understand (Figure 2 provides an instant explanation) but somewhat clumsy to define. Fix $(A,\mathcal{E})\in\mathcal{X}_1'$. Let $\mathscr{C}_1,\dots,\mathscr{C}_m$ be the connected components of $\mathrm{Graph}(A,\mathcal{E})$ that contain at least one vertex of A. For $i\in\{1,\dots,m\}$, let A_i be the set of vertices of A that intersect $\mathscr{C}_i$. Since $\mathscr{C}_i$ is a tree in which all vertices have degree at most d, there exists an isomorphism ψ_i between $\mathscr{C}_i$ and some connected subgraph of $\mathbb{T}^d$ (in fact, there are infinitely many such isomorphisms, but we choose one in some arbitrary way). Then $\xi(A,\mathcal{E}):=\sum_{i=1}^m\delta_{\psi_i(A_i)}$ is a herd configuration, and we let

$$\Psi(A,\mathcal{E}):=[\xi]\in\mathcal{X}_2.$$

It is now straightforward to verify that there exists a constant $C_f>0$ such that conditions (63) and (64) are satisfied with the choice

$$f(A,\mathcal{E}):=C_f\frac{(|A|+|\mathcal{E}|)^2}{n}.$$

We omit the details.

We now have all the ingredients to apply Lemma 4.1. Given an initial condition $(A,\mathcal{E})\in\mathcal{X}_1'$ for the exploration process (which will often, but not always, be equal to $(\{\bar{u}\},\varnothing)$), we can obtain the coupling $(\mathcal{A}_t,\mathcal{B}_t)_{t\geq0}$ started from $((A,\mathcal{E}),\Psi(A,\mathcal{E}))$ and satisfying the properties guaranteed by that lemma.

4.4. *Proof of Theorem 1.2.* For the rest of this section, we assume that $\lambda < \bar{\lambda}(\mathsf{v})$. By (61) it suffices to prove that, for $C > 0$ large enough, we have

$$n \cdot \mathbb{P}\big(\xi^{\bar{u}}_{C \log n} \neq \varnothing\big) \xrightarrow{n\to\infty} 0.$$

Let us explain our strategy to prove this. We take advantage of the coupling with the herds process from the previous section. The probability that the contact process, started from $\{\bar{u}\}$, survives until time $C \log n$ (with C some large constant), and moreover, the coupling remains good (meaning that $\mathcal{B}_t = \Psi(\mathcal{A}_t)$) for time $C' \log n$ (with $C' < C$ but still large) is $o(1/n)$, since this would imply survival of the herds process, which is subcritical, until $C' \log n$. However, the probability that the coupling turns bad before extinction is not $o(1/n)$, so we have to deal with that event. The most problematic case is that the coupling turns bad due to the exploration process finding an edge that causes the explored graph to no longer be a forest. In that case, apart from events of probability $o(1/n)$, this problematic edge is deleted after a short amount of time due to a switch (with no other problematic edges appearing in the meantime), and the explored region goes back to being a forest. At this moment when a forest reappears, we can start a brand new coupling between the exploration process (starting from its current state $(\xi_t, \mathscr{E}_t)$) and a herds process (starting from $\Psi(\xi_t, \mathscr{E}_t)$). Now, this second coupling also turning bad has too low probability (when we consider this event together with the already low probability of the breaking of the first attempt). It is also unlikely that this second coupling stays active for a long time without turning bad, for the same reason as for the first one.

The above explanation shows that the argument is naturally structured in three stages (of course, not all of them necessarily occur): the first coupling attempt, then the period until a problematic edge is removed, and then the second coupling attempt. We encapsulate Stages 2 and 3 in two lemmas in reverse order: Lemma 4.2 below deals with Stage 3, and Lemma 4.3 with Stage 2. Having these two lemmas in place, we are able to tell the full story from the beginning of Stage 1, concluding the proof.

LEMMA 4.2. *Let $(A, \mathcal{E}) \in \mathcal{X}_1'$, where $\mathcal{X}_1'$ is as in (69). Assume that $|A| + |\mathcal{E}| \le n^{1/6}$. Let $(\xi_t, \mathscr{E}_t)_{t\ge 0}$ be a contact process and exploration process started from $(\xi_0, \mathscr{E}_0) = (A, \mathcal{E})$. Then, letting τ denote the extinction time of the contact process and $\varphi = \varphi(\lambda, \mathsf{v})$ be the growth index of the herds process (as in (12)), for n large enough we have*

$$\mathbb{P}\left(\tau > \frac{2}{|\log \varphi|} \log n\right) \le \frac{1}{\sqrt{n}}. \tag{70}$$

PROOF. Let $(\mathcal{A}_t, \mathcal{B}_t)_{t\ge 0}$ be the coupling obtained from Lemma 4.1, started from $(\mathcal{A}_0, \mathcal{B}_0) = ((A, \mathcal{E}), \Psi(A, \mathcal{E}))$. Recalling from the statement of Lemma 4.1 that

$$\sigma := \inf\{t : \mathcal{B}_t \neq \Psi(\mathcal{A}_t)\}, \qquad T_a := \inf\{t : f(\mathcal{A}_t) \ge a\}$$

and abbreviating

$$t_* := \frac{2 \log n}{|\log \varphi|}, \qquad a_* := \frac{1}{4 t_* \sqrt{n}} = \frac{|\log \varphi|}{8 \sqrt{n} \log n},$$

we bound the probability on the left-hand side of (70) by

$$\mathbb{P}(\tau > t_*, \sigma > t_*) + \mathbb{P}(\sigma \le t_* \wedge T_{a_*}) + \mathbb{P}(T_{a_*} < \sigma \le t_*). \tag{71}$$

By Lemma 4.1 we have

$$\mathbb{P}(\sigma \le t_* \wedge T_{a_*}) \le 2 t_* a_* = \frac{1}{2\sqrt{n}}.$$

By the definition of σ, on the event $\{\tau > t_*, \sigma > t_*\}$ we have that $\mathcal{B}_{t_*}$ is not empty. Then $\mathbb{P}(\tau > t_*, \sigma > t_*)$ is smaller than the probability that a herds process started with fewer than $n^{1/6}$ particles is still alive by time t_*. By (6) and (17), we obtain

$$\mathbb{P}(\tau > t_*, \sigma > t_*) \le Cn^{1/6} \cdot \varphi^{t_*} = Cn^{1/6} \cdot n^{-2} < \frac{1}{4\sqrt{n}} \tag{72}$$

if n is large enough.

It remains to bound $\mathbb{P}(T_{a_*} < \sigma \le t_*)$. Recalling that $f(\mathcal{A}_t) = C_f(|\xi_t| + |\mathscr{E}_t|)^2/n$, if $T_{a_*} < \infty$, we have

$$|\xi_{T_{a_*}}| + |\mathscr{E}_{T_{a_*}}| \ge (na_*/C_f)^{1/2}.$$

Since $|\xi_0| + |\mathscr{E}_0| \le n^{1/6}$, we obtain that, for n large enough,

$$\text{on } \{T_{a_*} < \infty\}, \quad |\xi_{T_{a_*}}| + |\mathscr{E}_{T_{a_*}}| - \big(|\xi_0| + |\mathscr{E}_0|\big) \ge (na_*/C_f)^{1/2} - n^{1/6} > n^{1/5}.$$

Now, the process $(|\xi_t| + |\mathscr{E}_t|)$ only changes at times when $(\mathcal{A}_t) = (\xi_t, \mathscr{E}_t)$ changes. If $\mathcal{A}_t$ has a new infection appearing at time t, then $|\xi_t| + |\mathscr{E}_t|$ may increase by at most 2 at that time. On the other hand, if $\mathcal{A}_t$ performs a jump of any other kind, then $|\xi_t| + |\mathscr{E}_t|$ stays the same or decreases. Hence, for any t, we have

$$|\xi_t| + |\mathscr{E}_t| - \big(|\xi_0| + |\mathscr{E}_0|\big) \le 2\big|\{s \le t : \xi_s = \xi_{s-} + 1\}\big|.$$

Putting these observations together, we see that

$$\text{on } \{T_{a_*} < \sigma \le t_*\}, \quad \big|\{s < \sigma : \xi_s = \xi_{s-} + 1\}\big| \ge n^{1/5}/2.$$

Moreover, before time σ, whenever a new infection appears in $(\mathcal{A}_t)$, a new particle is also born in $(\mathcal{B}_t)$. Letting $\mathcal{N}_\infty$ denote the number of particles ever born in $(\mathcal{B}_t)$ (even after time σ), we obtain the bound

$$\mathbb{P}(T_{a_*} < \sigma \le t_*) \le \mathbb{P}\big(\mathcal{N}_\infty \ge n^{1/5}/2\big) \le \frac{\mathbb{E}[\mathcal{N}_\infty^p]}{(n^{1/5}/2)^p},$$

for any $p \ge 1$, by Markov's inequality. Recalling that $\mathcal{B}_0$ has at most $n^{1/6}$ particles and using Lemma 3.14 and Corollary 3.15, the right-hand side is smaller than

$$\frac{C_p(n^{1/6})^p}{(n^{1/5}/2)^p} = C'_p \cdot n^{-p/30},$$

for some constants $C_p, C'_p > 0$. Taking $p > 15$ and then n large enough, this is smaller than $\frac{1}{4\sqrt{n}}$, completing the proof. □

LEMMA 4.3. *There exist $\varepsilon > 0$ and $\delta > 0$ such that the following holds. Let $(A, \mathcal{E}) \in \mathcal{X}_1 \backslash \mathcal{X}'_1$ (so that* Graph$(A, \mathcal{E})$ *is not a forest). Assume that $|A| + |\mathcal{E}| \le n^\varepsilon$. Also, assume that there is an edge e in $\mathcal{E}$ such that $(A, \mathcal{E}\backslash\{e\}) \in \mathcal{X}'_1$, that is,* Graph$(A, \mathcal{E})$ *would become a forest if e were removed from $\mathcal{E}$. Letting $(\xi_t, \mathscr{E}_t)_{t\ge 0}$ denote the contact and exploration process started from $(A, \mathcal{E})$ and letting τ denote the extinction time of (ξ_t), for n large enough we have*

$$\mathbb{P}\left(\tau > \left(\delta + \frac{2}{|\log\varphi|}\right)\log n\right) < n^{-4\varepsilon}. \tag{73}$$

PROOF. Let U denote the time when the edge e disappears, due to being involved in a switch. The rate at which this happens equals $\upsilon_n = \frac{\mathsf{v}}{nd}$ times the number of other edges in the graph, which is $\frac{nd}{2} - 1$, times 2 (switches can be positive or negative). So this rate is $\mathsf{v}(1 - \frac{2}{nd})$. Hence, U has exponential distribution with parameter $\mathsf{v}(1 - \frac{1}{2nd})$.

Denote by B the event that:

- For all $t \in [0, U)$ we have $(\xi_t, \mathscr{E}_t \setminus \{e\}) \in \mathcal{X}_1'$; that is, it stays the case that the removal of e from the set of edges turns $\mathrm{Graph}(\xi_t, \mathscr{E}_t)$ into a forest.
- $\mathrm{Graph}(\xi_U, \mathscr{E}_U)$ is a forest.

We fix $\delta > 0$ and $\varepsilon > 0$ for now; their values will be chosen at the end of the proof. We define

$$T := \inf\{t : |\xi_t| + |\mathscr{E}_t| > ln^{1/6}\}.$$

We bound the probability in (73) by

$$\mathbb{P}(U > \delta \log n) \tag{74}$$

$$+ \mathbb{P}(U \le \delta \log n, T \le U) \tag{75}$$

$$+ \mathbb{P}(\{U \le \delta \log n, T > U\} \cap B^c) \tag{76}$$

$$+ \mathbb{P}\left(\{U \le \delta \log n, T > U\} \cap B \cap \left\{\tau > \left(\delta + \frac{2}{|\log \varphi|}\right) \log n\right\}\right). \tag{77}$$

We bound these four terms separately, starting with the first and last, which are the easiest. We have

$$\mathbb{P}(U > \delta \log n) = \exp\left\{-\mathsf{v}\left(1 - \frac{1}{2nd}\right)\delta \log n\right\} = n^{-\mathsf{v}(1-\frac{1}{2nd})\delta}.$$

Next, letting $B' := \{U \le \delta \log n, T > U\} \cap B$ and letting $(\mathcal{F}_t)_{t \ge 0}$ be the natural filtration of $(\xi_t, \mathscr{E}_t)$, the probability in (77) is

$$\begin{aligned}&\mathbb{P}\left(B' \cap \left\{\tau > \left(\delta + \frac{2}{|\log \varphi|}\right) \log n\right\}\right)\\ &\quad = \mathbb{E}\left[\mathbb{1}_{B'} \cdot \mathbb{E}\left[\tau > \left(\delta + \frac{2}{|\log \varphi|}\right) \log n | \mathcal{F}_U\right]\right] \le \mathbb{P}(B') \cdot \frac{1}{\sqrt{n}} \le \frac{1}{\sqrt{n}},\end{aligned}$$

where the first inequality follows from Lemma 4.2.

To bound (75), we note that $|\xi_t| + |\mathscr{E}_t|$ can increase by at most two units at a given jump time, and a jump that causes such an increase happens with rate at most $\lambda d |\xi_t|$. We can thus stochastically dominate $(|\xi_t| + |\mathscr{E}_t|)_{t \ge 0}$ by a pure-birth process $(Z_t)_{t \ge 0}$ on $\mathbb{N}$, which starts from $Z_0 = \lceil n^\varepsilon \rceil$ and jumps from k to $k+2$ with rate $\lambda d k$ (and has no other kind of jump). Then the probability in (75) is smaller than

$$\begin{aligned}&\mathbb{P}\Big(\max_{0 \le t \le \delta \log n} (|\xi_t| + |\mathscr{E}_t|) > n^{1/6}\Big)\\ &\quad \le \mathbb{P}\big(Z_{\delta \log n} > n^{1/6}\big) \le \frac{\mathbb{E}[Z_{\delta \log n}]}{n^{1/6}} = \frac{\lceil n^\varepsilon \rceil \cdot \exp\{2\lambda d \delta \log n\}}{n^{1/6}} \le 2n^{\varepsilon + 2\lambda d \delta - \frac{1}{6}}.\end{aligned}$$

We now turn to (76). For the process $(\xi_t, \mathscr{E}_t)_{t \ge 0}$, let us say that a "bad jump" is a jump time when either: (a) a contact process transmission occurs which causes the inclusion in the exploration process of an edge between two vertices that were already present in $\mathrm{Graph}(\xi_t, \mathscr{E}_t)$ or (b) a switch involving two edges that were already in $\mathscr{E}_t$. The point is that, as long as there are no bad jumps, it remains true that $\mathrm{Graph}(\xi_t, \mathscr{E}_t)$ would become a forest if e were removed, and in particular, if U is not a bad jump, then $\mathrm{Graph}(\xi_U, \mathscr{E}_U)$ is a forest. Hence, letting S denote the time at which the first bad jump occurs, we have

$$\mathbb{P}\big(\{U \le \delta \log n, T > U\} \cap B^c\big) \le \mathbb{P}(U \le \delta \log n, T > U, S \le U).$$

Now, let $\mathcal{R}_t$ denote the rate at which a bad jump occurs from the present state $(\xi_t, \mathscr{E}_t)$. It is straightforward to check that there is $C > 0$ such that $\mathcal{R}_t \le C\frac{(|\xi_t|+|\mathscr{E}_t|)^2}{n}$. Moreover, the process

$$Y_t := \mathbb{1}\{S \le t, S \le T\} - \int_0^{t\wedge S\wedge T} \mathcal{R}_s \,\mathrm{d}s, \quad t \ge 0$$

is a martingale. For a sketch of proof of this, fix $t \ge 0$. On the event $\{S \wedge T \le t\}$, for every $t' \ge t$, we have $\mathbb{1}\{S \le t', S \le T\} = \mathbb{1}\{S \le t, S \le T\}$ and $\int_{t\wedge S\wedge T}^{t'\wedge S\wedge T} \mathcal{R}_u \,\mathrm{d}u = 0$, so $Y_{t'} = Y_t$ almost surely, so $\mathbb{E}[Y_{t'}|(\xi_s, \mathscr{E}_s)_{0\le s\le t}] = Y_t$. On the other hand, on the event $\{S \wedge T > t\}$, we have $\mathbb{1}\{S \le t, S \le T\} = 0$, and the rate at which this indicator jumps to one from the current configuration at time t is $\mathcal{R}_t$; this is exactly compensated by the increase in the integral, showing that $\mathbb{E}[Y_{t+r}|(\xi_s, \mathscr{E}_s)_{0\le s\le t}] - Y_t = o(r)$ when $r \to 0+$. This shows that, in all cases, $\frac{\mathrm{d}}{\mathrm{d}r}\mathbb{E}[Y_{t+r}|(\xi_s, \mathscr{E}_s)_{0\le s\le t}]|_{r=0+} = 0$.

Then

$$0 = \mathbb{E}[Y_0] = \mathbb{E}[Y_{\delta\log n}] \ge \mathbb{P}(S \le \delta\log n, S \le T) - \delta\log n \cdot C\frac{(n^{1/6})^2}{n}.$$

Then we have

$$\mathbb{P}(U \le \delta\log n, T > U, S \le U) \le \mathbb{P}(S \le \delta\log n, S \le T) \le C\delta\log n \cdot n^{-2/3}.$$

Putting now all our bounds together, we have proved that the probability in (73) is smaller than

$$n^{-\mathrm{v}(1-\frac{1}{2nd})\delta} + n^{-1/2} + 2n^{\varepsilon+2\lambda d\delta-\frac{1}{6}} + C\delta\log n \cdot n^{-2/3}.$$

By first choosing δ small and then choosing ε much smaller, this expression is smaller than $n^{-4\varepsilon}$ when n is large enough. □

PROOF OF THEOREM 1.2. Let ε and δ be as in Lemma 4.3. Define

$$\beta := \frac{4}{|\log\varphi|} + \delta.$$

We will prove that

$$\mathbb{P}\big(\xi_{\beta\log n}^{V_n} \ne \varnothing\big) \xrightarrow{n\to\infty} 0.$$

By (61) this will follow from proving that

$$\lim_{n\to\infty} n \cdot \mathbb{P}\big(\xi_{\beta\log n}^{\bar{u}} \ne \varnothing\big) = 0,$$

where $\bar{u}$ is a deterministic vertex. In order to prove this, we take the coupling $(\mathcal{A}_t, \mathcal{B}_t)_{t\ge0}$ from Lemma 4.1, with $\mathcal{A}_0 = (\xi_0, \mathscr{E}_0) = (\{\bar{u}\}, \varnothing)$ and $\mathcal{B}_0 = \Psi(\mathcal{A}_0)$. Recall that $\sigma := \inf\{t : \mathcal{B}_t \ne \Psi(\mathcal{A}_t)\}$.

Let τ be the extinction time of (ξ_t) (which starts from $\{\bar{u}\}$), and also define

$$\beta_0 := \frac{2}{|\log\varphi|}$$

and

$$T := \inf\big\{t : |\xi_t| + |\mathscr{E}_t| \ge n^{\varepsilon}\big\}.$$

We bound

$$\mathbb{P}(\tau > \beta \log n)$$

$$\leq \mathbb{P}(\tau > \beta \log n, \sigma > \beta_0 \log n) \tag{78}$$

$$+ \mathbb{P}(\sigma \leq \beta_0 \log n, T \leq \sigma) \tag{79}$$

$$+ \mathbb{P}(\tau > \beta \log n, \sigma \leq \beta_0 \log n, T > \sigma). \tag{80}$$

To bound (78), we give the same argument as we have used to bound (72); here it gives

$$\mathbb{P}(\tau > \beta \log n, \sigma > \beta_0 \log n) \leq C\varphi^{\beta_0 \log n} < n^{-2}.$$

To bound (79), we observe again that $|\xi_t| + |\mathscr{E}_t|$ can only increase by 2 at any given jump time, and it only increases when there are contact transmissions generating new births. Moreover, before time σ, any time when there is a birth for $(\xi_t, \mathscr{E}_t)$, there is also a particle birth for $\mathscr{B}_t$. These considerations allow us to bound

$$\begin{aligned}
&\mathbb{P}(\sigma \leq \beta_0 \log n, T \leq \sigma) \\
&\leq \mathbb{P}(\sigma \leq \beta_0 \log n, |\{t \leq \sigma : |\xi_t| = |\xi_{t-}| + 1\}| \geq n^\varepsilon/2) \\
&\leq \mathbb{P}\Big(\sigma \leq \beta_0 \log n, |\{t < \sigma : |\xi_t| = |\xi_{t-}| + 1\}| \geq \frac{n^\varepsilon}{2} - 2\Big) \\
&\leq \mathbb{P}\Big((\mathscr{B}_t) \text{ has more than } \frac{n^\varepsilon}{2} - 2 \text{ births}\Big).
\end{aligned}$$

By Corollary 3.15 this is smaller than

$$C_p\Big(\frac{n^\varepsilon}{2} - 2\Big)^{-p}$$

for any $p \geq 1$. Hence, by taking $p > 1/(2\varepsilon)$ and n large enough, it is smaller than $1/n^2$.

We now turn to (80). Let us first bound, using Lemma 4.1,

$$\mathbb{P}(\sigma \leq \beta_0 \log n, T > \sigma) \leq C_f \cdot \frac{(n^\varepsilon)^2}{n} \cdot \beta_0 \log n = C_f \beta_0 n^{-1+2\varepsilon} \log n. \tag{81}$$

Letting $(\mathcal{F}_t)_{t\geq 0}$ be the natural filtration for $(\mathcal{A}_t, \mathcal{B}_t)_{t\geq 0}$, we write

$$\begin{aligned}
&\mathbb{P}(\tau > \beta \log n, \sigma \leq \beta_0 \log n, T > \sigma) \\
&\leq \mathbb{E}\big[\mathbb{1}\{\sigma \leq \beta_0 \log n, T > \sigma\} \cdot \mathbb{P}(\tau > \beta \log n | \mathcal{F}_\sigma)\big] \\
&= \mathbb{E}\big[\mathbb{1}\{\sigma \leq \beta_0 \log n, T > \sigma, \mathcal{A}_\sigma \notin \mathcal{X}_1'\} \cdot \mathbb{P}(\tau > \beta \log n | \mathcal{F}_\sigma)\big] \qquad (82) \\
&\quad + \mathbb{E}\big[\mathbb{1}\{\sigma \leq \beta_0 \log n, T > \sigma, \mathcal{A}_\sigma \in \mathcal{X}_1'\} \cdot \mathbb{P}(\tau > \beta \log n | \mathcal{F}_\sigma)\big]. \qquad (83)
\end{aligned}$$

To bound (82), we note that, on the event $\{\sigma \leq \beta_0 \log n, T > \sigma, \mathcal{A}_\sigma \notin \mathcal{X}_1'\}$, we have that $\mathcal{A}_\sigma = (\xi_\sigma, \mathscr{E}_\sigma)$ satisfies the assumptions of Lemma 4.3, and then that lemma implies that, on this event, $\mathbb{P}(\tau > \beta \log n | \mathcal{F}_\sigma) \leq n^{-4\varepsilon}$. Together with (81), this implies that (82) is smaller than

$$C_f \beta_0 n^{-1+2\varepsilon} \log n \cdot n^{-4\varepsilon} < n^{-1-\varepsilon}$$

if n is large enough.

Next, Lemma 4.2 implies that on $\{\sigma \leq \beta_0 \log n, T > \sigma, \mathcal{A}_\sigma \in \mathcal{X}_1'\}$, we have $\mathbb{P}(\tau > \beta \log n | \mathcal{F}_\sigma) < n^{-1/2}$; combining this with (81) shows that (83) is smaller than

$$C_f \beta_0 n^{-1+2\varepsilon} \log n \cdot n^{-1/2} < n^{-5/4}$$

for n large. This completes the proof. □

APPENDIX: PROOFS OF LEMMA 3.9 AND LEMMA 3.11

PROOF OF LEMMA 3.9. The proof is the same for the two functions, so we only treat the first. We use the simple bounds that come from comparison with a pure birth process,

$$\mathbb{E}[X_t|\Xi_0 = \delta_{A_{e,1}} + \delta_{A_{e,2}}] \le |A| \cdot e^{d\lambda t},$$

together with the expression (20), to obtain

$$\begin{aligned} g_{\mathsf{v}}(\xi, t) &\le e^{d\lambda t} \cdot \sum_A \xi(A) \cdot |A| \cdot \big|\{\text{active edges of } A\}\big| \\ &\le e^{d\lambda t} \cdot \Big(\sum_A \xi(A) \cdot |A|\Big)\Big(\sum_A \xi(A) \cdot \big|\{\text{active edges of } A\}\big|\Big) \\ &\le e^{d\lambda t} \cdot \big(X(\xi) + \mathscr{E}(\xi)\big)^2. \end{aligned}$$

The statement now readily follows from Corollary 2.3. □

The following preliminary result will allow us to perform the exchange of limit and expectation in (34).

LEMMA A.4. *There exist* $c_1, c_2 > 0$ (*depending on* λ, v, ε) *such that, for any* $0 \le t < t + s < T$, *on* $\{\tau_{\text{sep}} > t\}$, *we have*

$$\big|\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]\big| \le c_1\big(X(\mathcal{V}_t) + \mathscr{E}(\mathcal{V}_t)\big)^{c_2} \cdot s.$$

PROOF. We fix s, t, as in the statement, and let N denote the number of jumps of the process $(\mathcal{V}, \mathcal{W})$ in $[t, t+s]$. We have

$$\begin{aligned} &\mathbb{1}_{\{\tau_{\text{sep}} > t\}} \cdot \big|\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]\big| \\ &\quad \le \mathbb{1}_{\{\tau_{\text{sep}} > t\}} \cdot \big|\widehat{\mathbb{E}}\big[\mathbb{1}_{\{\tau_{\text{sep}} \le t+s, N=1\}} \cdot \widehat{\mathbb{E}}[\mathcal{Y}_T - \mathcal{X}_T|\mathcal{F}_{\tau_{\text{sep}}}]|\mathcal{F}_t\big]\big| \qquad (84) \\ &\qquad + \mathbb{1}_{\{\tau_{\text{sep}} > t\}} \cdot \big|\widehat{\mathbb{E}}\big[\mathbb{1}_{\{\tau_{\text{sep}} \le t+s, N\ge 2\}} \cdot \widehat{\mathbb{E}}[\mathcal{Y}_T - \mathcal{X}_T|\mathcal{F}_{\tau_{\text{sep}}}]|\mathcal{F}_t\big]\big|. \qquad (85) \end{aligned}$$

We will treat the terms (84) and (85) separately. For both, it will be useful to bound, using domination by a pure birth process,

$$\widehat{\mathbb{E}}[\mathcal{X}_T|\mathcal{F}_{\tau_{\text{sep}}}] \le e^{d\lambda T} X(\mathcal{V}_{\tau_{\text{sep}}}), \qquad \widehat{\mathbb{E}}[\mathcal{Y}_T|\mathcal{F}_{\tau_{\text{sep}}}] \le e^{d\lambda T} X(\mathcal{W}_{\tau_{\text{sep}}}) = e^{d\lambda T} X(\mathcal{V}_{\tau_{\text{sep}}}),$$

which gives

$$\big|\widehat{\mathbb{E}}[\mathcal{Y}_T - \mathcal{X}_T|\mathcal{F}_{\tau_{\text{sep}}}]\big| \le e^{d\lambda T} X(\mathcal{V}_{\tau_{\text{sep}}}). \qquad (86)$$

We start bounding (84). On the event $\{t < \tau_{\text{sep}} \le t+s, N = 1\}$, we have $X(\mathcal{V}_t) = X(\mathcal{W}_t) = X(\mathcal{V}_{\tau_{\text{sep}}}) = X(\mathcal{W}_{\tau_{\text{sep}}})$ (since in this event the only jump of $(\mathcal{V}, \mathcal{W})$ in $[t, t+s]$ is a split which causes the separation of the processes, so there is no change in the number of particles). Then, using (86), we obtain that (84) is smaller than

$$\mathbb{1}_{\{\tau_{\text{sep}} > t\}} \cdot e^{d\lambda T} X(\mathcal{V}_t) \cdot \widehat{\mathbb{P}}(\tau_{\text{sep}} \le t + s, N \ge 1|\mathcal{F}_t).$$

The rate with which the process jumps away from the state $(\mathcal{V}_t, \mathcal{W}_t)$ is at most

$$\mu_1(\mathcal{V}_t) := (d\lambda + 1)X(\mathcal{V}_t) + (\mathsf{v} + \varepsilon)\mathscr{E}(\mathcal{V}_t),$$

so the probability above is at most

$$1 - \exp\{-\mu_1(\mathcal{V}_t) \cdot s\} \le \mu_1(\mathcal{V}_t) \cdot s.$$

We have thus proved that (84) is bounded by the desired expression.

We now turn to (85). We start using (86) to bound (85) by

$$\begin{aligned}&\mathbb{1}_{\{\tau_{\mathrm{sep}}>t\}}\cdot e^{d\lambda T}\cdot\widehat{\mathbb{E}}\big[\mathbb{1}_{\{\tau_{\mathrm{sep}}\le t+s,N\ge 2\}}\cdot X(\mathcal{V}_{\tau_{\mathrm{sep}}})|\mathcal{F}_t\big]\\&\quad\le\mathbb{1}_{\{\tau_{\mathrm{sep}}>t\}}\cdot e^{d\lambda T}\cdot\widehat{\mathbb{E}}\Big[\mathbb{1}_{\{N\ge 2\}}\cdot\max_{u\in[t,T]}X(\mathcal{V}_u)|\mathcal{F}_t\Big].\end{aligned}$$

By Hölder's inequality the expectation on the right-hand side is smaller than

$$\mathbb{1}_{\{\tau_{\mathrm{sep}}>t\}}\cdot e^{d\lambda T}\cdot\widehat{\mathbb{P}}(N\ge 2|\mathcal{F}_t)^{2/3}\cdot\widehat{\mathbb{E}}\Big[\max_{u\in[t,T]}X(\mathcal{V}_u)^3|\mathcal{F}_t\Big]^{1/3}.\tag{87}$$

Corollary 2.3 and the Markov property imply that

$$\widehat{\mathbb{E}}\Big[\max_{u\in[t,T]}X(\mathcal{V}_u)^3|\mathcal{F}_t\Big]^{1/3}\le cX(\mathcal{V}_t).\tag{88}$$

Let

$$\mu_2(\mathcal{V}_t):=2(d\lambda+1)\big(X(\mathcal{V}_t)+1\big)+2(\mathsf{v}+\varepsilon)\big(\mathscr{E}(\mathcal{V}_t)+1\big).$$

In the event $\{\tau_{\mathrm{sep}}>t\}$, the process $(\mathcal{V},\mathcal{W})$ jumps away from $(\mathcal{V}_t,\mathcal{W}_t)$ with rate smaller than $\mu_1(\mathcal{V}_t)$ (as previously observed), and after performing a first jump, it performs a second jump with a rate that is smaller than $\mu_2(\mathcal{V}_t)$. Indeed, after the first jump, the number of particles or active edges of $\mathcal{V}$ and $\mathcal{W}$ increase by at most 1, and the factors 2 in the definition of $\mu_2(\mathcal{V}_t)$ account for the possibility that the first jump is the separation of the two processes. Letting $(Z_t)_{t\ge0}$ be a Poisson process with constant rate $\mu_2(\mathcal{V}_t)$, we bound (on the event $\{\tau_{\mathrm{sep}}>t\}$),

$$\begin{aligned}\widehat{\mathbb{P}}(N\ge 2|\mathcal{F}_t)\le\mathbb{P}(Z_s\ge 2)&=1-\mathrm{e}^{-\mu_2(\mathcal{V}_t)s}-\mu_2(\mathcal{V}_t)s\cdot\mathrm{e}^{-\mu_2(\mathcal{V}_t)s}\\&\le\mu_2(\mathcal{V}_t)s-\mu_2(\mathcal{V}_t)s\cdot\mathrm{e}^{-\mu_2(\mathcal{V}_t)s}\\&\le\big(\mu_2(\mathcal{V}_t)s\big)^2.\end{aligned}$$

Plugging this bound and (88) back in (87) gives the desired bound. □

PROOF OF LEMMA 3.11. Fix $t\in[0,T)$. For any $s\in(0,T-t]$, we have

$$\mathbb{1}\{\tau_{\mathrm{sep}}>t\}\cdot\frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]}{s}\le\mathbb{1}\{\tau_{\mathrm{sep}}>t\}\cdot c_1\big(X(\mathcal{V}_t)+\mathscr{E}(\mathcal{V}_t)\big)^{c_2}.$$

The random variable on the right-hand side is integrable by Corollary 2.3. This and the dominated convergence theorem justify the exchange of limit in (34). As explained before (34), this implies that

$$\begin{aligned}\lim_{s\to0+}\frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}]-\widehat{\mathbb{E}}[\mathcal{A}_t]}{s}&=\widehat{\mathbb{E}}\Big[\mathbb{1}\{\tau_{\mathrm{sep}}>t\}\cdot\lim_{s\to0+}\frac{\widehat{\mathbb{E}}[\mathcal{A}_{t+s}|\mathcal{F}_t]}{s}\Big]\\&=\varepsilon\cdot\widehat{\mathbb{E}}\big[\mathbb{1}\{\tau_{\mathrm{sep}}>t\}\cdot g_{\mathsf{v},\varepsilon}(\mathcal{V}_t,T-t)\big].\end{aligned}$$

Finally, any function $g:[0,\infty)\to\mathbb{R}$ which is continuous and has a continuous derivative from the right is necessarily differentiable (with derivative equal to the derivative from the right), so the proof is complete. □

REFERENCES

[1] ATHREYA, K. B. and NEY, P. E. (2004). *Branching Processes*. Dover, Mineola, NY. MR2047480

[2] BERGER, N., BORGS, C., CHAYES, J. T. and SABERI, A. (2005). On the spread of viruses on the Internet. In *Proceedings of the Sixteenth Annual ACM–SIAM Symposium on Discrete Algorithms* 301–310. ACM, New York. MR2298278

[3] BHAMIDI, S., NAM, D., NGUYEN, O. and SLY, A. (2021). Survival and extinction of epidemics on random graphs with general degree. *Ann. Probab.* **49** 244–286. MR4203338 https://doi.org/10.1214/20-AOP1451

[4] BOLLOBÁS, B. and BOLLOBÁS, B. (1998). *Random Graphs*. Springer, Berlin.

[5] CAN, V. H. (2017). Metastability for the contact process on the preferential attachment graph. *Internet Math.* 45. MR3683430

[6] CAN, V. H. (2019). Exponential extinction time of the contact process on rank-one inhomogeneous random graphs. *J. Theoret. Probab.* **32** 106–130. MR3908908 https://doi.org/10.1007/s10959-017-0786-9

[7] CAN, V. H. and SCHAPIRA, B. (2015). Metastability for the contact process on the configuration model with infinite mean degree. *Electron. J. Probab.* **20** no. 26, 22. MR3325096 https://doi.org/10.1214/EJP.v20-3859

[8] CASSANDRO, M., GALVES, A., OLIVIERI, E. and VARES, M. E. (1984). Metastable behavior of stochastic dynamics: A pathwise approach. *J. Stat. Phys.* **35** 603–634. MR0749840 https://doi.org/10.1007/BF01010826

[9] CHATTERJEE, S. and DURRETT, R. (2009). Contact processes on random graphs with power law degree distributions have critical value 0. *Ann. Probab.* **37** 2332–2356. MR2573560 https://doi.org/10.1214/09-AOP471

[10] COOPER, C., DYER, M. and GREENHILL, C. (2007). Sampling regular graphs and a peer-to-peer network. *Combin. Probab. Comput.* **16** 557–593. MR2334585 https://doi.org/10.1017/S0963548306007978

[11] CRANSTON, M., MOUNTFORD, T., MOURRAT, J.-C. and VALESIN, D. (2014). The contact process on finite homogeneous trees revisited. *ALEA Lat. Am. J. Probab. Math. Stat.* **11** 385–408. MR3249416

[12] DA SILVA, G. L. B., OLIVEIRA, R. I. and VALESIN, D. (2021). The contact process over a dynamical d-regular graph. To appear in *Ann. I.H.P. (B)*. arXiv preprint. Available at arXiv:2111.11757.

[13] DORT, L. and JACOB, E. (2023). Local weak limit of dynamical inhomogeneous random graphs. Available at arXiv:2303.17437.

[14] DURRETT, R. and LIU, X. F. (1988). The contact process on a finite set. *Ann. Probab.* **16** 1158–1173. MR0942760

[15] DURRETT, R. and SCHONMANN, R. H. (1988). The contact process on a finite set. II. *Ann. Probab.* **16** 1570–1583. MR0958203

[16] DURRETT, R., SCHONMANN, R. H. and TANAKA, N. I. (1989). The contact process on a finite set. III. The critical case. *Ann. Probab.* **17** 1303–1321. MR1048928

[17] HARRIS, S. C. and ROBERTS, M. I. (2017). The many-to-few lemma and multiple spines. *Ann. Inst. Henri Poincaré Probab. Stat.* **53** 226–242. MR3606740 https://doi.org/10.1214/15-AIHP714

[18] HILÁRIO, M., UNGARETTI, D., VALESIN, D. and VARES, M. E. (2022). Results on the contact process with dynamic edges or under renewals. *Electron. J. Probab.* **27** Paper No. 91, 31. MR4455874 https://doi.org/10.1214/22-ejp811

[19] HUANG, X. and DURRETT, R. (2020). The contact process on random graphs and Galton Watson trees. *ALEA Lat. Am. J. Probab. Math. Stat.* **17** 159–182. MR4057187 https://doi.org/10.30757/alea.v17-07

[20] JACOB, E., LINKER, A. and MÖRTERS, P. (2019). Metastability of the contact process on fast evolving scale-free networks. *Ann. Appl. Probab.* **29** 2654–2699. MR4019872 https://doi.org/10.1214/18-AAP1460

[21] JACOB, E., LINKER, A. and MÖRTERS, P. (2022). The contact process on dynamical scale-free networks. Available at arXiv:2206.01073.

[22] JACOB, E. and MÖRTERS, P. (2017). The contact process on scale-free networks evolving by vertex updating. *R. Soc. Open Sci.* **4** 170081, 14. MR3666161 https://doi.org/10.1098/rsos.170081

[23] LALLEY, S. and SU, W. (2017). Contact processes on random regular graphs. *Ann. Appl. Probab.* **27** 2061–2097. MR3693520 https://doi.org/10.1214/16-AAP1249

[24] LIGGETT, T. M. (1999). *Stochastic Interacting Systems*: *Contact*, *Voter and Exclusion Processes*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **324**. Springer, Berlin. MR1717346 https://doi.org/10.1007/978-3-662-03990-8

[25] LINKER, A., MITSCHE, D., SCHAPIRA, B. and VALESIN, D. (2021). The contact process on random hyperbolic graphs: Metastability and critical exponents. *Ann. Probab.* **49** 1480–1514. MR4255151 https://doi.org/10.1214/20-aop1489

[26] LINKER, A. and REMENIK, D. (2020). The contact process with dynamic edges on $\mathbb{Z}$. *Electron. J. Probab.* **25** Paper No. 80, 21. MR4125785 https://doi.org/10.1214/20-ejp480

[27] MOUNTFORD, T., MOURRAT, J.-C., VALESIN, D. and YAO, Q. (2016). Exponential extinction time of the contact process on finite graphs. *Stochastic Process. Appl.* **126** 1974–2013. MR3483744 https://doi.org/10.1016/j.spa.2016.01.001

[28] MOUNTFORD, T., VALESIN, D. and YAO, Q. (2013). Metastable densities for the contact process on power law random graphs. *Electron. J. Probab.* **18** No. 103, 36. MR3145050 https://doi.org/10.1214/EJP.v18-2512

[29] MOUNTFORD, T. S. (1993). A metastable result for the finite multidimensional contact process. *Canad. Math. Bull.* **36** 216–226. MR1222537 https://doi.org/10.4153/CMB-1993-031-3

[30] MOURRAT, J.-C. and VALESIN, D. (2016). Phase transition of the contact process on random regular graphs. *Electron. J. Probab.* **21** Paper No. 31, 17. MR3492935 https://doi.org/10.1214/16-EJP4476

[31] REMENIK, D. (2008). The contact process in a dynamic random environment. *Ann. Appl. Probab.* **18** 2392–2420. MR2474541 https://doi.org/10.1214/08-AAP528

[32] SCHAPIRA, B. and VALESIN, D. (2017). Extinction time for the contact process on general graphs. *Probab. Theory Related Fields* **169** 871–899. MR3719058 https://doi.org/10.1007/s00440-016-0742-0

[33] SCHONMANN, R. H. (1985). Metastability for the contact process. *J. Stat. Phys.* **41** 445–464. MR0814841 https://doi.org/10.1007/BF01009017

[34] SEILER, M. and STURM, A. (2023). Contact process in an evolving random environment. *Electron. J. Probab.* **28** Paper No. 1, 61. MR4651886 https://doi.org/10.1214/23-ejp1002

[35] SEILER, M. and STURM, A. (2023). Contact process on a dynamical long range percolation. *Electron. J. Probab.* **28** Paper No. 157, 36. MR4669748 https://doi.org/10.1214/23-ejp1042

[36] STACEY, A. (2001). The contact process on finite homogeneous trees. *Probab. Theory Related Fields* **121** 551–576. MR1872428 https://doi.org/10.1007/s004400100149